

ISSN 1257-1520

Les Après-midi de LAIRDIL

n° 14

**Le suivi des apprenant/es par les
systèmes numériques**

LAIRDIL

2009

*LAIRDIL - IUT Université Toulouse III
115 B route de Narbonne - 31 077 Toulouse Cédex 4
Tél.: 05 62 25 80 43 - Fax: 05 62 25 80 46
Courriel: lairdil@lairdil.org – <http://www.lairdil.org>*

Laboratoire Inter-Universitaire de Recherche en Didactique des Langues

Créé en 1989, LAIRDIL est un laboratoire inter-universitaire de recherche de l'IUT de l'Université Toulouse III. Il a pour objet la recherche en didactique et pédagogie des langues. La diffusion des résultats de cette recherche est une priorité.

Chaque année, LAIRDIL organise un cycle de séminaires-conférences sur des sujets de pédagogie ou de didactique susceptibles d'intéresser les enseignant/es d'anglais, voire d'autres langues. La conférence constitue la première partie d'une brochure sur ce thème. Les membres du laboratoire et d'autres personnes ajoutent leurs réflexions propres sur le thème abordé.

La journée d'études sur le suivi des apprenant/es en langues vivantes a eu lieu le 20 octobre 2006 conjointement, en visioconférence, à l'Université Toulouse III et à l'Université Paris-Dauphine. Elle a été organisée par Christine Vaillant-Sirdey, membre du laboratoire LAIRDIL et Alain Cazade.

Numéros disponibles des *Après-midi de LAIRDIL*

- *The Problems of Oral Testing. What Did you Say?*
- *Autonomous Learning of Vocabulary Through Extensive Reading*
- *Film, TV and Videotapes in EFL*
- *Aspects of Fluency and Accuracy*
- *Maximizing the Value of Jigsaw Activities*
- *Ten Top Principles in the Design of Vocabulary Materials*
- *Spécial recherche*
- *Questions d'articles: l'article scientifique*
- *La culture*
- *La pratique didactique du multimédia*
- *L'interaction orale*
- *Stratégies d'apprentissage*
- *La communication interculturelle*

Numéros disponibles des *Cahiers Pédagogiques de LAIRDIL*

- *Vocabulaire technique et apprentissage des langues de spécialité*
- *Apprentissage de l'anglais technique en autonomie guidée: Application au génie chimique*
- *Poèmes pour la classe de langues/Poems for the English class*
- *Jeux pour la classe d'anglais / Games for the English*
- *Nouvelles pour la classe d'anglais / Short stories for the English class*
- *Les homographes: un outil pour aborder la phonétique + Dictionnaire vocal LAIRDIL*
- *Les métiers de la forme: glossaire anglais-français*

Rédacteur en chef: **Alain Cazade**

Responsable d'édition: **Nicole Décuré**

©février 2009

Sommaire

Suivi apprenant: éléments d’une problématique	7
<i>Alain Cazade</i>	
Approches psychométriques de l’évaluation à l’aide de test.....	19
<i>Marianne Mavel, Pascal Bessonneau, Christophe Lalanne</i>	
Médiation, suivi et tutorat en ligne: approche théorique, perspectives de recherche	45
<i>Jean-Claude Bertin, Jean-Paul Narcy-Combes, Patrick Gravé</i>	
Suivi de l’apprenant à distance: plaidoyer pour un <i>feedback</i> rationnel d’accompagnement conatif et évaluatif.....	63
<i>Jean-Christophe Coquilhat</i>	
Conception d’un outil expérimental de suivi apprenant en langues: le cas de “SUIVIX”	93
<i>Alain Cazade</i>	
Moodle et le suivi des apprenants en langues vivantes Olivier Catteau, présentation commentée par Christine Vaillant	117

Suivi apprenant: éléments d'une problématique

Alain Cazade, Professeur, Université Paris Dauphine

Introduction

L'enseignement à distance est, aujourd'hui, en forte progression dans tous les cadres institutionnels ou/et privés de formation. Il existe de nombreux environnements matériels et surtout logiciels pour exploiter ce qui apparaît, à divers égards, comme un véritable filon. Parmi les applications les plus souvent citées, on peut mentionner *WebCT-BlackBoard*, *FirstClass*, *Learning Space*, *Ingénium*, *Moodle*, *Dokeos*, etc. À l'origine de la mise en place de tels environnements et de l'engouement qui l'accompagne, on trouve toutes sortes de motivations, plus ou moins nobles et désintéressées. L'insistance omniprésente dans le camp des instances décisionnaires sur les avantages économiques qui en découlent ne peut qu'amener à se demander si les avantages au plan pédagogique entraînent une conviction de même nature chez les principaux utilisateurs, enseignants et apprenants. Il semble évident que l'introduction de tels outils là où des formations équivalentes n'existaient pas ne peut qu'apporter un plus. Pourtant, on doit bien constater que, bien souvent, ce qui est visé est essentiellement la disparition ou, au mieux, la progressive diminution d'un accompagnement présentiel, toujours coûteux à maints égards, au profit d'activités non présentesielles, jugées à court terme moins onéreuses car seulement intermédiées par une relation homme-machine dont "l'efficacité" ne peut être, notamment dans un apprentissage en sciences humaines et donc à long terme, qu'inégalement convaincante. La question est tout particulièrement délicate à apprécier dès lors qu'il s'agit de formations en langues vivantes où, pour ne prendre qu'un aspect parmi d'autres, le développement et l'accompagnement du travail oral, surtout en phase d'entraînement à l'interaction, posent des questions aiguës dues à la prise en compte de formulations/productions progressives, d'hésitations, de reconstructions, de gestion de bribes d'expression significatives, tous éléments fondamentaux dans l'apprentissage d'une langue car inhérents à la nature même des processus qui le caractérisent.

De nombreux sigles apparaissent dans ce domaine. Cela tient au fait que les aboutissements de ce thème de recherche omniprésent touchent d'innombrables secteurs de l'activité d'apprentissage: l'apprenant lui-même, l'enseignement et la formation, les concepteurs de produit, les didactiseurs, le domaine comportemental en général, la psychologie cognitive en particulier, les neurosciences, sans oublier le contexte administratif – pour n'en citer que quelques-uns. Parmi ces sigles, l'un deux mérite une attention particulière: l'EIAH pour "Environnement Informatique d'Apprentissage Humain". Ce sigle est également le titre d'une conférence annuelle, résultant de la fusion de deux précédentes conférences "Hypermédias et Apprentissages" et "Journées Environnements Interactifs pour l'Apprentissage par Ordinateur".

Par ailleurs, pour déborder l'ancien sigle EAO (enseignement aidé par ordinateur), puis EIAO (enseignement intelligemment aidé par ordinateur), une nouvelle approche demande à être ciblée, permettant de tenir compte des outils de communication et d'interaction informatisés à distance (le mot distance étant pris à la fois aux sens physique et temporel). C'est ainsi qu'apparaissent désormais l'AMO ("enseignement médiatisé par ordinateur") et plus récemment: l'EAMD, pour "Environnements d'Apprentissage Médiatisés et Distribués", et même WELL (*Web Enhanced Language Learning*) afin de souligner la portabilité des modules d'enseignement concernés.

Des attentes fortes mais encore mal définies

Le développement de tous ces environnements de formation à distance entraîne nécessairement des attentes fortes en termes de gestion du "suivi apprenant". Force est de constater que les réponses données à ces attentes ne sont pas toujours à la hauteur des espérances qu'on peut avoir.

En fait, il est intéressant de constater que le thème précis du "suivi apprenant" est en forte progression sur la planète Internet. Si ce n'est pas encore une préoccupation fondamentale, une recherche dans ce domaine portant sur cette chaîne exacte de caractères proposait en septembre 2007 seulement une vingtaine de réponses. Aujourd'hui, en novembre 2008, on obtient une série de plus de 1200 contacts possibles. Le moteur *Webcorp* coordonné avec *Google* en ciblant le domaine des sciences humaines montre la même progression exponentielle. Avec un outil comme *Kartoo* (<http://www.kartoo.com/>), dont le fonctionnement et surtout les affichages rappellent ceux que l'on obtient avec des logiciels de "*mind mapping*", une recherche ciblant une combinatoire des termes "suivi" "apprenant" "dispositif" "enseignement" produit environ 5500 points de référence. Pour obtenir des réponses plus avancées, on peut encore consulter quelques sites spécialisés, comme "Thot" ("Nouvelles de la formation à distance", <http://www.thot.cursus.edu/>) ou encore "*Advanced Distributed Learning*" (<http://www.adlnet.gov/index.aspx>). On peut aussi rechercher sur *e-learning*: <http://www.brandon-hall.com/>.

Les réponses obtenues (*hits*) portent le plus souvent sur des fiches de suivi proposées dans divers types de formation, beaucoup moins souvent sur la problématique permettant de les établir.

La mise en place d'un dispositif permettant de suivre un apprenant, à distance ou même en présentiel, peut s'appuyer sur un nombre considérable de paramètres pratiquement impossibles à tous intégrer dans un système rationnel. Par rationnel, on entendra un système distinguant de manière motivée reconnaissable un certain nombre de priorités suivant des axes d'approche variables.

Priorités variables

Parmi les axes privilégiés de suivi dans un environnement FOAD (Formation ouverte à distance) figurent en bonne place les types de ressources auxquels les apprenants ont accès, comme le redisent Monique Grandbastien et Lahcen Oubahssi

dans un article de juillet 2007¹. Ceci est vrai d'ailleurs depuis les premières solutions de plateformes FOAD, lesquelles ambitionnaient déjà de "mettre des contenus de formation à disposition des apprenants et leur offraient diverses fonctionnalités de communication, entre eux ou avec leurs enseignants" (*ibid.*). Pour répondre depuis lors aux besoins d'organiser cette approche, "[des] langages de modélisation pédagogique sont apparus et par ailleurs diverses spécifications sont aussi proposées pour tenter de standardiser les échanges dans le domaine de la formation en ligne." (*ibid.*). Les auteurs de l'article classent ainsi les propositions existant "selon trois types: ceux qui modélisent les ressources (ARIADNE, CanCore, Dublincore, LOM, etc.), ceux qui modélisent les activités en particulier l'activité d'apprentissage (EML, IMS LD, etc.) et ceux qui modélisent d'autres éléments comme les compétences et l'apprenant (PAPI, IMS LIP, etc.)". Cette simplification de la diversité est séduisante mais ne saurait répondre, et les auteurs cités en sont bien conscients, qu'à une partie du problème relativement aisée d'accès.

En fait, à tout niveau et quels que soient les axes d'approches privilégiés, en plus de la nature des outils et ressources divers mis à disposition, les priorités divergent encore si on part

- d'un point de vue administratif,
- d'un point de vue pédagogique,
- du point de vue de la recherche.

Chacune de ces trois subdivisions premières varie elle-même suivant, à tout le moins, deux points d'ancrage:

- le point de vue enseignant,
- le point de vue apprenant.

Mais cette vue dichotomique est encore trop simplificatrice. Essayons de détailler un peu plus avant ces subdivisions.

- *Sur le plan administratif*: il y aura lieu de mettre en avant l'observation de la gestion des travaux *apprenants*, orientation, bilans successifs, réorientations éventuelles etc., mais aussi la prise en compte des activités *enseignantes* (types d'interventions, horaires, etc.) afin de suivre et régler leurs activités, les mettre en contact avec les cohortes apprenantes voulues, les rétribuer comme il sied, etc.
- *Sur le plan pédagogique*: l'enseignant voudra pouvoir...
 - faciliter l'accès des apprenants à des outils et ressources divers, comme suggéré ci-dessus, bénéficier d'interfaces intermédiaires de type "base de données", documentées, illustrées graphiquement, catégorisées, toutes fonctionnalités permettant d'aider les procédures de choix nécessaires;
 - observer les réactions de l'apprenant en situation d'apprentissage, avoir les outils techniques lui permettant de les analyser en vue d'une remédiation didactique immédiate et/ou différée (autocorrection/autobilan pour l'apprenant et autre pour l'enseignant), avec des retombées possibles sur le plan de l'amélioration des outils déjà en place;
 - savoir – ce qui est peut-être le plus difficile à réaliser avec un accompagnement automatisé – quand il *ne faut pas* intervenir, a) afin de laisser l'apprenant aborder les

¹ <http://www.revue-eti.net/document.php?id=1476>.

questions rencontrées avec suffisamment de temps pour que les étapes cognitives successives nécessaires à l'avancement de sa compréhension puissent se mettre en place et s'enchaîner, *b)* afin surtout de favoriser la prise de responsabilité individuelle de l'apprenant dans son parcours et la prise de conscience de son autogestion apprenante.

- *Sur le plan de la recherche:* on pourra vouloir se donner les moyens d'un recul théorisant, avec possibilité d'identification de différents profils apprenants, de stratégies d'apprentissage récurrentes, de pratiques métadidactiques majoritaire, de modèles de remédiation possibles; il sera intéressant également – c'est précisément l'optique que je favorise personnellement – d'envisager les retombées de tous ces aspects sur la conception d'outils appropriés permettant d'optimiser l'accompagnement apprenant.

Cadre ↓	Type de suivi ↓					
Administratif	plus axé sur le quantitatif =====	sommatif	suivi simple	statique	court terme	pratique
Pédagogique apprenant et enseignant	plus de qualitatif que de quantitatif =====	formatif	suivi plus élaboré	dynamique réactif et récuratif – reméditatif	Immédiat et différé	pratique et théorique sur le plan linguistique
Recherche	Qualitatif - quantitatif pour faire apparaître le qualitatif =====	Sommatif et formatif	suivi beaucoup plus élaboré	dynamique réactif et récuratif – projectif	non immédiat mais bouclages nécessaires	théorique linguistique et didactique

Tableau 1

Les priorités ne sont ainsi pas les mêmes suivant les cadres d'appréciation où l'on se place et on pourrait les distinguer sommairement en se référant au tableau 1. S'il est évident qu'il faut viser au bout du processus une approche sommative, il est tout autant nécessaire de préserver à tout moment une approche formative, et c'est d'ailleurs sur ce plan que les problèmes se posent avec le plus de difficultés si l'on veut intégrer cette approche de façon adéquatement variable en fonction des étapes d'apprentissage et des divers intervenants concernés. L'auto-évaluation continue pour l'apprenant est un élément clé pour qu'il puisse savoir où il en est, ce qu'il doit faire pour faire mieux ou/et autrement si nécessaire, pour qu'il puisse assumer sa prise de responsabilité et l'autogestion effective – et peut-être efficace – de son parcours. De même pour l'enseignant qui doit pouvoir influencer, aider, optimiser les parcours tout au long de leurs déroulements. Sans oublier l'institution qui veut

pouvoir évaluer les systèmes en place, enseignants et apprenants, pour améliorer ce qui peut l'être, déplacer, recadrer, etc., avant de mieux relancer la "machine" de la formation en fin de cycle.

Quels objets pour un suivi intelligent?

Afin de suivre les parcours des apprenants, il devient nécessaire de pouvoir avoir recours à des modélisations portant sur les "objets" travaillés. Ces modélisations multiples essayent de discerner un certain nombre de paramètres susceptibles de ramener simultanément à l'étude des différences. De telles modélisations touchent:

- les items ressources utilisables et utilisées,
- les modes et modalités d'utilisation,

mais aussi:

- les savoirs et compétences visées,
- les comportements et actions des apprenants.

Du fait de la multiplicité des secteurs concernés, ces modélisations sont on ne peut plus délicates à mettre en place, surtout si l'on veut pouvoir échanger et partager entre chercheurs les résultats de ses expérimentations.

Pour donner une idée de la multiplicité des paramètres qu'on doit pouvoir étudier pour suivre ces comportements et actions, et donc en ne prenant qu'une approche ciblant l'apprenant et en ne prenant que quelques axes d'analyse, on pourra chercher à étudier l'interdépendance des items suivants:

- les *didactèmes*, ou unités minimales d'apprentissage visées,
- les *productèmes*, ou unités minimales de productions apprenantes,
- les *actèmes*, ou actions minimales décelables chez l'apprenant.

Options linguistiques

Encore faut-il s'entendre, si l'on accepte ces axes d'intérêt, sur le sens et les réalités qu'on place derrière de tels vocables, chacun d'eux ramenant aux querelles ou options divergentes que connaissent les chercheurs dans chacun des domaines afférents. Ces questions sont fondamentales et resurgissent à tous les niveaux du suivi apprenant, autant dans les phases formatives que sommatives, et donc dans la mise au point de tout processus évaluatif.

Pour ne prendre que les termes didactèmes et productèmes, on imaginera aisément les difficultés de compréhension et d'échange de résultats qui naîtront d'une approche énonciative ou non au départ, approche chère à la conception française de la linguistique, qui, malgré de notables évolutions dans les années récentes, ne parvient malgré tout pas à susciter l'assentiment des chercheurs du monde anglophone. Faudra-t-il, pour rejoindre une vision Chomskyenne et plus agréable au monde anglo-saxon, favoriser une approche générativiste et considérer l'expression langagière, pour simplifier, comme pouvant être décrite, et donc reproduite, par une sorte de système mathématique décomposant des liens d'arborescences impliquées entre tous les items utilisés? Faudra-t-il encore chercher ici et là les traces d'une grammaire universelle, en gardant présentes à l'esprit les différentes étapes que Chomsky a pu

discerner dans ce concept? Faudra-t-il partir de l'idée de son *Language Acquisition Device* (LAD), en considérant qu'un tel dispositif peut pré-structurer les éléments rencontrés au fil de l'apprentissage, ou bien, pour aller à grands pas dans ces visions évolutives de l'apprentissage des langues, faudra-t-il privilégier l'échange avec l'environnement cognitif proposé par le cadre d'apprentissage et l'adapter au vu des appropriations progressives de l'apprenant, en gardant l'optique que Bruner privilégiait? En partant d'un autre point de vue, des difficultés au moins aussi nombreuses surgiront si l'on privilégie dans la performance langagière la notion de compétence ou celle de connaissance (même si l'une ramène à l'autre). Faudra-t-il s'intéresser prioritairement aux traces d'interlangue décelables dans les productions apprenantes, afin d'y repérer les traces de règles inférées depuis la langue première et éviter qu'elles ne perturbent l'approche de la langue cible ou bien faudra-t-il considérer que ces systématisations approchantes sont autant d'émergences d'une stratégie métadidactique qui ne demande qu'à être valorisée, fournissant divers points de départ pour un échange productif (*feedback*) avec l'apprenant en vue de favoriser ce stade négocié de l'apprentissage que les didacticiens appellent *uptake*?

Sans vouloir pousser trop loin ce résumé un peu indigeste de conceptions passablement concurrentes, on percevra que l'axe d'étude portant sur le suivi du parcours apprenant différera grandement en fonction de tels choix et qu'on pourra difficilement comparer les comportements apprenants (stratégies métadidactiques ou/et "performances") qui auront été différemment décrits suivant tel ou tel modèle.

Des réponses à certaines des questions posées ci-dessus découleront les choix d'actèmes à retenir comme jalons comportementaux significatifs. Les mouvements plus ou moins erratiques de la souris sur la table, les allées et venues du curseur dans un écran de saisie de réponse (avec notamment les rectifications d'écriture que cela souligne) pourront être considérés, dans certains cas, comme devant être suivis mais trop lourds à gérer et traiter valablement dans beaucoup d'autres. En revanche les mouvements navigationnels au sein d'un module didactique, ou parmi plusieurs applications (recherche d'aide via un dictionnaire, d'une information aperçue dans un écran exploratoire pour aider à la production d'une réponse dans tel ou tel exercice), parmi plusieurs sites internet visités, par exemple, dans le cadre d'une *Webquest*², seront la plupart du temps pris en compte. De même, la notion temps (temps de réponse, temps d'hésitation, temps entre plusieurs actèmes, etc.) sera pratiquement toujours riche de sens.

Pistes et axes à explorer

Cette multiplicité d'approches ne rebute pourtant pas la recherche dans le secteur du suivi apprenant. Parmi les pistes qui semblent le plus souvent explorées, l'axe sémantique se révèle particulièrement intéressant.

² Une "*webquest*" est une sorte de rallye utilisant l'internet pour que les apprenants puissent quérir depuis divers sites une série d'informations permettant de construire, en fin de parcours, tel ou tel objet d'apprentissage, exposé, rapport, reportage, etc.

Axe sémantique

Gilbert Paquette³ décrit ainsi le “Le référencement sémantique des objets”:

Sans une représentation des connaissances, l’environnement d’apprentissage ne peut fournir qu’une faible assistance à ses utilisateurs. Un générateur de portail à base d’objets de connaissance doit contenir, en plus d’un éditeur de la structure des activités, un éditeur de la structure cognitive permettant d’associer connaissances et compétences aux activités et aux ressources de la structure pédagogique. L’utilisation de ces éditeurs clefs par l’ingénieur pédagogique permet de rendre ces structures visibles au moment de la diffusion. Les connaissances et les compétences visées servent à orienter la démarche de l’apprenant et le soutien apporté par le formateur. Elles fournissent également une base par rapport à laquelle l’apprenant peut évaluer ses progrès, permettant également l’évaluation formative ou évaluative des apprentissages par un formateur.

Paquette développe le concept de “portail à base d’objets de connaissance”, qui doit orienter l’apprenant vers une structure d’activités possibles autour d’une série de ressources pédagogiques mais qui doit également offrir une ouverture sur la structure cognitive des objets proposés, et donc tenter de relier les descripteurs orientés “connaissances” et “compétences” aux ressources accessibles et aux activités qui sont liées à celle-ci.

L’organisation de ces ressources dans des référentiels et leur référencement au moyen de métadonnées, d’ontologies ou de modèles des connaissances et des compétences font du concept “d’objet de connaissances” l’élément central d’un portail d’apprentissage. Ce faisant, je tracerai l’évolution des recherches, des développements et des applications que nous menons au Centre de recherche LICEF-CIRTA depuis sa création en 1992. (*ibid.*)

Axe praxématique

Cet axe semble également prometteur. Il est encore peu exploré dans ce type de recherche mais il promet déjà d’être une source très riche d’enseignements. Rappelons que la praxématique est centrée sur l’analyse de la production du sens en langage. En linguistique anthropologique, c’est une branche de recherche qui se veut à la fois “réaliste”, au sens objectuel du terme, et qui vise la dynamique liée aux processus d’échanges. L’idée fondatrice à la base de cette démarche est que les représentations linguistiques naissent dans la nature même et le développement de la *praxis*. Parmi ses pistes de recherche privilégiées, le chercheur en praxématique s’intéresse tout particulièrement à la description et l’analyse de corpus fabriqués à partir d’écrits et d’échanges oraux; l’étiquetage (*tagging*) qui accompagne ces corpus met ainsi fortement en valeur les liens possibles avec les contextes de production de ces corpus.

³ Directeur du réseau de recherche LORNET – qui tient la chaire de recherche en ingénierie cognitive et éducative (CICE) – Centre LICEF-CIRT) – Conférence ouverture aux “Premières rencontres jeunes chercheurs en EIAH” : INT d’Évry 11-12 mai 2006. http://events.it-sudparis.eu/rjc_eiah2006/conference.

Modélisation des items, des apprenants et des parcours – T.R.I.⁴

Il est impossible de ne pas mentionner ici un ressort qu'on doit désormais considérer comme nécessaire dans toute activité apprenante liée à des outils numériques: l'adaptativité. Cet aspect d'un environnement apprenant informatisé lui confère une attractivité irremplaçable et une véritable identité. Sans ce ressort déterminant, les dispositifs d'apprentissage à base de TICE, tout spécialement en langues, ne peuvent prétendre qu'à être des adaptations plus ou moins réussies, agréables et innovantes, des environnements classiques de travail utilisés bien avant l'arrivée de l'informatique.

L'adaptativité de l'environnement d'apprentissage, et donc sa réactivité au comportement particulier d'un apprenant donné, implique de s'intéresser à ce qu'on appelle la "théorie des réponses aux items" ou TRI⁵. La TRI est plus souvent évoquée en relation avec la modélisation des procédures de test qu'en lien avec la gestion des parcours apprenants mais chacun comprendra que ces champs de recherche sont liés puisque, aussi bien au moment de concevoir un parcours véritablement adaptatif, qu'au moment d'apprécier les réactions de chaque apprenant dans ce parcours, il faudra bien être capable d'intégrer dans l'éventail des déroulements anticipés, non seulement un descriptif des objets présentés et des liens qui peuvent les relier les uns aux autres, mais également les aptitudes cognitives et comportementales (profils auditif, visuel, capacité à maintenir l'attention, etc.) de l'apprenant.

Après les années 60, la théorie classique des tests, les travaux de Spearman, Thurstone, Guilford, etc., se heurtaient à diverses difficultés pour être maintenus comme bases de travail productives alors que, parallèlement, le besoin de recours à des tests d'intelligence, de connaissances, en divers domaines incluant d'ailleurs celui des langues, étaient de plus en plus demandés par le monde du travail. La constatation que le postulat d'un facteur d'"homogénéité de la variance" était marqué par un manque de réalisme flagrant s'imposait à toute recherche sérieuse en la matière. De même, les travaux voulant intégrer l'indépendance des items et l'indépendance des sujets posaient des difficultés de modélisation de plus en plus évidentes et pourtant incontournables aux modèles existant alors, plus déterministes que probabilistes. Les modèles mis en place dans les années 50, ne parvenaient pas à dépasser les caractères ou "traits" dépendant de chaque groupe testé. Des "traits latents" (traits de compétences, de capacité, d'aptitude, etc.) propres à tel ou tel individu dans chaque groupe témoin devaient pouvoir être pris en compte. L'élément "chance" dans le choix des réponses à un ensemble de questions posées⁶ devait également être intégré aux modèles testés. La théorie des réponses aux items ou *Item Response Theory* en anglais (IRT), l'intégration de calculs probabilistes, voulaient se

⁴ Voir, pour ce qui concerne la présentation des problèmes de gestion de l'attention: <http://www.irdp.ch/edumetrie/tri.htm>. Voir aussi: <http://www.psy.ulaval.ca/~pepin/tifmanuel.pdf>.

⁵ Cf. article produit par les collègues travaillant au CIEP dans ce même volume.

⁶ Les questionnements à base de QCM, et donc le nombre limité de réponses possibles attendues, entraînent nécessairement la prise en compte du fait que l'individu testé coche la bonne réponse au hasard. Notons qu'un autre problème reste lié à bon nombre de tests encore proposés aujourd'hui, puisque, de par la formulation même des questions, certains individus testés parviennent à deviner – ou plutôt déduire – la réponse attendue.

rapprocher d'une modélisation plus fidèle des évaluations tenant plus complètement compte de l'hétérogénéité des individus testés, en intégrant notamment les difficultés liées à la gestion de l'attention. Elle continuait à postuler l'unidimensionnalité du trait mais laissait possible le recours à la multi dimensionnalité. Le modèle mathématique de Rasch est souvent – car l'un des plus simples en la matière – évoqué en lien avec cette théorie: il vise à modéliser la relation entre le “trait latent” de l'individu testé et la probabilité qu'il réponde correctement à un item-question donné.

Une norme en constant développement: SCORM

Le sigle SCORM (*Sharable Content Object Reference Model*) correspond à un modèle, développé à l'origine aux États-Unis, en lien avec l'univers du ministère de la défense⁷, comme un prolongement des travaux du consortium ADLI (*Advanced Distributed Learning Initiative*). Ce modèle fait également suite au travail de l'AICC (*Aviation Industry CBT Committee*) et des demandes de Boeing. Le sigle CBT, correspondant à *Computer-Based Training* (ou formation aidée par l'ordinateur), montre bien que le ressort initial de ces normalisations est étroitement lié aux exigences de l'informatique sur le monde de l'apprentissage.

Pour revenir à la norme SCORM, celle-ci demande que les objets pédagogiques créés respectent un ensemble de caractéristiques et spécificités compatibles entre divers supports ou/et environnements d'apprentissages, dans tous les compartiments d'enseignement possibles, en langues y compris.

D'autres sigles existent qu'il serait trop long de développer ici mais qui montrent bien que les acteurs du secteur de la formation aussi bien dans le monde du travail que dans le milieu “académique” comprennent bien qu'il faut tenir compte, dès le moment de la conception, aussi bien du matériel que des espaces logiciels, de besoins plus ou moins communs d'échange, d'efficacité, de rentabilité. On pourra citer ainsi le sigle LMS, ou *Learning Management System*, (également décliné en LCMS, ou *Learning Content Management System*) et son pendant, LOM, ou *Learning Object Metadata*, le premier étant orienté vers le dispositif, au sens matériel et organisationnel du terme, le deuxième étant plus porté vers la description et donc la gestion des objets intégrés dans le processus de formation.

Pour détailler rapidement le modèle SCORM, précisons qu'il demande à tout dispositif de formation et à tout outil et concepteur d'outil, d'intégrer divers paramètres, devant chacun répondre à un certain nombre d'exigences en matière:

- d'*accessibilité*: localement ou, le plus souvent, à distance, via Internet, via des plateformes électroniques d'enseignement, de façon à permettre l'interrogation et l'administration au sens large des objets et sujets gérés par le système;
- d'*adaptabilité*: au gré des exigences liées à tel ou tel type de formation, d'apprenants, d'objet d'apprentissage;
- de *durabilité*: une autre forme d'adaptabilité, en quelque sorte, concernée par l'évolution des systèmes, sur le plan technologique mais aussi sur le plan pédagogique et donc conceptuel;

⁷ On sait que la dynamique et le sens de l'efficacité de la sphère militaire ont souvent accéléré les choses en matière de formation et d'enseignement.

- d'*interopérabilité*: ce point reprend celui de l'accessibilité sous un angle logistique essentiellement concerné par les échanges entre outils différents, logiciels, plateformes, modules hébergés par un site donné mais faisant ponctuellement appel à un autre site pour qu'une mutualisation des éléments déjà développés ailleurs puisse être effective;
- de *réutilisabilité*: ce point développe le précédent et vise à faciliter l'intégration des modules extérieurs dans des parcours et applications multiples; l'idée est que chacun ne soit pas obligé de tout refaire à chaque fois, là où il se trouve, alors que d'autres ont déjà donné une réponse satisfaisante à telle ou telle question posée.

Comme toute norme qui veut embrasser la diversité du monde, celle-ci ne recueille pas l'enthousiasme de la majorité des acteurs concernés du domaine, ce qui est peut-être tout particulièrement inévitable dans le domaine de l'éducation. Elle est toutefois vécue comme une nécessité, appelée donc à évoluer – ce qu'elle a déjà constamment fait depuis son origine pourtant récente.

Conclusion

On voit bien que les champs d'application et donc de recherche concernés par l'étude d'un parcours apprenant sont multiples et demandent à s'ouvrir sur des pans de la recherche qui dépassent allègrement les limites habituellement posées par de nombreux écrits portant sur la pédagogie. Et rien n'a encore été dit de la prise en compte de la phonologie, de l'interaction sociale (cf. *pair work*, *peer working*, parfois appelés en français "échafaudage mutualisé" ou "procédure d'étayage"). Tous ces domaines sont plus riches les uns que les autres et pourtant inter-dépendants. Ces constatations expliquent pourquoi un sourire parfois ironique pointe sur le visage de leurs spécialistes lorsqu'ils découvrent que le commerce propose à tel ou tel centre de langue un certain nombre d'outils, de logiciels pédagogiques, de progiciels d'administration, revendiquant un intérêt qualifié pour la gestion des ressources, des utilisateurs, apprenants et enseignants, et même des évaluations, en mode individuel ou en groupe, certains prétendant donner à eux seuls réponse à l'ensemble de ces questions.

Bibliographie

BARBOT, M.-J. & T. LANCIEN (dir.). 2003. *Médiation, médiatisation et apprentissages. Notions en Questions 7*. ENS Éditions.

BEACCO, J-C., M. DI GIURA, O. GUEYE, P. TEUTSCH & F. TAUZER-SABATELLI. 1999. *Enseigner à distance le français comme langue étrangère: Principes pour l'élaboration d'un environnement d'apprentissage numérique et interactif intégré: CROISIÈRES*. Deuxièmes Entretiens Internationaux sur l'Enseignement à Distance, 1 et 2 décembre 1999, CNED Poitiers Futuroscope, décembre 1999. 209-217.

BOURDET J.-F. & P. TEUTSCH. 2000. Définition d'un profil d'apprenant en situation d'auto-évaluation. *ALSIC 5* (vol 3, n°1). 125-136.

- DESPRÉS, C. 2001. *Modélisation et conception d'un environnement de suivi pédagogique synchrone d'activités d'apprentissage à distance*. Thèse de doctorat, Université du Maine, Le Mans.
- GUEYE, O. & P. TEUTSCH. 2001. Visualisation of Learner's Profile in Distance Training Situation. ICDE'2001, 20th World Conference on Open Learning and Distance Education, Dusseldorf, 1-5 April 2001.
- JACQUINOT, G. 1999. Le tutorat: pièce maîtresse et pourtant parent pauvre des systèmes et dispositifs de formation à distance. Deuxièmes Entretiens Internationaux du CNED, Poitiers Futuroscope, décembre 1999.
- BLAIS J.-G., BERTRAND R., RAÎCHE G. 2004. *Modèles de mesure: L'apport de la théorie des réponses aux items*. Presses de l'Université du Québec.
- KOPER, R. 2003. *Combining re-usable learning, resources and services to pedagogical purposeful units of learning*. A. Littlejohn (dir.). *Reusing Online Resources: A Sustainable Approach to eLearning*. Kogan Page. 46-59.
- LAPERROUSAZ, C., P. LEROUX, & P. TEUTSCH. 2005. *Perception of individual activities in a group activity through qualitative information about the group dynamics*. Web Intelligence, Compiègne. 732-738.
- OUBAHSSI, L. 2005. *Conception de plates-formes logicielles pour la formation à distance, présentant des propriétés d'adaptabilité à différentes catégories d'utilisateurs et d'interopérabilité avec d'autres environnements logiciels*. Thèse de Doctorat en informatique, Université René Descartes-Paris V.
- OUBAHSSI, L. & M. GRANDBASTIEN. 2006. *From learner information packages to student models: Which continuum?* The 8th International Conference on Intelligent Tutoring Systems. Jhongli, Taiwan, June 26-30, 2006.
- OUBAHSSI, L. & M. GRANDBASTIEN. 2007. Une généralisation du modèle d'activité d'IMS LD pour les systèmes de e-formation. *e-TI - la revue électronique des technologies* 4. <http://www.revue-eti.netdocument.php?id=1476>.
- PAQUETTE, G. 2002. *Modélisation des connaissances et des compétences*. Presses de l'Université du Québec.
- PAQUETTE, G. 2005. Apprentissage sur l'Internet: des plateformes aux portails à base d'objets de connaissance. Pierre, S. (dir.). *Innovations et tendances en technologies de formation et d'apprentissage*. Presses de l'école polytechnique de Montréal. 1-30.
<http://www.licef.teluq.quebec.ca/gp/docs/Chapitre%20DIVA.doc>.
http://www.licef.teluq.quebec.ca/gp/fr/publications/campus_virtuel.htm.
- PAQUETTE, G., I. ROSCA, S. MIHAILA & A. MASMOUDI. 2006. TELOS, a service-oriented framework to support learning and knowledge Management. PIERRE, S. (dir.). *E-Learning Networked Environments and Architectures: a Knowledge Processing Perspective*. Springer-Verlag.

PERNIN, J.P. & A. LEJEUNE. 2004. Dispositifs d'apprentissage instrumentés par les technologies: vers une ingénierie centrée sur les scénarios. Actes du Colloque TICE octobre 2004, Université de Technologie de Compiègne. 407-414.

TEUTSCH, P., X. DUBOURG & O. GUEYE. 2001. Modèles d'interaction pour CROISIERES, dispositif de formation à distance. Poster aux journées EIAO 2001 dans le cadre de la conférence ASTI. Actes publiées dans la Revue *STE* (Sciences et Techniques Éducatives) 8 : 1-2, 2001. 168-170.

TRICOT, J.-M. & M. MESBAH. 2000. Un modèle de réponses aux items. Propriétés et comparaison de groupes de traitement en épidémiologie. *Revue de Statistique Appliquée* 48 : 4. 29-39.

http://archive.numdam.org/ARCHIVE/RSA/RSA_2000__48_4/RSA_2000__48_4_29_0/RSA_2000__48_4_29_0.pdf.

Sitographie concernant certaines normes utilisées pour cerner la notion de suivi apprenant

IMS Global Consortium

<http://www.imsglobal.org>

Pôle Conseil Business Interactif. Étude des outils de gestion de ressources numériques pour l'enseignement.

<http://www.educnet.education.fr/lcms/index.php>

IEEE P1484.1/D9, Draft Standard for Learning Technology - Learning Technology Systems Architecture (LTSA).

<http://ltsc.ieee.org>.

ARIADNE, Alliance of Remote Instructional Authoring & Distribution Networks for Europe.

<http://www.ariadne-eu.org>

Canadian Core Learning Resource Metadata Specification.

<http://www.cancore.ca>

Dublin Core. Metadata Initiative

<http://dublincore.org/>

ADL - SCORM

<http://www.adlnet.org>

LOM

<http://ltsc.ieee.org/wg12>

Approches psychométriques de l'évaluation à l'aide de tests

*Marianne Mavel*¹, *Pascal Bessonneau*, *Christophe Lalanne*
Centre International d'Études Pédagogiques (CIEP)

Introduction

Le ministère de l'Éducation nationale, de l'enseignement et de la recherche a confié au CIEP, Centre International d'Études Pédagogiques, établissement public national à caractère administratif sous tutelle de ce ministère, "la responsabilité de l'organisation hors de France des examens institués par le ministère de l'Éducation nationale pour évaluer l'enseignement du français langue étrangère" (article 2 du décret 2000 – 1017 du 12 octobre 2000).

Dans cette optique, le CIEP assure la conception et la gestion pédagogique et administrative de diplômes officiels destinés à tous les publics non francophones: le DELF (diplôme d'étude de langue française), le DALF (diplôme approfondi en langue française) et depuis peu le DILF (diplôme initial de langue française). Il gère également le test TCF (test de connaissance en français). Ces certifications évaluent les compétences langagières des candidat/es en français. Elles sont harmonisées sur le cadre européen commun de référence des langues (CECR), c'est-à-dire qu'elles sont calibrées et étalonnées sur des normes européennes et internationales. Le CECR a été élaboré par la division des politiques linguistiques du Conseil de l'Europe. Il s'agit d'un outil de référence européen pour l'évaluation, l'enseignement et l'apprentissage des langues. On retrouve également en annexe les six niveaux communs de référence établis par ALTE (Association of language testers in Europe – association des centres d'évaluation en langues en Europe). La qualité des certifications proposées par le CIEP est toujours vérifiée par une expertise pédagogique et, dans le cas du TCF et du DELF, une expertise psychométrique des épreuves de compréhension et de maîtrise des structures de la langue est aussi réalisée.

La psychométrie regroupe un ensemble de méthodes dont l'objectif est de mesurer de manière équitable et rigoureuse des capacités humaines qui ne sont pas directement observables (traits latents) telles que les compétences intellectuelles ou les connaissances en français, par exemple, (voir Bernier et Pietrulewicz, 1997). L'origine de la psychométrie remonte à la fin du 19^{ème} siècle notamment avec les travaux de Wundt, Galton et Cattell. Cette discipline s'est développée au début du 20^{ème} siècle avec la mise au point par Binet et Simon en 1905 de l'échelle métrique de l'intelligence. Cet outil permet d'évaluer le niveau mental d'un enfant en prenant compte de son âge. Les instruments de mesure utilisés pour évaluer ces capacités

¹ Cellule qualité et expertises - Psychométrie, département évaluation et certifications, Centre International d'Études Pédagogiques (CIEP), 1 avenue Léon-Journault, 92318 Sèvres cedex. Contact : mavel@ciep.fr.

humaines doivent posséder un certain nombre de qualités fondamentales: l'objectivité, la validité et la fidélité de la mesure. Ils prennent des formes très variées telles que des tests de connaissance, des questionnaires ou des échelles. Un test est un instrument d'évaluation dont l'objectif est de mesurer une faculté essentiellement cognitive telle que la compétence langagière. Un test se compose de plusieurs items dont les réponses sont binaires ou multiples. Pour quantifier les compétences individuelles, on peut appliquer un modèle mathématique de mesure sur les données recueillies à l'aide de l'instrument de mesure. Les constructeurs de tests ont souvent recours à deux modèles: la théorie classique des tests (TCT) et la théorie de réponse à l'item (IRT). La TCT fournit des informations sur la qualité fonctionnelle des items d'un test et l'IRT permet de donner aux candidat/es des scores fidèles et précis reflétant bien leurs positions par rapport au trait latent mesuré.

Cet article traitera de l'analyse psychométrique des items d'un test. La première partie sera consacrée au cadre général de la construction d'un instrument de mesure et visera à souligner les qualités fondamentales d'un instrument de mesure et la TCT. La deuxième partie développera le concept des modèles de l'IRT. La dernière partie présentera des exemples d'items du DELF analysés à l'aide de ces deux modèles de mesure.

I. Cadre général de la construction d'un instrument de mesure

1.1. Le trait latent

Un instrument de mesure cherche à mesurer une caractéristique chez un individu qui n'est pas directement mesurable, on appelle cette caractéristique un trait latent ou le construit de l'instrument. Par exemple, dans le cadre de l'évaluation des langues, la compétence langagière est assimilée à un trait latent.

1.2. L'objectivité

Un instrument de mesure est considéré comme étant objectif si les résultats qu'il fournit ne dépendent pas du correcteur ou de l'évaluateur qui l'utilise. Un individu devrait donc obtenir le même résultat à une épreuve quelle que soit la personne qui a corrigé sa copie. L'objectivité peut être assimilée à la fidélité inter-juges (inter-correcteurs). Dans le cadre de la correction d'épreuves de production écrite, par exemple, l'utilisation de grilles de correction standardisées et de manuels de correction permettent de limiter les biais classiques de correction (effet de halo, de compensation) et fournit un cadre de référence pour les correcteurs/rices. Il existe différentes méthodes pour s'assurer de leur degré d'objectivité: le calcul de coefficients de corrélation ou de concordance tel que le Kappa de Cohen, ou l'application d'un modèle de réponse à l'item, par exemple le modèle de Rasch à facettes multiples.

1.3. La fidélité

La fidélité ou fiabilité désigne la capacité d'un instrument de mesure à appréhender avec précision les caractéristiques qu'il est censé mesurer. Un

instrument de mesure est considéré fidèle si le résultat qu'il produit est reproductible. En théorie, un individu auquel on aura appliqué plusieurs fois le même instrument de mesure obtiendra à chaque fois le même score. L'analyse de la fidélité permet de s'assurer de l'équité des différentes versions d'un instrument de mesure. Il existe différentes méthodes pour évaluer la fidélité: les tests parallèles, le test-retest, la consistance interne du test (méthode des moitiés, alpha de Cronbach).

La méthode des tests parallèles donne une estimation du degré d'équivalence entre deux versions d'un même test composées d'items différents. Elle consiste à calculer la corrélation entre les scores obtenus aux deux versions du test qui auront été passées par les mêmes individus. La difficulté consiste à s'assurer de l'équivalence du contenu et du construit des deux versions.

La méthode du test-retest permet de mesurer la stabilité du test dans le temps. La même version du test est appliquée au même groupe de candidat/es à deux moments différents plus ou moins espacés dans le temps. Tout comme la méthode des tests parallèles, on calcule la corrélation entre les deux passations du test. Le principal problème de cette méthode est l'effet d'apprentissage chez les candidat/es dû au fait qu'ils/elles repassent la même version du test.

La méthode la plus souvent utilisée pour estimer la fidélité consiste à évaluer la consistance interne du test (ou son homogénéité) car elle ne nécessite qu'une seule administration du test et qu'elle s'applique aussi bien aux items dichotomiques² qu'aux items polytomiques³. Évaluer la consistance interne d'un test consiste à déterminer dans quelle mesure les items le composant évaluent la caractéristique (le trait latent) que le test doit mesurer. Comme il est impossible de déterminer la fidélité d'un instrument de mesure à partir d'une seule version de celui-ci, on obtient, en estimant la consistance interne du test, la limite inférieure de la fidélité et on sous-estime la fidélité de l'instrument. On juge qu'elle est satisfaisante si le coefficient de fidélité est assez élevé (alpha de Cronbach > 0,70).

1.4. La validité

La validité d'un instrument de mesure désigne son aptitude à bien mesurer ce qu'il est censé mesurer et uniquement cela. On distingue trois types de validité: la validité de contenu, la validité de construit et la validité prédictive. La vérification des deux premières repose essentiellement sur le jugement professionnel d'experts.

Un test est valide du point de vue de son *contenu* s'il est représentatif en tant qu'échantillon du domaine de comportement mesuré. On détermine la représentativité du test en s'assurant que ses contenus sont pertinents et couvrent bien le domaine que le test est censé mesurer. Dans le cadre de l'évaluation des compétences langagières, le CECR fournit un ensemble de recommandations pour les concepteurs de tests.

² Un item dichotomique est un item binaire, il est soit réussi, soit échoué. Le score de cet item ne peut alors prendre que deux valeurs: 0 pour l'échec et 1 pour le succès par exemple.

³ Un item polytomique est un item dont le score prend plus de deux valeurs. Le score varie entre 0 et une valeur supérieure à 1 appelée score maximal de l'item. Par exemple, un item noté sur deux points pourra être codé de la manière suivante: 0, l'item est échoué; 1, l'item est partiellement réussi; 2, l'item est complètement réussi.

En ce qui concerne la validité de *construit*, il s'agit d'établir jusqu'à quel point le test fournit une mesure adéquate du construit théorique qui est mesuré (utilisation d'un ensemble de méthodes). Cette validation s'établit à partir d'une accumulation progressive de données qui viennent appuyer l'hypothèse selon laquelle "le test mesure bien ce qu'il prétend mesurer".

La validité *prédictive* s'applique lorsque l'on cherche à connaître le degré auquel les scores d'un test permettent de prédire la caractéristique étudiée. Par exemple, les tests d'aptitudes scolaires permettent de prédire les chances de réussite à l'école.

1.5. La théorie classique des tests

La théorie classique des tests (TCT), appelée aussi théorie du "score vrai" a été développée au début du 20^{ème} siècle (Dickes *et al.*, 1994). Elle est utilisée essentiellement pour détecter un certain nombre de dysfonctionnements concernant les composantes du test (les items) et à estimer la fidélité de l'instrument de mesure. Elle permet de calculer un ensemble d'indicateurs permettant de juger de la qualité du test: le score brut (score total⁴), la difficulté de l'item, le pouvoir discriminant de l'item et l'indice de consistance interne du test. Le modèle de mesure utilisé est un modèle linéaire. Cette théorie est basée sur la décomposition des scores observés en un score vrai et une erreur de mesure. La relation fondamentale postulée dans ce cadre théorique est alors: score observé = score vrai + erreur de mesure.

Le score observé d'une personne est le score qu'elle a obtenu en passant le test. On peut concevoir le score vrai d'un individu comme étant le score qu'il obtiendrait à un test parfait, c'est-à-dire un test qui mesure sans erreur. Dans le cas d'un test parfait, le score observé serait égal au score vrai. Cependant, en pratique, il existe toujours une marge d'erreur. On peut aussi exprimer le score vrai comme étant le score moyen d'une personne qui serait évaluée indéfiniment avec le même instrument de mesure. Le score vrai et l'erreur de mesure sont inconnus et l'on suppose l'erreur de mesure aléatoire.

1.5.1. Analyse des items

L'analyse des items consiste à étudier leurs qualités fonctionnelles à l'aide de deux indicateurs statistiques: l'indice de difficulté et le pouvoir discriminant. L'interprétation de ces indicateurs est facilitée par une analyse graphique. Cette analyse permet également de détecter un certain nombre de dysfonctionnements: des items trop faciles ou trop difficiles, des réponses ambiguës, des distracteurs non choisis. L'ensemble de ces informations est évidemment très utile lors de l'élaboration d'un test.

Indice de difficulté d'un item

L'indice de difficulté d'un item calculé sur un échantillon donné est une estimation de la difficulté de l'item. L'indice de difficulté d'un item polytomique est égal au score moyen relatif des candidat/es à cet item (score moyen des candidat/es

⁴ Le score total est le nombre total de points qu'a obtenu le/la candidat/e au test.

divisé par le score maximal de l'item). Dans le cas dichotomique, il s'agit plus simplement de la proportion de candidat/es ayant répondu correctement à l'item. Cet indice varie entre 0 et 1. Plus la valeur de l'indice est élevée, plus l'item est facile.

Indice de discrimination d'un item

Le pouvoir discriminant d'un item est sa capacité à différencier les candidat/es en fonction d'un critère donné, par exemple leur niveau de compétence. Il existe plusieurs types d'indices pour évaluer le pouvoir discriminant d'un item: la corrélation entre la réponse à l'item et le score total au test (item-test), la corrélation entre la réponse à l'item et le score au test sans cet item (item-reste) et pour les items à choix multiple, les corrélations entre les distracteurs et le score au test.

L'indice de discrimination est un indice d'homogénéité; il indique dans quelle mesure un item contribue à la détermination du score total. Cet indice varie entre -1 et 1. S'il est proche de zéro, l'item ne permet pas de différencier les candidat/es selon leur compétence. S'il est négatif, l'item présente alors une anomalie: il est mieux réussi par les candidat/es qui ont obtenu un score total faible.

Notons que l'utilisation de la corrélation item-test présente un inconvénient lié au fait que le score à l'item est compris dans le score total du test. Ainsi, la corrélation entre la réponse à l'item et le score total peut être positive alors que l'item ne contribue pas à augmenter le score total des candidat/es en fonction de leur compétence. Le coefficient de corrélation item-reste permet de corriger ce problème. On note que la différence entre la corrélation item-test et la corrélation item-reste est assez élevée quand un test est composé d'un petit nombre d'items.

En pratique, dans le cas des items dichotomiques, le coefficient de corrélation point-biserial, noté r_{pbis} , est très souvent utilisé comme indice de discrimination. Dans le cas d'un test composé d'items polytomiques, on utilise le coefficient de corrélation de Bravais-Pearson.

Analyse graphique

L'analyse graphique d'un item est très utile car elle permet de visualiser la difficulté et le pouvoir discriminant de l'item. Les figures 1a et 1b sont les représentations graphiques des réponses à deux items à choix multiples dont la clé⁵ est A (B et C sont les distracteurs⁶). Nous pouvons lire sur ces graphiques la proportion de réponses correctes et incorrectes en fonction de quatre groupes de niveaux, de taille à peu près équivalente. Ces groupes de candidat/es sont ordonnés par le score total (score brut): le groupe 1 est ainsi composé des candidat/es dont les scores totaux sont les plus faibles et le groupe 4 des candidat/es dont les scores totaux sont les plus élevés. Pour chaque groupe de niveau, la proportion de réponses correctes (réponse A) est calculée (ici en pourcentages), ainsi que les proportions de réponses incorrectes B et C. On retrouve ces proportions en ordonnées sur les graphiques de la figure 1.

⁵ La clé est la bonne réponse à l'item.

⁶ Les distracteurs sont les réponses incorrectes d'un l'item à choix multiple.

Dans le cas de la figure 1a, le taux de réponses correctes (courbe A) augmente fortement en fonction du niveau des candidat/es: on passe d'un taux de réussite de 20 % pour le groupe 1 à un taux de 80% pour le groupe 4. À l'inverse, la proportion de réponses incorrectes (courbes B et C) diminue en fonction du niveau des candidat/es. Ainsi, lorsque l'item permet de bien distinguer les candidat/es en fonction de leur niveau, on dira que l'item est discriminant. Le coefficient de corrélation point biserial de cet item est bien supérieur à 0; il vaut 0,50 ce qui indique que cet item a un fort pouvoir discriminant et confirme ce que l'on a observé graphiquement. Concernant la difficulté de l'item, elle semble plutôt moyenne car les candidat/es de bon niveau réussissent en moyenne très bien cet item contrairement aux candidat/es dont le niveau est plus faible: le taux de réussite des groupes 3 et 4 est supérieur à 50 % et celui des groupes 1 et 2 est inférieur à 30 %.

L'item de la figure 1b est un item très facile, son taux de réussite stagne autour de 95 % quel que soit le groupe de niveau. Cet item ne permet donc pas de différencier les candidat/es selon leur groupe de niveau, ce que l'on vérifie aisément à l'aide de l'indice de discrimination (cf. 1.5.1. Analyse des items, § Indice de discrimination d'un item), le coefficient de corrélation point-biserial, qui vaut 0,08 (valeur très proche de 0).

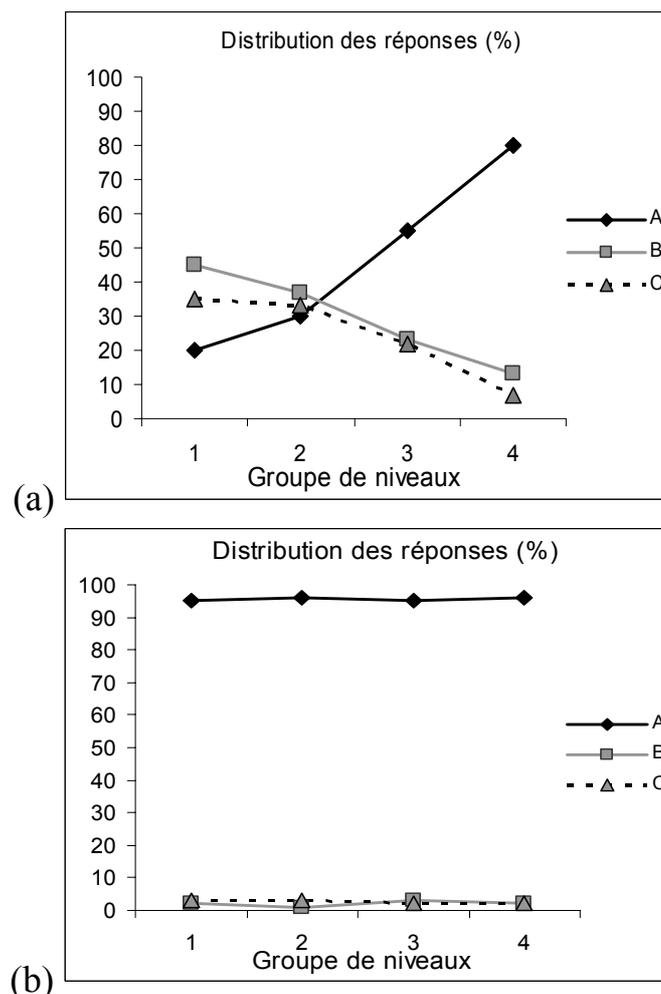


Figure 1 – Graphique des proportions de réponses A, B et C des candidat/es pour un item donné.

1.5.2. Fidélité – alpha de Cronbach

L'alpha de Cronbach⁷ est un indice permettant de mesurer la consistance interne d'un test. Il permet de déterminer dans quelle mesure l'ensemble des items évalue bien la caractéristique que le test doit mesurer. La valeur de cet indice dépend à la fois de la longueur du test (nombre d'items) et de l'homogénéité des items (déterminée à partir des inter-corrélations entre items). Il est à noter qu'une façon d'augmenter la consistance interne d'un test est de rajouter des items.

L'alpha de Cronbach varie entre 0 et 1: plus sa valeur est proche de 1, plus le test est consistant. Comme nous l'avons dit plus haut (cf. § 1.3.), cet indice n'est pas une estimation de la fidélité, il s'agit de l'estimation de la valeur minimale que pourrait atteindre la fidélité. En règle générale, on juge qu'elle est satisfaisante si l'alpha de Cronbach est supérieur à 0,70.

1.6. Les limites du modèle

La théorie classique des tests se heurte, dans la pratique, à deux problèmes. Tout d'abord, elle fournit des résultats dépendants de l'échantillon d'individus. Ainsi, un item considéré comme un item facile pour un échantillon d'individus, pourrait apparaître difficile si l'on changeait l'échantillon. Il suffit de penser à un item dont la difficulté serait estimée chez des élèves de lycée puis chez des élèves de collège. De même, un exercice de conjugaison sur le plus-que-parfait sera plus facilement réussi par un groupe d'élèves qui a étudié ce temps en classe que par un autre groupe qui ne l'a jamais étudié.

D'autre part, il apparaît délicat de comparer les résultats obtenus à deux tests contenant des items de difficultés différentes. Par exemple, si deux candidat/es passent deux tests différents, il sera difficile d'affirmer que le/la premier/e candidat/e a une compétence deux fois plus élevée si son score est deux fois plus élevé que celui du/de la deuxième candidat/e. Il s'agit là d'un problème lié au fait que la TCT repose sur l'utilisation du score total, qui n'est rien d'autre qu'un rang (tel/le candidat/e a obtenu plus de bonnes réponses que tel/le autre) et ne possède pas de métrique propre (l'écart entre le nombre de bonnes réponses ne peut être interprété de manière directe et absolue).

Malgré ces inconvénients, cette théorie continue d'être très largement utilisée par les chercheurs/ses et les concepteurs/rices de tests car elle offre de très bons outils pour l'analyse statistique des items. La reconnaissance des limites de ce modèle a conduit à de nouveaux développements et à l'élaboration de nouveaux modèles psychométriques, notamment les modèles issus de la théorie de réponse à l'item (TRI) dont l'objectif est d'estimer les caractéristiques des items (par exemple la difficulté, le pouvoir discriminant) ainsi que le niveau de compétence des individus. La partie suivante se propose de présenter le concept de l'IRT et différents modèles de réponse à l'item.

⁷ Le coefficient alpha de Cronbach est aussi appelé KR20 quand il est utilisé avec des items dichotomiques.

II. Les modèles de réponse à l'item

Les modèles de réponse à l'item font suite aux travaux de deux grandes écoles, l'une danoise et l'autre américaine composées de mathématicien/nes et de psychométricien/nes des années 60 (Rasch, 1960; Birnbaum, 1968). Nous décrivons leur intérêt par rapport à la théorie classique des tests (pour une présentation plus détaillée, consulter Embretson et Reise, 2000). Puis nous décrivons le plus simple d'entre eux, le modèle de Rasch, avant d'évoquer le cas des modèles de réponse à l'item plus complexes (nombre de paramètres plus élevés, utilisation d'items polytomiques).

À la différence de la théorie classique des tests, les modèles de réponse à l'item vont permettre d'obtenir des paramètres d'item indépendants de ceux de la population sur laquelle ils ont été obtenus et inversement: ils possèdent ce que l'on appelle la qualité d'*objectivité spécifique*. Par ailleurs ils vont permettre d'obtenir conjointement une mesure de difficulté (pour les items) et d'habileté (pour les individus), interprétable sur la même échelle assimilée à une mesure continue du trait latent (la compétence linguistique par exemple). Ceci est valable dans le cas où l'on ne considère qu'une seule dimension (cas unidimensionnel) mais peut être généralisé au cas où plusieurs dimensions sont susceptibles de rendre compte des performances des individus (cas multidimensionnel). Par souci de clarté, nous nous limiterons volontairement à l'étude du cas unidimensionnel.

2.1. Présentation générale

Les modèles de réponse à l'item sont avant tout des modèles probabilistes, mais également des modèles qui permettent de fournir une mesure objective de la performance des individus, en relation avec une théorie clairement circonscrite tant dans son domaine que dans les moyens qui permettent de la vérifier.

L'essence même de ce type de modèle repose sur l'utilisation non pas de la réponse de l'individu à un item donné mais plutôt de la probabilité de réponse associée à cet item. Généralement, cette probabilité est la probabilité que le/la candidat/e donne une réponse correcte, mais l'on peut très bien modéliser la probabilité que le/la candidat/e choisisse l'une des réponses incorrectes (distracteur). Il convient dès lors de noter que la relation entre le trait latent et la réponse des candidat/es n'est plus considérée comme linéaire, à la différence de la théorie classique des tests (cf. §1.5). Cette probabilité de réponse est modélisée par ce que l'on appelle la courbe caractéristique de l'item (CCI): celle-ci représente alors la probabilité d'une réponse correcte à un item donné en fonction du niveau présumé (habileté) du/de la candidat/e.

2.2. Le modèle de Rasch

Le modèle de Rasch est le modèle de réponse à l'item le plus simple. Il permet de traiter exclusivement les réponses dichotomiques (codées vrai ou faux, par exemple), et se limite à un seul paramètre: la difficulté de l'item. On suppose également que les réponses s'expliquent par un seul trait latent (hypothèse d'unidimensionnalité) et que les réponses sont indépendantes les unes des autres. En

d'autres termes, l'ensemble des items qui composent le test permet de mesurer la même dimension, et la réponse d'un/e candidat/e à un item donné n'est pas influencée par sa réponse à un autre item. Nous reviendrons sur ces deux postulats de base après avoir brièvement décrit le principe général de fonctionnement du modèle de Rasch.

2.2.1. Caractérisation d'un item sur une échelle de mesure

Comme on l'a indiqué plus haut, le lien entre le trait latent et le score observé peut être modélisé par la courbe caractéristique de l'item (Figure 2). Plus précisément, le modèle de Rasch vise à modéliser l'écart entre l'habileté de l'individu et la difficulté de l'item. Rappelons que ces deux mesures sont exprimées sur la même échelle (exprimée le plus généralement en unités logit, par construction du modèle statistique) qui est supposée mesurer de façon continue le trait latent étudié.

Le recours à un tel modèle permet un *calibrage* qui conduit à une échelle commune aux individus et aux items. Tous les items mesurant une compétence particulière peuvent être positionnés le long d'une échelle, leurs positions et leurs espacements étant déterminés par le niveau de difficulté auquel ils correspondent. La réussite d'une personne à une partie de ces items peut être exprimée au travers d'une valeur qui correspond à un point quelque part sur cette échelle. Les items situés à gauche de la position de cette personne sont des items plus faciles pour lesquels elle a une probabilité de réussite supérieure à 50 %. Les items situés à droite sont des items plus difficiles pour lesquels la personne a une probabilité de réussite inférieure à 50 %. La difficulté d'un item est ainsi définie comme la valeur sur le trait latent pour laquelle l'individu a 50% de chance de répondre correctement.

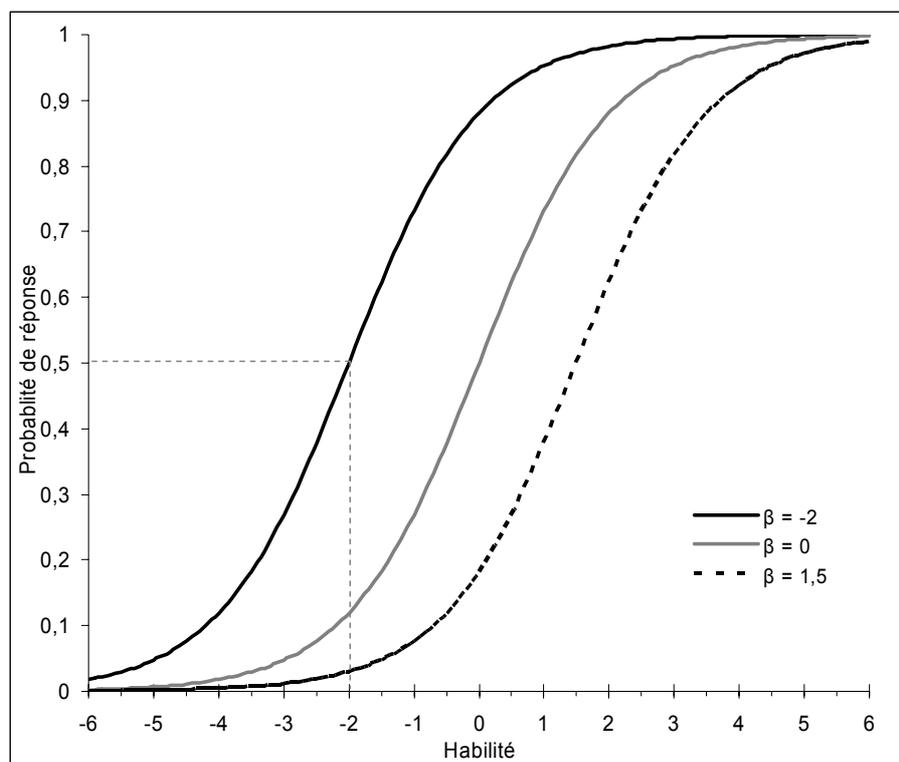


Figure 2 – Courbes caractéristiques de trois items de difficulté croissante

Comme on peut le voir dans la figure 2, la probabilité pour un/e candidat/e de répondre correctement (en ordonnées) augmente avec son habileté (en abscisses). Lorsque l'habileté est très inférieure à la difficulté, la probabilité de trouver la bonne réponse est pratiquement nulle. Inversement lorsque l'habileté est très supérieure à la difficulté, la probabilité de trouver la bonne réponse est proche de 1.

Ce modèle permet par conséquent d'ordonner les items en fonction de leur difficulté et d'évaluer le niveau des candidat/es en fonction des réponses correctes qu'ils/elles ont fournies. Naturellement, l'estimation de l'ensemble des paramètres définissant le modèle de Rasch pour un test donné (habileté de chacun des individus et difficulté de chacun des items) est une procédure assez lourde qui est, fort heureusement, confiée aux ordinateurs.

Ce type de modèle permet de fournir une indication précise concernant les caractéristiques intrinsèques d'un item, et ce de manière indépendante de l'échantillon d'individus auxquels les items ont été administrés. Chaque item apporte une certaine quantité d'information et celle-ci est maximale quand la difficulté est égale à l'habileté. On peut montrer que l'erreur de mesure, quant à elle, est inversement proportionnelle à la racine carrée de l'information. En d'autres termes, lorsque l'on cherche à situer précisément le niveau d'un/e candidat/e, on gagnera d'autant plus d'information en utilisant des items dont la difficulté est proche de son habileté réelle.

Étant donné que les items sont supposés indépendants, on va pouvoir additionner l'information apportée par chacun des items et calculer la courbe d'information du test. Celle-ci donne une idée de la gamme d'habiletés que l'on est susceptible d'évaluer de manière suffisamment précise avec le test considéré. Un test linguistique composé d'items de compréhension écrite ou orale de difficulté relativement peu élevée permettra de bien évaluer des candidat/es débutant/es dans cette langue. À l'inverse, un test linguistique similaire, composé uniquement d'items relativement difficiles, serait peu approprié pour évaluer ce même public de candidat/es: l'écart trop important entre le niveau réel des candidat/es (estimé au travers de leur habileté) et la difficulté des items rendrait très imprécis leur positionnement relatif sur une même échelle de mesure.

2.2.2. Hypothèses

Comme on l'a indiqué, le modèle de Rasch comme la plupart des modèles statistiques repose sur des hypothèses fortes: l'unidimensionnalité et l'indépendance locale

L'*unidimensionnalité* signifie qu'il n'y a qu'un seul et unique trait latent qui explique les réponses des individus. Cette hypothèse est difficile à vérifier en pratique. De nombreuses méthodes dans la littérature sont proposées. On trouvera des indices tels que l'indice de reproductibilité de Guttman ou l'alpha de Cronbach (cf. §1.5.2). Des tests statistiques ont également été créés (DIMTEST de Stout, par exemple). Les analyses factorielles et les analyses en composantes principales sont également largement utilisées. La plupart des méthodes développées ont pour objet,

sinon de prouver l'unidimensionnalité, du moins de démontrer la prédominance d'un seul et unique trait latent qui soit à même d'expliquer les performances observées.

La seconde hypothèse fondamentale est l'*indépendance locale*. Cette hypothèse stipule que, pour un niveau d'habileté donné, les réponses aux items sont indépendantes. En d'autres termes, pour l'ensemble des candidat/es possédant le même niveau de compétence, la réponse à un item n'a aucune influence sur la réponse à un autre item.

Ces deux hypothèses sont liées dans le sens où si le test est unidimensionnel alors l'hypothèse d'indépendance locale est vérifiée. Mais généralement la réciproque n'est pas vraie.

On notera enfin qu'une troisième hypothèse est également mise en avant par certains psychométriciens: il s'agit de la constance de la discrimination, c'est-à-dire de la capacité d'un item à bien discriminer les candidat/es selon leurs compétences et ce quelle que soit leur habileté. Ce point sera développé plus avant lors de la présentation des modèles plus complexes.

2.2.3. Propriétés

On a vu qu'une des propriétés d'un instrument de mesure était sa capacité à délivrer des scores reproductibles (cf. §1.3). Cette caractéristique est appelée propriété d'*objectivité spécifique*. Elle permet de comparer deux personnes indépendamment de l'instrument de mesure (avec des tests différents). L'une des conséquences de cette propriété est que pour un ensemble d'individus d'habileté différente, les difficultés des items demeurent les mêmes.

Par ailleurs, le score total possède ce que l'on appelle la propriété d'*exhaustivité*: toute l'information sur le trait latent est contenue dans le score. Ceci signifie que lorsque l'on considère qu'il n'y a qu'une seule dimension latente qui sous-tend les performances observées, la connaissance du score observé chez un ensemble de candidat/es suffit entièrement pour caractériser cette dimension et, par là, attribuer un niveau d'habileté propre à chaque personne et un niveau de difficulté propre à chaque item. Cette propriété est vraie pour les modèles de la famille du modèle de Rasch. Cette propriété est capitale pour certaines méthodes d'estimation des paramètres. Elle a, en revanche, une conséquence qui n'est pas intuitive en première approche. Deux candidat/es qui ont un même score global mais qui auront répondu correctement à des items différents auront la même estimation d'habileté. Si l'on exclut les scores parfaits (ensemble de réponses correctes ou incorrectes), on peut estimer 79 habiletés différentes à partir d'un test composé de 80 items. À chaque ensemble de réponses correctes correspond ainsi une habileté unique, nonobstant le pattern de réponses.

2.2.4. Adéquation entre les données réelles et les données modélisées

L'utilisation d'un modèle de réponse à l'item vise à fournir une mesure avec le moins d'erreurs possible (il s'agit du point de vue du statisticien), ou de manière équivalente à fournir un ensemble de prédictions ne s'écartant pas trop du modèle théorique supposé dès le départ (on parle de *construct*, et il s'agit plutôt du point de vue de l'expert en pédagogie). Il apparaît donc crucial de vérifier l'adéquation entre les données réelles et les données modélisées.

À l'image des tests portant sur la vérification de l'unidimensionnalité du test, les indices déterminant l'adéquation entre les données réelles et les données modélisées sont très nombreux. Ils dépendent étroitement du logiciel utilisé. Seuls deux indices simples et couramment employés seront abordés: l'*oufit*⁸ et l'*infit*⁹. Ces indices sont calculés à partir de ce que l'on appelle les résidus ou écarts entre données observées et données prédites par le modèle: les résidus sont ainsi estimés par la différence entre la réponse observée (binaire dans le cas du modèle Rasch qui permet de modéliser des réponses dichotomiques) et la probabilité de réponse du/de la candidat/e à l'item. L'*oufit* permet de détecter les réponses improbables des individus, par exemple un individu de niveau élevé qui échouerait à un item facile. L'*infit* est pondéré par l'information fournie par l'item; ceci permet de minimiser l'information apportée par une réponse inattendue de la part d'un/e candidat/e pour lequel l'item est inadapté. Par exemple, un item trop facile auquel un/e candidat/e de bon niveau échouerait serait moins pris en compte dans l'évaluation du niveau du/de la candidat/e. D'une certaine manière, ces indicateurs permettent de "filtrer" les données aberrantes qui peuvent résulter, par exemple, d'une erreur d'inattention de la part du/de la candidat/e au moment de fournir sa réponse.

Ces indices prennent la valeur 1 si l'ajustement est "parfait". On retiendra que les valeurs que doivent prendre ces indices pour indiquer un bon ajustement sont empiriques et varient selon les auteur/es. Par consensus, un bon ajustement est associé à des valeurs d'*oufit* situées entre 0,5 et 1,5 et à des valeurs d'*infit* entre 0,8 et 1,2.

2.3. Modèles plus complexes

Le modèle de Rasch est très simple. Il est notamment très réducteur concernant le comportement de l'item qui n'est caractérisé que par un paramètre de difficulté. En outre, il ne s'applique qu'aux items dichotomiques. Pour ces raisons, d'autres modèles plus complexes ont été développés dont quelques exemples sont présentés dans les paragraphes suivants.

2.3.1. Modèles à 2 et 3 paramètres

Dans le cas du modèle à 2 paramètres, appelé encore modèle de Birnbaum, un paramètre supplémentaire est ajouté: il s'agit de la discrimination. Ce paramètre va permettre de moduler la pente de la courbe caractéristique de l'item (Figure 3), c'est-à-dire la capacité de l'item à discriminer plus finement des individus possédant des niveaux d'habileté proches. Dans ce cas, plus la pente est élevée (c'est-à-dire plus la partie centrale de la courbe apparaît orientée verticalement), plus l'item est discriminant.

Une des objections couramment faite au modèle de Rasch est la constance de la discrimination, usuellement fixée à 1. En effet la discrimination est généralement corrélée positivement au coefficient point bisérial utilisé dans la Théorie classique

⁸ Moyenne non pondérée des carrés des résidus.

⁹ Moyenne pondérée des carrés des résidus.

des tests (§1.5.1). En pratique, il est pratiquement impossible d’obtenir une grande homogénéité des coefficients des points bisériaux des items. C’est pour cette raison que certains auteurs préconisent l’utilisation de ce modèle plutôt que du modèle de Rasch.

Dans le modèle à 3 paramètres, on ajoute un paramètre parfois appelé “*guessing*”, “conjoncture” ou “pseudo-chance”, et qui constitue une manière de rendre compte de la probabilité pour le/la candidat/e de deviner une bonne réponse parmi plusieurs réponses possibles. Dans ce cas, même pour les plus bas niveaux de l’échelle d’habileté (vers la gauche, sur l’axe des abscisses), c’est-à-dire pour les candidat/es dont le niveau de compétence est le plus faible, la probabilité de répondre correctement n’est pas nulle, quelle que soit la difficulté de l’item.

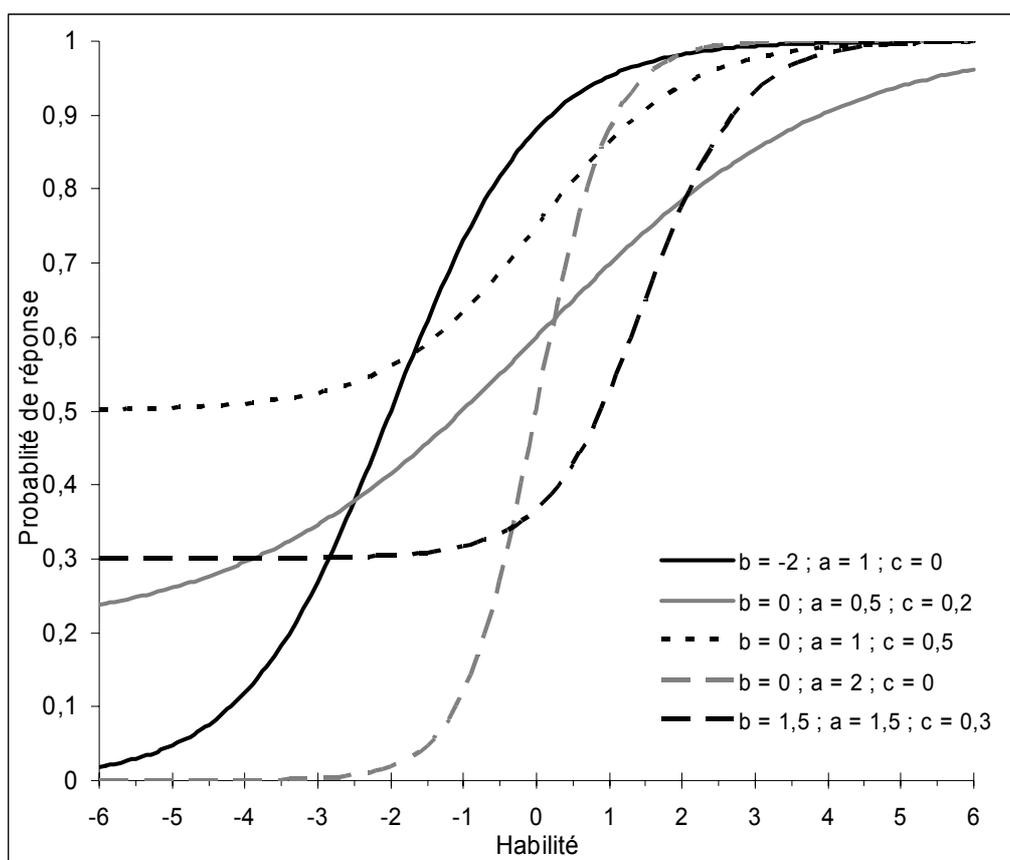


Figure 3 – Modèle de Birnbaum: exemples de courbes caractéristiques d’items
 Le paramètre b correspond à la difficulté de l’item, le paramètre a à la discrimination et le paramètre c au paramètre de “pseudo-chance”.

Avec ce type de modèles, on perd la propriété d’exhaustivité du score. Il n’est plus possible de calculer l’habileté simplement à partir des difficultés des items et du score des candidat/es. Il est alors nécessaire, pour chaque test, de calculer systématiquement, par une procédure assez lourde en temps de calcul, l’habileté du/de la candidat/e à partir de son “motif” de réponse. Enfin, il est nécessaire d’avoir un nombre plus important de réponses pour estimer les paramètres du modèle: de 600 à 800 individus au minimum pour le modèle à 2 paramètres et plus de 1000 individus

pour le modèle à trois paramètres, alors que pour utiliser le modèle de Rasch on peut se contenter de 250 individus.

2.3.2. Modèle pour items polytomiques

Dans le cas où le modèle est polytomique, la réponse à l'item n'est plus binaire et plus de deux réponses sont possibles. Le modèle de Rasch a donc été adapté pour ce type de situation, notamment quand la réponse est graduée, c'est-à-dire pour des échelles dite échelles de Lickert. Les réponses observées sur une échelle de ce type à 4 modalités de réponse peuvent être codées de la manière suivante: faux (0), faible (1), moyen (2), élevé (3), exact (4). Les modèles permettant de prendre en compte ce type de données dépendent de la façon d'interpréter le score associé aux réponses des candidat/es en relation avec le codage de cette échelle.

Le seul modèle qui sera évoqué ici est le modèle à crédit partiel (Masters, 1982) dont on trouvera un modèle d'application dans la partie suivante. Il est applicable lorsque les réponses sont ordonnées ou graduées. Contrairement à d'autres types de modèles élaborés pour traiter ce type de données (par exemple, les réponses graduées de Samejima, 1969), le modèle à crédit partiel est dérivé du modèle de Rasch et en conserve donc les propriétés: objectivité spécifique et exhaustivité du score. Tout comme dans le cas du modèle de Rasch, une difficulté est associée à chacune des modalités de réponse. Cette difficulté correspond au point pour lequel une réponse devient majoritaire par rapport à l'autre. En revanche, les courbes caractéristiques des items s'interprètent différemment car elles indiquent désormais la probabilité de fournir l'une des modalités de réponse (conditionnellement aux autres choix de réponse possibles) en fonction de l'habileté.

2.4. De la Théorie classique des tests aux modèles de réponse à l'item

Le modèle de réponse à l'item vient en complément de la théorie classique des tests. Si cette dernière permet de pointer les dysfonctionnements des items, elle ne permet pas de mesurer à la fois l'habileté des candidat/es et la difficulté des items. Or la relation qui lie ces deux grandeurs est la base même du comportement de l'individu. C'est parce qu'un/e candidat/e possède un bon niveau dans une langue étrangère, par exemple, qu'il/elle est à même de répondre correctement à des items dont le degré de difficulté est relativement peu élevé. C'est donc bien cette possibilité de modéliser la dépendance intrinsèque entre niveau d'habileté d'un individu et niveau de difficulté des items (dans le modèle de Rasch, cette dépendance ou relation fonctionnelle est objectivée par l'écart entre ces deux grandeurs) qui rend les modèles de réponse à l'item si attractifs et incomparablement plus utiles que la Théorie classique des tests lorsque l'on souhaite positionner des candidat/es les un/es par rapport aux autres à l'aide d'un test standardisé.

En revanche, comme tout modèle statistique, c'est-à-dire comme toute tentative de réduction de la réalité à un ensemble de processus objectivables au travers de mesures physiques, ce type de modèle possède des limitations. On peut en mentionner deux principales. La première concerne l'adéquation du modèle aux données: afin de fournir des estimations précises du niveau d'habileté, il semble évident que le modèle doit permettre de fournir des prédictions correctes et "se

conformer” aux données observées. En second lieu, les hypothèses d’application du modèle doivent être vérifiées. En effet, comment pourrait-on affirmer qu’un test mesure bien ce qu’il est censé mesurer, au travers de cet outil de mesure objectif que constitue le modèle de réponse à l’item, si en réalité les réponses observées ne peuvent être expliquées qu’à l’aide d’un faisceau de traits latents (violation du principe d’unidimensionnalité)?

III. DELF

Le DELF, Diplôme d’études en langue française et le DALF, Diplôme approfondi de langue française sont des diplômes du Ministère de l’éducation nationale, de l’enseignement supérieur et de la recherche. Ils s’adressent à des publics non francophones souhaitant faire valider leur connaissance du français. Ils existent depuis 1985 (arrêté du 22 mai 1985) et depuis 2005, ils sont harmonisés sur les 6 niveaux de l’échelle de compétence en langue définie dans le CECR (arrêté du 7 juillet 2005). Le DELF se compose de 4 diplômes indépendants correspondant aux niveaux A1, A2, B1 et B2 du CECR et le DALF de deux diplômes de niveaux C1 et C2¹⁰. Pour chaque niveau du DELF, quatre compétences sont évaluées: la compréhension orale, la production orale, la compréhension des écrits et la production écrite. Il existe pour le DELF une version “tous publics” et une version “junior et scolaire” destinée à des candidat/es scolarisé/es dans le secondaire. Régulièrement, des analyses psychométriques sont réalisées sur des épreuves de compréhension afin d’étudier la qualité fonctionnelle des items. Les psychométricien/nes étudient aussi le degré d’objectivité des correcteurs/trices pour certaines épreuves de production écrite. L’objectif à long terme est de pouvoir justifier de l’équité des épreuves du DELF.

Les épreuves de compréhension écrite et orale prennent le format d’un test composé de plusieurs exercices. Chaque exercice comprend plusieurs items portant sur un document ou un bloc de documents. La structure d’une épreuve est standardisée par le nombre, le type et le barème des exercices. Suivant le document sur lequel porte l’exercice, les concepteurs/trices d’items peuvent choisir le nombre et la forme des items. Les items peuvent prendre des formes assez variées: questions à choix multiples, questions ouvertes et exercices d’appariement.

Nous présentons, dans cette partie, l’analyse psychométrique de quatre items. Ces items proviennent d’une épreuve de compréhension écrite d’un sujet DELF niveau B2 version “scolaire et junior”. Les items 1 et 2 portent sur le document 1 et les items 3 et 4 sur le document 2 (Figure 5). Les items 1 et 3 sont deux questions à choix multiples composées d’une clé C et de deux distracteurs A et B. Les items 2 et 4 sont deux questions ouvertes, le premier est noté sur 3 points et le deuxième sur 2 points. Nous commencerons par étudier les qualités fonctionnelles des items avec la TCT et puis nous terminerons par le modèle PCM (cf. 2.3.2. Modèle pour items

¹⁰ A1, niveau élémentaire; A2, niveau élémentaire avancé; B1, niveau intermédiaire; B2, niveau intermédiaire avancé; C1, niveau supérieur; C2, niveau supérieur avancé.

polytomiques, § 0) avec lequel nous estimerons le niveau de difficulté des items et étudierons la qualité de la structure interne des items polytomiques. Ce modèle de réponse à l'item est utilisé car le DELF est composé d'items dichotomiques et polytomiques dont les modalités sont ordonnées (Bode, 2004; Linacre, 2004). Les modalités d'un item correspondent dans le cas du DELF aux différents scores qu'un individu peut obtenir à l'item. Elles sont ordonnées par ordre croissant de difficulté: un individu qui obtient 2 points à un item est censé avoir un niveau de compétence supérieur à celui qui n'obtiendra qu'1 point au même item.

Document 1: SPORTIFS EN HERBE

Qu'il court, nage, saute, tape dans un ballon ou pédale, l'enfant tire de multiples bénéfices – tant sur le plan physique que mental – de la pratique d'un sport. Ce dernier lui donne de l'assurance, modèle son corps, forge son esprit, facilite sa socialisation, développe son goût de l'effort. Et avec les diverses possibilités qui sont aujourd'hui proposées, il est rare qu'un enfant ne trouve pas une activité qui lui convienne. Le rôle des parents est ici primordial mais délicat: inciter sans forcer et savoir encourager l'athlète en herbe qui fait preuve de certaines qualités sans pour autant le lancer dans la course aux médailles.[...]

Contrairement à certaines rumeurs tenaces, le sport ne fait pas grandir. Et il est aujourd'hui prouvé qu'il ne bloque pas non plus la croissance. En revanche, certaines activités, comme la gymnastique, pratiquées de façon intensive peuvent ralentir cette dernière. Mais toutes les adorables petites gymnastes alliant grâce, souplesse et énergie finiront par avoir une taille parfaitement normale, seulement avec quelques années de retard. [...] Leur squelette est-il plus fragile pour autant? Pour répondre à cette question, le Dr Michel Binder a une explication très imagée: "Chez les adultes, les os peuvent être comparés à des murs de béton dans lesquels sont fixées des ficelles soutenant les muscles. Lorsque l'on tire trop fort sur la ficelle, elle casse. L'adulte se fait ainsi des claquages, des elongations ou des tendinites. Pour un enfant, les os sont plutôt des murs de plâtre dans lesquels sont enfoncés des clous, c'est-à-dire des cartilages de croissance, qui retiennent les ficelles. Lorsque l'on tire trop fort sur la ficelle, ce sont les clous qui cèdent. La maladie de la sur-sollicitation chronique du surentraînement chez l'enfant est essentiellement liée au cartilage de croissance. On peut ainsi observer des fractures du cartilage ou des arrachements des points d'ancrage." D'où la nécessité de respecter la douleur, qui est le signal d'alarme le plus efficace.

Côté alimentation, il faut savoir que les besoins énergétiques de l'enfant – sportif ou non – sont voisins de ceux de l'adulte. À 5 ans, l'apport doit être en effet de 1 500 kilocalories par jour. À 9 ans, de 2 200, chez les garçons comme chez les filles. À l'adolescence, les premiers peuvent manger quotidiennement jusqu'à 3 000 kilocalories, les secondes 2 400. Et l'activité physique accroît les besoins, en fonction de son intensité. L'alimentation doit évidemment être équilibrée, à la fois au cours de chacun des trois repas – or 10 à 15 % des enfants et adolescents ne prennent pas de petit déjeuner – et sur l'ensemble de la journée. Mais, pendant la croissance, la

consommation de protéines doit être assez importante et, plus encore chez le jeune sportif car elles jouent un rôle essentiel dans la “construction et l'édification” de l'organisme. [...] Et, comme tout sportif, l'enfant doit boire suffisamment, notamment pendant et après l'effort. D'autant plus qu'il transpire souvent beaucoup. Or c'est rarement le cas. “La peur de ne pas savoir où uriner à un âge où l'on devient très pudique l'incite à limiter sa consommation d'eau”, regrettent de nombreux médecins du sport. [...] Pour conserver “une âme saine dans un corps sain”, les enfants pratiquant un sport doivent donc avoir une bonne discipline de vie et alimentaire.

Mais il ne faut pas oublier que cet édifice est fragile. Notamment au moment de l'adolescence, quand l'opposition devient la règle, les poussées de croissance perturbent le geste sportif – au risque de dégoûter les plus “accros” – et l'attrait pour le sexe opposé déplace les centres d'intérêt.

Anne Jeanblanc, *Le Point*, 17-06-03

Item 1 (1 point)

Lorsqu'ils s'entraînent trop, les enfants:

- A. risquent moins de se blesser que les adultes car ils sont plus souples.
- B. se blessent plus souvent mais moins gravement que les adultes.
- C. se font des blessures de nature différente de celles des adultes.

Item 2 (3 points)

Quels facteurs sont déterminants dans l'évaluation des besoins caloriques de l'enfant?

L'âge (1 point), le sexe (1 point), l'intensité de l'activité sportive (1 point).

Document 2: SURDOUÉS: TROP INTELLIGENTS POUR ÊTRE HEUREUX

Depuis qu'elle a créé l'Association française pour les enfants précoces (Afepe) en 1993, Sophie Côte a reçu toutes sortes de sollicitations des médias. Elle est d'ailleurs habituée à y répondre favorablement, afin de faire connaître au grand public la cause de ces petits qui sont “trop tout”. Trop éveillés, trop curieux, trop exigeants, trop angoissés, et parfois trop mauvais en classe, où ils peuvent s'ennuyer à périr sous le regard exaspéré des enseignants. [...]

Jean-Charles Terrassier est encore plus ancien dans le métier. Ce psychologue clinicien de Nice a, dès 1971, exploré le fonctionnement de ces têtes drôlement faites qui représentent tout de même plus de 2 % de la population. Lui qui s'est battu pour faire admettre l'existence des surdoués, pour encourager leur détection et éviter du même coup leur marginalisation a écrit des livres, multiplié les colloques, les interviews dans les journaux. Mais il n'en est pas revenu quand une grande chaîne de télévision publique a décidé de bâtir une émission autour d'un enfant qui serait testé “en direct” et dont les téléspectateurs connaîtraient le QI juste avant le générique de fin, au terme d'un suspense savamment orchestré...

Encore ignoré par l'institution scolaire, toujours nié par les tenants d'une psychanalyse pure et dure, l'enfant précoce, depuis quelque temps, est l'objet de

fantasmes collectifs de plus en plus tenaces. Dans une société où la performance est devenue une valeur en soi et l'enfant idéal un objet de désir narcissique, le mythe du petit génie envahit les têtes [...]

De très rares collèges publics et des établissements privés en nombre plus important ont ouvert des classes spécifiques pour les enfants précoces, qui ont la possibilité de faire deux années en une. Ainsi du collège privé Saint-Louis, au Mans, où les professeurs, recrutés sur la base du volontariat, reçoivent une formation spécifique. Cet établissement, qui poursuit l'expérience depuis huit ans sans faire de tapage, doit refuser la plupart des demandes. "Pour l'entrée en quatrième, nous avons 10 places et 200 demandes", explique le directeur, avant de préciser que la plupart de ces enfants sont, d'une manière ou d'une autre, en difficulté scolaire, et qu'il ne pratique donc aucune sorte d'élitisme qui améliorerait ses résultats.

Mais, à côté d'institutions comme celle-ci, intéressées par l'épanouissement des enfants et non par leurs performances, d'autres ont vu dans cette population très spéciale un vivier propre à améliorer leurs résultats au brevet et au bac, tout en se situant sur un créneau porteur. [...]

Dans l'imagerie populaire, l'enfant surdoué n'est pas à plaindre, c'est presque un nanti. Personne ne voit ses fragilités, et éventuellement ses souffrances. Résultat, plus d'un enfant sur trois n'a pas son baccalauréat, parce qu'il est rejeté du système scolaire avant le lycée [...] Faute de recherches, faute de solutions pédagogiques adaptées, les surdoués risquent de demeurer ces objets de fantasme qui inspirent même les sectes, puisque des mouvements comme les Enfants indigo, avec des arguments à dormir debout, parviennent à convaincre des parents désemparés que leurs enfants sont des envoyés d'une autre civilisation et qu'il convient de les élever "autrement", loin des psys. Entre propositions délirantes et indifférence des institutions, les familles ont bien du mérite à se frayer un chemin vers la simple reconnaissance d'une particularité qui fascine et indispose.

Sophie Coignard, *Le Point*, 17-03-05

Item 3 (1 point)

Vrai ou faux? Cochez la case correspondante:

Les professeurs qui travaillent dans ces classes spécifiques reçoivent un salaire plus élevé.

- A. Vrai.
- B. Faux.
- C. On ne sait pas.

Item 4 (2 points)

Que signifie Sophie Coignard lorsqu'elle écrit: "Dans l'imagerie populaire, l'enfant surdoué n'est pas à plaindre, c'est presque un nanti."?

Pour la plupart des gens (1 point), l'enfant surdoué a une situation enviable (1 point).

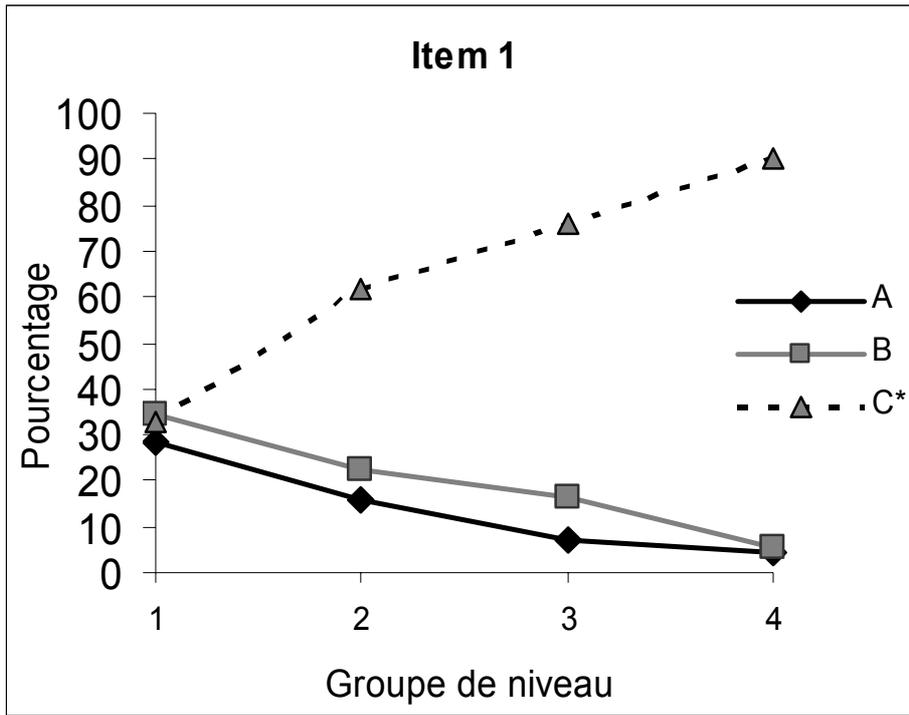
Figure 4 – Documents et items extraits d'une épreuve du DELF junior et scolaire niveau B2

3.1.1. Théorie classique des tests

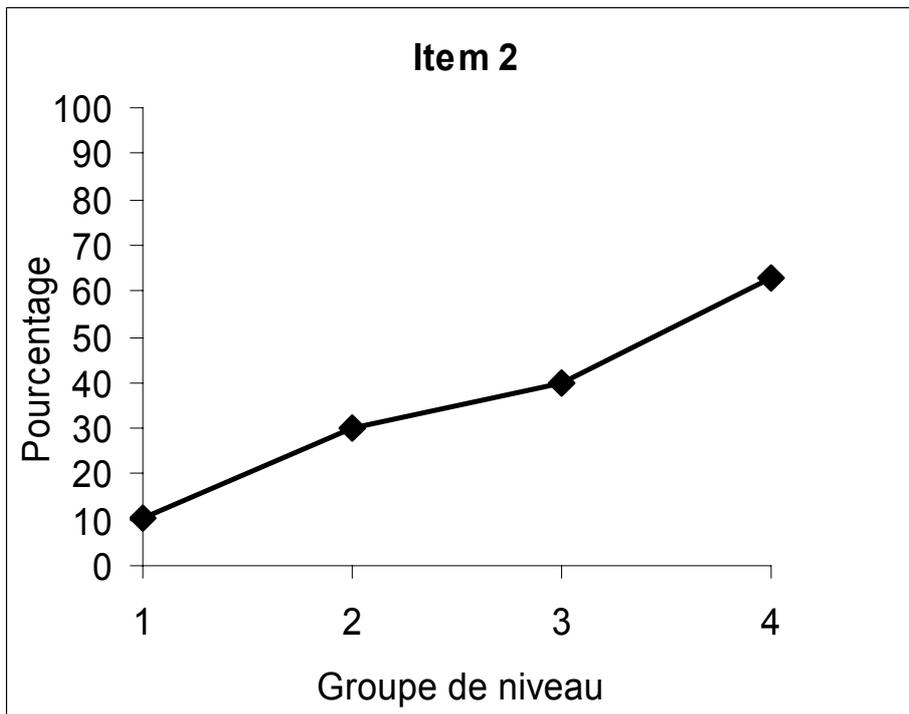
Les items 1 et 3 sont assez faciles, ils sont respectivement réussis par 66 % et 62 % des candidat/es (Tableau 1). Les deux autres items sont assez difficiles, le taux de réussite est de 37 % pour l’item 2 et 30 % pour l’item 4. Les deux questions ouvertes sont donc plus difficiles. Les indices de discrimination item-reste des items 1, 2 et 4 sont supérieurs à 0,30 et les “indices de discrimination” des distracteurs de l’item 1 sont négatifs (Tableau 1), ce qui suggère que ces items permettent de bien distinguer les candidat/es suivant leurs performances au test. Leur pouvoir discriminant est observable aussi graphiquement (Figure 6): le taux de réponses correctes augmente fortement en fonction du niveau des candidat/es. Par exemple pour l’item 1, on passe d’un taux de réussite de 35% pour le groupe 1 à un taux de 90 % pour le groupe 4 (Figure 6.a, courbe C) et les proportions de réponses incorrectes (courbes B et C) diminuent en fonction du niveau des candidat/es. En ce qui concerne l’item 3, il présente un certains nombre de dysfonctionnements. Son indice de discrimination item-reste est très proche de zéro, il est égal à 0,06 (Tableau 1), l’item ne permet donc pas de différencier les candidat/es selon leur compétence. Le graphique des proportions de réponses des candidat/es (Figure 5c) illustre bien ce problème: le taux de réponse correcte n’augmente pas continuellement du groupe 1 au groupe 4. Il passe de 60 % pour le groupe 1 à 50 % pour le groupe 2 alors que ce dernier devrait réussir cet item plus facilement que le groupe 1. Ensuite, le taux de réussite augmente jusqu’au groupe 4 et atteint les 71 %. En ce qui concerne les distracteurs, les proportions de réponses incorrectes, B et C, ne diminuent pas en fonction du niveau des candidat/es et le distracteur B est plus souvent choisi que le distracteur C (Tableau 1, figure 4): à la question “Vrai ou faux? Les professeurs qui travaillent dans ces classes spécifiques reçoivent un salaire plus élevé.”, 33 % des candidat/es répondent que c’est faux (distracteur B) et 5 % pense que c’est vrai (distracteur C). Les qualités fonctionnelles des items 1, 2 et 4 sont très bonnes, par contre l’item 3 ne fonctionne pas correctement, il ne sera donc pas intégré à la suite de l’analyse.

Item	Clé	Pourcentages			Score maximal	Score moyen	Difficulté	Indices de discrimination				
		A	B	C				Item-test	Item-reste	A	B	C
1	C	13	19	66	1	0,66	66%	0,43	0,37	-0,26	-0,18	0,37
2					3	1,11	37%	0,48	0,32			
3	C	5	33	62	1	0,62	62%	0,13	0,06	-0,10	0	0,06
4					2	0,60	30%	0,46	0,35			

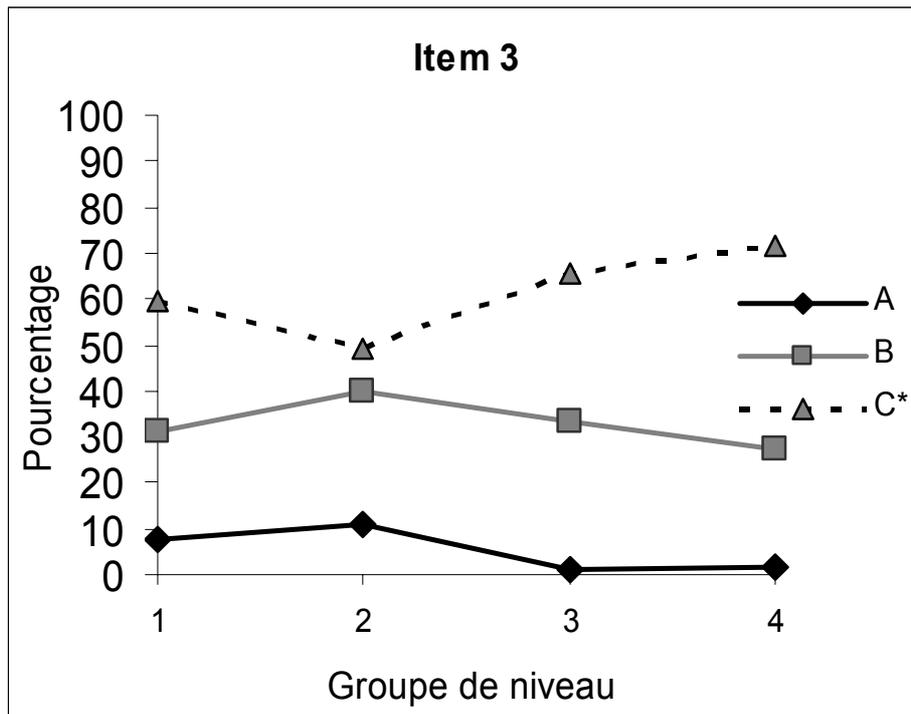
Tableau 2. Résultats statistiques concernant les items obtenus avec la TCT



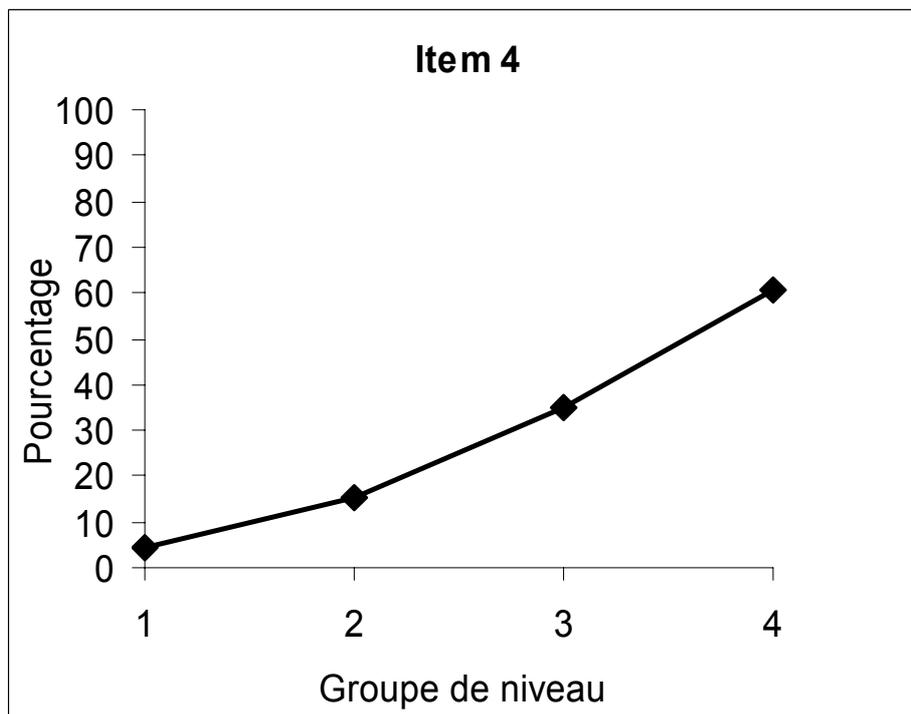
(a)



(b)



(c)



(d)

Figures 5 – Distribution des réponses des candidat/es à un item en fonction de leur niveau

3.1.2. Modèle à crédit partiel

Pour estimer le niveau de difficulté des items, nous utilisons le modèle à crédit partiel (PCM) présenté dans la section précédente (§2.3.2). Comme on l'a vu, les modèles de réponses à l'item ont l'avantage, contrairement à la TCT, de fournir des estimations indépendantes de l'échantillon sur lequel on travaille. La difficulté des

items est estimée en logit (unité de mesure), et varie généralement entre -3 et +3 (plus la valeur de la difficulté est élevée plus l'item est jugé difficile). La compétence des candidat/es est aussi estimée en logit sur la même échelle que les items.

On constate que l'ordre de difficulté des items est identique à celui obtenu précédemment avec la TCT: l'item 1 est nettement plus facile que les items 2 et 4, l'item 4 est un peu plus difficile que l'item 2 (Tableau 2). L'erreur de mesure de la difficulté est très faible pour les items 2 et 4, un peu plus élevée pour l'item 1: l'estimation de la difficulté est donc plus précise pour les items 2 et 4. L'infit et l'outfit des trois items sont très proches de 1, ce qui indique que le modèle permet de prédire correctement les réponses des candidat/es pour ces items.

Les items 2 et 4 étant polytomiques, nous analysons chacune de leurs modalités. Les modalités d'un item sont censées être ordonnées, elles indiquent le degré de réussite à l'item (plus leur valeur est élevée, plus la réponse à l'item est correcte). Si elles ne sont pas ordonnées, l'item ne peut pas correctement évaluer le trait latent, ici la compétence langagière. Pour l'item 2, la compétence moyenne des candidat/es (exprimée en logit) augmente en fonction de la valeur de la modalité (Tableau 3), ce n'est pas le cas pour l'item 4, elle est de 0.78 logit pour la modalité 1 et de 0.71 pour la modalité 2. Les candidat/es qui ont eu 1 point à cet item ont en moyenne un niveau légèrement supérieur à celles/ceux qui ont eu 2 points. Cette anomalie peut indiquer que les modalités de l'item 4 ne sont pas ordonnées, et que cet item ne permet pas d'évaluer correctement la compétence langagière des candidat/es. Notons que leur compétence moyenne pour la modalité 1 est calculée sur un petit nombre de candidat/es, 15 sur 257. Pour l'item 2, l'infit et l'outfit associés à la modalité 1 sont élevés, 1,68 et 1,88, ce qui indique que l'on observe beaucoup de réponses improbables pour cette modalité. Par exemple, un individu qui aurait dû avoir 2 points à cet item n'en n'a eu qu'un. Pour finir l'analyse des modalités des items, nous étudions les seuils de transition¹¹ entre les modalités voisines, c'est-à-dire entre les modalités 0 et 1 (seuil 1), entre les modalités 1 et 2 (seuil 2), et ainsi de suite. Le seuil de transition entre les modalités 0 et 1 d'un item correspond au niveau de compétence d'un/e candidat/e qui aurait autant de chance d'obtenir l'une ou l'autre de ces modalités. Il en va de même pour les autres seuils. Les seuils de transition des items 2 et 4 ne sont pas ordonnés par ordre croissant en fonction des valeurs des modalités (Tableau 3). Cela n'implique pas que leurs modalités soient désordonnées, mais qu'une ou plusieurs modalités ont une probabilité d'être observées toujours inférieure aux autres modalités. Les courbes caractéristiques, CCI (Figure 6), de ces items permettent de voir rapidement de quelles modalités il s'agit. Les CCI illustrent le lien qui existe entre la probabilité d'observer une modalité et le trait latent mesuré (i.e. la compétence langagière des candidat/es). Les courbes des modalités 1 et 2 de l'item 2 se situent toujours en dessous des courbes des modalités 0 et 3 : leurs probabilités d'être observées est donc toujours inférieure à celle des modalités 0 et 3. Ainsi la probabilité qu'un candidat obtienne 0 ou 3 points à l'item 2 est plus forte que la probabilité qu'il obtienne 1 ou 2 points. La courbe de la modalité 1 de l'item 4 est elle aussi située en dessous des courbes des modalités 0 et 2. Étant donné que les

¹¹ Les seuils de transition sont des paramètres du modèle à crédit partiel.

items 2 et 4 sont assez difficiles, la modalité 0 est en effet très souvent observée : 49% des candidats ont eu 0 points à l’item 2 et 67% ont eu 0 points à l’item 4 (Tableau 3).

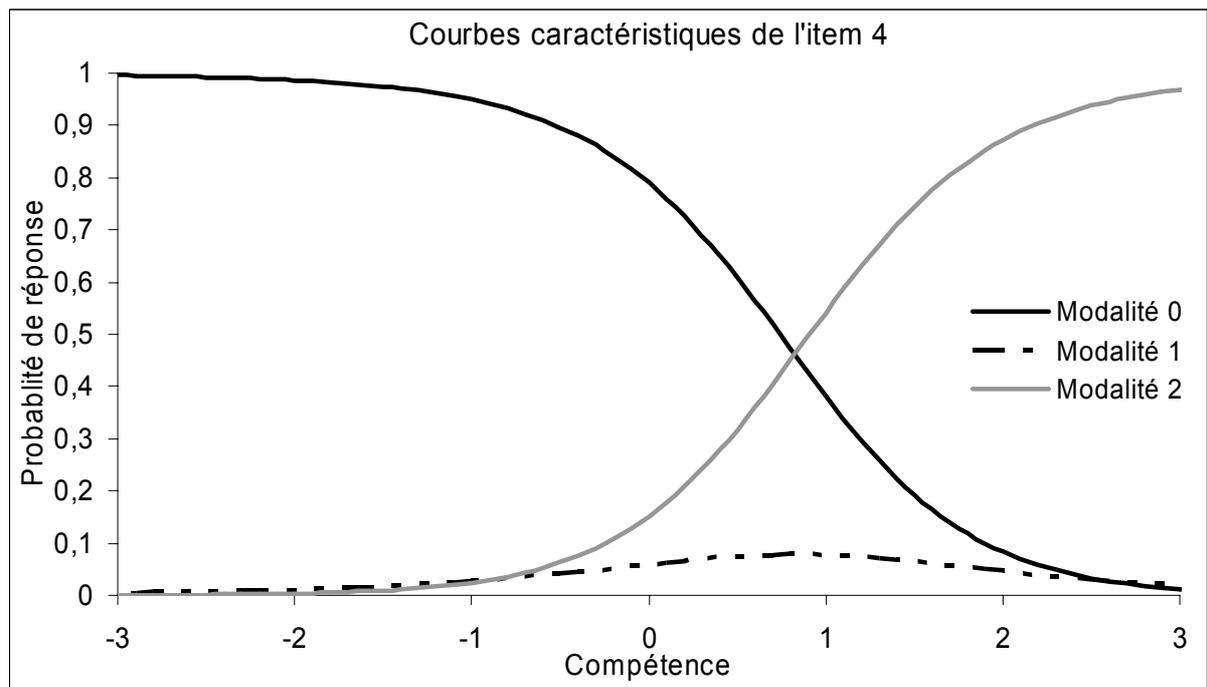
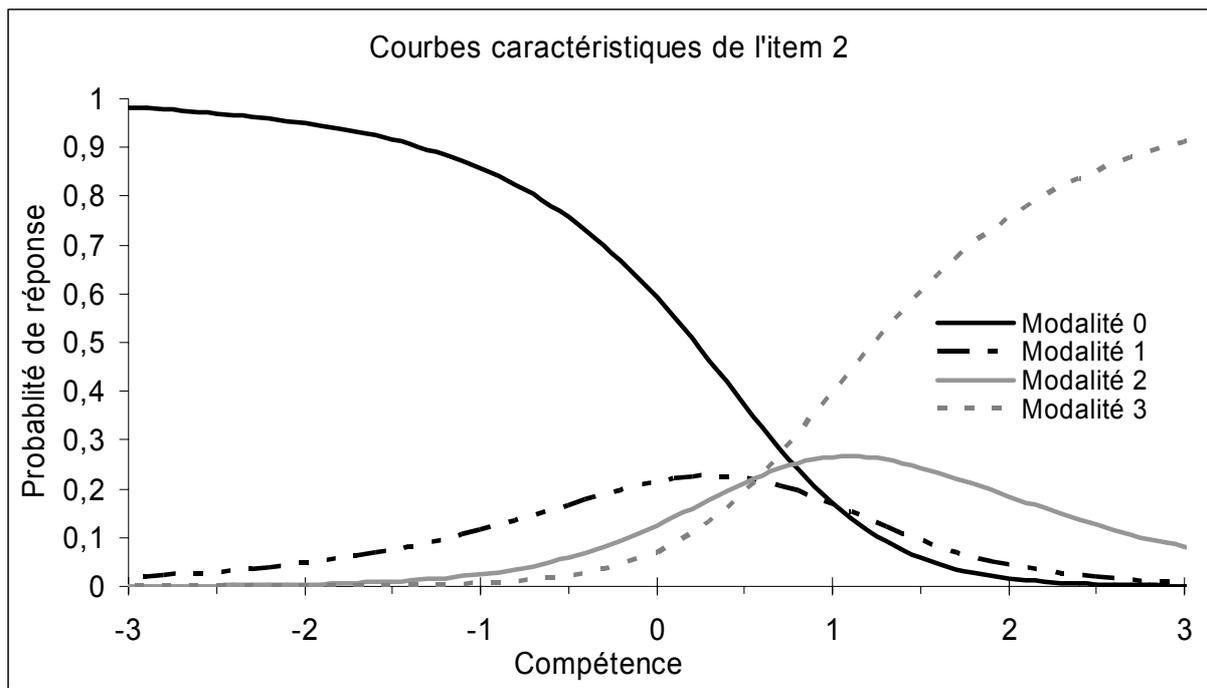
En conclusion, l’analyse psychométrique des items nous a permis d’étudier leur fonctionnement. L’item 3 ne permet pas de distinguer les candidat/es en fonction de leur compétence langagière, c’est pourquoi nous n’avons pas estimé sa difficulté avec le modèle de réponse à l’item PCM. L’item 4 pourrait ne pas évaluer correctement le niveau de compétence des candidat/es car celles/ceux qui réussissent partiellement l’item ont en moyenne un niveau légèrement supérieur à ceux/celles qui le réussissent complètement. L’item 2 fonctionne à peu près correctement, mais nous observons, parmi les candidat/es qui ont eu 1 point, beaucoup d’individus pour lesquels le modèle n’avait pas prévu qu’ils obtiennent ce score. Quant à l’item 1, il ne montre aucun dysfonctionnement.

Item	Mesure de la difficulté	Erreur de mesure de la difficulté	Infit	Outfit
1	-0,58	0,14	0,94	0,92
2	0,72	0,06	1,05	1
4	0,82	0,08	1,04	0,91

Tableau 2. Résultats statistiques des items obtenus avec le modèle PCM

Item	Modalité	Compétence moyenne	Infit	Outfit	Mesure du seuil de transition	Effectif	Pourcentage
2	0	-0.16	1.02	0.97	Aucun	125	49 %
	1	0.11	1.68	1.88	1.02	46	18 %
	2	0.63	0.75	0.48	0.54	40	16 %
	3	0.86	1.07	1.00	0.59	46	18 %
4	0	-0.07	1.01	0.99	Aucun	172	67 %
	1	0.78	0.91	0.47	2.61	15	6 %
	2	0.71	1.06	1.06	-0.96	70	27 %

Tableau 3. Résultats statistiques sur les modalités des items obtenus avec le modèle PCM



Figures 6 – Courbes caractéristiques pour les items polytomiques du DELF

Conclusion

La psychométrie fournit un cadre méthodologique solide et apporte un éclairage particulier dans le domaine des sciences de l'éducation à l'aide de tests standardisés (pour une revue, Laveault et Grégoire, 2002). L'identification des dysfonctionnements que peuvent présenter les éléments constitutifs (les items) d'un instrument de mesure et leur remédiation contribuent grandement à améliorer la qualité des tests proposés, que ce soit dans le domaine de l'éducation ou plus généralement de la psychologie

appliquée. Notons toutefois que le résultat à un test ne donne pas une mesure absolue d'une capacité ou d'un trait de personnalité, comme une balance attribue à chacun un poids, mais que celui-ci permet une comparaison de la performance d'un individu avec les performances des individus auxquels il est légitimement comparable. Ainsi, puisque par définition la dimension latente (la compétence en langues, par exemple) n'est pas directement observable, il faut se donner les moyens, d'une part, d'assurer une mesure aussi précise que possible de cette dimension à l'aide d'une ou de plusieurs variables manifestes, et, d'autre part, de pouvoir évaluer les candidat/es, non pas nécessairement de manière absolue mais en tous les cas de manière équitable.

Les différentes techniques exposées dans cet article permettent d'assurer la fiabilité et la validité d'un test, au travers de sa standardisation et de son étalonnage, dans un souci constant d'assurer la qualité des items qui le composent. Le travail fourni par les experts du CIEP participe de cette volonté de délivrer des examens et des certifications correspondant aux standards d'élaboration d'un outil d'évaluation. Les processus de conception et de gestion de ces outils d'évaluation obéissent tous aux standards minimums établis par les membres de ALTE (Association of Language Testers in Europe) et sont évalués périodiquement par des experts qualitatifs dans le cadre de la certification ISO 9001: 2000 délivrée au Test de connaissance du français.

Bibliographie

BERNIER JEAN-JACQUES & BOGDAN PIETRULEWICZ. 1997. *La psychométrie. Traité de mesure appliquée*. Gaëtan Morin.

PAUL DICKES, JOCELYNE TOURNOIS, ANDRÉ FLIELLER & JEAN-LUC KOP. 1994. *La psychométrie. Théories et méthodes de la mesure en psychologie*. PUF, Le psychologue.

JOHN M. LINACRE. 2004. Rasch Model Estimation: Further Topics. EVERETT V. SMITH JR. & RICHARD M. SMITH (dir.). *Introduction to Rasch Measurement*. Jam Press. 48-72.

RITA K. BODE. 2004. Partial Credit Model and Pivot Anchoring. EVERETT V. SMITH JR. & RICHARD M. SMITH (dir.). *Introduction to Rasch Measurement*. Jam Press. 279-295.

BIRNBAUM, A. 1968. Some latent trait models and their use in inferring an examinee's ability. LORD, F.M. & M.R.NOVICK. *Statistical theories of mental tests scores*. Addison-Wesley.

EMBRETSON, S.E. & S.P REISE. 2000. *Item response theory for psychologists*. Laurence Erlbaum Associates.

LAVEAULT D. & J.GRÉGOIRE. 2002. *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. De Boeck Université.

MASTERS, G.N. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47. 149-174.

RASCH, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press.

SAMEJIMA. 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 17, Monograph Supplement.

Remerciements

Les auteur/es souhaitent remercier Patrick Riba, responsable du bureau DELF-DALF, Ingrid Jouette et Auréliane Baptiste, chargées de programme, pour leur aimable autorisation à diffuser et exploiter certains items composant les épreuves du DELF (avant leur révision finale).

Médiation, suivi et tutorat en ligne: approche théorique, perspectives de recherche

Jean-Claude Bertin, Professeur à l'université du Havre, UMR IDEES 6228 – CIRTAI

Jean-Paul Narcy-Combes, Professeur à l'université de Paris 3, DILTEC

Patrick Gravé, Maître de Conférences à l'université du Havre, UMR IDEES 6228 – CIRTAI

Le suivi des apprenants dans le cadre de formations ouvertes et à distance représente un domaine de recherche encore peu abordé dans la littérature du fait de la courte histoire qui caractérise ces dispositifs de formation. Si le concept de tutorat a fait l'objet d'études déjà poussées dans des contextes aussi divers que l'enseignement/apprentissage en présentiel dans des contextes généraux, universitaires ou de formation professionnelle et continue (Prévost 1994, Barnier 2001, Bezsonoff 2000, Blanc 2000), et en ce qui concerne l'apport des nouvelles technologies (Annoot 1996, 1998, D'Halluin 2001), la notion de suivi pédagogique dans le cadre de la FOAD pose encore nombre de questions, dont la moindre n'est pas celle de sa définition. Les publications spécialisées, pour l'essentiel, ne remontent pas au-delà des années 2000 (Hubbard 1999, 2000, Boullier 2000, Glickman 2002, Collectif du Moulin 2002, pour n'en citer que quelques-unes). Elles abordent le problème du tutorat à distance principalement sous l'angle du transfert des connaissances acquises dans un contexte en présentiel vers celui de la distance, ou sous celui de la capacité de la technologie à mettre en œuvre des dispositifs de tutorat dits "intelligents" (colloque EIFAD 2004).

La richesse et la variété des expérimentations, liées au rythme rapide de l'innovation technologique, laisse cependant peu de place au recul théorique alors même que certains chercheurs (De Lièvre *et al* 2004) mettent en exergue le manque criant de modèles sur lesquels appuyer la recherche.

C'est pour tenter d'apporter des éléments de réponse à ces questions que les auteurs de cet article ont commencé, à partir de 2005 (Bertin et Narcy-Combes 2005) à envisager la question du suivi de l'apprenant sous l'angle de la théorie.

Le présent article trouve son origine dans cette conférence commune, complétée par plusieurs communications ultérieures, ainsi que dans les travaux de l'équipe DILTEC (Paris 3) et de l'axe "TIC" de l'UMR IDEES 6228 – CIRTAI (Le Havre)¹. Il reprend en particulier les grandes lignes de la communication présentée conjointement au Congrès de la SAES 2007 à Avignon. Après un rappel du modèle théorique de référence, on présentera successivement le cadre pédagogique en place à Paris 3 (Master 2 recherche et en doctorat en sciences du langage et didactique des langues) puis les grandes lignes d'un programme de recherche intégrant une dimension technologique

¹ En particulier la collaboration avec Emmanuelle Annoot et Patrick Gravé (cf. bibliographie).

(collaboration avec le laboratoire LITIS, Le Havre-Rouen) et visant à tester les conditions d'un tutorat à distance s'appuyant sur une analyse dynamique du suivi des apprenants.

I. Médiation, tutorat, suivi et distance

La question essentielle qui sous-tend la recherche entreprise pourrait se résumer de la manière suivante: comment est-il envisageable d'aider l'apprenant dans un dispositif d'apprentissage à distance afin de répondre, au moins partiellement, aux objectifs suivants?

- *Réduire le nombre d'abandons*: les formations à distance souffrent d'un taux d'abandon élevé pour lesquelles les explications avancées varient, mais au nombre desquelles figure de manière significative l'isolement (physique, psychologique, pédagogique, etc.) dans lequel se trouve plongé l'apprenant.
- *Apporter des réponses rapides* aux questions de l'apprenant, réduire le délai entre l'activité et le *feedback*, améliorer les conditions de la rétroaction didactique.
- *Faciliter l'apprentissage*, en attirant l'attention de l'apprenant sur les processus cognitifs mis en œuvre dans les activités, ses stratégies d'apprentissage (approche métacognitive).
- *Offrir un suivi et un étayage personnalisés*, afin de répondre aux besoins de tutorat de chaque apprenant.
- *Développer l'autonomie* de l'apprenant ("apprendre à apprendre").

Ces objectifs rejoignent ceux cités de manière générale pour l'activité de tutorat en mode présentiel. Une réponse fréquente consiste à s'appuyer sur les apports du socioconstructivisme de Vigotsky et Bruner, en particulier sur les concepts de Zone Proximale de Développement et d'étayage pédagogique, considéré comme l'une des formes (évolutive et transitoire) de la médiation. La distinction entre "enseignant" et "tuteur" apparaît ainsi en attribuant à l'un ou l'autre des fonctions particulières, sans pour autant préciser si cette différence implique ou non des acteurs différents: selon le contexte, l'enseignant peut prendre en charge la fonction tutorale en partie ou en totalité, ou cette dernière peut être partagée par des acteurs distincts (humains, ou impliquant une médiation technologique)².

"Enseignant" et "tuteur" participent tous deux de ce que nous appellerons la médiation pédagogique, c'est-à-dire le processus d'interaction cognitive entre l'enseignant/tuteur et l'apprenant, qui influe sur la nature des interactions entre l'apprenant et la langue (fig. 1).

Ce premier niveau de médiation peut se décliner selon plusieurs modalités:

² Nous n'entrons pas, dans le cadre de cet article, dans le détail des rôles sous-tendus par la notion de médiation, de suivi et de tutorat, et renvoyons le lecteur aux définitions que nous en avons proposé ailleurs (en particulier Nancy-Combes et Bertin 2007 a et b, 2008).

- la médiation langagière est relative au choix de l'input: elle correspond au degré d'adéquation entre ce que Daniel Coste (1981) appelle la "langue usitée" et la "langue enseignée"³;
- la médiation pédagogique intervient également sur le plan du choix des activités proposées / conseillées / imposées (selon le cas) à l'apprenant;
- Le suivi pédagogique, enfin, permet à l'enseignant d'intervenir sur les choix stratégiques de l'apprenant par le biais du processus d'information / rétroaction sur l'activité d'apprentissage.

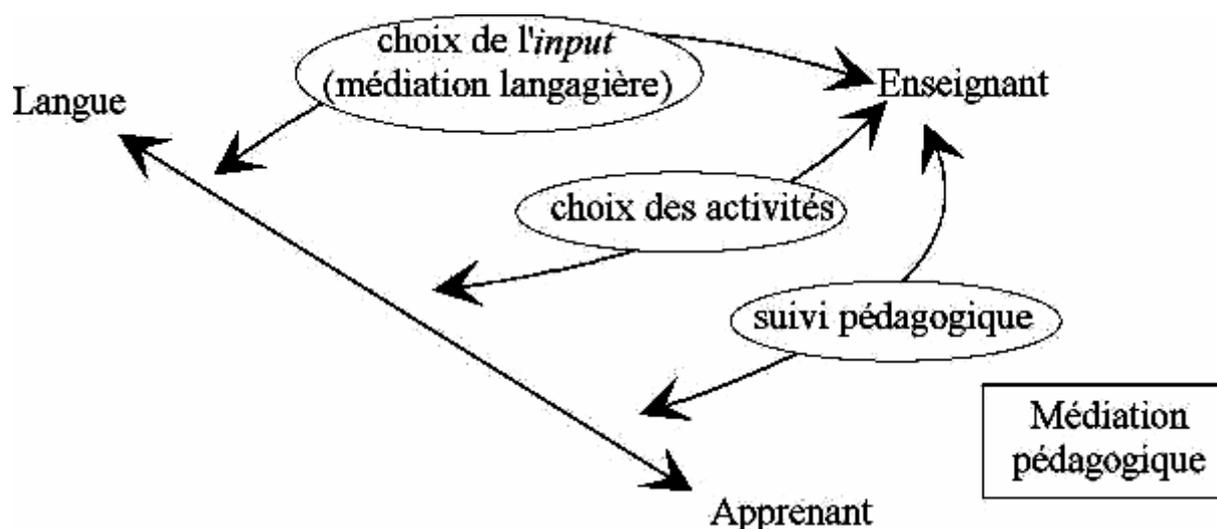


Figure 1. Premier niveau de médiation: la médiation pédagogique

Cette représentation simplifiée vise principalement à mettre en évidence la tri-polarisation (langue – apprenant – enseignant) qui caractérise la plupart des tentatives de modélisation de la situation pédagogique (Houssaye 1988, Legendre 1988, Carré et al 1997) et le système d'interactions qui relie les trois sommets de ce triangle.

³ Rappelons que Daniel Coste effectue la distinction suivante entre quatre niveaux de langue: la *langue usitée* - LU: "tout ce qui relève de l'usage effectif d'une langue naturelle par des locuteurs qui la possèdent"; la *langue décrite* - LD: "ce dont rendent compte les analyses et études, théories et modèles des linguistes et autres observateurs de l'usage des langues"; la *langue enseignée* - LE: "ce qui, dans une perspective didactique, est sélectionné et présenté par un auteur de cours et/ou un enseignant à des élèves"; la *langue apprise* - LA: "ce qui est effectivement maîtrisé par les apprenants au cours et à l'issue de l'apprentissage". Selon la perspective didactique, la relation entre ces quatre types de langue peut s'inverser:

- LU > LE > LD > LA, dans une perspective d'enseignement fondée sur un processus de réduction: "on ne décrit jamais qu'une partie de ce qui est usité; on n'enseigne jamais qu'une partie de ce qui est décrit; on n'apprend jamais qu'une partie de ce qui est enseigné."
- LA > LE > LD > LU, dans une perspective d'apprentissage: "on apprend plus que ce qui est enseigné; on enseigne plus que ce qui est décrit; on décrit plus que ce qui est usité." (Coste 1981: 33-37).

Le recours à la technologie introduit un second niveau de médiation dans la mesure où elle s'insère au sein du réseau d'interactions généré par le processus de médiation pédagogique, allant parfois jusqu'à les masquer (fig. 2). La nature de ce second niveau de médiation est une fonction directe de la volonté didactique qui anime l'enseignant à l'origine des choix technologiques: elle résulte de la manière dont la technologie est systématiquement ou non questionnée par le concepteur du dispositif et/ou des matériaux pédagogiques et peut se ranger le long d'un continuum allant du simple gadget répondant à un phénomène de mode au dispositif de formation élaboré prenant en compte l'ensemble des composantes de ce dispositif (humaines, technologiques, institutionnelles, etc.) pour repenser tout à la fois les modalités de l'interaction avec l'input et la nature des activités didactiques⁴.

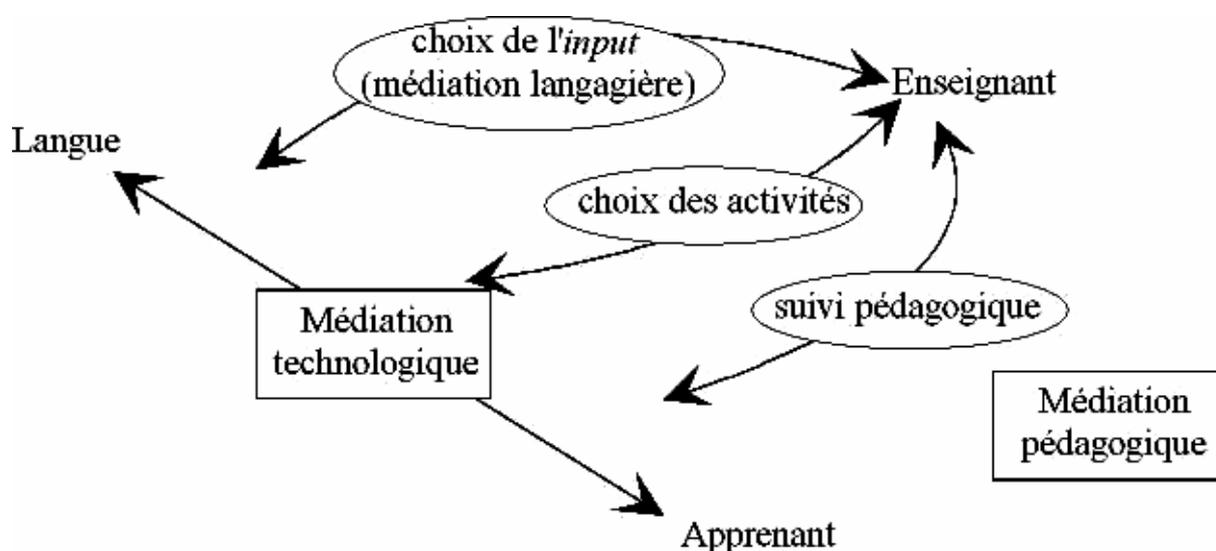


Figure 2. Second niveau de médiation: la médiation technologique

La médiation technologique intervient tout à la fois dans des situations d'apprentissage en présentiel, avec laboratoire multimédia, et dans le cadre des formations ouvertes et à distance utilisant les réseaux informatiques. Dans ce dernier cas, toutefois, l'intervention de la distance modifie sensiblement la nature des médiations pédagogique et technologique que nous venons d'esquisser, et constitue un troisième niveau de médiation (fig. 3).

⁴ Cf., en particulier, Bertin, Annoot, Gravé 2005.

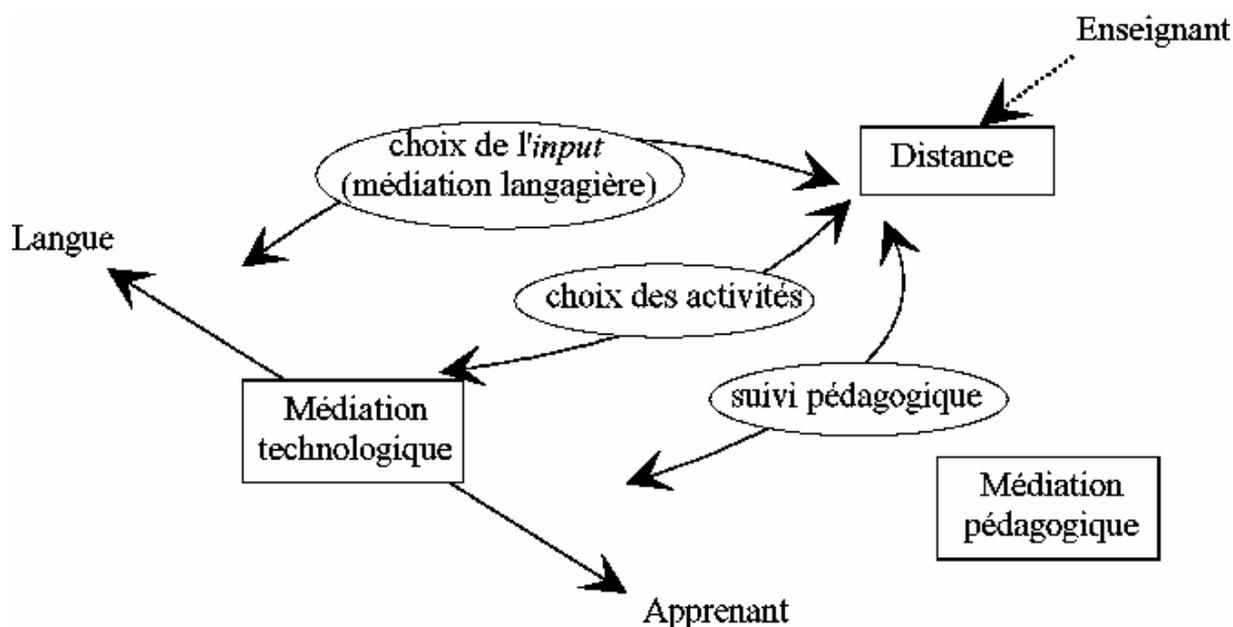


Figure 3. Troisième niveau de médiation: l'impact de la distance

La virtualité introduite par l'irruption conjointe de la technologie et de la distance repose différemment (radicalement?) la question de la nature du suivi, de ses objectifs et de ses modalités: puisque l'humain ne saurait faire face à la diversité des situations générées par la FOAD (distance physique mais également temporelle – (a)synchronisme des interactions), l'hypothèse d'une médiation technologique plus large, intégrant les notions de suivi et de tutorat, se trouve relancée. Cette situation inédite⁵ semble également pouvoir mettre en question certains positionnements épistémologiques qui fondent la vision de l'apprentissage des langues en présentiel. Il convient donc, pour le chercheur, de tenter à la fois de réduire ces incertitudes théoriques et de repenser les rôles respectifs des acteurs humains et de la technologie.

C'est le sens de la présente réflexion, qui se positionne dans une perspective systémique relevant de la pensée complexe d'Edgar Morin pour lequel un système se définit comme une unité globale organisée d'interrelations entre éléments, actions et individus.

La notion de suivi peut, en ce sens, s'inscrire comme l'une des composantes d'un dispositif de formation, conçu comme un système dans lequel chaque élément interagit et modifie les autres. Rappelons qu'un dispositif "se caractérise toujours par un ensemble: ensemble d'éléments [...] mis en œuvre pour une intervention précise" (Montandon 2002: 17). Un dispositif d'enseignement/apprentissage des langues peut donc se concevoir comme "un ensemble de procédures diverses d'enseignement/apprentissage, incluant moyens et supports, construit en fonction d'un public (et éventuellement d'une

⁵ Nous nous référons ici en particulier à l'intégration des réseaux et de la communication médiatisée par ordinateur, qui refonde totalement le concept de formation à distance "papier/courrier".

institution), d'objectifs et de conditions de travail particuliers. Les buts d'un dispositif sont de répondre au mieux à des demandes variées et d'individualiser ce travail grâce à la flexibilité du système ainsi qu'à la variété des supports (utilisation des TICE) et des modalités (travail personnel, tutorat, séances en groupe)" (Pothier 2003: 81).

Cette perspective, qui conjugue le modèle d'ergonomie didactique proposé par J.C. Bertin (2001) et le modèle des macro/micro tâches de J.P. Narcy-Combes (2005), nous permet de figurer un tel dispositif sous la forme ci-dessous (fig.4).

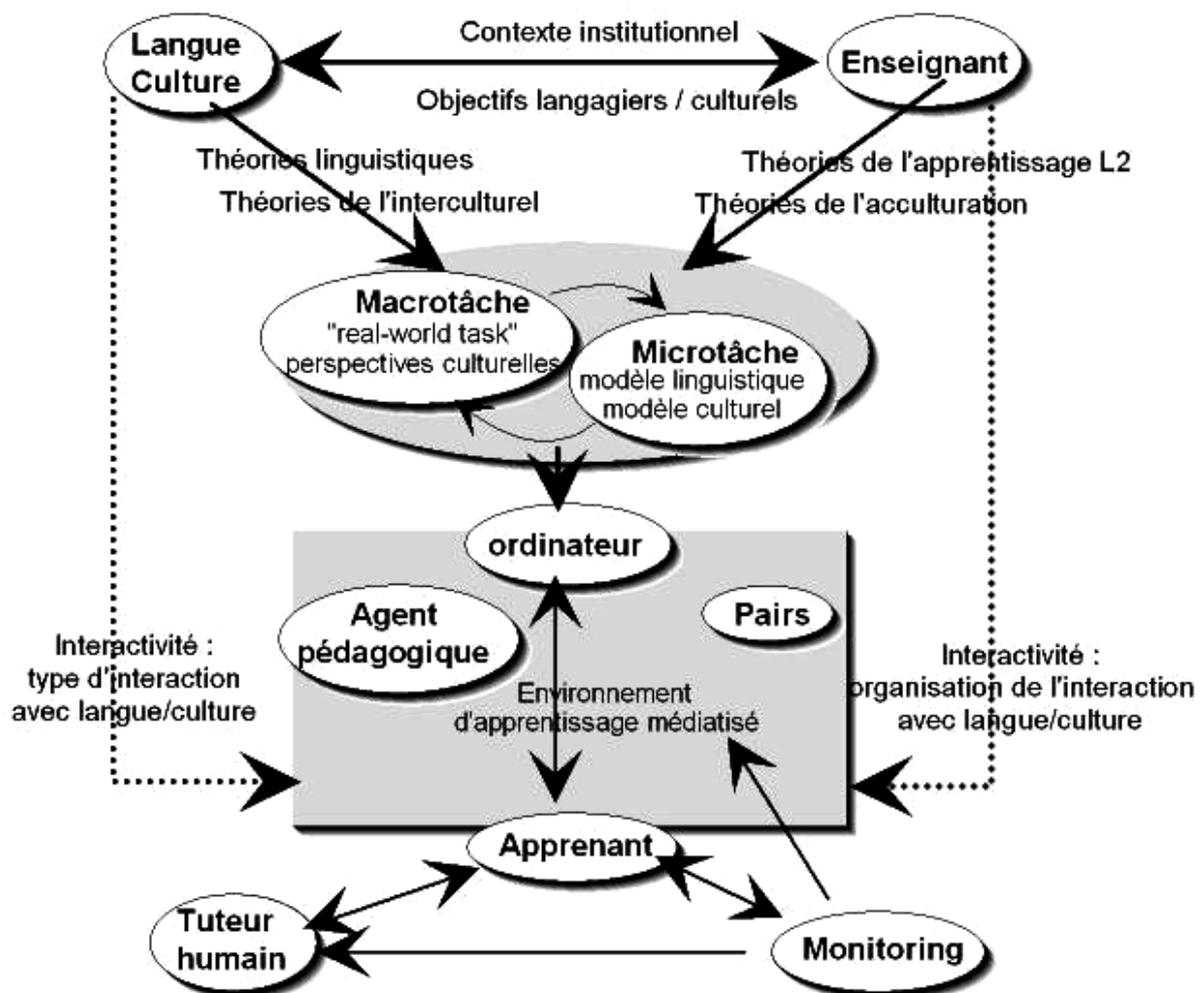


Figure. 4. Modèle systémique de dispositif de formation intégrant technologie et distance

L'intérêt d'un tel modèle est tout à la fois de repérer les composantes du dispositif et de montrer comment chacune s'inscrit dans un système complexe d'interactions intégrant les composantes institutionnelles et pédagogiques. Pour ce qui nous concerne ici, la notion de suivi s'inscrit dans un double système de relations interpersonnelles (régulation de la médiation pédagogique) et

institutionnelles (formation d'objectifs, explicitation des besoins,⁶ etc.). Le suivi apparaît bien comme l'un des éléments clés de la régulation du dispositif, orienté à la fois vers l'apprenant, l'enseignant et l'institution.

Ce modèle ne figure pas l'ensemble des flèches annonçant des interactions: outre qu'une volonté d'exhaustivité en rendrait la lecture plus ardue, il convient de souligner la difficulté du chercheur à préciser la nature exacte de ces interactions. Ainsi, il est possible sans grand risque, au regard des observations déjà menées, d'affirmer l'existence d'interactions entre les différentes composantes de l'environnement virtuel d'apprentissage, l'absence de données fiables et de recul, d'une part, la grande diversité des dispositifs existants, d'autre part, rendent encore impossible d'en définir les modalités en détails.

La difficulté pour le chercheur, dès lors, réside dans l'obligation dans laquelle il se trouve de combiner une approche théorique non encore totalement validée et les nécessités du terrain. C'est la démarche que nous nous attacherons maintenant à décrire.

II. Démarche d'expérimentation

Le travail de recherche entrepris conjointement entre l'équipe du DILTEC (Paris 3) et l'UMR IDEES – CIRTAI (Le Havre) consiste à combiner le système de tâches proposées dans le cadre des enseignements de Master 2 recherche et de doctorat en sciences du langage et didactique des langues (Paris 3) et le projet visant à transférer vers le domaine de la FOAD un système expérimental de détection des risques mis au point par les chercheurs du laboratoire LITIS (Universités Le Havre et Rouen). Un exemple de pratique éclairera nos propos et mettra en valeur les enjeux de notre réflexion.

Une UE de communication scientifique orale en anglais a été mise en place à Paris 3, à destination des étudiants de Master 2 recherche et de doctorat en sciences du langage et didactique des langues (FLE largement). L'objectif en est d'amener les étudiants à se sentir à l'aise dans la communication scientifique orale en anglais. L'approche retenue consiste à laisser l'apprenant choisir et organiser son travail (individuellement ou en binôme, en lien avec le tuteur, en fonction de ses besoins) pour produire un exposé complet. Les étapes proposées sont le résultat de recherches préalables sur les techniques de travail qui permettent de déclencher les processus d'apprentissage de l'anglais tout en accomplissant une macro-tâche (cf. Narcy-Combes 2005 et 2007).

Après une séance de présentation, la première étape consiste à demander aux étudiants de lire un article puis de le transformer en série de titres, sous-titres, en-têtes de paragraphes. Le plan est alors envoyé à l'enseignant qui relève les problèmes à résoudre et organise une première rencontre. La seconde étape

⁶ La récente thèse d'Anne-France Beaufile (2007) s'attache à montrer comment l'explicitation des besoins des apprenants résulte à la fois d'une analyse a priori et de la prise en compte du suivi des apprenants au cours du processus de formation.

consiste à effectuer une communication d'entraînement à partir de ce plan et à l'envoyer à l'enseignant, qui relève les problèmes à résoudre et peut proposer des micro-tâches (spécificité de la langue du domaine, l'expression écrite, etc.), en ligne dans notre cas du fait de l'absence de centre de ressources. Au cours de la troisième étape, les étudiants doivent préparer un authentique exposé en procédant de la même manière, à distance. Selon leurs besoins, ils envoient au tuteur le plan, puis un enregistrement d'une répétition ou de leur présentation *Powerpoint*. L'enseignant relève les problèmes à résoudre et propose une nouvelle série de micro-tâches en ligne, touchant à l'expression orale, mais également aux techniques de présentation et d'argumentation en communication scientifique orale. Des séances d'exposés en groupes n'excédant pas 12 étudiants sont alors organisées. Lors de la cinquième étape, les étudiants font l'exposé devant le groupe. L'enseignant cherche à évaluer si les problèmes ont été résolus. Les étudiants doivent tenir compte de leurs points faibles pour se préparer à un second exposé.

Seules 4 séances ont donc lieu en commun et, si l'enseignant n'en anime que 8 au maximum, il doit encore assurer le suivi des tâches qu'il reçoit en étant sélectif. Son travail consiste essentiellement à organiser, suivre et donner du *feedback*. L'ensemble du travail peut se synthétiser comme le présente le tableau ci-dessous (fig. 5).

Synthèse du module
<p>CSOA: UE de communication scientifique orale en anglais (Master 2 recherche et doctorat en sciences du langage et didactique des langues).</p> <ul style="list-style-type: none"> - séance de présentation puis 5 étapes (deux fois: deux exposés/semestre). - 1. Lire un article. - 2. Le transformer en série de titres, sous-titres, en-têtes de paragraphes, et l'envoyer à l'enseignant (relève les problèmes à résoudre). - Première rencontre de bilan (suggestion éventuelle de micro-tâches, décomposition du travail). - 3. Diaporama envoyé à l'enseignant, - 4. Communication d'entraînement à partir du diaporama envoyée à l'enseignant (conseille et suggère micro tâches, en ligne dans notre cas, concernant la spécificité de la langue du domaine, l'expression écrite, etc.). - 5. Authentique exposé en groupes n'excédant pas 12 étudiants. - Hypothèse: chacune des étapes conduit à l'acquisition et le fait d'agir (<i>Learning by doing</i>) facilite des procédures semi-implicites. Les étudiants s'investissent si les attentes sont réalistes et l'environnement

Figure 5. Synthèse de l'organisation du module CSOA

Comment, dans ce contexte, concevoir un dispositif prenant en compte à la fois la distance et les besoins des apprenants?

Nous distinguons trois manières de décliner cette aide:

- une aide opérationnelle, relative à la manière de se situer dans le dispositif, d'utiliser l'environnement à distance, d'effectuer les micro-tâches, etc.;
- une aide disciplinaire, offrant une expertise dans le domaine d'apprentissage (la langue);
- une aide cognitive / métacognitive, visant à favoriser les processus d'apprentissage et à développer/améliorer les stratégies des apprenants.

Deux grandes questions de fond se posent, qui conditionneront l'expérimentation visant à combiner l'humain et la technologie au sein d'un dispositif intégrant un suivi à distance.

- Est-il possible de concevoir un environnement innovant combinant homme et machine?
- Si oui (ce qui constitue l'hypothèse du programme de recherche que nous présentons), quels rôles définir pour la fonction tutorale et comment imaginer répartir ces rôles entre l'homme et la machine?

Une autre manière de poser ces questions consiste à savoir si une médiatisation du tutorat est faisable, d'une part, souhaitable, de l'autre, la seconde question dépendant largement de la réponse à la première.

Deux étapes de réflexion se dégagent alors, qui conditionneront l'approche technologique qui suivra:

- clarifier les fonctions tutorales en partant à la fois du souhaitable (référence aux théories) et des usages (comment les acteurs se sont-ils appropriés les dispositifs existants?);
- en fonction des horizons ouverts par la recherche, imaginer la manière de répartir les rôles entre l'homme et la machine de manière à les rendre cohérents au sein d'un système hybride.

Un début de réflexion collective, qui sera complétée par un recueil de données à l'automne 2007, nous permet de concevoir comment scénariser le travail et répartir le suivi entre un agent informatique (souvent appelé « tuteur machine »), le tuteur humain et les pairs.

Étape 1: Lecture

Travail du tuteur machine: suivre les techniques de lecture pour signaler tout comportement de lecture inapproprié et donner les conseils adaptés.

Étape 2: Plan

Travail du tuteur machine: mesurer l'efficacité des techniques de travail, vérifier si les titres, sous-titres, etc., sont en forme de titres et conformes aux normes orthographiques et grammaticales, voir s'il y a emploi adéquat du copier-coller, et des correcteurs automatiques. La machine doit diriger l'étudiant sur des outils adéquats: dictionnaire, concordancier, grammaire, suggestions de micro-tâches.

Étape 3: Document Powerpoint

Travail du tuteur machine: mêmes fonctions qu'en étape 2, vérification de la langue des schémas, signaler d'avance les mots, les phrases des commentaires qui seront difficiles à prononcer pour que l'étudiant se rende dans une base de données, ou un concordancier oral, qui lui en donne la prononciation, suggérer des micro-tâches).

Étape 4. Powerpoint et premier oral

Travail du tuteur humain: écoute, survol pour vérifier les plus gros problèmes, lecture, rythme, intonation, débit, etc.

Étape 5. Exposé

Tuteur humain et étayage des pairs: langue, techniques de présentation et domaines de référence.

III. La part de la technologie – une utopie accessible

La conception d'un "agent pédagogique", susceptible de prendre en charge une partie des fonctions tutorales, prend appui sur le concept de *teaching agent* initialement présenté par Philipp Hubbard (1999, 2000), pour qui un tel agent doit réunir plusieurs caractéristiques élémentaires:

- une présence physique et une personnalité,
- une expertise dans le domaine de référence,
- une aptitude à l'individualisation,
- une aptitude à initier l'apprentissage.

L'introduction de cet agent au sein de la relation tutorale implique de prévoir et de modéliser (au moins sur les principes) un nouveau système de relations entre les acteurs, schématisé ci-dessous.

Il s'agit avant tout de seconder le tuteur humain et d'apporter une réponse à la fois à la gestion du suivi de groupes d'apprenants importants et à celui du délai de réaction acceptable pour l'apprenant (problème de la temporalité dans les dispositifs de FOAD). Dans son principe, l'agent pédagogique se nourrit des données transmises par le système de suivi (*monitoring*) pour interagir avec l'apprenant et le tuteur humain, selon un processus de décision qui tient compte à la fois de la capacité de la technologie à fournir une réponse appropriée, de l'aptitude du tuteur humain à répondre dans un délai raisonnable et selon l'activité considérée aux sollicitations et aux actions des apprenants, ainsi que des modalités du *feedback* (proaction / rétroaction).

S'il faut aussi intégrer la variété des rôles dévolus au tuteur, qu'il convient de bien définir et de gérer, la conception d'un système hybride homme-machine se présente toutefois comme une hypothèse de travail envisageable, car la

recherche en informatique ouvre de son côté de nouvelles pistes que le didacticien va chercher à s'approprier.

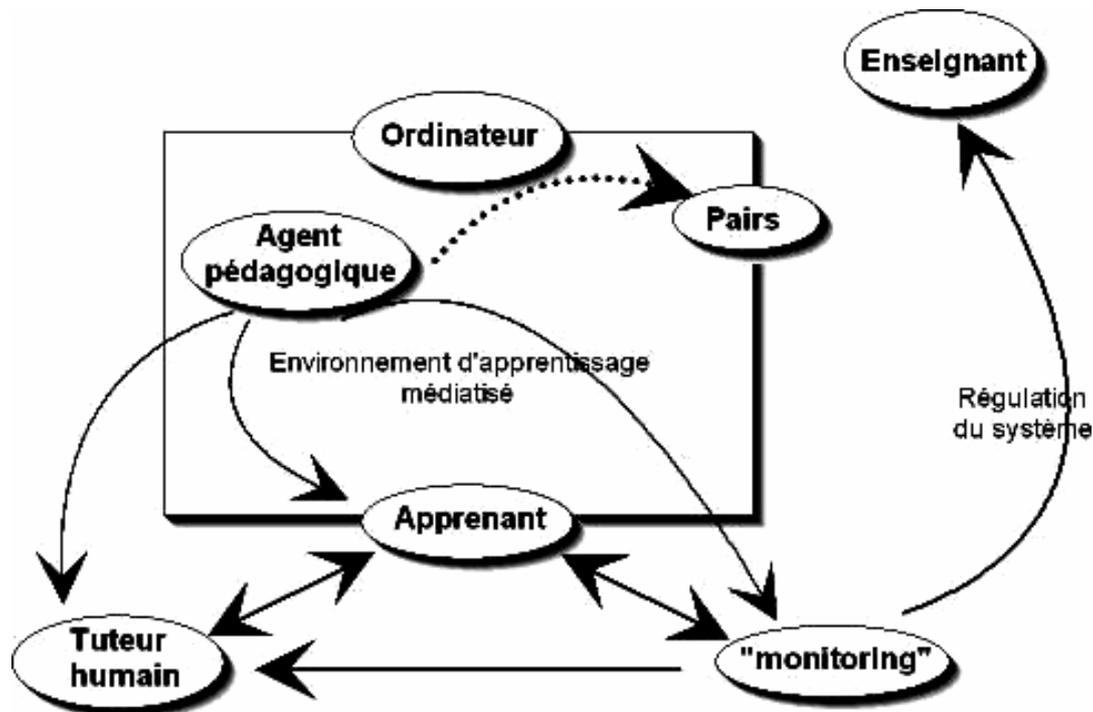


Figure 6: Jeu d'interactions entre l'agent pédagogique et le dispositif de FOAD

Il en est ainsi du concept de système multi-agents vers lesquels s'orientent les spécialistes de l'intelligence artificielle, dont tire profit l'une des applications, en expérimentation au laboratoire havro-rouennais LITIS, pour développer un système de détection du risque dans des situations dynamiques. Ces situations sont caractérisées par le fait qu'elles continuent d'évoluer en même temps que le système informatique en analyse les données et réagit.

Sous bien des aspects, la situation d'apprentissage peut être assimilée à une telle situation:

- le suivi pédagogique (*monitoring*) est bien un processus dynamique, dont les éléments doivent être collectés et traités, en même temps que l'apprenant travaille sur l'écran;
- la notion de "détection du risque" doit être interprétée, dans notre cas, comme la détection d'activités ou de techniques en contradiction avec ce que l'on connaît des processus d'acquisition et d'apprentissage;
- les scénarios doivent correspondre aux tâches et aux séquences recommandées pour la situation donnée;
- l'intervention du système (*follow-up* et *feedback*) doit être assez rapide pour que l'apprenant et/ou le tuteur puisse bénéficier utilement de cette aide.

Mais si la similitude est tentante, le transfert de notions didactiques telles que celles proposées par Hubbard, sous-tendues par l'approche théorique exposée précédemment, ne peut se concevoir comme une simple traduction ligne à ligne des concepts vers le champ informatique. Une telle traduction

d'ordre global, de concept à concept, a débouché pour l'agent informatique sur les caractéristiques suivantes:

- une interface homme-machine adaptable et intelligente;
- une base de connaissances contenant les profils d'apprenant, les contenus à acquérir, l'expertise disciplinaire requise, les données de suivi de l'activité de l'apprenant⁷.
- une base de scénarios (ensemble de modèles de stratégies d'apprentissage correspondant aux situations rencontrées);
- un ensemble d'attributs pour le système: autonomie, proactivité, dynamisme, adaptabilité, ouverture.

Cette description caractérise précisément les concepts informatiques de "système complexe" et de "système multi-agents" d'aide à la décision en situation dynamique. L'hypothèse que nous formulons est qu'il est envisageable de construire une ontologie d'un domaine simplifié d'apprentissage, tel que celui mis en place à Paris 3 et qui sera décliné pour des enseignements à l'université du Havre, et d'intégrer dans le dispositif de formation une part de suivi informatique.

La *figure 7* illustre l'architecture en trois couches du système multi-agents sur lequel repose le volet technologique de notre approche.

Sans entrer dans des détails trop techniques pour le domaine qui nous intéresse ici, l'originalité du système se caractérise par une architecture multicouche, dont chacune des trois couches interagit en permanence avec les autres.

Dans la *1^{ère} couche*, les agents factuels utilisent les données de traces (*tracking data*: enregistrement en temps réel des activités de l'apprenant) afin de fournir une représentation de la situation et de définir des traits significatifs de l'activité de l'apprenant. Le traitement des données de traces par les agents factuels est orienté selon les spécificités de l'objet d'apprentissage, précisément pour pouvoir inférer les éléments significatifs.

La *2^{ème} couche* est un passage intermédiaire au cours duquel les différents agents factuels sont comparés et regroupés en *clusters* afin de souligner les traits saillants de l'activité de l'apprenant.

Dans la *3^{ème} couche*, l'activité de l'apprenant est comparée aux scénarios de la base de données et selon leurs similarités, un agent prédictif produira la réponse appropriée à l'apprenant.

⁷ Ce que les informaticiens nomment "l'ontologie du domaine".

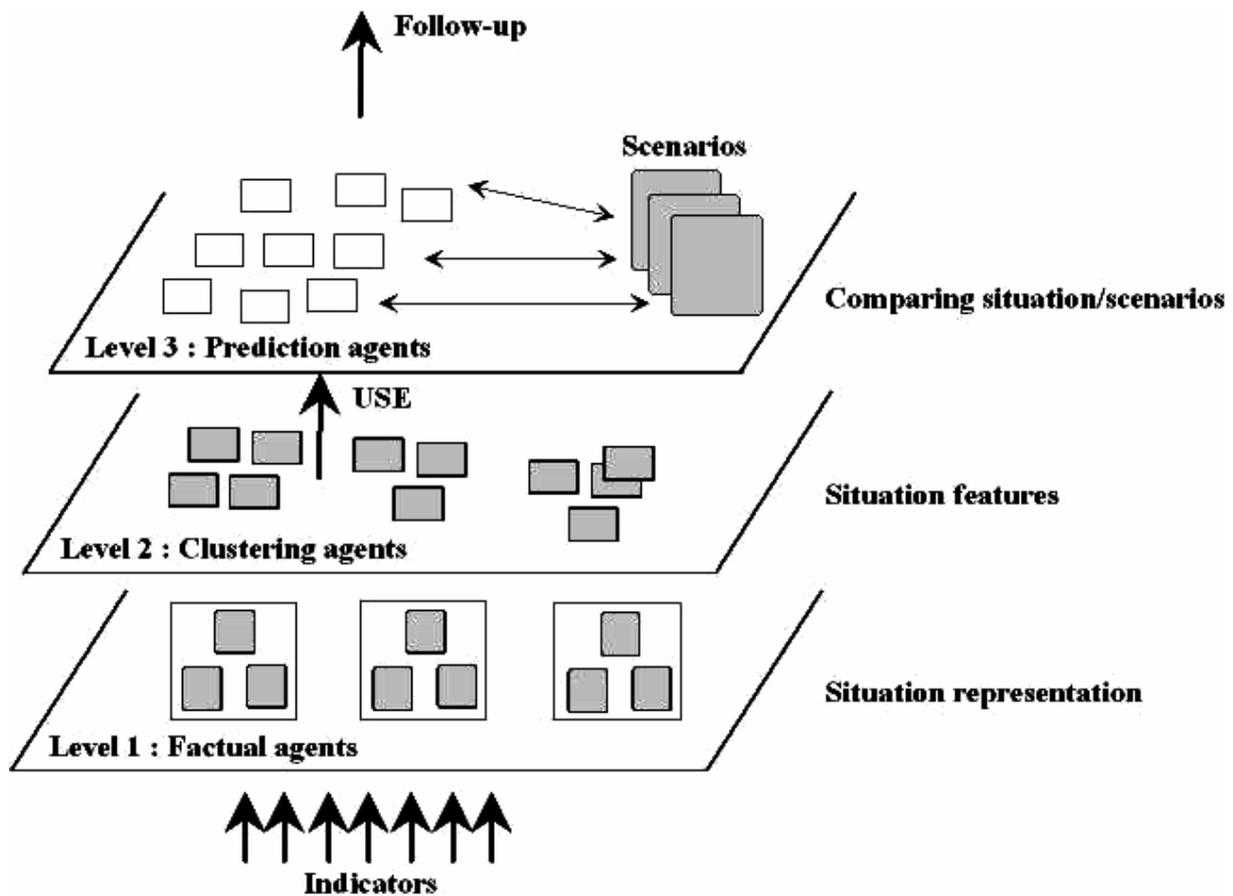


Figure. 7 : Représentation schématique du système multiagent en trois couches⁸

La forme et la nature de la réponse finale (rétroaction vers l'apprenant) reste à formaliser grâce à une définition des modèles cognitifs appropriés, des styles d'apprentissage et des spécificités de l'objet d'apprentissage.

L'intérêt de ce système expérimental tient toutefois à la possibilité qu'il offre de se dispenser de la traditionnelle linéarité des algorithmes informatiques, d'une part, et de modèles théoriques trop rigides et nécessairement inadaptés (De Lièvre *et al* 2004). Malgré cette flexibilité, il reste encore à réduire un certain nombre d'incertitudes théoriques liées à l'apprentissage: comment assurer, par exemple, une automatisation des processus visés par les micro-tâches, fondées sur une approche cognitive, et leur transfert vers la macro-tâche, qui repose sur une vision socioconstructiviste de l'apprentissage? Quelles définitions fiables des diverses fonctions tutorales proposer pour les modèles et scénarios requis pour le fonctionnement du système multi-agent?

L'une des difficultés tient à la complexité et la diversité de ces rôles et à leurs interactions à l'intérieur d'un système hybride homme-machine. La répartition entre les acteurs humains et la technologie résultera de la capacité de cette dernière à prendre en charge une partie de la responsabilité du tutorat. Si l'on accepte la définition de l'apprentissage comme le résultat de décisions

⁸ D'après Person *et al* (2006) et Bertin, Annoot & Gravé (2007).

individuelles sur la connaissance, la représentation de cette connaissance et les stratégies pour la faire évoluer, alors le suivi de l'apprenant peut tout à fait être envisagé sous l'angle d'un processus plus large d'aide à la décision.

Nous pouvons désormais proposer une représentation des interactions visées entre l'agent pédagogique et le dispositif de FOAD (fig. 8)⁹.

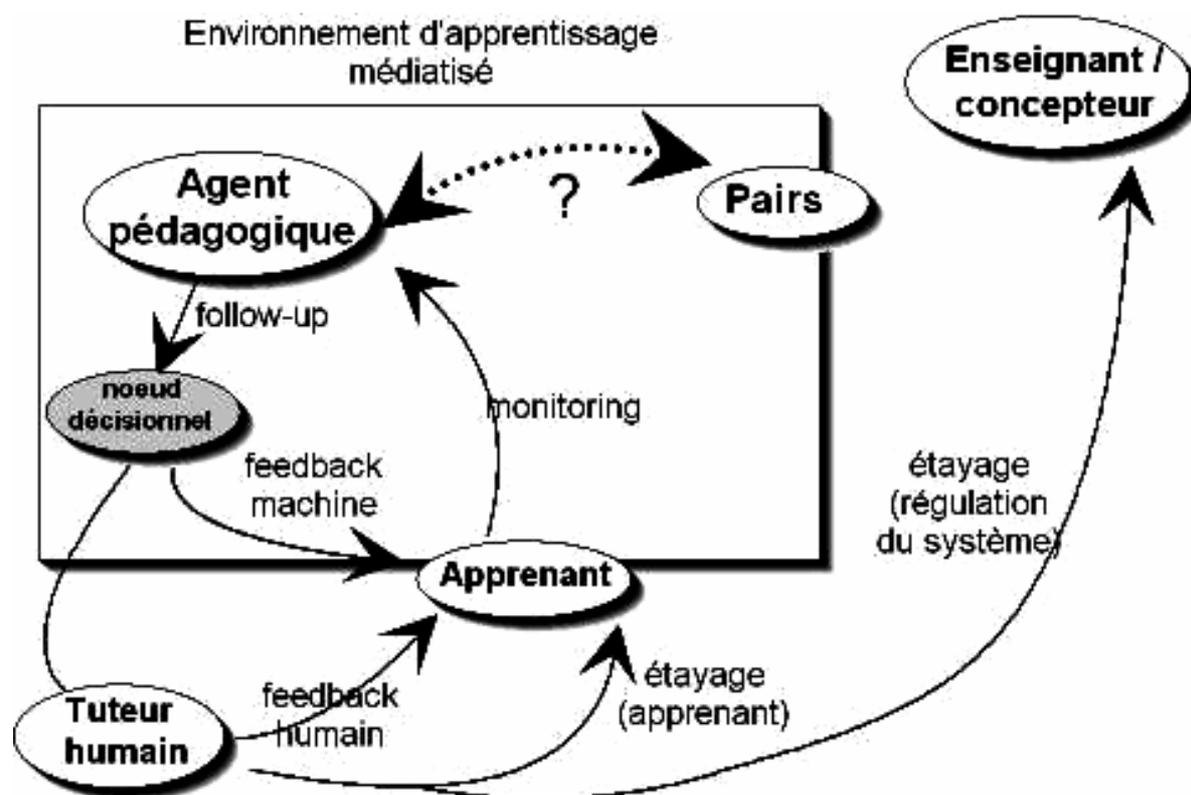


Figure 8: Modèle d'agent pédagogique pour la FOAD

Le modèle d'agent pédagogique qui découle de l'approche théorique que nous avons esquissée se fonde sur un dispositif de suivi pédagogique en dynamique (*monitoring*) qui enregistre la trace de l'activité de l'apprenant à travers une série d'indicateurs prédéfinis et évalue le risque de développement d'attitudes défavorables à l'apprentissage. Le traitement des données par le système multi-agent déterminera la nature du suivi ultérieur (nœud décisionnel): quel type de réponse produire et quel acteur est le plus apte à fournir la réponse appropriée (rétroaction par la machine vers l'apprenant lorsque c'est faisable, avertissement au tuteur humain dans tous les autres cas).

Le tuteur humain aura également la charge d'assurer l'étayage pédagogique, soit par un conseil direct à l'apprenant, soit par l'adaptation de la tâche ou du matériau au niveau de compétence révélé par les données de suivi (régulation du système).

⁹ Nous avons choisi de conserver la terminologie anglo-saxonne là où elle nous semble plus précise.

La question de savoir si le suivi des relations entre l'apprenant et ses pairs doit être intégré à ce dispositif fait encore débat: bien que techniquement faisable, son absence peut contribuer à préserver un espace de liberté dans cet environnement médiatisé, à l'intérieur duquel les apprenants peuvent interagir sans interférence de l'enseignant ou de l'institution. La question fondamentale et éthique devient alors: le suivi doit-il intégrer tous les aspects sociaux de la situation d'apprentissage? Que dire par ailleurs d'un espace d'apprentissage non contrôlé et donc sans régulation possible en cas de conflit dans cet espace? "Big Brother" ou socioconstructivisme totalement assumé? Ces délicates questions restent ouvertes.

Conclusion

L'approche théorique du suivi dans des situations de formation à distance, telle que nous cherchons à la construire, peut conduire à l'utopie ou à l'innovation. Rappelons que nous avons délibérément écarté de notre démarche le "tout-technologique", qui ne nous semble aucunement susceptible de répondre aux questions posées. La machine reste un outil au service de l'homme et c'est bien dans ce sens qu'il faut comprendre la notion de dispositif homme-machine. L'intérêt de notre démarche, pensons-nous, est d'ordre épistémologique: poser le cadre d'une réflexion demandant à être complétée par des recherches-actions visant à préciser la nature des interactions entre les composantes du dispositif. Ceci n'est pas chose simple, dans la mesure où, selon les principes de la pensée complexe de Morin, il est difficile de se limiter au principe cartésien de séparabilité¹⁰ pour une approche systémique des phénomènes: ce qui compte est *l'aller-retour incessant entre certitudes et incertitudes, entre l'élémentaire et le global, entre le séparable et l'inséparable* (Morin & Le Moigne 1999: 265). Ceci implique que la recherche dans le domaine qui nous occupe ici ne saurait uniquement être atomisée entre des équipes ou des projets par trop indépendants: l'initiative de RANACLES d'aborder ces questions de manière collégiale permettra sans doute d'avancer dans la compréhension des mécanismes du suivi et de mieux appréhender les conditions dans lesquelles une forme innovante de tutorat à distance pourra voir le jour.

Bien que de nouvelles voies soient ouvertes par la recherche en informatique, bien des questions restent en suspens avant de pouvoir envisager une intégration cohérente de la technologie dans les dispositifs de formation à distance.

- Comment faciliter les mécanismes de décision de nature cognitive ou métacognitive dans un environnement médiatisé?
- Quelles modalités imaginer pour un système de tutorat à distance? La littérature existante (EIFAD, 2004) suggère que le tuteur humain semble

¹⁰ La capacité de décomposer un phénomène en éléments simples pour mieux l'étudier.

mieux adapté à des modalités de suivi proactives, anticipant l'activité de l'apprenant, tandis que la technologie serait plus indiquée sur le mode réactif. Notre objectif d'intégrer les mécanismes d'étayage pédagogique dans cet environnement ferait porter notre choix vers un agent pédagogique proactif.

- Si on pense l'apprentissage comme un processus dynamique et évolutif, largement influencé par des facteurs individuels et culturels, comment peut-on espérer présenter un modèle fiable de la cognition?
- Quels éléments peuvent être formalisés pour fournir des modèles de travail utilisables par le système informatique: modèle cognitif, styles d'apprentissage, profils des apprenants, spécificités de la discipline enseignée, différents rôles du tuteur?
- Comment définir quand le *feedback* doit avoir lieu mais aussi comment déterminer le choix entre la rétroaction par l'humain ou par la machine?

Autant de questions pour orienter des recherches à venir, mais le défi en vaut la peine car y répondre permettra de mettre en place des dispositifs hybrides authentiquement intégrés et respectueux des pratiques des apprenants comme des enseignants.

Bibliographie

ANNOOT, EMMANUELLE. 1996. *Les formateurs face aux nouvelles technologies: le sens du changement*. Ophrys. 50-68.

ANNOOT, EMMANUELLE. 1998. Tutorat et ressources éducatives: la question étudiante. *Perspectives Documentaires en Éducation*.

BARNIER, GÉRARD. 2001. *Le tutorat dans l'enseignement et la formation*. L'Harmattan.

BEAUFILS, ANNE-FRANCE. 2007. *L'explicitation des besoins des adultes en formation à l'anglais en entreprise – une expérience au Portugal*, Thèse de Doctorat, Paris 3.

BERTIN, JEAN-CLAUDE. 2001. *Des outils pour des langues - Multimédia et apprentissage des langues*. Ellipses.

BERTIN, JEAN-CLAUDE, EMMANUELLE ANNOOT & PATRICK GRAVÉ. 2005. *Quelles médiations dans les formations à distance avec les nouvelles technologies dans l'Enseignement Supérieur?* CPER Pôle SHS, Universités du Havre et de Rouen, 2002-2004. <http://bertinjc.free.fr/Rapport%20IRSHS%202005.pdf>.

BERTIN, JEAN-CLAUDE & PATRICK GRAVÉ. 2006a. How to devise a cognitive agent for distance language learning. CALICO 2006, University of Hawaii at Manoa.

- BERTIN, JEAN-CLAUDE & PATRICK GRAVÉ. 2006b. Cognitive agents for distance learning. 3rd meeting of the global U8 Consortium's Education Committee, Université du Havre.
- BERTIN, JEAN-CLAUDE & PATRICK GRAVÉ. 2006. Devising a cognitive agent for distance language learning. CALL 2006: *How are we doing? CALL and Monitoring the Learner*, University of Antwerp.
- BERTIN, JEAN-CLAUDE & JEAN-PAUL NARCY-COMBES. 2005. Un suivi pour qui? Pour quoi? Conférence plénière au Congrès de RANACLES, Université de Caen. <http://www.unicaen.fr/unicaen/service/labolangues/ranacles/programme-samedi.php>.
- BERTIN, JEAN-CLAUDE, EMMANUELLE ANNOOT & PATRICK GRAVE. 2007. *Agents pédagogiques et agents cognitifs dans les systèmes médiatisés d'apprentissage à distance*, CPER Région Haute-Normandie – IRSHS. Rapport de recherche.
- BERTIN, JEAN-CLAUDE & JEAN-PAUL NARCY-COMBES. 2007. Les coulisses du tutorat en ligne: acteurs et machinerie. Congrès de la SAES 2007: "L'envers du Décor", Avignon.
- BEZSONOFF, CATHERINE. 2000. *Pratique de la formation: du discours à la réalité*. Les Éditions d'Organisation.
- BLANC MARION. 2000. Les nouveaux modèles de formation. Samier Henry (dir.) *L'université virtuelle. Les Cahiers du numérique 2*. 115-128.
- BOULLIER, DOMINIQUE. 2000. La loi du support: leçons de trois ans d'enseignement numérique à distance. Samier Henry (dir.). *L'université virtuelle, Les Cahiers du numérique 2*. 145-172.
- CARRÉ, PHILIPPE, ANDRÉ MOISAN & DANIEL POISSON. 1997. *L'autoformation: psychopédagogie, ingénierie, sociologie*. PUF.
- COLLECTIF DU MOULIN. 2002. Intégrer des dispositifs de formation ouverts et à distance, qui ont été conçus "ailleurs", dans des établissements d'enseignement supérieur. Campus numérique français 2000. Projet n° 15: texte issu de la conférence du Moulin, accessible en ligne: <http://ressources.algora.org/reperes/comprendre/moulin/index.asp>.
- COSTE, DANIEL. 1981. Spéculations sur la relation langue écrite - langue enseignée en classe. Cortès J. (dir.) *Une introduction à la recherche en didactique des langues*. Didier. 33-44.,
- DEMAIZIÈRE, F. & J.-P. NARCY-COMBES. 2005. Méthodologie de la recherche didactique: nativisation, tâches et TIC. *ALSIC 8*. 45-64. (http://alsic.u-strasbg.fr/v08/demaiziere/alsic_v08_14-rec8.htm)
- DE LIÈVRE BRUNO, DEPOVER CHRISTIAN, DILLENBOURG PIERRE. 2004. Au nom du "tuteur système" et du "tuteur humain". Colloque EIFAD, Poitiers.

- D'HALLUIN, CHANTAL (dir.) 2001. Usages d'un environnement médiatisé pour l'apprentissage coopératif. *Les Cahiers d'études du CUEEP* 43.
- GLICKMAN, VIVIANE. 2002. *Des cours par correspondance au "e-learning"*. PUF, Éducation et formation.
- HOUSSAYE, JEAN. 1988 *Le triangle pédagogique*. Peter Lang.
- HUBBARD, PHILLIP 1999. Teaching agents in CALL tutorials. CALICO 1999, Oxford, Ohio, USA.
- HUBBARD, PHILLIP. 2000. Taming teaching agents, meaning technologies and participatory dramas. CALICO 2000, University of Arizona, Tucson, USA.
- LEGENDRE, R. 1988 *Dictionnaire actuel de l'éducation*, Paris, Montréal: Larousse.
- MONTANDON, CHRISTIANE. 2002. *Approches systémiques des dispositifs pédagogiques – enjeux et méthodes*. L'Harmattan.
- MORIN, EDGAR. 1994. *La complexité humaine*. Flammarion.
- MORIN, EDGAR & JEAN-LOUIS LE MOIGNE. 1999. *L'intelligence de la complexité*. L'Harmattan.
- NARCY-COMBES, JEAN-PAUL 2005. *Didactique des Langues et TIC: vers une recherche-action responsable*. Ophrys.
- NARCY-COMBES, JEAN-PAUL & JEAN-CLAUDE BERTIN. 2007a. Monitoring the learner – Who? Why and what for? CALL 2006, *How are we doing? CALL and Monitoring the Learner*, University of Antwerp, publié dans *Computer Assisted Language Learning* 20 : 5. 443-457.
- NARCY-COMBES, JEAN-PAUL & JEAN-CLAUDE BERTIN. 2007b. Monitoring the Learner and Intercultural Awareness in a Computer-mediated Language Learning Environment. 7^{ème} Colloque des Usages des Nouvelles Technologies dans l'Enseignement des Langues Étrangères (UNTELE 2007).
- NARCY-COMBES, JEAN-PAUL & MARIE-FRANÇOISE. 2007. La tâche comme moyen d'optimiser l'enseignement/apprentissage en anglais pour les disciplines autres dans le contexte universitaire français. *Le français dans le monde – Recherches et applications* 42: *Langue et travail*. Clé international/FIPF.LFDM.
- PERSON, P., H. BOUKACHOUR, M. COLETTA, T. GALINHO & F. SERIN. 2006. Data Representation Layer in a MultiAgent Decision Support System. LITIS, *MultiAgent and Grid Systems – an International Journal* 2 (special issue).
- POTHIER, MAGUY. 2003. *Multimédias, dispositifs d'apprentissage et acquisition des langues*. Ophrys.
- PREVOST, HERVE. 1994. *L'individualisation de la formation: Autonomie et/ou socialisation*. Chronique sociale.

Suivi de l'apprenant à distance: Plaidoyer pour un *feedback* rationnel d'accompagnement conatif et évaluatif

*Jean-Christophe Coquilhat*¹

Notre propos va tenter d'établir des passerelles entre les concepts présidant à la mise en œuvre d'un dispositif de Formation Ouverte et à Distance (FOAD) et la réalité du *feedback* dans sa réalisation concrète. Notre conception du suivi de l'apprenant est celle d'un suivi prodigué tout au long du processus de formation en fonction des productions écrites ou orales que l'enseignant reçoit. Mais il ne s'agit pas d'exposer l'apprenant à des retours dont le contenu se contenterait d'un corrigé sur la forme. À notre sens, un *feedback* constructif doit pouvoir montrer son utilité dans le réinvestissement, par l'apprenant, de conseils prodigués dans une approche empathique où la notion d'erreur s'estompe au profit de celle d'une performance énoncée, trace d'une interlangue en évolution. Pour que ce suivi puisse opérer de manière optimale, il semble nécessaire de résoudre un certain nombre de problèmes conceptuels. Le premier touche à la distance elle-même et impose un renversement de posture impliquant une proximité inversement proportionnelle à la distance temporelle et géographique. Cela passe par l'instauration d'une présence sociale et de l'instauration d'un dialogue, au-delà des réalités imposées par le décalage du mode asynchrone. Le second problème tient dans la nature de cet accompagnement. Si celui-ci doit, comme c'est le cas dans notre recherche, s'exercer uniquement à travers le *feedback*, il convient de se poser la question de son efficacité et de sa légitimité. Le Cadre Européen de Référence pour les Langues (CECRL) peut-il rendre compte d'un réel état des compétences en se surimposant à une évaluation plus traditionnelle évaluant une performance? Telles sont les pistes que nous allons aborder dans cet article.

¹ Jean-Christophe Coquilhat est professeur Agrégé à l'Université Bordeaux 1 dont il dirige le Centre de Langues, l'Espace Omega. Sa thèse de doctorat en recherche-action et recherche-développement porte sur la création d'un dispositif d'enseignement-apprentissage de l'anglais faisant appel au monde du logiciel libre (Linux) et à une approche tutorale orientée sur un profilage et une correction fonctionnelle des productions alloglottes à l'écrit et à l'oral.

I - Mode asynchrone et *feedback* conatif²: éléments d'une logique de conciliation

“Il n’y a [...] pas la distance à vaincre ou à supprimer, il y a la distance dans sa pluralité à assumer”, nous dit Michel Bernard (1999 : 193). Nous aurions même tendance à affirmer qu’il n’y a *que* de la distance. Dans la mesure où les êtres ne se confondent ni de manière physique, ni de manière psychique, les relations sociales se régulent forcément par les liens qui s’instaurent entre les êtres à divers degrés de proximité. Qu’elle soit de type volontaire (consciente ou inconsciente) ou médiatique (critique ou dialectique) la distance *est*, de fait. Selon la théorie de la proxémie (Hall, 1971) il existe des schémas culturels qui visent à contrôler et réguler les distances interpersonnelles afin de gérer les différents degrés de proximité et d’implication relationnelles.

On peut dégager quatre distances, elles-mêmes modulables selon le degré de proximité: la distance intime, la distance personnelle, la distance sociale et la distance publique (Hall, 1971 : 147). Dans les rapports entre tuteur et apprenant ce sont les distances sociale (rapports professionnels et sociaux) et publique (rapport de l’orateur à son public) qui représentent les modes les plus conformes à la relation pédagogique médiée³.

La distance publique, selon Hall (*loc. cit.*: 155) implique une élaboration particulière du vocabulaire, de la grammaire et de la syntaxe. Il s’agit d’un style formel compatible avec la structuration signifiante de la création didactisée à distribuer. C’est donc une régulation sociale collective et unidirectionnelle qui a aussi pour but de légitimer un système – et donc l’institution qui le porte – par rapport aux acteurs qui le composent.

La distance sociale aura tendance à séparer les individus (Hall, *loc. cit.*: 154) et à leur donner une latitude dans l’instauration ou non d’un dialogue. On peut voir dans cette distance l’incarnation de la relation privilégiée entre tuteur et apprenant. Dans ce rapport individualisant, la richesse des échanges sera fonction à la fois de la volonté et des besoins de l’apprenant, et de l’investissement du tuteur à travers l’instauration d’une présence sociale.

1. La présence sociale en mode asynchrone

La présence sociale pourrait se définir comme le degré perçu de la présence d’autrui dans l’interaction (Short *et al.*, 1976) en termes de sociabilité, sensibilité, chaleur et relation humaine, c’est-à-dire dans les domaines conatifs⁴

² A comprendre dans son sens le plus large, c'est-à-dire qui fait référence à l'expérience, au vécu personnel et à la motivation. Chez l'enfant il s'agira de la prise en considération des interactions entre les dimensions cognitives et affectivo-comportementales.

³ Joseph Rézeau (2001 : 204-205) souligne l’existence, dans la recherche à défaut du dictionnaire, du verbe “médié” qui semble plus euphonique que “médiater”.

⁴ Les indicateurs de la présence sociale sont définis par Rourke *et al.* (2001) selon trois catégories: l’affectivité, l’interaction et la cohésion. White (2003 : 114) met l’accent sur l’empathie et la motivation.

de l'enseignement-apprentissage. Or, les réalisations de la présence sociale ne se conçoivent pas identiquement dans les communications à distance selon que l'on utilise le mode synchrone ou le mode asynchrone. Selon Walther (1992 : 52) "asynchronous online environment reduces a communicator's ability to establish his or her own presence or to perceive others as being present". Russo insiste sur ce même point.

A focus on tasks and deadlines (especially in the context of self-directed asynchronous courses where students must manage their investment of time) is likely to enhance this effect (Russo *et al.*, 2004 : 218).

D'où non seulement un besoin, mais aussi un devoir, d'instaurer dans la durée une relation interpersonnelle forte et immédiate (*construct of immediacy*, Andersen, 1979 : 544) à l'instar de la situation de classe en présentiel (Ellis, 2000 : 264). Cette notion a été formalisée par Shin en tant que "présence transactionnelle" où l'apprenant "senses the availability of, and connectedness with, teachers, peers, and the institution" (Shin, 2003 : 69). L'immédiateté du temps de réponse est ainsi définie comme une fonction de la distance psychologique que l'apprenant perçoit entre sa personne et l'objectif communicationnel, et donc comme corrélat au bon déroulement perçu des échanges pédagogiques. Or, cette dynamique globale de l'interaction est, par la nature même des caractéristiques du mode asynchrone, moins évidente à mettre en œuvre qu'en mode synchrone. Dans la temporalité synchrone, la mise en présence en temps réel par le dispositif de médiation agit comme un substitut spatial et vecteur de rapprochement. Au niveau de la gestion de la parole, les échanges prennent une fonction de séquenciation et de marquage du travail de formulation. En contrepoint de cet avantage apparent, cette structuration des échanges autour d'une substance sonore encadrée intervient dans l'ici et maintenant, et ses incidences cognitives sont celles d'une mise en pratique (élaboration, formulation, pauses) de la production discursive dans le réemploi immédiat. Le temps du mode asynchrone rompt avec la continuité de la parole car son but est d'élaborer, à travers le traitement mental du langage en construction, des formulations étayées sur un rythme distendu. Les environnements CMAL (*Computer-Mediated Asynchronous Learning*) défendent l'idée d'un contexte propice à la réflexion et à la délibération que l'on ne trouve pas dans l'immédiateté du mode synchrone. La permanence du dialogue et ses traces enregistrées permettent un marquage cognitif qui va au-delà des interactions parfois superficielles du mode synchrone et prolonge les échanges hors de l'adéquation temporelle. Ainsi, les apprenants sont vus comme des "problem solvers rather than memorizers of facts" (Collins *et al.*, 1996 : 3). Les termes "careful deliberation over course content" (Weasenforth *et al.*, 2002 : 59), "self-pacing and self-directed learning" (Benbunan-Fich *et al.*, 1999 : 411) évoquent avec insistance la dimension structurante de la temporalité spécifique du mode asynchrone.

Il n'est évidemment pas question d'opposer ces deux modes qui ont leurs avantages respectifs, mais plutôt de montrer que leur réalisation temporelle implique une réflexion sur les objectifs poursuivis, et non une juxtaposition de nature. Dans un cas (la synchronie) le *feedback* réciproque entre les participants en interaction rend compte implicitement ou explicitement de la mise en pratique langagière et permet de reformuler, de réorganiser la pensée en actes et d'obtenir un guidage en fonction des besoins immédiats. Dans l'autre cas (l'asynchronie), le décalage temporel entre émission et réception autorise "la perte de temps nécessaire, le tâtonnement" (Aumont *et al.*, 1992 : 104) propices à l'élaboration du message, sans pour autant remettre en cause le degré de proximité entre tuteur et apprenant (présence médiée, filiation des échanges). Cela légitime selon nous le recours à ce mode particulier d'un point de vue pédagogique, à condition qu'un véritable dialogue s'instaure entre tuteur et apprenant.

2. La relation apprenant-tuteur en mode asynchrone. Quelles conditions pour un dialogue?

Un dialogue est-il possible en mode asynchrone? Le dialogue n'est-il pas plutôt d'essence interactive dans l'ici et maintenant? Le fait que deux locuteurs inscrivent un échange dans la synchronie interdit-il pour autant que l'on puisse parler de dialogue en mode asynchrone? Tels sont les points de questionnement que l'on peut envisager dans la dimension du tutorat à distance. Il existe en fait un glissement de sens relativement commun de l'acceptation du terme. Hormis les sens restreints impliquant un échange dyadique⁵, on peut, par extension, exprimer une valeur dialogique à un discours non dialogal. Le dialogisme, en tant que "dialogisation" interne ou "dialogue cristallisé" (Ducrot, 1980 : 50) ne nécessite pas un destinataire dans l'échange mais un co-énonciateur qui soit plutôt inscrit dans la réalité du texte ou du discours. On peut ainsi se référer à Bakhtine quand il déclare: "on peut comprendre le mot "dialogue" dans un sens élargi, c'est-à-dire non seulement comme l'échange à haute voix et impliquant des individus placés face à face, mais tout échange verbal, de quelque type que ce soit [...] Toute énonciation, quelque signifiante et complète qu'elle soit par elle-même, ne constitue qu'une fraction d'un courant de communication verbale interrompu" (Bakhtine, 1977 : 136). Ainsi, il n'y a pas langage parce qu'il y a langage entre des locuteurs, mais parce que cette activité est une réflexion qui dépend nécessairement de cette possibilité. Il est impossible de ne pas être en situation de communication. Donc, toute production monologale est dialogue en son principe car elle est déterminée par la somme des productions antérieures mais aussi par la projection sur un interlocuteur potentiel.

⁵ La différence entre le suffixe "dia-" - signifiant "à travers" – et "di-" signifiant "deux" permet une sorte de confusion dont on pourrait sortir en utilisant des néologismes comme "dilogue", "trilogue" ou "polylogue" (Kerbrat-Orecchioni *et al.*, 1995).

Ainsi, le discours peut être monologal (ou monogéré dans la mesure où il est le fait d'un seul locuteur) mais dialogique dans la mesure où il met en scène un énonciateur et un co-énonciateur (Jeanneret, 1999). Les dialogues se situant hors du cadre de la conversation naturelle comme les dialogues dramatiques, cinématographiques ou encore les dialogues de manuels scolaires, sont à même d'atteindre d'autres buts que l'aboutissement à un consensus, par un formalisme construit dont les visées peuvent être d'ordre artistique, littéraire ou pédagogique.

Si l'on considère que la dimension dialogique ne s'opère que dans un vis-à-vis où la distance temporelle est exclue, le dialogue asynchrone rejoint l'échange de type épistolaire. Si, à l'extrême opposé, et à l'instar d'un dialogue entre acteurs de théâtre où le dialogue est préexistant et sans cesse rejoué et incarné, on considère que la dimension interactionnelle met en jeu des personnages actants interagissant et inter-réagissant dans le temps, on peut entrevoir un mode dialogique différé dont les effets peuvent rejoindre ceux induits par un dialogue en synchronie par des voies différentes. La vivacité, la multiplicité et la spontanéité de l'échange dialogique synchrone se trouvent ainsi supplantés par la profondeur et la qualité de l'argumentation du mode asynchrone. Si l'on rejoint les axes de la médiologie défendue par Régis Debray (2000), le point de vue de la communication est synchronique, (c'est celui de l'humanité dans l'actualité des vivants), alors que le point de vue de la transmission est diachronique. Sous cet angle, la mise en présence d'un savoir et d'un récepteur ne suffit pas pour que l'information s'incarne dans le récepteur, et la dimension synchronique semble n'apporter que les conditions de la communication. Néanmoins, si dans une approche dite "communicative" le fait communicatif en lui-même représente l'objet de la transmission (il s'agit ici de compétences et non de savoirs), la synchronie est aussi d'essence transmissive. Les deux modes semblent à l'évidence contenir en eux les éléments transmissifs que poursuit la FOAD. Dans le cadre de notre recherche, ce "dialogue" s'instaure à travers des navettes évaluatives et constitue donc un *feedback* devant respecter à la fois des règles empathiques au niveau socio-cognitif et des règles d'harmonisation au niveau institutionnel.

3. Les voies de l'empathie par l'approche socio-cognitive

L'approche sociocognitive tend à faire prendre en compte l'interaction existant entre le comportement d'un apprenant, ses caractéristiques propres et l'environnement dans lequel il évolue. Cette interaction énonce le concept de déterminisme réciproque. En effet, chaque composante du triangle environnement-comportement-caractéristiques individuelles est à la fois influencée par les deux autres et opère une influence sur les deux autres. Les degrés d'influence mutuelle font souvent l'objet de débats tant il est difficile de pouvoir mesurer efficacement le "poids" de l'environnement sur des

comportements eux-mêmes induits par des caractéristiques individuelles particulières.

Ainsi, le comportement d'un apprenant à distance pourra s'expliquer par des modèles environnementaux influençant l'individu (contexte familial, urbain, télévisuel) au niveau de ses valeurs mais aussi par des caractéristiques individuelles portées sur l'autonomie et le besoin de suivi. L'apprentissage des langues et les facteurs motivationnels qui y sont attachés reposent sur une dynamique de ce type.

Le traitement cognitif que l'apprenant fait des événements auxquels il est associé est le point de départ de sa motivation. Bandura (1986) dégage quelques aptitudes communes à tous les êtres humains à des degrés bien sûr différents:

- la capacité de se représenter et d'interpréter leur environnement grâce à des systèmes symboliques comme les langages parlé et écrit. (conceptualiser, théoriser et solutionner);
- la capacité de se référer au passé et d'anticiper le futur (dynamique motivationnelle);
- la capacité d'observer les autres et d'en tirer des conclusions (apprentissage par mimétisme);
- la capacité de s'autoréguler en fonction des évaluations de situations (contrôle et modification).

L'apprenant est donc porteur, si l'on en croit ces constats sociocognitifs, des capacités nécessaires à une participation active et responsable dans son apprentissage. Si le dispositif se doit d'offrir un environnement favorable à l'apprentissage, il ne peut se substituer à l'apprenant pour assumer l'acte d'apprendre. En revanche, il peut offrir un point de vue différent sur l'erreur et modifier la perception de l'apprenant quant à ses propres compétences.

Il semble que l'apprenant doive, dans cette optique, cesser de considérer l'erreur comme l'écart plus ou moins répréhensible entre sa performance et une norme langagière difficilement identifiable par lui-même. Dans le contexte éducatif, l'apprenant est amené à rendre compte du degré d'acquisition de nouvelles connaissances, ce qui peut représenter une certaine incertitude quant à la correspondance entre ce qui est attendu et ce qui est produit.

Le concept de perception de sa compétence devrait sans doute aussi être dissocié de celui de l'attente de résultats. La perception de sa compétence est l'évaluation qu'une personne fait de sa capacité de réussir, alors que l'attente de résultats est l'évaluation qu'elle fait de ses possibilités. Ainsi, un dispositif de FOAD peut être conçu et considéré comme tout à fait apte (capable) à induire un enseignement correct d'une matière mais dans le même temps être perçu comme incapable de permettre la réalisation d'objectifs. De même, un apprenant peut se sentir incapable d'effectuer une tâche donnée alors qu'il en a les capacités car la perception qu'un apprenant a de sa compétence peut exercer une influence néfaste sur son engagement cognitif.

Idéalement, le suivi de l'apprenant devrait réinstaurer la perception des compétences comme un déterminant de l'activité de l'apprenant. Ce suivi aurait idéalement pour conséquence de permettre à l'apprenant de s'investir à plusieurs niveaux.

- Dans ses choix: le choix qu'un apprenant fait d'accepter l'activité d'apprentissage ou de la refuser par des stratégies d'évitement, de retardement, de négociation, etc.
- Dans sa persévérance: durée dans le travail, temps consacré aux "devoirs", qualité de l'activité,
- Dans son engagement cognitif: ce terme se définit comme l'utilisation par l'apprenant de stratégies d'apprentissage et de stratégies d'autorégulation lorsqu'il accomplit une activité.

Si les stratégies d'apprentissage sont au cœur de l'engagement cognitif de l'apprenant, le *feedback* agit comme instance régulatrice de la perception de la performance. C'est ce que Bandura appelle le Sentiment d'Efficacité Personnelle.

4. Feedback et appropriation langagière

Le *feedback* peut être considéré comme le vecteur principal du processus d'autonomisation puisqu'il remplit quatre fonctions primordiales: "il rassure, il informe, il corrige et il aide" (Paquelin, 2000). Ceci rejoint la vision de Bruner sur le caractère auto-adaptatif du cycle d'apprentissage (Bruner 1982 : 313). Dans cette optique, la fonction qui rassure l'apprenant permettra d'influer sur le filtre affectif/motivationnel dans une dynamique d'encadrement corroborant ou infirmant la voie suivie par l'apprenant dans sa performance. Ce type de *feedback* doit avoir pour conséquence un encouragement à progresser ou à corriger le tir. Quand on connaît l'impact positif ou négatif de l'effet Pygmalion⁶, il demande (exige) de la part de l'enseignant-tuteur les précautions psychologiques d'usage.

En second lieu, le *feedback* se doit d'informer l'apprenant sur son positionnement au sein de l'apprentissage qu'il suit. Dans un environnement où le chemin est unilatéralement négocié par l'apprenant, l'enseignant-tuteur peut, sans doute par ses qualités d'expertise et de recul, réorienter l'apprenant sur des chemins de remédiation, d'engagement vers un niveau plus élevé ou même d'amélioration d'une tâche en imposant un recommencement.

Ces informations sont à même de toucher les domaines grammaticaux et lexicaux, mais aussi phonétiques et prosodiques, sans oublier le fond et le sens. Ici, le type de *feedback* s'affirme comme résolument critérié et recouvre sans doute les fonctions de l'évaluation de type formatif.

L'aide est bien entendu l'objectif vers lequel convergent les facteurs précités. Cela signifie que ces derniers doivent être perçus dans leur globalité

⁶ Terme introduit par Rosenthal pour désigner l'influence que les attentes de l'enseignant (au même titre que ses préjugés parfois) peuvent avoir sur le comportement de réussite des élèves.

comme concourant à une progression et non à un quelconque jugement de type sommatif, auquel cas ce nouvel *input* que représente le *feedback* pourrait se réduire à un blocage supplémentaire (cacophonie cognitive). Le *feedback* réflexif semble donc lié à l'acceptation par l'apprenant d'une critique instituant un changement de son état pour le cycle suivant (que cela soit orienté vers le haut, vers le bas, ou même encore vers la reconduction d'une performance à des fins de consolidation).

Le *feedback* humain semble rester le garant du lien enseignant-apprenant à travers l'éloignement. Si l'interface peut proposer des exercices automatisés (le plus souvent de type structural), l'état actuel des développements en intelligence artificielle interdit une correction synchrone des productions de l'apprenant à l'écrit comme à l'oral, sur le fond comme dans la forme. En tout état de cause, le *feedback* semble à la fois devoir jouer ce rôle de suppléance à un recul épistémique défaillant de l'apprenant et de ciment dans la construction de l'interlangue. Au sein de cette interlangue, un lien acquisitionnel fort semble positionner l'apprenant dans un triangle impliquant les mécanismes de la mémoire, les tâches à effectuer et la langue à apprendre.

Dans le domaine de la FOAD, peu d'études existent sur l'impact du *feedback*, surtout en mode asynchrone (Bangs, 2002 ; Felix, 2002) et elles se divisent entre des études de la performance et des études sur l'interaction. Il en ressort que le *feedback* métalinguistique (Yang *et al.*, 1999 ; Nagata, 1997) est plus efficace qu'un *feedback* de correction d'erreurs traditionnel car les apprenants intègrent les explications fournies comme un élément d'apprentissage supplémentaire susceptible d'être réinvesti. D'un point de vue didactique, le *feedback* semble être rattaché à la fois à la théorie interactionniste de l'acquisition du langage (Gass, 1997 ; Long, 1996) et à la notion de *noticing* (Schmidt, 1990). Selon l'hypothèse interactionniste, l'interaction par la négociation du sens facilite l'acquisition notamment dans le domaine du *feedback* correctif. Dans notre optique, cela incite à respecter les points suivants:

- L'*input* est susceptible de devenir *intake* s'il est repéré (*noticed*).
- L'apprenant est susceptible de remarquer une forme linguistique par l'interaction (négociation de la forme).
- Les formes linguistiques devraient être appréhendées à la fois au niveau syntaxique et sémantique afin de constituer un *intake*.
- Le but du *feedback* interactionnel est de permettre à l'apprenant de mesurer l'écart entre *input* (message correctif reçu) et *output* (production langagière de l'apprenant).

Si l'on se réfère au cadre de la distance de transaction de Moore (1993)⁷, le tutorat d'encadrement agira comme l'instauration d'un dialogue visant à pallier la structure plus ou moins rigide du dispositif, et donc à minimiser la distance transactionnelle. En cela, le tutorat agit comme vecteur de réinvestissement, de

⁷ C'est-à-dire, pour faire simple, à l'espace psychologique et communicationnel qui sépare l'enseignant de l'apprenant dans une situation d'enseignement-apprentissage à distance.

reconstruction sur le plan verbal dans une logique sémantique mais aussi méthodologique de ce que l'apprenant ne maîtrise pas.

Mais il serait louable que l'intentionnalité de la relation quadripolaire évoquée *supra* se pare d'objectifs d'autonomisation allant au-delà de la simple assistance, c'est-à-dire que le tutorat d'accompagnement relationnel soit le préalable à un tutorat d'accompagnement cognitif.

5. Le Cadre Européen de Référence pour les Langues comme instrument de mesure de l'interlangue

L'interlangue renvoie au rapport de l'objet didactique à la complexité de l'appréhension individuelle mais aussi, d'un point de vue descriptif, à des systèmes transitoires, des états intermédiaires d'une langue en cours d'apprentissage dont les processus dynamiques (externes pour la régulation et internes pour la cognition) sont soumis à l'idéologie descriptive du chercheur.

Selon Wendt (2000 : 115) on peut relier le concept de stratégie aux théories de l'action (cf. Leontiev) en considérant la stratégie comme une suite d'opérations et en remplaçant la notion de résolution de problèmes par celle "d'action de langue". Dès lors, on peut envisager une conception bifide de stratégies d'apprentissage et de stratégies de communication, elles-mêmes liées aux quatre compétences. Ces processus cognitifs d'élaboration de stratégies menant à une conscience langagière de l'interlangue, peuvent parfaitement être recouverts par la dénomination de "conscience réflexive de l'interlangue". Cette vision correspond parfaitement avec celle d'un enseignement à distance puisqu'elle sépare les actions de langue suivant les buts à atteindre, eux-mêmes tributaires des tâches à effectuer.

Ainsi, les stratégies de communication peuvent aussi participer de l'acquisition à travers la sollicitation d'aides, le renforcement d'acquis par leur mise en œuvre, la correction ou l'incompréhension de l'interlocuteur, la mise en parallèle de la langue source et de la langue cible. Inversement, les phases d'apprentissage, même les plus simples, sont (ou devraient être dans l'absolu) engagées dans une optique de communication. Un contexte d'apprentissage en approche communicative nourrira la "valeur communicative potentielle" (Gaonac'h 1991 : 181) même si celle-ci peut rester différée. La langue se définit ainsi non seulement par ses caractéristiques structurelles mais aussi selon des critères réciproques d'intelligibilité au sein d'une communauté d'utilisateurs. Dans le cas de l'apprentissage institutionnel, l'interlangue ne renvoie pas à une communauté partageant un sociolecte commun, puisqu'il n'y a que des interlangues individuelles, mais à un espace pédagogique dont le dénominateur commun serait la langue cible. L'évaluation d'après les critères du Cadre Européen de Références pour les Langues (CECRL) devrait permettre de donner un instantané des compétences distribuées au sein de l'interlangue, et constituer un complément éclairant et socialement légitime au *feedback* d'accompagnement. Mais ces critères sont-ils fiables dans leur nature

évaluative? Il apparaît que, malgré sa complexité et son exhaustivité, le CECRL se heurte à des critiques de deux ordres. Une critique que l'on peut qualifier de méthodologique et une critique associée à un certain flou en ce qui concerne la langue de spécialité.

Puren déplore que le Cadre, s'il reconnaît l'absence de consensus suffisamment solide dans la recherche pour se fonder sur une quelconque théorie d'apprentissage, justifie son ouverture éclectique par des arguments négatifs et ne développe pas de réflexion systématique sur les relations entre pratiques de l'éclectisme et théories de la complexité. Le Cadre, selon cet argument, évite une approche didactologique prenant en compte des points de vue déontologique, épistémologique et idéologique, courant ainsi le risque de "dérives d'instrumentalisation au service de pratiques formatives autoritaires et d'harmonisations institutionnelles forcées" (Puren, 2006 : 74). En outre, Puren pointe un "décalage impressionnant entre d'un côté la précision des niveaux de compétence en langues et leurs descripteurs, et une réflexion méthodologique aussi ambiguë qu'inachevée". (*op. cit.*: 79).

En ce qui concerne la langue de spécialité, le Cadre définit des domaines particuliers d'action des acteurs sociaux dans la mobilisation de compétences langagières pour la réalisation de buts communicatifs. Ces domaines sont: le domaine éducatif, le domaine professionnel, le domaine public et le domaine personnel. Dans le domaine professionnel, le Cadre ne se décompose pas en descripteurs portant sur l'objet central de la profession ou sur les conditions de pilotage de l'activité. Michel Petit (2007) relève que, en termes de discours, il n'y a pas d'évocation de discours interne spécialisé, de discours hétérospécialisé ni de discours d'interface fonctionnelle. En somme, les descripteurs du Cadre ne traitent pas spécifiquement de la langue de spécialité mais intègrent sa pratique aux usages ordinaires. L'utilité opérationnelle des descripteurs prête le flan à une marge d'interprétation: quelle différence existe-t-il entre "comprendre un texte" et "lire sans effort"?

Au niveau phonologique, bien que le Cadre insiste sur cette dimension en de nombreuses occasions (c'est d'ailleurs plutôt le terme "phonétique" qui revient le plus souvent), il ne fournit qu'un seul tableau de descripteurs (p.92 du Cadre) en donnant pour instruction: "Les utilisateurs du Cadre de référence envisageront et expliciteront selon le cas, les aptitudes phonologiques nouvelles exigées par l'apprenant, l'importance relative des sons et de la prosodie, si l'exactitude phonétique et l'aisance constituent un objectif d'apprentissage immédiat ou à plus long terme" (*ibid.*).

Même si cela peut se comprendre dans la mesure où l'approche consensuelle du Cadre favorise les compétences pragmatiques sur le fond plutôt que sur la forme (peu importe en somme que le locuteur ait un accent étranger dès lors qu'il manie la langue correctement), il semble que le travail d'évaluation ne puisse faire l'impasse sur ces aspects cruciaux de l'expression orale puisque le but affiché reste de fournir à l'apprenant un feedback constructif

sur sa performance autant sur le fond que sur la forme. Enfin, le besoin d'adaptation institutionnel (note chiffrée de contrôle continu) et technique (dispositif de FOAD) font que le Cadre doit rester un pôle de référence en arrière-plan, c'est-à-dire un outil de cadrage, balisant la mise en place d'outils d'évaluation spécifiques.

II - Le dispositif mis en place à l'Université Bordeaux 1

1. Principes d'un dispositif hybride

On peut considérer le cœur du dispositif en relation avec le potentiel d'action qu'il induit et l'appréhender comme une co-construction à achever. C'est une approche modulaire qui permet d'envisager l'ajustement du dispositif au gré du décalage qui s'instaure nécessairement entre l'organisation planifiée proactive et la "potentialisation des situations" (Paquelin, 2005 : 4). Cela implique l'existence d'un schème d'usage transformateur actualisé par le sujet en situation dans son interaction avec le dispositif, hors du champ de la prescription. Ainsi, en fonction de l'appréhension par l'apprenant de ce que l'on attend de lui (objectifs, tâches prescrites), il projettera une organisation unique d'emploi du dispositif qui induira des conduites d'appropriation, de rejet, de substitution de moyens, d'exploitation des ressources, de mise en place de routines d'apprentissage éminemment personnelles. L'exploitation des artefacts et les recours aux aides, la navigation et la gestion du temps, évoquent une auto-régulation environnementale, passage de l'instrumentalisation à l'instrumentation dans ce que Proulx (2002 : 180) appelle des "trajectoires d'usages"⁸ et que nous préférons appeler "la négociation de la distance de transaction". Il n'existe pas de scénario ou de synopsis pré-établi de manière figée, ni même de planification, mais plutôt une trame en filigrane par rapport à laquelle l'utilisateur choisira de construire un cheminement non linéaire en fonction de ses besoins et capacités. Ce cheminement évoque la "trajectance", "démarche par laquelle la personne produit son chemin faisant en situation, dans le temps et en tenant compte de cette perspective "bifilaire" où s'entrecroisent et s'entrelacent trajet et trajectoire" (Bernard, 1999 : 162).

Nous allons reprendre ici la distinction que Puren (2006 : 77) effectue dans la modélisation des tâches et des actions conjointement à la distribution des rôles d'utilisateur ou d'apprenant en l'adaptant à la FOAD. Le dispositif asynchrone mis en place à l'Université Bordeaux I se décomposait d'un point de vue didactique de la manière suivante.

a. Objectif: la société comme domaine de réalisation des actions langagières.

⁸ "Les parcours singuliers que les individus empruntent à travers la constellation d'objets communicationnels passés, présents ou émergents qui leur sont offerts et qui constituent un environnement informationnel et cognitif privilégié dans l'élaboration de leurs pratiques d'information et de communication".

b. Moyens:

- *À distance*: la FOAD comme domaine de réalisation des tâches.
 - L'interface comme lieu de conception d'actions (transposition et intégration didactique)
 - L'interface comme lieu d'action (réalisation de tâches).
 - L'interface comme lieu d'entraînement et de préparation aux compétences langagières et culturelles à mettre en œuvre dans un contexte futur d'interaction sociale (simulations orales et écrites).
 - Le feedback comme lieu de remédiations et de reconstructions (feedback en cours de parcours ou après production).

- *En présentiel*

- La classe comme lieu de réalisation artificiel d'actions de langue.
- La classe comme lieu synchrone de remédiations.

D'un point de vue technique, le dispositif s'organisait, en amont, par la mise à disposition en ligne ou sur cédérom hybride de contenus organisés selon des briques pédagogiques adaptatives. En ce qui concerne l'ingénierie de formation, l'effort a porté sur la création d'un environnement de travail répondant à une quadruple contrainte:

- l'adaptabilité au plus grand nombre de systèmes informatiques existants,
- une orientation *open source* économiquement viable,
- l'intégration de contenus organisés en briques pédagogiques adaptatives,
- un suivi de l'apprenant constructif et harmonisé.

2. Mise en œuvre

Cet environnement de travail consiste en une distribution Linux Slackware transformée et adaptée à la mise à distance d'un enseignement-apprentissage de l'anglais. Il est issu des expérimentations effectuées les deux années précédentes avec l'interface IDEM LAB (Interface Didactique d'Enseignements Multimédia en Laboratoire). Open Lads signifie à la fois "Outil Pédagogique d'Enseignement Numérique en Langue Anglaise de Spécialité" et fait référence à l'ouverture (Open) et à la communauté d'apprentissage. Dans le système sont intégrés: une présentation, des instructions, un calendrier, des adresses de contact, un point sur l'évaluation et le Cadre Européen Commun de Référence en Langues, et les ressources pédagogiques organisées selon un mode adaptatif particulier. Les briques pédagogiques adaptatives (Figure 1) ont pour but de prendre en compte la variabilité des styles cognitifs, des compétences communicatives et de l'interlangue de chacun à travers l'implémentation d'un mécanisme didactique de perception des formes qui puisse permettre aux apprenants de bâtir des associations vers la langue cible. Ainsi, des constituants subordonnés visant à la particularisation, la définition ou l'analyse, des éléments évaluatifs ou phatiques, des items paraphrastiques (reformulation ou redite) ou d'exposition (précisions contextuelles) se trouvent insérés au sein d'un agrégat didactique afin d'en faire un réel objet d'apprentissage, voire d'acquisition

relative. Dans ce sens, et sans chercher nécessairement à provoquer une hiérarchisation des contenus sémantiques à partir de leur fonction dans l'organisation argumentative, on peut raisonnablement espérer un impact dans l'utilisation d'éléments au gré d'un besoin de la structure discursive.

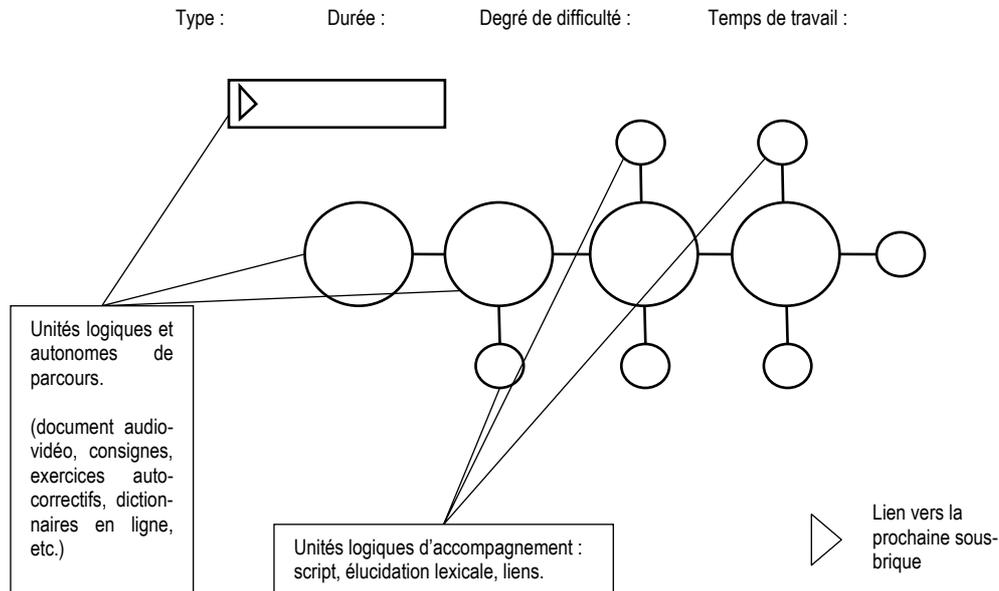


Figure 1: Structure graphique d'un élément de brique pédagogique adaptative

Au niveau du travail demandé, les tâches à accomplir étaient de deux ordres. Soit il s'agissait de celles proposées par l'interface, et donc graduées par un système d'étoiles indiquant le niveau de difficulté de la tâche. Les étudiants pouvaient s'entraîner sur des exercices ou des tâches de bas niveau (*). Au final, ils devaient respecter dans leur choix une tâche de production langagière d'un minimum de deux étoiles. Soit, second cas, l'étudiant décidait de se créer lui-même sa tâche en abordant un sujet de son choix (basé sur un document proposé) et en respectant les modèles fournis dans l'interface.

En termes de production, la tâche demandée aux étudiants n'impliquait pas de planification polygérée, comme dans un dialogue, où le contrôle est assuré tour à tour par les divers coproducteurs du discours. Il n'y avait pas d'interaction à réguler, et la cohésion textuelle n'avait pas à être assurée par des références à des paramètres de production inhérents à des interlocuteurs. Dans l'organisation générale des productions, la structuration est un élément important puisqu'elle articule les procédures de représentation langagière du contexte à celles de l'organisation communicative de la chaîne textuelle (la textualisation, c'est-à-dire l'organisation discursive mettant en jeu les opérations de connexité, de cohésion et de modalisation, reste un élément primordial). Au niveau de la correction de la langue, les erreurs concernant la grammaire plus que les tournures idiomatiques ou simplement l'organisation de la phrase de type anglo-saxon, demeurent des éléments à corriger impérativement, mais aussi à expliquer.

On se trouvait donc face à des productions langagières où les opérations étaient, de fait, liées à la planification et à l'organisation de la progression argumentative. Donc, ancrage autonome, tournures impersonnelles (dues sans doute à un "apprentissage" préalable par cœur), voix passive et un emploi fréquent des temps et des modaux. Les correcteurs avaient ainsi à gérer la correction de productions langagières sur plusieurs niveaux.

En aval de la mise à disposition des contenus, il était nécessaire de structurer le *feedback* selon trois points clé. Le *feedback* devait être:

- *contingent*: congruent avec la tâche demandée et la production;
- *spécifique*: adapté aux traits caractéristiques de chaque production;
- *immédiat*: afin de ne pas casser le cycle en induisant un temps de réaction trop long entre production et correction;

et décliné selon une triple évaluation:

- *feedback* écrit et oral (évaluation selon les descripteurs du CECRL);
- *feedback* écrit et oral sur les productions corrigées et améliorées (évaluation selon les descripteurs du Cadre);
- présentation orale sur le thème en exposé et *feedback* immédiat sur la performance.

3. Contraintes techniques

D'un point de vue technique, il existe globalement trois sortes d'outils susceptibles d'être utilisés dans la correction de productions écrites:

- des logiciels dédiés comme *Mark-in*⁹,
- le module de correction d'un traitement de texte (mode révision),
- l'utilisation d'un *tablet-pc* qui numérise la graphie manuscrite sur le document dactylographié.

Pour l'oral, certaines universités ont développé des outils de type client-courrier intégrant dans le mode fenêtré du courrier une piste son associée au contenu du message. Mais le principe d'association d'un fichier son à un fichier correctif ne pose pas de problème avec les outils classiques de *voice-mail*.

Dans ces cas de figure, il semble que cela soit plutôt l'utilisation que l'on fait de ces outils qui importe que la valeur ajoutée de leur spécificité. Il semble dès lors nécessaire de respecter les principes de présence sociale à travers ces outils pour incarner au mieux un tutorat à même d'accompagner les apprenants dans leur parcours.

Les choix ont été dictés, dans le cas qui nous intéresse, par les constats suivants.

- *Mark-in* n'agit que sur des fichiers rtf ou html alors que les étudiants utilisent des traitements de texte.
- le *tablet-pc*, hormis la dimension financière, peut reproduire une écriture difficilement lisible.
- Word est un logiciel propriétaire que tout le monde ne possède pas.

⁹ <http://www.cict.co.uk/software/markin/index.htm>.

C'est pourtant ce dernier choix qui a été opéré dans la mesure où l'on peut accepter une grande variété de fichiers (rtf, odt, txt, html, etc.) et les convertir en format pdf si nécessaire.

4. Modes de correction

La mise en place de ce système hybride (20 heures éq. TD + 4h30 à distance) ne permettait pas de rentrer avec les enseignants dans des considérations complexes d'uniformisation et d'harmonisation de leurs corrections en fonction de critères spécifiques d'analyse du discours. Au préalable, il a été nécessaire d'effectuer une analyse des modes de correction des enseignants afin de déterminer si ce contexte particulier pouvait induire des courants convergents au niveau du *feedback*. L'observation des différents modes de correction montre une convergence vers une utilisation pragmatique de l'objet à corriger, sans doute induit par un objectif d'efficacité conjugué à la répétitivité de la tâche corrective. Ainsi, pour l'écrit, le texte a servi de support pour l'utilisation variée de marques de correction: double barré, couleur rouge (associée parfois à des caractères gras) et commentaires dans les cases de correction. Tous ont aussi associé un commentaire sur le fonds et des conseils méthodologiques sur la manière de réorganiser l'argumentation, soit dans le corps du courriel, soit en bas de page du fichier Word.

À l'issue de l'analyse des modes de correction des enseignants, des règles d'harmonisation des corrections ont été définies selon les axes suivants (*Annexe 1a et 1b*).

Dans le corps du texte:

- texte remplacé en rouge,
- texte barré (superflu),
- omissions,
- surlignage pour la prononciation.

En marge du texte:

- correction nécessitant une mise au point grammaticale en français; explications claires et détaillées.
- mises au point sur le style, les erreurs de dictionnaire et les emplois *abusifs non admis dans la langue*.

À la suite du texte:

- commentaire global sur la longueur, le niveau de langue, l'approche méthodologique et la structure;
- conseils pour la seconde version; encouragements.

Évaluation d'après les descripteurs du CECRL: items à cocher ou non selon les compétences développées (possibilité de deux tableaux consécutifs pour rendre compte d'un niveau intermédiaire (B2+, C1+): les items du premier tableau sont tous cochés et ceux du second, partiellement seulement). L'évaluation finale en notation chiffrée sanctionne l'effort et la prise en compte des conseils prodigués à partir du niveau de performance de départ, ce qui respecte la

variété des degrés de compétence des étudiants, reflète une évaluation formative plutôt que normative et agit comme facteur d'encouragement (*Annexe 2*).

Évaluation de la performance oralisée: fichier son d'accompagnement: commentaire suivi de la performance orale, correction des éléments phonétiques et prosodiques, exemples de prononciation, exemples d'oralisation sur un paragraphe (*Annexe 3*).

À l'oral, la technique a sans doute pesé plus lourd car peu d'enseignants on renvoyé des fichiers sonores. En revanche, tous ont renvoyé des fichiers texte¹⁰ portant un regard critique à la fois sur les mêmes aspects communicatifs et langagiers que les textes, mais aussi sur l'accentuation, voire la prosodie. Les codes de couleur rouge sur les syllabes accentuées ont été la règle générale. On retrouve les mêmes aspects qu'à l'écrit dans la mesure où il s'agit pour la plupart de commentaires lus face à un micro. À cela, rien de répréhensible (c'était admis depuis le départ) puisque l'interface asynchrone n'avait pas pour objectif d'enregistrer de l'oral sur un mode dialogique ou conversationnel, mais d'entraîner dans ce but (articulation et argumentation conjointe). En revanche, un commentaire audio envoyé à l'étudiant reste un objectif pour la phase suivante, à la fois dans un souci de cohérence avec l'exercice mais aussi dans une perspective d'abolition de la distance par la proximité pédagogique évaluative.

L'objectif de ces navettes évaluatives ne correspond pas à l'illusion que l'on pourrait avoir de considérer que le travail effectué sur un semestre suffirait à faire évoluer le niveau langagier des étudiants de manière significative. Plus prosaïquement, l'objectif principal consistait à délivrer des repères individualisés susceptibles de développer une réelle conscience de l'interlangue en mutation:

- des repères dans les domaines lexicaux, syntaxiques, morfo-syntaxiques et phonologiques;
- des repères permettant aux étudiants de se positionner par rapport aux descripteurs du cadre correspondant aux productions monologiques alloglottes (le but avéré étant de passer à une phase interactionnelle authentique);
- des repères méthodologiques dans le travail asynchrone encadré à la fois dans leur parcours au sein des briques pédagogiques, des tâches proposées, et des sujets abordés, mais aussi dans les registres de compréhension orale et d'appropriation de la langue de spécialité;
- enfin, des repères permettant de mieux appréhender un parcours hybride entre la responsabilisation induite par l'autonomie guidée et l'effort participatif dû au travail interactionnel en classe.

¹⁰ Certains avec des liens hypertextes sur des fichiers son offrant la prononciation correcte. Quelques exemples de productions et des *feedbacks* associés à l'écrit et à l'oral sont disponibles à l'adresse suivante : http://www.english-site.com/Tests_Openlads.

5. Analyse des productions

L'observation des modèles corrigés permet de dégager quelques grandes lignes sur le niveau général des étudiants de Master 1 d'Informatique. Tout d'abord, d'un point de vue méthodologique, et quand la langue cible ne fait pas défaut en termes de moyens langagiers, la grande majorité des étudiants sait allier une compétence énonciative (cohérence par rapport au contexte situationnel du document de travail) et une compétence textuelle (cohésion interne de la production du point de vue de l'analyse et du commentaire). Au niveau des sous-groupes que Bronckart (1985) dégage dans les opérations langagières de structuration, les opérations de repérage de types discursif, déictique voire historique, de même que la planification argumentative ne font pas défaut à de très rares exceptions près. Les étudiants de ce niveau ont visiblement appris, à travers leur parcours, à aborder les thèmes d'étude de manière ordonnée et à argumenter de manière construite. Au niveau du lexique de spécialité, il n'y a pas de problème de réemploi, du fait que la cohérence narrative en fait l'objet d'une utilisation impérative, mais aussi qu'il s'agit d'un vocabulaire utilisé dans leur filière de spécialité. En revanche, la correction de la langue générale reste sujette à bien des défauts: la structure des phrases produites mélange, pour l'immense majorité, des expressions apprises, des phrases correctes, mais calquées sur le français, des phrases incorrectes, calquées aussi sur le français, du lexique issu de recherches dans le dictionnaire de manière plus ou moins heureuse.

6. L'évaluation orale présentielle post-FOAD

L'instance finale du suivi individuel devait se formaliser dans une présentation orale suivie d'une interaction avec les membres du groupe classe. Pour l'enseignant, il était nécessaire d'évaluer les retombées du travail préparatoire encadré à distance et d'utiliser pour cela un outil cohérent à la fois avec une notation institutionnelle (contrôle continu), avec les descripteurs du CECRL (prise de parole en continu et interaction) et qui puisse faire l'objet d'un retour sur la performance (document donné à l'étudiant à l'issue de sa performance).

Nous avons pour cela essayé de critérier les éléments entrant en jeu dans cet exercice particulier de l'exposé-entretien en prenant en compte les dimensions de communication, d'organisation, de compréhension de la problématique, de compétence langagière et d'interaction. L'ensemble des points accordés suivant la nomenclature des items représentait 40 points (*Annexe 4*). Le lien avec le CECRL est expliqué dans la partie droite du document (qui est en fait un cadre dédié à la prise de note de l'enseignant pendant l'exposé afin d'assurer un *feedback* oral "à chaud" après la performance) et était purement indicatif car, même si nous avons élaboré cette grille les yeux rivés sur les descripteurs du CECRL, nous ne pouvons affirmer qu'il y ait une correspondance absolue avec eux dans la mesure où la performance fait l'objet

d'une préparation en amont. La performance reflète donc, dans un domaine particulier, sur un sujet connu (cf. le problème de la langue de spécialité soulevé par Petit, *supra*), un état de la compétence qui ne couvre pas la généralité globale des sujets abordables. L'indication donnée (note inférieure à 10 = A2, note entre 10 et 13 = B1, note entre 14 et 17 = B2, note supérieure à 18 = C2) permettait de donner une indication globale sur le degré de compétence atteint durant l'exposé sans pour autant avoir une valeur de certification. Nous avons tenu à ce que les étudiants le sachent.

Nous avons tenu, dans l'élaboration de la fiche, à répartir les points équitablement entre, d'une part, la compétence de communication (autonomie, attitude, adaptation à l'auditoire, etc., sur 10 points) et les qualités d'organisation (8 points) pour un total de 18 points et, d'autre part, la réalisation concrète des actes de langage (15 points) et l'interaction (4 points) pour un total de 19 points, le travail préparatoire et le recul de l'étudiant dans son effort de synthèse étant doté de 3 points.

Une dernière question demeure. Quel est l'avis de l'apprenant sur ce suivi particulier? Existe-t-il des différences de perception suivant les niveaux dégagés par les critères du CECRL? Nous avons sélectionné quelques éléments d'un long questionnaire distribué aux étudiants à l'issue de leur formation sur leur vécu socio-pédagogique.

7. Impressions des apprenants

Ce questionnaire a été rempli de manière anonyme afin que la question de la sincérité ne soit pas de mise. En revanche, nous avons demandé aux étudiants de déposer leurs feuilles remplies sur des chemises correspondant aux niveaux du CECRL qu'ils avaient obtenus à l'issue de leur formation. Cela nous a permis de faire un tri révélateur à bien des égards.

a. *Le recours à l'évaluation selon les critères du Cadre Européen de Référence en Langues constitue surtout (une seule réponse):*

	B1	B2	C1
- un instrument de vérification de l'appropriation des connaissances.	39%	33.5 %	16.5%
- un instrument de constatation du chemin à parcourir pour s'améliorer.	60%	50%	72.5%
- une approche globalement réductrice.	0%	12.5%	5.5%
- un gadget dont on pourrait se passer.	1%	4%	5.5%

La perception des critères du CECRL semble rendre compte d'une approche reconnaissant sa légitimité sur l'ensemble des étudiants. On peut sans doute expliquer que 12,5% des niveaux B2 y trouvent "une approche globalement réductrice" par le fait que leur niveau de compétence (qui aspire à tendre vers le C1 qu'ils sentent proche) est suffisamment élevé pour percevoir

un décalage entre leur perception personnelle de leur compétence et celle donnée par notre adaptation (sans doute critiquable) des critères du CECRL. Cela ne s'étend pas aux C1 qui considèrent que le niveau C2 est proche du bilinguisme et donc pratiquement inatteignable.

L'immense majorité des étudiants semble consciente que, dans notre optique, le CECRL n'est pas un instrument de vérification des connaissances ou de compétences, mais un instrument permettant d'assurer un suivi de l'interlangue en actes. Nous pensons, à cet égard, avoir réussi à faire partager notre vision de l'adaptation et de l'utilisation du CECRL dans cet objectif.

b. *La correction de vos travaux écrits vous a-t-elle paru offrir de réels outils d'amélioration (conseils, règles, approfondissements)?*

	B1	B2	C1
- Oui.	65%	58.5%	55.5%
- Assez.	35%	25%	44.5%
- Peu.	0%	16.5%	0%
- Non.	0%	0%	0%

c. *La correction de vos travaux oraux vous a-t-elle paru offrir de réels outils d'amélioration (phonétique, accentuation, rythme)?*

	B1	B2	C1
- Oui.	72.5%	62.5%	66.5%
- Assez.	27.5%	29%	28%
- Peu.	0%	8.5%	5.5%
- Non.	0%	0%	0%

Les réponses à ces deux questions doivent être analysées conjointement. Il est intéressant de constater que le *feedback* sur l'oral représente des éléments de satisfaction supérieurs à celui de l'écrit. Dans notre plaidoyer, nous avons insisté sur la dimension empathique du *feedback*, notamment à travers le dialogue. Il semble que les étudiants aient été plus sensibles aux contenus des fichiers audio accompagnant le retour sur leurs performances oralisées. À travers une voix s'adressant à eux et leur parlant de manière conviviale, il ont sans doute mesuré le degré zéro de la distance de transaction, et donc apprécié cette instance d'individualisation conative moins impersonnelle que la trace écrite. On notera une quasi égalité chez les étudiants estampillés C1 entre un "oui" et un "assez", ce qui semble indiquer une variation de perception entre le *feedback* productif considéré comme indispensable, et le *feedback* en tant qu'élément parmi d'autres (notamment les ressources didactisées du dispositif).

Enfin, nous ne pouvons que nous féliciter de l'unanimité des réponses positives sur l'utilité du *feedback* même s'il ne s'agit que d'éléments de perception dont la subjectivité peut être un élément déformant. Nous y voyons néanmoins la transmission, chez l'étudiant, du sentiment que le travail à distance

ne s'arrête pas à l'utilisation de l'interface et que l' "après" est de nature autant, sinon plus, importante que la première phase d'apprentissage.

d. *Considérez-vous que les corrections étaient surchargées?*

	B1	B2	C1
- Oui.	15%	0%	0%
- Assez.	55%	0%	0%
- Peu.	15%	16.5%	11%
- Non.	15%	83.5%	89%

On voit ici une répartition logique entre les niveaux par rapport à la dimension quantitative du *feedback*. Le niveau B1 a fait l'objet d'un *feedback* adapté à des productions dont les éléments langagiers, méthodologiques et phonologiques nécessitaient une profusion de commentaires. Cela s'estompe pour les niveaux B2 et C1 pour lesquels le *feedback* n'a été qu'un accompagnement indicatif par petites touches.

e. *Auriez-vous désiré des critiques et des commentaires supplémentaires?*

	B1	B2	C1
- Oui.	50%	42%	44.5%
- Non.	50%	58%	45.5%

En relation avec la question précédente, ces réponses appellent un commentaire permettant de relever un paradoxe. Tous les niveaux se partagent entre un besoin de commentaires supplémentaires ou un statu quo par rapport à la situation vécue. Cela semble indiquer que parmi les 100% d'étudiants C1 estimant que les corrections n'étaient pas surchargées, la moitié d'entre eux (question 4) aurait voulu un *feedback* plus complet, et qu'une partie des 70% de B1 estimant les corrections surchargées en auraient voulu encore plus.

L'élément satisfaisant, en l'occurrence, est que l'ensemble des étudiants estime n'en avoir pas trop eu. Si cela avait été le cas, on aurait pu se poser la question de la correspondance des niveaux aux échelles du CECRL: une profusion de correction pour un C1 indiquerait que ce n'est pas un C1.

f. *Les critères d'évaluations des travaux écrits (grille) étaient-ils assez explicites?*

	B1	B2	C1
- Oui.	90%	83.5%	45.5%
- Non.	10%	16.5%	44.5%

g. *Les critères d'évaluations des travaux oraux (grille) étaient-ils assez explicites?*

	B1	B2	C1
- Oui.	80%	75%	61%
- Non.	20%	25%	39%

Comme pour les réponses aux questions 2 et 3, les réponses 6 et 7 méritent un traitement conjoint. On retrouve la préférence pour le feedback oral chez les étudiants niveau C1, alors que la tendance s'inverse légèrement pour les étudiants B1 et B2. La position des étudiants C1 est relativement étonnante puisque le CECRL, nous l'avons vu, ne donne pas de grille très précise concernant les éléments phonologiques et prosodiques. Peut-être que les C1 ont aussi pris en compte l'évaluation au sens large, c'est-à-dire les critères de correction du professeur sur les éléments de la chaîne parlée.

h. *À l'issue des corrections, vous sentez-vous capable de refaire vos productions en vraiment mieux?*

	B1	B2	C1
- Oui.	85%	71%	72%
- Non.	15%	29%	28%

La réponse quasi unanime et également répartie à cette question pourrait suffire à légitimer notre dispositif en tant qu'instrument préparatoire à un réinvestissement langagier à travers le développement de compétences langagières spécifiques. Pourtant, un petit tiers des B2-C1, et 15% des B1 n'envisagent pas les choses sous cet angle. On peut considérer le manque d'assurance de certains B1 comme élément de réponse. Mais en ce qui concerne les utilisateurs indépendants et expérimentés, cela pose un certain nombre de questions, notamment sur la nature adaptative de nos briques pédagogiques et sur la valeur du *feedback* proposé. Ces éléments montrent en outre un paradoxe avec l'ensemble des réponses aux questions 2 et 3. Il existe donc sans doute, pour certains, un vide transitionnel entre la reconnaissance de l'utilité du *feedback* (questions 2 et 3) et les conséquences opérationnelles attendues à partir de ces retours.

i. *Pensez-vous que l'ouverture d'un forum en anglais aurait apporté quelque chose à votre parcours dans le cadre de l'expérimentation Open Lads?*

	B1	B2	C1
- Oui.	59.5%	58.5%	44.5%
- Non.	40.5%	41.5%	45.5%

Nous avons envisagé, dans notre suivi, la possibilité d'utiliser un forum (expérience d'ailleurs engagée à la fin de notre recherche, sans grand succès)

afin de ne pas seulement circonscrire ce suivi à l'unique exploitation des navettes évaluatives. Les avis sont ici partagés et en légère contradiction avec la réalité de l'expérimentation menée, il est vrai l'année suivante avec d'autres étudiants, dans le cadre de notre dispositif remanié. Aux vues de la proportion entre présentiel et mode à distance (cf. II.4), nous pensons que l'utilisation du forum de discussion augmenterait en fonction de l'importance du nombre d'heures dévolues au travail à distance.

j. *Selon vous, l'évaluation aurait dû être accompagnée d'un barème comme pour un examen.*

	B1	B2	C1
- Oui.	20%	29%	44.5%
- Non.	80%	71%	45%

Il y a ici une différence claire entre les B1-B2 et les C1 au niveau de la perception de la pression institutionnelle. Les B1-B2 semblent, dans leur immense majorité, considérer le dispositif et le suivi comme des outils de formation et rejeter les barèmes d'évaluation dans le champ institutionnel et administratif de l'exigence d'un contrôle continu. Les C1 sont d'ailleurs partagés également sur ce point. On peut sans doute expliquer ce plus grand pourcentage par le fait que, à la question 1, 72.5% d'entre eux n'envisagent les descripteurs du CECRL que comme une balise de progression.

En ce qui concerne la place de l'enseignant(e) dans ce dispositif, vous l'avez plutôt trouvé(e) (barrer les mentions inutiles):

a. Proche des étudiants, à l'écoute	B1 90%	B2 91.5%	C1 83%
b. investi(e) dans votre parcours	B1 90%	B2 91.5%	C1 78%
c. assurant un bon suivi	B1 90%	B2 83%	C1 91.5%
d. surchargé(e) de travail	B1 50%	B2 29%	C1 44.5%
e. distant(e)	B1 0%	B2 4%	C1 5.5%
f. peu investi(e)	B1 0%	B2 0%	C1 0%
g. tantôt proche et tantôt distant(e)	B1 50%	B2 29%	C1 22%
h. n'assurant pas un bon suivi	B1 0%	B2 0%	C1 0%
i. ayant peu de travail	B1 0%	B2 0%	C1 0%
j. s'investissant assez peu pour vous	B1 0%	B2 0%	C1 0%

La dernière question porte sur la perception de la relation apprenant-tuteur, et montre d'étonnants points de convergence au sein de la répartition par niveaux de compétence. Un bon point est délivré au niveau de la proximité (a) et de l'écoute (présence sociale) de même que dans la perception du suivi et de l'investissement de l'enseignant-tuteur dans le parcours (b et c). Ces éléments sont corroborés par les réponses aux questions miroir (e, f, g, h, i et j) qui donnent un pourcentage très faible sur la distance de l'enseignant et son manque d'investissement. Les réponses à la question g sont intéressantes car elles

montrent une conscience de la distance interpersonnelle à la fois dans la FOAD mais aussi en présentiel. La moitié des étudiants B1 est concernée par cette impression, ce qui semble rendre compte d'un point de vue fréquentiel de la nature parfois distendue des relations, et d'un point de vue qualitatif de la qualité de la présence sociale du tuteur.

Un autre aspect intéressant est celui-ci: les enseignants n'apparaissent pas surchargés de travail, ce qui atteste que l'envers du décor, c'est-à-dire le nombre conséquent d'heures passées à gérer le *feedback* individualisé, sans compter les préparations spécifiques des cours en présentiel, n'est que peu perçu par les étudiants. L'investissement pédagogique permettant d'assurer un suivi individualisé de l'apprenant dans un dispositif hybride est pourtant relativement important et terriblement chronophage. Les étudiants semblent avoir intégré que cela n'a, en soi, rien d'exceptionnel d'un point de vue pédagogique. Mais du point de vue des moyens, cela implique une mise en œuvre sans commune mesure avec ceux dévolus à un simple enseignement en classe. L'université française devrait peut-être commencer à en tenir compte.

Conclusion

Les éléments recueillis après dépouillement des réponses des étudiants semblent indiquer que nous sommes sur la bonne voie. Ce "plaidoyer pour un *feedback* rationnel d'accompagnement conatif et évaluatif" tend à vouloir montrer, à travers la relation que nous établissons entre orientations théoriques et mise en œuvre pratique, que l'on peut obtenir des résultats satisfaisants en termes d'effets escomptés (motivation de l'apprenant) et de représentations (cf. résultats du questionnaire).

Nous restons conscient de la perfectibilité de cette approche, notamment au regard de la fiabilité de nos adaptations du CECRL et dans une gestion de l'hétérogénéité des publics qui pourrait être plus exhaustive, et nous n'avons d'ailleurs pas vocation à élaborer un "modèle" de suivi de l'apprenant en formation à distance. Notre contexte institutionnel a néanmoins eu une utilité certaine dans la mesure où la dimension "abandon" ou "décrochage" en cours de formation n'a jamais été de mise, les étudiants de Master étant suffisamment mûrs et motivés pour acquérir tous leurs crédits ECTS nécessaires à l'obtention de leur diplôme. Simplement, nous espérons que le partage de cette expérience menée à l'Espace Omega de l'Université Bordeaux 1 pourra inspirer d'autres développements fructueux.

Annexe 1

a. Principes de correction utilisant le module correctif de Word.

<p>Actually, I think private companies should not be a part of the University system. First, it contributes in making disparities between students, on one hand we can find rich students who can afford such expensive studies and on the other hand, [there is students from the working] class who cannot afford it.</p> <p>What's more, in my opinion, the standards of private education are worse than the public education. The fact that, private universities are run like businesses, therefore they have to keep surlignage pour la prononciation by doing their best to keep rich students into their institution. For example, bringing down the general difficulty for having a better success rate per year makes student's parents luring, to get their wallet out for their child. Moreover, most of the time, students who come from private high school have a better chance of texte remplacé en rouge going into a reputable firm instead of other students who has more deserved to get a good job.</p> <p>To my mind, the fact that these firms offer young students a vocational training does not help them because these texte barré (superflu) firms make these students what they need .As a result if these students were fired, they will not have the general training needed for X company.</p> <p>and Indians; they just have an idea of what is it! Instead of X pour omission people, who have their own computer since their childhood, young Indian can only access to computers by shared time when it is possible.</p>	<p>Correction nécessitant une mise au point grammaticale en français. Explications claires et détaillées.</p>
	<p>Mises au point sur le style, les erreurs de dictionnaire et les emplois abusifs non admis dans la langue.</p>

- ➔ Commentaire global sur la longueur, le niveau de langue, l'approche méthodologique et la structure.
- ➔ Conseils pour la seconde version. Encouragements.

b. Copie d'écran d'un exemple de texte corrigé.

Before beginning **To start with** I would like to **precise- delineate** the subject; because I think that this question is **nearly to be quite stupid, asked like this if asked like this, verges on stupidity**. I'm sure a good answer must *be based on take into account the* public's age even if it's neither mentioned in the video (~~but TV news reports are not known for showing comprehensive reports~~) nor in this question. I'm convinced that people agree that machines are not the best way ~~for handwriting to be learnt by~~ **to teach handwriting to pupils**, that's why this statement is a *rotten* (too rude. Use a dictionary and find a more suitable word for a written text) summary (done by a person with a quite poor argument, this explains that?). **produced by ... with quite a poor argument.**

Now, I'm going to explain **Here is** my point of view: machines can greatly improve the teaching content (so they must be used by schools or countries **who which** can offer ~~them~~ such an expensive material). Their purpose is not (for the moment) to replace teachers, so the "human interaction" is not affected; it's only a new item that can be used to access a more complete knowledge. No one **thinkS** that books affect the "human interaction", and a computer can be seen — (and used) — just like a "super" book : containing millions of pages, dealing with all imaginable subjects, moreover it can be *daily updated* **daily** and people can interact with it. In the report, machines are seen like this and not as a **teacher substitute substitute for teachers**, so the human interaction would *stay remain* the same as now: with our school bag full of heavy, incomplete and always out-of-date books.

As a conclusion, I would like to add that for students, even more for scientific **students**; it's a very important thing to have an easy access to all **this these** resources and for the teachers to use them to demonstrate. For them, it's the most useful item to illustrate their lessons; indeed, it's easier to show the Linux code on a screen than to write it (that could take long...) on a blackboard, for example.

In a nutshell, **let me point out** that machines are extremely useful and have an enormous potential; ; sorry for bad **philosopherS** and sociologists, who are **always** trying to find theories that would completely change the world, but the

Commentaire [E01] : News est invariable en anglais et donc toujours au singulier : ex. what is the news today. / the news was terrible. Dans votre texte pour exprimer un jugement d'ordre général, il vous faut une forme pluriel, il faut donc 'tricher' en associant **reports** à **news** pour produire un nom composé qui puisse prendre la forme pluriel.

Commentaire [E02] : C'est une habitude bien française que de penser son texte de phrases entre parenthèses. L'anglais ne le fait pas, ou beaucoup moins. N'en abusez donc pas. À la place, l'anglais utilise des virgules ou le trait long, comme ceci — ou cela

Commentaire [E03] : Fb : 'them' ne réfère à aucun nom pluriel cité juste avant. Il vaut mieux soit changer de verbe ... which can offer such expensive materials, soit ne pas donner de complément au vb offer : ... which can offer such expensive materials

Commentaire [E04] : voir E02. Dans ce cas où placer 'for the moment' ? Début de phrase ou après teachers?

Commentaire [E05] : Syntaxe. Le début de phrase 'it's a very important thing' est trop long pour garder le balancement. Il vaut mieux écrire 'and teachers can use them for demonstrations ...' ou 'as for teachers, they can use them ...'

Commentaire [E06] : ponctuation erronée. Il ne faut pas de virgule ici.

Annexe 2

Évaluation d'après les descripteurs du cadre: items à cocher ou non (colonne de droite) par l'enseignant selon les compétences développées.

Level B2 - Writing	
Can write a clear and detailed text (composition, report or text of a presentation) on various topics related to one's field of interest.	<input type="checkbox"/>
Can write summaries of articles on topics of specific interest.	<input type="checkbox"/>
Can summarise information from different sources and media.	<input type="checkbox"/>
Can discuss or debate a topic in a composition, giving reasons for or against a specific point of view and explaining the advantages and disadvantages of different alternatives.	<input type="checkbox"/>
Can systematically discuss a topic, emphasising decisive points and including supporting details.	<input type="checkbox"/>
Can express a viewpoint and can briefly list positive and negative points.	<input type="checkbox"/>
Can write a fairly long and easily comprehensible text.	<input type="checkbox"/>
Can make a statement in favour of or against a point of view.	<input type="checkbox"/>
Has a relatively large vocabulary.	<input type="checkbox"/>

If you have over 80% of the points ticked, you have probably reached Level B2.

Note: Possibilité de deux tableaux consécutifs pour rendre compte d'un niveau intermédiaire (B2+, C1+): les d'items du premier tableau sont tous cochés, et ceux du second, seulement partiellement.

Annexe 3

Évaluation de la performance oralisée: Grille d'évaluation selon le CECRL.



Oral expression: competences in phonology and phonetics.

		Level	A1 A2	B1	B2	C1	C2
A	Aptitude for delivering expressive phrases (fluency).		1	2	3	4	5
B	Word stress (authenticity, intelligibility).		1	2	3	4	5
C	Rhythm (speech delivery).		1	2	3	4	5
D	Intonation (fall-rise).		1	2	3	4	5

La grille est suivie d'un fichier son d'accompagnement: commentaire suivi de la performance orale, correction des éléments phonétiques et prosodiques, exemples de prononciation, exemples d'oralisation sur un paragraphe.

Annexe 4

① COMMUNICATION

parle à partir de (1) rien (0,5) notes (0) texte rédigé
autonomie (1) consulte ses notes (0,5) récite par cœur (0) lit son texte
regarde (1) le public (0) sa feuille (0) l'enseignant (0) l'horizon ou ailleurs
débit (1) naturel (0,5) un peu rapide / lent (0) beaucoup trop rapide / lent
ton (1) convaincant (0,5) neutre (0) monocorde
volume voix (1) clair (0,5) un peu trop bas (0) inaudible
attitude (1) naturelle (0,5) rigide (0) surexcitée

adaptation à l'auditoire (1) excellente (0,5) correcte (0) faible ou inexistante

utilisation de supports : lesquels

qualité du support : (2) complémentaire de l'exposé
(1) cohérent mais trop important par rapport à l'exposé
(0) surchargé / confus
(0) est le double de l'exposé

La partie "communication" englobe la performance orale (savoir être) et au niveau du support, un savoir-faire didactique (transposition, adaptation, intégration).

② ORGANISATION

introduction : (2) claire (1) peu claire (0) inexistante
plan : (2) clair (1) confus (0) inexistant
conclusion : (2) ouverte (1) expédiée (0) inexistante
transitions, enchaînements : (2) variés et judicieux
(1) banals
(0) inexistants

La partie "organisation" porte sur le savoir-faire performatif et vise à évaluer le lien entre méthodologie et performance.

③ COMPREHENSION DE L'ARTICLE, DU SUPPORT DE DEPART / DE LA PROBLEMATIQUE DU SUJET.

(3) excellente (2) bonne (1) correcte (0) faible

La troisième partie de la grille d'évaluation concerne le recul épistémique.

④ LANGUE

phonétique : (3) excellente (2) bonne (1) correcte (0) faible
accentuation : (3) excellente (2) bonne (1) correcte (0) faible
intonation : (3) excellente (2) bonne (1) correcte (0) faible
correction grammaticale : (3) excellente (2) bonne (1) correcte (0) faible
richesse lexicale : (3) excellente (2) bonne (1) correcte (0) faible

Pour la langue, on se réfère à la page 96 du CECRL et on évalue à la fois la gestion articulatoire du message et les savoirs et savoir-faire communicatifs.

⑤ **QUESTIONS / REPONSES**

compréhension des questions : (2) excellente (1) correcte (0) faible
réponses : (2) claires et pertinentes (1) brèves mais correctes (0) confuses
(0) superficielles (0) répète ce qui a été dit

La partie 5 vise la dimension interactionnelle.

⑥ **IMPRESSIONS GENERALES / REMARQUES / NOTES**

NOTE / 40 → NOTE / 20 Echelle niveau CECRL :

Et la dernière partie sert à opérer un bilan à la fois qualitatif et chiffré de la performance. Il semble important de préciser que la partie droite de la fiche sert à la prise de note de l'enseignant pendant l'exposé.

Bibliographie

ANDERSEN, JANIS. 1979. Teacher immediacy as a predictor of teaching effectiveness. *Communication Yearbook* 3. 543-559.

AUMONT, BERNADETTE & PIERRE-MARIE MESNIER. 1992. *L'acte d'apprendre*. Presses Universitaires de France.

BAKHTINE, MIKHAIL. 1977. *Le marxisme et la philosophie du langage*. Minuit.

BANDURA, ALBERT. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall.

BANGS, PAUL. 2002. Why is feedback dying of starvation? Let's try to revive it. *Eurocall 2002*. Finlande: Jyväskylä.

BENBUNAN-FICH, RAQUEL & STARR ROXANNE HILTZ. 1999. Impact of asynchronous learning networks on individual and group problem solving: a field experiment. *Group Decision and Negotiation* 8. 409-426.

BERNARD, MICHEL. 1999. *Penser la mise à distance en formation*. L'Harmattan.

BRUNER, JÉRÔME. 1982. The nature of adult-infant transaction. VON CRANACH MARIO & HARRE ROM. *The analysis of Action*. Cambridge University Press.

COLLINS, MAURI & ZANE BERGE. 1996. Facilitating interaction in computer-mediated online courses. <http://www.emoderators.com/moderators/flcc.html>.

- CONSEIL DE L'EUROPE. 2001. *Cadre commun de référence pour les langues. Apprendre, Enseigner, Évaluer*. Didier.
- DEBRAY, RÉGIS. 2000. *Introduction à la médiologie*. Presses Universitaires de France.
- DUCROT, OSWALD. 1980. *Les échelles argumentatives*. Minuit.
- ELLIS, KATHLEEN. 2000. Perceived teacher confirmation: the development and validation of an instrument and two studies of the relationship to cognitive and effective learning". *Human Communication Research*, 26 : 2. 264-291.
- FELIX, USCHI. 2002. Teaching language online. Deconstructing myths. *Eurocall 2002*. Finlande: Jyväskylä.
- GAONAC'H, DANIEL. 1991. *Théories d'apprentissage et acquisition d'une langue étrangère*. Didier.
- GASS, SUSAN. 1997. *Input, Interaction and the Second Language Learner*. Lawrence Erlbaum Associates.
- HALL, EDWARD. 1971. *La Dimension Cachée*. Le Seuil.
- JEANNERET, THÉRÈSE. 1999. *La coénonciation en français. Approches discursive, conversationnelle et syntaxique*. Peter Lang, coll. Sciences pour la communication.
- LONG, MICHAEL. 1996. The role of linguistic environment in second language acquisition. RITCHIE, WILLIAM & TEJ BHATIA (dir.) *Handbook of Second Language Acquisition*. Academic press. 413-468.
- MOORE, MICHAEL. 1993. Theory of transactional distance. KEEGAN, DESMOND (dir.) *Theoretical principles of distance education*. Routledge. 22-38.
- NAGATA, NORIKO. 1997. An experimental comparison of deductive and inductive feedback generated by a simple parser. *System*, 25 : 4. 515-534.
- PAQUELIN, DIDIER. 2000. Quelques règles ergonomiques pour l'analyse et la conception d'applications multimédias interactives. <http://robin.blondet.free.fr/ipm/reglesconception.html>.
- PAQUELIN, DIDIER. 2005. Planification versus potentialisation. De la structuration des contenus à la structuration de la contenance. Colloque SIF. <http://sif2005.mshparisnord.org/pdf/Pauelin.pdf>
- PETIT, MICHEL. 2007. Les descripteurs du Cadre: quelle conception de la langue de spécialité? HARAMBOURE, FRANÇOISE, MICHEL PETIT & CHRISTOPHE CHAMBOST (dir.). *Enseigner les langues vivantes: enjeux contemporains*. Bordeaux, IUFM d'Aquitaine / Université Bordeaux 2. 14-29.
- PROULX, SERGE. 2002. Trajectoires d'usages des technologies de communication: les formes d'appropriation d'une culture numérique comme

enjeu d'une société du savoir. *Annales des télécommunications* 57 : 3-4. 180-189.

PUREN, CHRISTIAN. 2006. Le Cadre Européen Commun de Référence et la réflexion méthodologique en didactique des langues-cultures: un chantier à reprendre. *Synergies Europe* 1. 74-80.

RUSSO, TRACY & SCOTT CAMPBELL. 2004. Perceptions of mediated presence in an asynchronous online course: interplay of communication behaviors and medium. *Distance Education* 25 : 2. 215-232.

SHIN, NAMIN. 2003. Transactional presence as a critical predictor of success in distance learning. *Distance Education* 24 : 1. 69-86.

SCHMIDT, RICHARD. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11 : 2. 129-158.

SHORT, JOHN, EDERYN WILLIAMS & BRUCE CHRISTIE. 1976. *The Social Psychology of Telecommunications*. Wiley.

WALTHER, JOSEPH. 1992. Interpersonal effects in computer-mediated interaction: a relational perspective. *Communication Research* 19. 52-90.

WEASENFORTH, DONALD, SIGRUN BIESENBACH-LUCAS & CHRISTINE MELONI. 2002. Realizing constructivist objectives through collaborative technologies: threaded discussions. *Language Learning and Technology* 6 : 3. 58-86.

WENDT, MICHAEL. 2000. Stratégies de l'apprenant et stratégies d'apprentissage: recherche au service des chercheurs. HOLTZER, GISELE & MICHAEL WENDT (dir.), *Stratégie de l'apprenant: une notion en mouvement*. Université de Franche-Comté. 111-126

YANG, YU-FENG. 2005. Situated learning as a framework for CALL research. EGBERT, JOY & GINA MIKEL PETRIE (dir.) *CALL Research Perspectives*. Lawrence Erlbaum Associates. 155-168.

Conception d'un outil expérimental de suivi apprenant en langues. Le cas de "SUIVIX"

Alain Cazade, Professeur, Université Paris Dauphine

Ce travail entrepris sur le thème du suivi apprenant est parti d'un constat, celui que les produits disponibles sur le marché permettant de suivre le parcours de mes apprenants en langues, en présentiel ou à distance, ne me donnaient que des éléments trop partiels et surtout insuffisamment parlants pour que je puisse rapidement savoir s'il me fallait intervenir pour les aider, les soutenir, pour éviter qu'ils continuent dans une voie qui semblait peu productive, pour rectifier une difficulté, pour les accompagner de manière efficace.

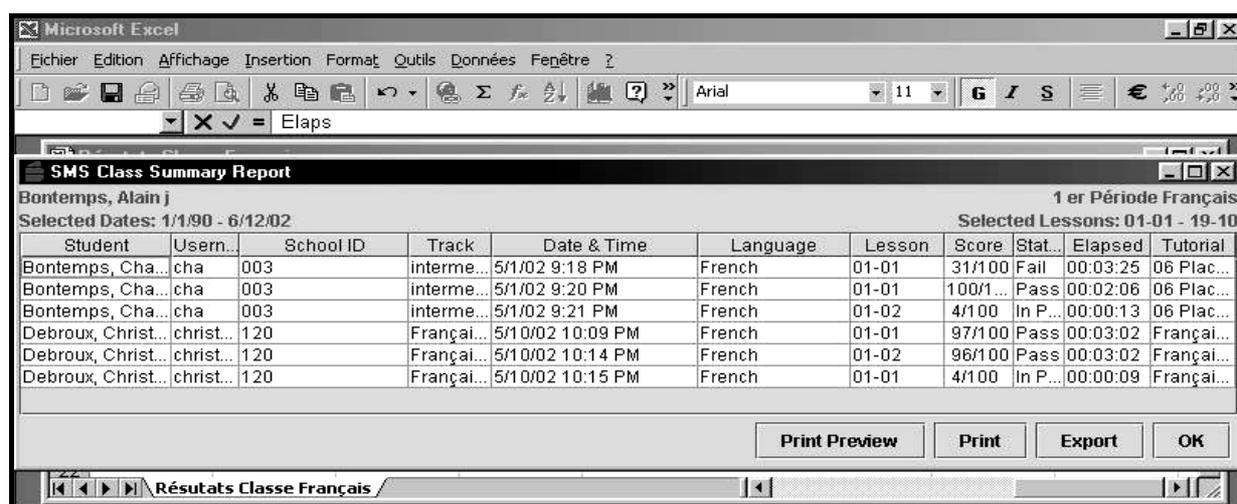
I. Considérations générales

Quelques exemples de suivis logiciels

En examinant les outils disponibles, on constate que les items choisis pour rendre compte de l'activité apprenante sont trop axés sur le temps passé (durée d'activité), sur les éléments travaillés, mais pas assez sur la progression, sur le cheminement, sur les erreurs retravaillées etc.

Je ne mentionnerai ici que quelques exemples sommaires, à titre purement indicatif. Mon but n'est naturellement pas de proposer un état des lieux exhaustif, une revue détaillée de l'existant. Il faudrait, pour cela, aborder autant les logiciels ciblant telle ou telle activité pédagogique que les logiciels qu'on pourrait appeler "intégrateurs", de type plateforme d'enseignement à distance.

Un outil très présent sur le marché, *Rosetta Stone* (Fairfield), propose un service de suivi apprenant dont l'écran suivant donne une petite idée.



The screenshot shows a Microsoft Excel spreadsheet titled "SMS Class Summary Report". The report is for "Bontemps, Alain j" and "1 er Période Français". It covers "Selected Dates: 1/1/90 - 6/12/02" and "Selected Lessons: 01-01 - 19-10". The data is presented in a table with the following columns: Student, Usern..., School ID, Track, Date & Time, Language, Lesson, Score, Stat..., Elapsed, and Tutorial. The table contains six rows of data.

Student	Usern...	School ID	Track	Date & Time	Language	Lesson	Score	Stat...	Elapsed	Tutorial
Bontemps, Cha...	cha	003	interme...	5/1/02 9:18 PM	French	01-01	31/100	Fail	00:03:25	06 Plac...
Bontemps, Cha...	cha	003	interme...	5/1/02 9:20 PM	French	01-01	100/1...	Pass	00:02:06	06 Plac...
Bontemps, Cha...	cha	003	interme...	5/1/02 9:21 PM	French	01-02	4/100	In P...	00:00:13	06 Plac...
Debroux, Christ...	christ...	120	Françai...	5/10/02 10:09 PM	French	01-01	97/100	Pass	00:03:02	Françai...
Debroux, Christ...	christ...	120	Françai...	5/10/02 10:14 PM	French	01-02	96/100	Pass	00:03:02	Françai...
Debroux, Christ...	christ...	120	Françai...	5/10/02 10:15 PM	French	01-01	4/100	In P...	00:00:09	Françai...

Figure 3 – Suivi de Rosetta Stone

On constate que sont essentiellement mis en valeur dans cet affichage les temps de travail et les scores correspondant à chaque utilisateur.

Les quelques écrans suivants, tirés d'une présentation de *Tell Me More* (Auralog, version 2001), montrent que, malgré la volonté louable de détailler le travail de l'apprenant suivant un certain nombre d'axes d'approches qui se veulent complémentaires, l'accent est davantage mis sur le nombre des items travaillés, des données temporelles, en résumé: sur des données qui tournent plus autour du quantitatif que du qualitatif. La présentation d'un certain nombre de pourcentages de réussite mériterait d'être affinée.

Suivi de l'élève : M... B

Parcours : **Vocabulary (UK) 3** Date : **Cumul**

Table de progression | Synthèse

Cours : **Abordés** Activités : **Abordées**

	Fait	Correct	Score	Durée
A car trip				
Association image/mot	100%	87%	7.00	00:01:17
Association de mots	100%	100%	2.00	00:01:05
On the motorway				
Mots croisés : traduction	20%	0%		00:03:36
At the airport				
Rédaction	100%	0%	0.59	00:06:07

Figure 4 – Suivi de *Tell Me More* – Pourcentages

7

Enregistrements sonores de l'élève : M... B...

Enregistrement	Date	Note
Anglais britannique		
Arriving at the villa		
Prononciation de phrases		
Where's the bath plug?	29/01/01 09:36:39	3
Where's the bath plug?	29/01/01 09:36:44	5
What have you done?	29/01/01 09:36:50	1
There we go. The beginning of our holiday...!	29/01/01 09:37:01	5
I'll fix it.	29/01/01 09:37:06	5
Yes, it does look funny!	29/01/01 09:37:13	3
Prononciation de mots		
always	29/01/01 09:38:46	3
amazing	29/01/01 09:38:52	2
broomstick	29/01/01 09:40:39	1
broomstick	29/01/01 09:40:23	2
draining	29/01/01 09:41:12	4
dresser	29/01/01 09:41:38	3

Figure 5 - Suivi de *Tell Me More* - Détail des enregistrements sonores

L'intérêt de la figure 3 est de montrer qu'on a la possibilité de vérifier la production orale de l'apprenant suivi. Néanmoins, étant donné qu'on doit assurer le suivi de nombreux étudiants et qu'on est de ce fait amené à naviguer parmi de nombreux cas de figures, phrases à prononcer diverses, contextes de production variables, il serait nécessaire, pour pouvoir valablement apprécier ce type de production orale, de replacer l'audition de l'item testé dans son contexte complet afin de vérifier: si l'ensemble des paramètres accompagnant la production orale attendue ont bien été pris en compte dans la production obtenue, si le ton de la réponse à un stimulus s'accorde bien au ton du stimulus lui-même, si les aspects visuels, écrits et autres sont correctement intégrés à cette production.

The screenshot shows a web browser window titled 'TutorTools Exportation - Microsoft Internet Explorer'. The address bar shows 'C:\Mes Documents\Résultats.htm'. The main content area displays the following information:

Elève : M... B...
Parcours : Vocabulary (UK) 3
Date : Cumul

Cours	Activité	Total	Fait	Correct	Score	Durée
A car trip	Association image/mot	8	8	7	7.00	00:01:17
A car trip	Association de mots	2	2	2	2.00	00:01:05
On the motorway	Mots croisés : traduction	10	2	0	0.00	00:03:36
At the airport	Dialogue : découverte	16	0	0	0.00	00:00:00

The status bar at the bottom indicates 'Terminé' and 'Poste de travail'.

Figure 6 – Suivi de *Tell Me More* - Mode cumulatif

Il est intéressant de noter que le logiciel propose pour chaque apprenant une approche double, par le détail (figures 2, 3, 5) mais aussi en mode cumulatif (figure 4).

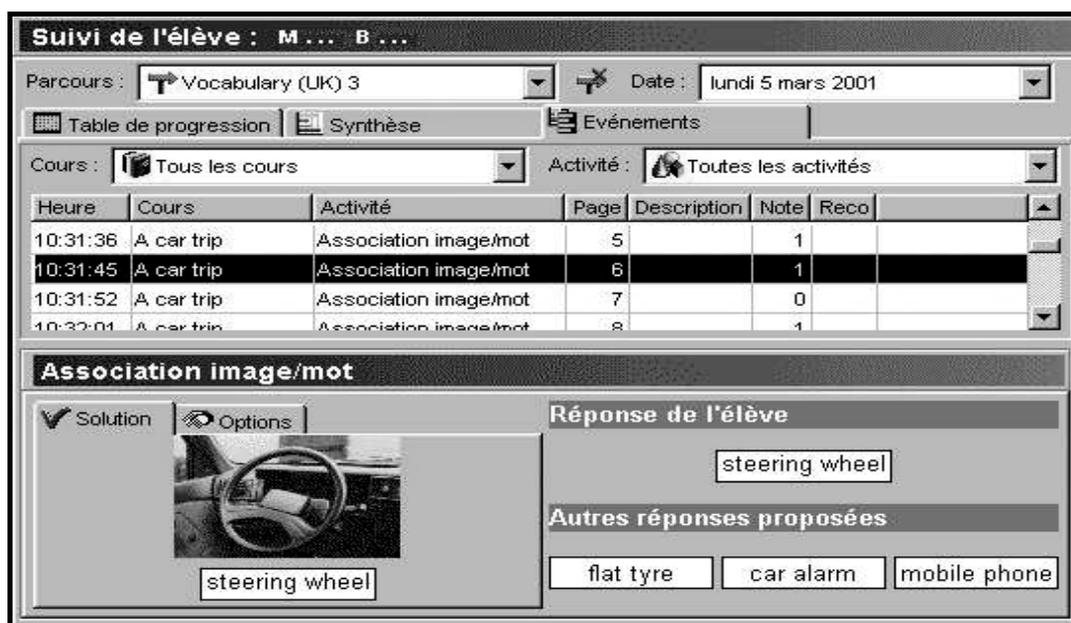


Figure 7 – Suivi de *Tell Me More* - Contextualisation d'un item

Dans la figure 5, on note la volonté appréciable de préciser la nature d'un item testé dans un QCM visuel par une image et par le rappel des "distracteurs" (ou leurres) proposés. Une notation temporelle (l'heure à laquelle l'élément a été travaillé) permet de re-situer sommairement l'item dans un parcours. Néanmoins, on comprend vite que, pour l'enseignant-tuteur qui prendra connaissance de cet écran, il sera assez difficile de juger du contexte de progression dans lequel la réponse proposée a été produite. Or le recours à ce type de contexte, avec les hésitations et les tâtonnements que cela inclut, est indispensable pour avoir les moyens d'apprécier correctement la dynamique dans laquelle la démarche apprenante se développe. Il est clair que l'élément décrit ici – même s'il est plus richement présenté que dans d'autres produits, il faut le noter au passage – reste trop parcellairement appréciable.

On pourrait de la même manière commenter quelques écrans auxquels le produit "WebCT" (Blackboard) donne accès¹. Sans pouvoir les analyser en détail, comme ils le mériteraient pourtant, on voit que, comme avec *Tell Me More*, en cochant telle ou telle case à l'intérieur de sub-divisions accessibles par un jeu d'onglets, il est possible de passer d'une vision globale à une vision détaillée. Malgré les réels efforts de lisibilité² dans la présentation des divers éléments sélectionnés pour rendre compte du travail de chacun, la multiplicité des items choisis, pourtant bien nécessaire pour que le suivi soit suffisamment fin, ne pourra qu'être plus ou moins indigeste pour l'enseignant qui doit au bout du compte suivre ainsi un grand nombre d'apprenants.

¹ Écrans de la version WebCT-Vista, proposés durant notre séminaire du 20 octobre 2006, sur le suivi apprenant par Nicolas Klotz, Société BlackBoard – WebCT.

² Une lisibilité de bonne qualité, dont les saisies d'écran reproduites dans cet article ne rendent pas suffisamment compte pour des raisons d'économie d'espace.

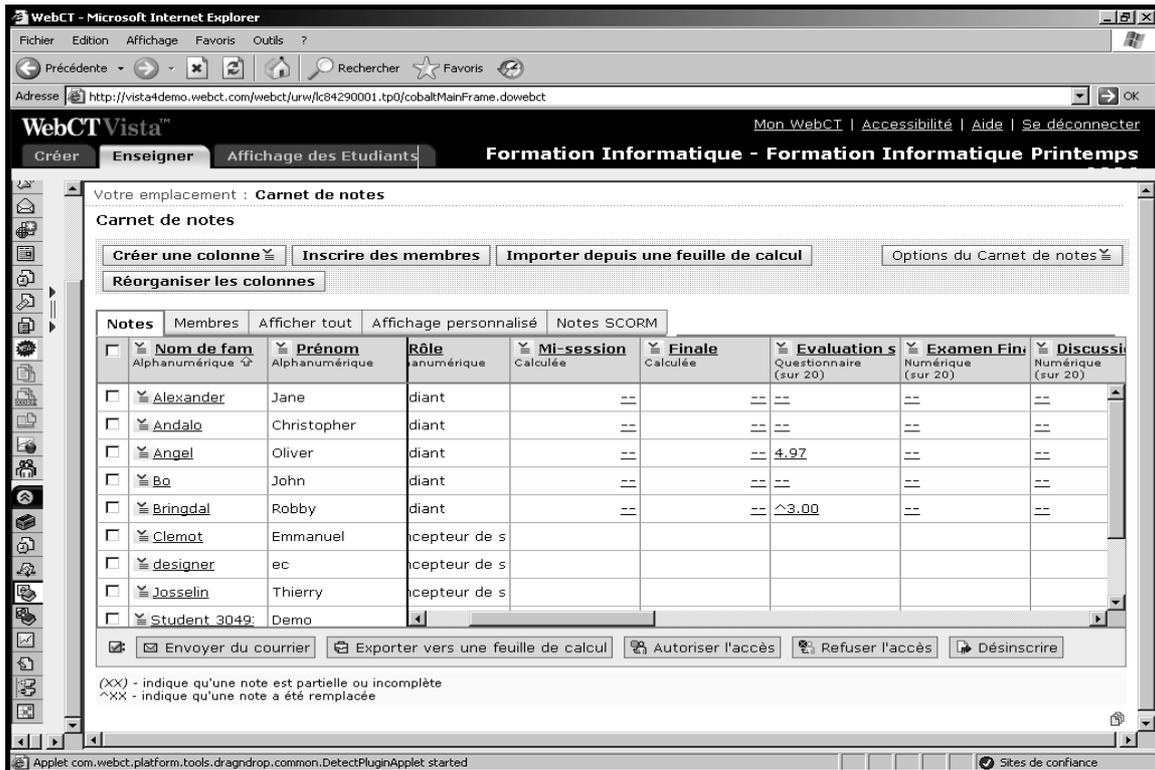


Figure 8 - WebCT - Suivi global des notes de l'étudiant

La superposition visuelle de divers écrans permet à la fois de garder assez facilement en tête la nature de sa position dans l'arborescence d'analyse offerte par le produit, mais oblige également, dans cet outil comme dans d'autres, à savoir jongler parmi tous ces affichages cumulatifs ou/et successifs, en gardant bien en tête le fil conducteur de l'approche diagnostique qu'on voulait privilégier en commençant à fouiller dans les nombreuses données accessibles.

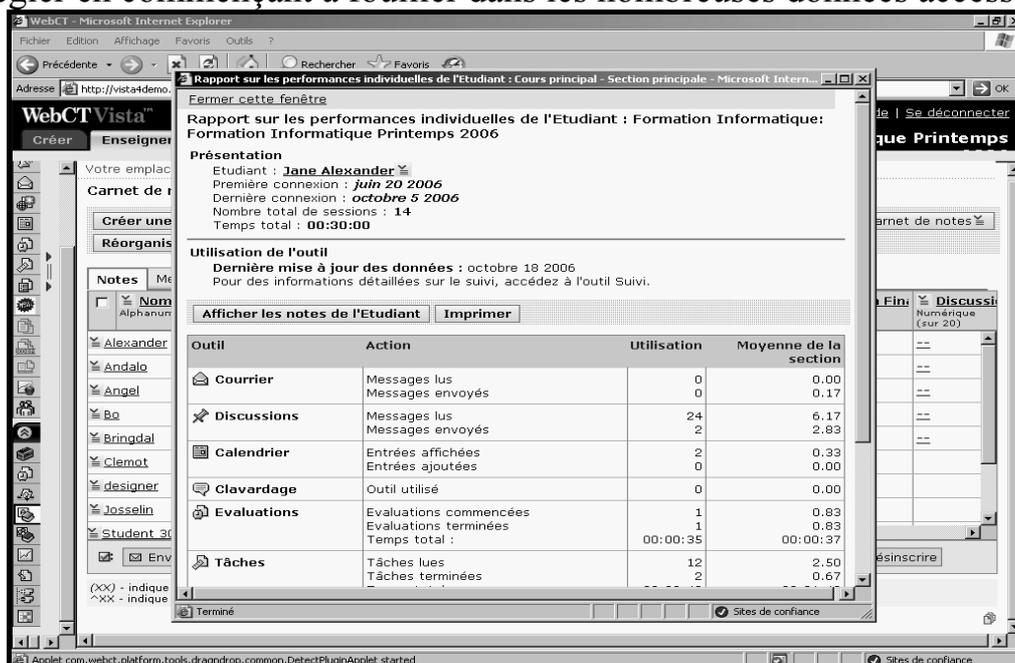


Figure 9 - WebCT: Suivi individuel détaillé de l'étudiant

La “quadrature du cercle” du suivi pédagogique: précision, flexibilité, souplesse et lisibilité

Ces quelques exemples montrent bien le problème quasiment insoluble que pose la mise en place d'un suivi apprenant digne de ce nom, tout spécialement en langues.

Précision

Du point de vue de l'enseignant, il faut donner la possibilité de retrouver, de fouiller, d'analyser au plus près le détail d'un parcours, lequel est caractérisé par un ensemble d'actions diverses (actèmes), de productions (productèmes), mais aussi de “silences comportementaux”, dont on doit pouvoir décider de façon justifiée si ce sont des périodes de réflexion, d'inaction après une période de saturation, d'indifférence, etc. Le contexte de scrutation d'un environnement de suivi devrait pouvoir faire apparaître un certain nombre d'indices permettant d'opter dans un sens ou un autre. Or, ce type de choix d'interprétation du comportement apprenant est crucial pour savoir si on doit intervenir pour aider, ou au contraire laisser à l'apprenant le temps de murir sa propre réflexion, sans l'interrompre, afin qu'il puisse aboutir à une décision dont il pourra revendiquer, consciemment ou non, la paternité, la responsabilité, elle-même synonyme d'appropriation, d'acquisition. On conçoit plus facilement qu'il sera possible d'interpréter un ensemble d'actes, leur fréquence, leurs cibles et supports de réalisations. En revanche, il est toujours plus difficile d'interpréter le vide, l'absence de signe. Pourtant, à titre d'exemple: un “silence comportemental” encadré en amont et en aval par une période d'intense activité pourrait suggérer la réflexion ou le repos avant un nouvel effort; l'abandon d'une activité en cours, en plein milieu de celle-ci, pour faire tout autre chose, pour aborder un autre module, pourrait suggérer la saturation, etc.

Flexibilité

Le but est de pouvoir affiner l'accompagnement – plutôt formatif donc – du travail effectué, en suivant une combinaison associée de paramètres, afin d'éclairer telle ou telle activité ou compétence. Le but est également de rendre possible une appréciation globale, autant en termes formatifs que sommatifs. Cela veut dire qu'un choix élargi de telles associations de paramètres, de mises en relation de divers facteurs d'appréciation, aisément permutable, doit être rendu accessible à l'enseignant, afin de pouvoir jouer de l'une ou/et de l'autre successivement.

Souplesse

Il est clair que le risque de se noyer parmi tous ces paramètres et combinaisons d'appréciation est grand et menace d'aboutir au résultat inverse de ce qui était espéré, par découragement ou encore par peur de ne pas savoir tirer profit des fonctionnalités proposées. Une certaine souplesse – synonyme ici de

facilité et de rapidité de passage d'un mode de lecture à un autre – doit caractériser un environnement de suivi apprenant réellement utilisable dans la pratique et non pas réservé à un seul noyau de spécialistes avertis. Quelques choix de combinaisons pré-organisées, dont l'expérience aura montré qu'elles sont associées à un nombre majoritairement reconnaissable de profils, devrait pouvoir être proposé et aisément accessible. Cette souplesse doit être de plus en plus présente à mesure que l'utilisateur est moins spécialiste: chercheur en didactique, enseignant, tuteur ou simple accompagnateur-moniteur.

Lisibilité

À tous les niveaux, aussi bien pour celui qui se déclare spécialiste que pour celui qui ne peut, ou ne veut pas, prétendre à s'investir trop profondément dans l'analyse des mécanismes dont il est question ici, il faut pouvoir viser la lisibilité des résultats, afin de pouvoir envisager avec rapidité et facilité une sorte de diagnostic sur le parcours examiné et la décision à prendre en conséquence. Là encore, facilité et rapidité, effets de zoom avant et arrière, représentations graphiques synthétiques, ne peuvent qu'aider à la lisibilité, elle-même devant être accompagnée, pour les mêmes raisons que celles qui ont été évoquées à propos de la souplesse, d'un certain nombre de modes pré-réglés de lecture, pour aider à interpréter les cas de figure rencontrés.

Ce qui précède vaut essentiellement pour le chercheur en didactique, l'enseignant, l'accompagnateur, mais il ne faut pas oublier qu'une facette importante d'un environnement de suivi apprenant doit également viser l'utilisation de tout ou partie de ses fonctionnalités par l'apprenant lui-même, afin qu'il puisse prendre conscience de sa progression, de tendances favorables ou non dans son évolution, ceci devant lui permettre de faire divers bilans de son travail et, partant, de mieux orienter la suite de son parcours. Ce type de décision est de nature responsabilisante, et est la plupart du temps considérée comme pouvant favoriser ses processus d'acquisition. On sait l'importance de l'auto-évaluation dans un parcours d'autonomie guidée.

II. SUIVIX: choix, caractéristiques, développement

Les lignes qui suivent ne prétendent évidemment pas présenter un outil capable de répondre à toutes les exigences envisagées dans cet article et ailleurs en matière de suivi apprenant. L'outil que je développe progressivement, SUIVIX, que j'ai déjà eu l'occasion de présenter durant le 44^{ème} congrès de la SAES en mai 2004, à l'université de Versailles-Saint-Quentin, mais qui continue d'évoluer à l'usage, ne visait au départ qu'à essayer de donner réponse à un certain nombre de questions concrètes que je me posais afin de suivre le travail

de mes étudiants lorsqu'ils travaillent avec un logiciel: HELP YOURSELF³, que j'ai conçu et réalisé il y a quelques années maintenant et qui continue, lui aussi, à évoluer. Au fil des semaines et d'interrogations nouvelles, SUIVIX se voit enrichi de nouvelles possibilités pour rendre compte plus finement de l'avancée du travail de nos étudiants. Il s'agit, pour aller vite et en résumant, d'une part de permettre de savoir où en est leur travail, que ce soit en présentiel, dans une salle de langues ou bien à distance (physique ou/et temporelle), mais il s'agit aussi, d'autre part, de leur fournir des éléments pour qu'eux-mêmes puissent, au moins partiellement, s'auto-évaluer.

Confiance dans le système et maintien de la motivation apprenante

La confiance dans la fiabilité d'un système d'apprentissage est un élément majeur dans le processus de maintien d'un certain niveau de motivation et donc de persévérance dans le travail chez l'apprenant. De fait, à l'usage, un des aspects de cet outil qui s'avère important – peut-être même primordial, dans mon cadre de travail – est de pouvoir renforcer ce sentiment de confiance grâce à l'outil de suivi mis en place. Pour illustrer rapidement ce point, je préciserai que cet outil me permet notamment de couper court à diverses tentatives de verbalisation insuffisamment claires de leur part pour m'expliquer une difficulté rencontrée, les conditions de son apparition et leur impossibilité d'y répondre valablement. Il me permet aussi à l'occasion de répondre de manière objective, factuelle, à certaines incompréhensions concernant les refus éventuels par le système de telle ou telle réponse qu'ils ont proposée. Si on peut montrer à ses étudiants l'enchaînement de leurs tentatives successives, mettre le doigt sur les raisons du refus de telle ou telle production écrite par le système d'évaluation, faire apparaître devant leurs yeux qu'au cours de leurs tâtonnements ils ont proposé divers items réponses qui étaient intéressants et qu'il faudrait intégrer aux futures analyses de réponses permises par le module⁴, si l'outil de suivi permet tout cela, alors un lien de confiance dans l'environnement d'apprentissage proposé ne peut qu'en sortir renforcé. L'aspect "confiance" se révèle primordial, il valorise autant l'apprenant que l'enseignant dans le processus co-constructif en marche et conforte la démarche didactique dans son ensemble.

Un outil bâti à partir de HELP YOURSELF

Pour pouvoir bénéficier de ces ouvertures, SUIVIX s'articule sur le système de fabrication automatique de fichiers traces prévu par HELP YOURSELF. Il m'est possible, dans un environnement bâti autour de macro-

³ HELP YOURSELF, écrit en langage Logiix (*Infoaccess*), outil auteur gérant les piles hypertextuelles internes et externes, les fichiers multimédia sons, vidéos, etc.

⁴ HELP YOURSELF intègre en effet diverses possibilités d'analyse de réponse, en plus d'autres fonctionnalités de choix de niveau de difficulté, de possibilités d'aide, de "degré de ludicité", etc.

procédures de consultation et de sélections imbriquées, de suivre au centième de seconde près les diverses activités des apprenants, toutes les étapes successives de leurs parcours, l'enchaînement des écrans consultés, les demandes d'aides, lexicales et autres, les appels à d'éventuels sites Internet proposés pour compléter les ressources informatives intégrées à chaque module, l'activation de tous les "boutons" ou liens divers faisant exécuter les fichiers multimédias proposés, la progression des scores, le temps passé sur chaque item travaillé avant de passer au suivant (et donc l'évolution du taux d'activité de chacun), les temps de réflexion au moment d'entrer une réponse dans un questionnaire donné, mais aussi entre les actions, etc.

Il se trouve qu'il m'est toujours possible, même en 2008, d'utiliser HELP YOURSELF, malgré les évolutions de l'environnement informatique en général et de Windows en particulier, depuis ses premières versions dans la fin des années 90. Nombre de ses fonctionnalités (multimédias notamment, mais aussi tout ce qui concerne la gestion du temps, car les processeurs sont de plus en plus rapides), prenant appui sur le cœur même du système opérationnel (OS) de Microsoft, les fonctionnalités du logiciel ont pu bénéficier, à chaque nouvelle mouture, à la fois de ce qu'on appelle la "compatibilité ascendante" (i.e. la nécessité pour un système de permettre à d'anciens produits de continuer à tourner), mais aussi des multiples progrès et enrichissements de l'OS.

Quelques principes de base

En écrivant cet article, je constate une fois de plus, étant souvent amené à communiquer sur des travaux en recherche-développement et la mise au point d'outils multimédias, qu'il est vain d'espérer rendre compte avec clarté et exhaustivité du fonctionnement d'un outil, ou même des raisons précises pour lesquelles telle ou telle fonction a été introduite dans un module. Dans un environnement multimédia réellement appuyé sur le traitement dynamique de données et sur un enrichissement graphique qui donne sens à aux traces forcément statiques des mouvements progressifs d'un parcours, seule la manipulation en présentiel des objets (ou au minimum leur démonstration) permet de se faire une idée de ce dont on parle. Essayons pourtant d'aborder quelques principes et fonctionnalités de base de l'outil dont il est question.

Deux directions majeures permettent de viser la nature et la qualité d'un travail derrière des données factuelles et quantitatives: une approche liée à la matière des données et celle liée à la manière dont celles-ci sont utilisées, même si ces éléments sont intimement liés et essentiellement inter-significatifs.

Matière

- *nature des items rencontrés*, des ressources utilisées, choix de modules, aides diverses, contacts externes, sites internet, vidéos, sons, etc.
 - permet l'analyse des choix de parcours opérés, des récurrences ou absences significatives d'items, de la navigation parmi les modules et items proposés, etc.

- *nature des items produits* (écrites, orales, interactions, etc.)
 - permet l'analyse des contenus produits, de leur nombre, dans un contexte donné, selon un rythme de progression, des scores éventuellement obtenus, du nombre d'essais, etc.

Manière

- *durées*
 - de connexion, de sessions de travail, de passations lors de tests, etc.
 - données assez faciles à maîtriser – mais données le plus souvent insuffisamment traitées,
- *fréquence*
 - calculée à partir des retours sur des catégories d'items données; par ex. quantités et durées de phases d'écoute plus que de production, attentes entre actions successives, attente avant de produire une "réponse" à une question donnée, etc.
 - assez facile à maîtriser - mais données le plus souvent insuffisamment traitées,
- *intensité*
 - calculée dans un laps de temps donné à partir du nombre d'items abordés et d'actions effectuées; par ex. ratio du nb de questions abordées par rapport au nb total d'items d'un parcours, insistance sur une question donnée ou au contraire abandon rapide après peu d'essais infructueux, taux de réussite et taux d'échec dans un éventail donné d'actions attendues, etc.
 - approche assez délicate à cerner, à définir
 - très rarement - sinon jamais - prise en compte habituellement
 - données à traiter très nombreuses pourtant!

Les problèmes essentiels consistent à choisir parmi les diverses données de traçage celles qui sont les plus significatives, et surtout comment les combiner entre elles pour qu'elles "parlent" davantage ou/et plus clairement.

Présentation de l'outil et de quelques-unes de ses fonctionnalités

Le simple affichage séquentiel des données concernant les angles d'approche mentionnés ci-dessus – données parfois sur des milliers de lignes, et surtout de nature disparates, et dont les deux tableaux suivants donnent une petite idée – pose un véritable défi d'interprétation à l'enseignant-tuteur-accompagnateur, même en phase de travail asynchrone. Pour faciliter la lecture, ne figurent pas dans les tableaux suivants de nombreux types de données complémentaires.

G r	Nom	Module [écran] – bouton	Date	Item question	Signe	Masque après	Essais	Score	Réponse	B o n u s	Temps interm.	Temps réponse
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:48:57	would have to buy	••	— ha — — —	1 10	105 105	had bought	1	0:00:04	00:23,8
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:01	would have to buy	•-	— ha — — —	2 10	105 105	s	1	0:00:05	00:02,5
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:06	would have to buy	••	— d ha — — —	3 10	105 105	d	1	0:00:22	00:04,3
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:28	would have to buy	••	would have — —	4 10	105 105	would have bought	1	0:00:08	00:19,6
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:36	would have to buy	••	would have — b—	5 10	105 105	bought	1	0:00:11	00:06,3
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:47	would have to buy	••	would have — buy	6 10	105 105	buy	1	0:00:10	00:10,8
L 1	Benoît	UNREALTE [3] -61069	Jan 22 08 15:49:57	would have to buy	+	would have to buy	7 10	110 110	to	1	0:00:11	00:09,5

Tableau 3 - Séquence de travail sur la concordance des temps

L'enchaînement ci-dessus montre une courte succession de réponses, travaillant en mode de restitution de réponse (••: de manière fructueuse ou •-: infructueuse), qui progressent petit à petit, 7 essais sur 10 prévus, sans appel de dictionnaire, d'aide grammaticale ou autre mouvement intermédiaire, vers la réponse attendue (+), l'attente entre chaque tentative variant de 4 à 22 secondes, et le temps de réflexion avant de produire chaque réponse posée variant lui de 2,5 secondes à 23,8 secondes.

Le tableau suivant montre une autre courte séquence portant sur un travail de reconnaissance de phonèmes (cf. signes dans la colonne "Réponse": + correctement ou – incorrectement reconnus).

Gr	Nom	Module [écran] – bouton	Date	Écran / Item	Pt scruté	Réponse	Essais	Score	Bonus	Tps interm.
M1	François	Phonetic [17]	Apr 12 07 11:41:16	=>écran n°17	Ex.3bis: mixed pairs					0:00:34
M1	François	Phonetic [17] msq qc9798	Apr 12 07 11:41:50		±2 4	əʊ-D-D+əʊ+	1 8	215 225	0	0:00:45
M1	François	Phonetic [17] msq qc9798	Apr 12 07 11:42:35		±3 4	ɔ:+ɔ:-D+əʊ+	2 8	215 225	0	0:00:40
M1	François	Phonetic [17] msq qc9798	Apr 12 07 11:43:15		±3 4	ɔ:+aʊ-D+əʊ+	3 8	215 225	0	0:00:24
M1	François	Phonetic [17] msq qc9798	Apr 12 07 11:43:39		±3 4	ɔ:+D-D+əʊ+	4 8	215 225	0	0:00:04
M1	François	Phonetic [16]	Apr 12 07 11:43:43	=> écran n°16	Ex.3: min. pairs 2					0:00:06
M1	François	Phonetic [17]	Apr 12 07 11:43:49	=> écran n°17	Ex. 3bis: mixed pairs					0:00:10
M1	François	Phonetic [17] msq qc9798	Apr 12 07 11:43:59		1	ɔ:+əʊ+D+əʊ+	5 8	220 230	0	0:00:32
M1	François	Phonetic [17] snd	Apr 12 07 11:44:31	Voy.wav	ɑ:	<u>Snd 7064</u>				0:00:29
M1	François	Phonetic [17] msq qc9890	Apr 12 07 11:45:00		1	əʊ+aʊ+D+ɑ:+	1 8	225 235	0	0:00:05
M1	François	Phonetic [17] snd	Apr 12 07 11:45:05	Diph.wav	əʊ	<u>Snd 7068</u>				0:00:35
M1	François	Phonetic [17] msq qc9982	Apr 12 07 11:45:40		1	əʊ+aʊ+ɑ:+D+	1 8	230 240	0	0:00:00

Tableau 4 - Séquence de travail - initiation phonétique

Avec un peu d'habitude, on finit par se fabriquer des réflexes de jugement mais la phase d'appropriation des données et de leurs modes de lecture reste longue et quelque peu fastidieuse. Il fallait donc compléter le déluge de données accessibles par des représentations synthétiques et simplifiées: graphiques donc! Une possibilité d'ajouter à ces graphiques de progression, souvent apparemment très irréguliers, est de leur ajouter des courbes de tendances (grâce à des fonctions, choisies dans le cas présent de type logarithmique), ce qui permet de faire apparaître une progression plus ou moins positive, négative ou stationnaire dans l'activité, dans l'intensité, dans les scores etc., pour justifier le silence du tuteur-accompagnateur ou une nécessaire intervention.

La copie d'écran suivante donne un aperçu de la façon dont ces données sont présentées dans SUIVIX, accompagnés de "boutons fonctions" permettant de traiter, sélectionner, analyser, etc. de manière préparée et automatisée diverses parties du "fichier traces" utilisé.

Nouveau SUIVI	RAFFRAICHIR	Bilan Fichier	Bilan Général	Complet	Jouer MM	Tout	+	**	...	@?	Dic	IPA	Décaler A	Décaler X
Ajouter suivi	EFFACER	Bilan Séance	Noms seuls	CHOIX SONS	Active Bouton	Sélect	1+	**	!	☐	Rec	gr. Tps	Essais+Tps	Essais+TpsRép.
Total: 8538		Tps Tot: 48:21:20		SsTot.: 201		Tps Part: 0:58:03		%		02,35%		gr. Score	Score+Tps	Score+TpsRép.
Moy.Tot: 0:00:20		Moy. Part: 0:00:17		Ecartype: 0,00033788		Ecartype: 0,00022248				gr. Tries		Ecrans	TpsInt.+TpsRép.	
C:\HELP\YOU\FIC\cazade\07_08\complet.txt														
Grp.	Names	FILE	BOUT.	PAGE	DATE	Question item:	sign	Mask after	tries:	Score	learner's answer:	Bonus	Durées	Tps_Rep
L1	A__Benoît	UNREALTE[4]-63244			Jan 22 2008 16:24:20	would be able to speak and write	**	would be able —	1 18	225 240	would be able	0	0:00:16	00:16,2
L1	A__Benoît	UNREALTE[4]-63244			Jan 22 2008 16:24:36	would be able to speak and write	+	would be able to speak and write	2 18	230 245	to speak and write	0	0:00:12	00:15,1
L1	A__Benoît	UNREALTE[4]-63319			Jan 22 2008 16:24:48	if you spent	+	if you spent	1 8	235 250	if you spent	0	0:00:02	00:11,3
L1	A__Benoît	UNREALTE[5]			Jan 22 2008 16:24:50	=>Frame n°5							0:00:32	00:00,0
L1	A__Benoît	UNREALTE[5]-65282			Jan 22 2008 16:25:22	they had come	**	— — — — —	1 8	235 250	them coming	0	0:00:04	00:14,5
L1	A__Benoît	UNREALTE[0]-65282			Jan 22 2008 16:25:26	they had come	...	A I D E	2 8	235 250		0	0:00:32	00:00,0
L1	A__Benoît	UNREALTE[5]-65282			Jan 22 2008 16:25:58	they had come	+	they had come	2 8	240 255	they had come	0	0:00:09	00:29,0
L1	A__Benoît	UNREALTE[5]-65357			Jan 22 2008 16:26:07	he had known	**	— — — — —	1 8	240 255	knew	0	0:00:15	00:05,6
L1	A__Benoît	UNREALTE[5]-65357			Jan 22 2008 16:26:22	he had known	**	— had known	2 8	240 255	had known	0	0:00:12	00:13,9
L1	A__Benoît	UNREALTE[5]-65357			Jan 22 2008 16:26:34	he had known	+	he had known	3 8	245 260	he	0	0:00:11	00:11,3

Figure 10 - Logiciel SUIVIX - Vision globale

Afin de comprendre un peu mieux les choix d’affichage opérés, précisons que le tableau suivant (reproduit de la copie d’écran ci-dessus) présente une série de chiffres, de pourcentages, d’écarts types, etc. permettant de situer une partie du profil de la sélection opérée – ici l’étudiant A – Benoît – dans le groupe auquel il appartient.

Total	Temps total	Sous total	Temps partiel	%
8538	48:21:20	201	0:58:03	02,35%
Moy. Tot:	0:00:20	Moy. Part.	0:00:17	
Écart type:	0,00033788	Écart type:	0,00022248	

Tableau 5 - Profilage chiffré d'un étudiant dans un groupe donné

La série d’actions traitées (201 sur 8538 pour le groupe), a été effectuée en 58 minutes et 3 secondes (sur un total de 48 heures, 21 minutes et 20 secondes). On constate que la moyenne de temps d’attente entre les diverses actions effectuées par l’étudiant est de 17 secondes, ce qui semble montrer qu’il est légèrement “plus actif” que l’ensemble du groupe, la moyenne correspondante de tous les étudiants ayant travaillé durant la même séance étant de 20 secondes. Son intensité de travail peut être jugée également légèrement supérieure à celle de l’ensemble du groupe puisque l’écart type de temps d’attente entre les diverses action, pondéré avec celui des temps de réponses en phase de questionnement, est légèrement moindre que l’écart type correspondant à l’ensemble des membres du groupe dont il fait partie. Il est évident que la consultation de tels chiffres n’est pas suffisamment parlante, surtout en présentiel. Même pour les deux seules dernières données, la présentation d’un graphique sera plus lisible.

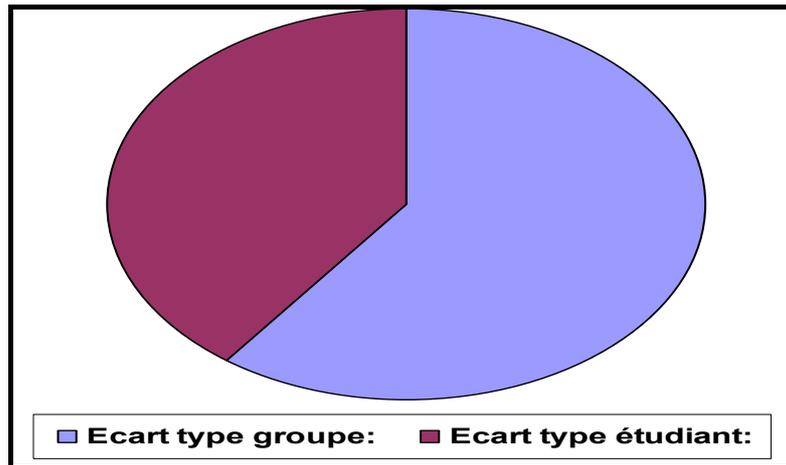


Figure 11 - écart type de temps d'attente entre les diverses action

Nouveau SUIVI	RAFRAICHIR	Bilan Fichier	Bilan Général	Complet	Jouer MM
Ajouter suivi	EFFACER	Bilan Séance	Noms seuls	CHOIX SONS	Active Bouton

Figure 12 - SUIVIX: Quelques "boutons - actions"

Descriptif des boutons proposés (par colonnes, de gauche à droite)

<i>Nouveau Suivi</i>	permet de choisir un fichier individuel ou l'état complet d'un groupe ou d'un groupe de groupes
<i>Ajouter Suivi</i>	permet d'ajouter un fichier individuel à une série déjà affichée
<i>Rafraîchir</i>	permet de rafraîchir les données au cours d'un accompagnement de travail en présentiel
<i>Effacer</i>	permet d'effacer les traçages choisis
<i>Bilan fichier</i>	à chaque fois qu'un étudiant ferme un module, un traçage du bilan d'utilisation du module est inséré dans le fichier
<i>Bilan séance</i>	permet d'accéder au bilan général d'une séance donnée
<i>Bilan général</i>	idem pour tous les étudiants faisant partie d'un groupe
<i>Noms seuls</i>	permet d'exclure les bilans des traçages affichés
<i>Complet</i>	permet de fabriquer puis de consulter le traçage complet d'un groupe d'étudiants.
<i>Choix sons</i>	permet de choisir les enregistrements des étudiants dans le cadre d'un travail en simulation d'interaction orale

Les deux derniers boutons (à droite) permettent à l'enseignant de replacer visuellement l'item tracé dans le contexte exact dans lequel l'apprenant a travaillé, avec tout son accompagnement sonore et visuel. En pratique, cela signifie que le logiciel de travail est lancé, ouvert avec le module voulu, à la page correspondante, le bouton est activé, mis en inversion vidéo, la ressource multimédia est activée, etc.

- Jouer MM* permet, après avoir placé le curseur sur une ligne quelconque du traçage, de lancer l'exécution d'un fichier multimédia utilisé par l'apprenant (son, vidéo etc.)
- Active bouton* permet, après avoir placé le curseur sur une ligne quelconque du traçage, de lancer Help Yourself et d'exécuter le bouton qui avait été activé par l'étudiant et qui est désigné dans la ligne de traçage.

Les boutons suivants permettent d'automatiser, d'affiner, de cumuler etc. les procédures de sélections de traces.



Figure 13 - SUIVIX: quelques "boutons sélections"

- | | |
|-------------|---|
| Tout | permet d'annuler une sélection précédente |
| + | ne sélectionne que les éléments de réponse acceptés par le module et donc considérés comme "bonnes réponses", pour un individu, un groupe, etc. |
| •• | ne sélectionne que les éléments de réponse ayant justifié un travail de restitution/démasquage partiel dans les questions posées |
| ... | ne sélectionne que les demandes de démasquage du nombre de lettres et de mots masqués dans des questionnements à trous |
| @? | ne sélectionne que les traçages correspondant à des demandes d'aide proposées par le logiciel |
| 🎙️ | ne sélectionne que les demandes d'activations d'éléments multimédias |
| IPA | sélectionne les traces liées à l'activation de boutons correspondant aux phonèmes travaillés dans les écrans portant sur l'apprentissage de la prononciation (IPA = <i>international phonetic alphabet</i>) |
| Sélect | le curseur étant positionné sur une réponse longue donnée, permet de sélectionner les traces contenant les éléments qu'on désignera – cela permet de cibler telle ou telle difficulté d'expression, des questions portant sur des points particuliers, etc., <i>en mode concordancé</i> |
| 1+ | sélectionne les traces correspondant à des réponses considérées comme bonnes par le logiciel dès le premier essai de réponse |
| •- | sélectionne les affichages de type "jeu du pendu" dont le résultat n'aura pas été complet |
| ! | sélectionne les demandes d'aides |
| □ | sélectionne les traces correspondant à des abandons dans un questionnement donné |
| Rec | sélectionne les traces correspondant aux enregistrements sonores effectués par les étudiants (et permet de les activer par le bouton MM) |



sélectionne les traces correspondant aux boutons sonores activés par l'étudiant

L'utilisation combinée de plusieurs de ces boutons permet d'affiner ou/et d'élargir à volonté l'espace d'analyse isolée ou comparative, portant sur un individu, un individu dans un groupe, plusieurs groupes entre eux, sur tel type de réponse attendue, tel type d'action, tel type de résultats, de "performance" observés ailleurs, etc.

D'autres "boutons fonctions" permettent de lancer le traitement automatique d'une sélection multi-critériée effectuée précédemment grâce aux boutons mentionnés ci-dessus, et de fabriquer à partir de cette sélection divers graphiques simples ou cumulés, avec l'ajout de courbes de tendances (voir graphes 13 et 14 ci-dessous).

gr. Tps	Essais+Tps	Essais+TpsRép.
gr. Score	Score+Tps	Score+TpsRép.
gr. Tries	Ecrans	TpsInt.+TpsRép.

Figure 14 - Quelques boutons pour fabriquer des graphes cumulés

Une utilisation répétée, avec divers groupes d'étudiants, permet désormais de privilégier telles combinaisons qui semblent plus parlantes que d'autres en présentiel, et d'autres plus fines en phase asynchrone.

Les combinaisons qui semblent les plus souvent pertinentes, nécessairement combinées avec diverses sélections précédemment annoncées, correspondent aux "boutons fonctions" suivants: temps intermédiaires, temps de réflexion avant réponse, scores, nombre d'essais, écrans parcourus. Ces boutons-fonctions permettent des combinaisons internes et donc de nouveaux affichages graphiques cumulés (voir certains boutons en haut et à droite de la copie d'écran en figure 8).

Ces procédures me permettent de mettre en valeur les critères déjà énumérés supra visant *matière* et *manière*, et facilitent notamment l'observation:

- de l'impact du suivi enseignant et apprenant sur le processus apprenant,
- des ressorts de motivations: diversité des approches du thème étudié, des modes d'activité, gradations de difficultés, de ludicité, etc.,

On trouvera ci-après quelques exemples de traitements et de prises de diagnostic possibles. Hélas, le présent tirage en noir et blanc ne permet pas d'identifier clairement les données.

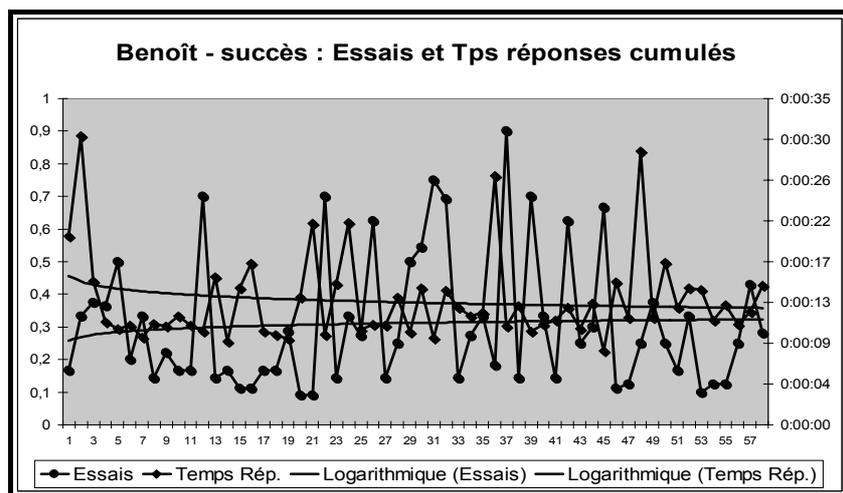


Figure 15

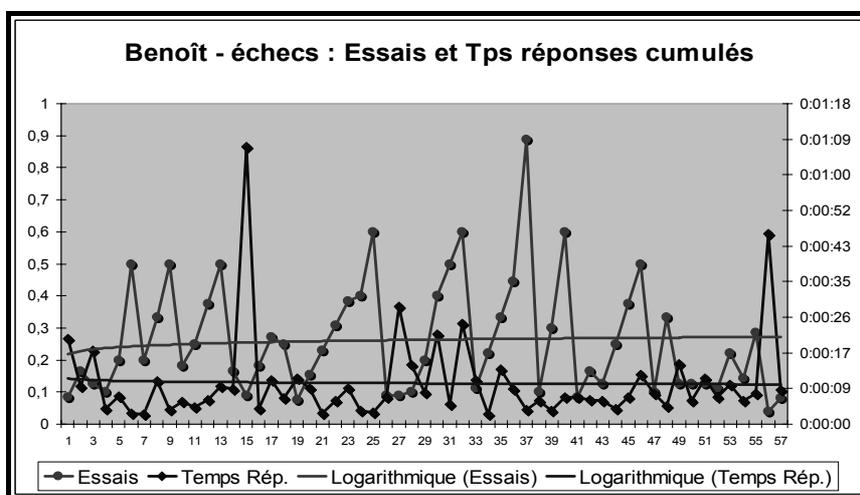


Figure 16

La comparaison de ces deux graphes permet à l'enseignant de comprendre – et peut-être de faire valoir auprès de l'apprenant – la corrélation entre trois paramètres: le temps de réflexion (temps de réponse), le nombre d'essais tentés et le taux de réussite. Plus il prend le temps de réfléchir meilleur est son taux de réussite aux questions posées. En effet, la courbe de tendance varie entre 13 et 16 secondes dans le premier tableau, et entre 9 et 10 secondes dans le deuxième alors que le nombre d'essais est sensiblement égal – la courbe de tendance varie entre 2 et 3 dans les deux cas,

Benoît fait partie des “bons”, il a l'habitude que les choses avancent vite et bien durant son parcours et, pourtant, lui aussi peut gagner à ne pas sous-estimer l'importance du temps de la réflexion.

Travail sur l'évaluation des temps de réflexion

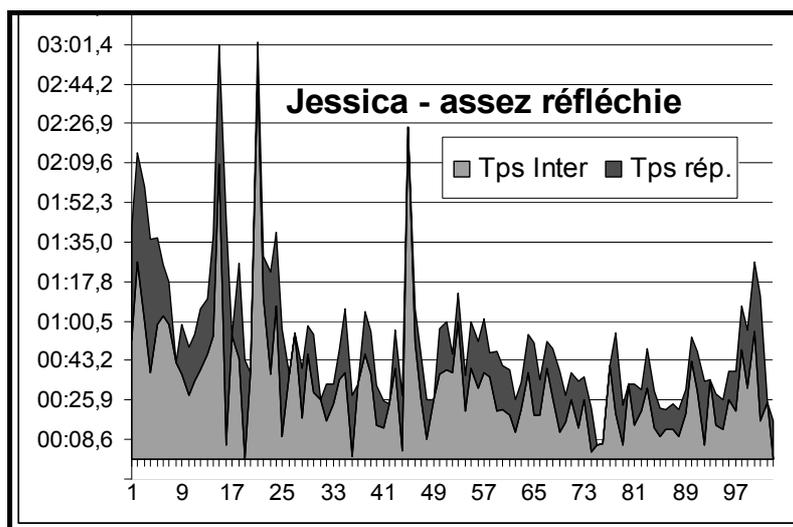


Figure 17 – Jessica

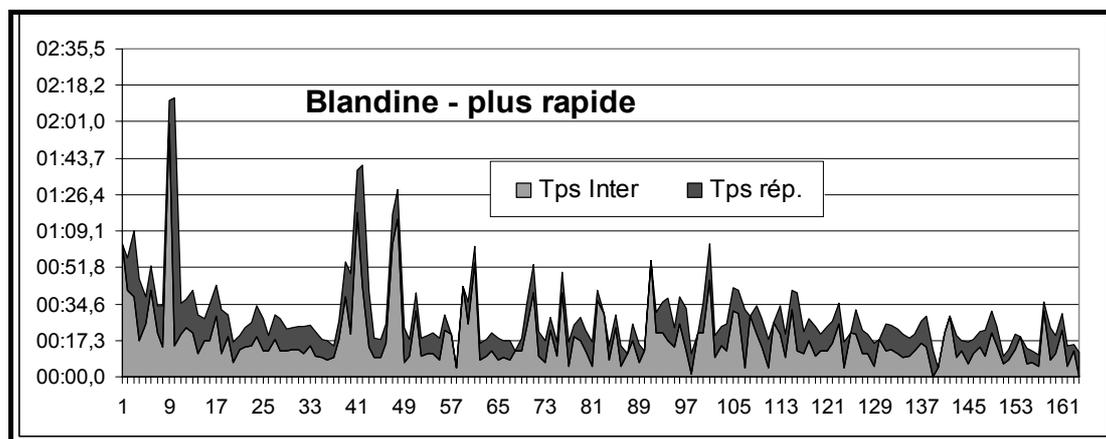
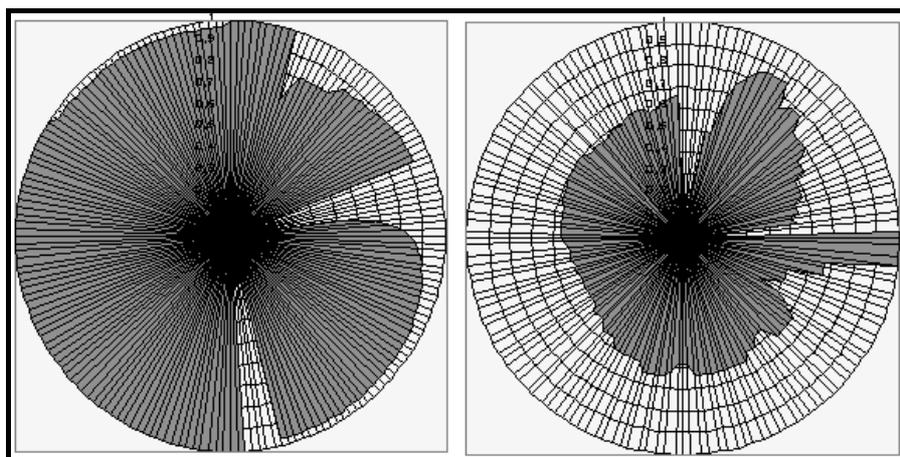


Figure 18 – Blandine

Il apparaît ici – et c'est là le point qui peut aider l'enseignant à savoir quand et qui aider – Blandine (fig. 16) prend beaucoup moins de temps que Jessica (fig.15) avant de répondre (partie basse de chaque graphe) et hésite moins avant de terminer la formulation de sa réponse. Cela lui permet d'avoir un parcours plus actif, plus "intense" dans le même temps de travail, d'être peut-être, au moins sous cet angle, plus "efficace". Son taux de réussite se trouve d'ailleurs être, parallèlement, bien meilleur que celui de Jessica.

Dans les graphes suivants, on a sélectionné uniquement des items questions-réponses.

Graphes "Scores"



Rahim

Figure 19

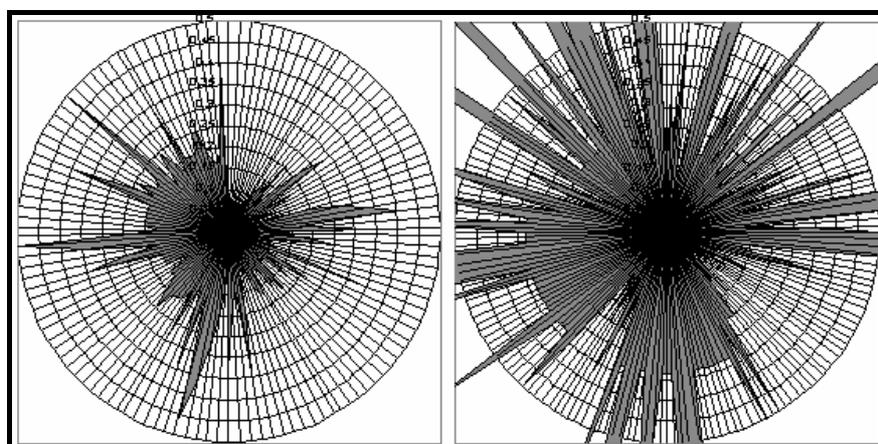
Benoît

L'affichage visuel des scores de Rahim et Benoît (le graphe se lit en partant de 12h et en suivant le sens des aiguilles d'une montre) font apparaître chez le premier moins de réussite en début de parcours qu'à la fin de celui-ci, ou le graphe atteint presque la limite extérieure du cercle, signe de "taux plein"; Benoît, quant lui, parvient beaucoup moins bien, dans le même parcours, à s'écarter suffisamment de la partie centrale du graphe, signe de taux à peine moyen de réussite.

Graphes "essais" (fig. 18)

La limite du cercle correspond à la moitié du maximum d'essais prévus pour répondre aux items questions correspondants.

Camille, pour répondre aux questions qu'elle a choisies, a utilisé bien moins d'essais que Farid, qui dépasse souvent la moitié du nombre d'essais alloués. Il avance moins vite, avec plus de difficultés, etc. Si les scores de Camille restent corrects, il faut certainement aider Farid davantage que Camille.



Camille

Figure 20

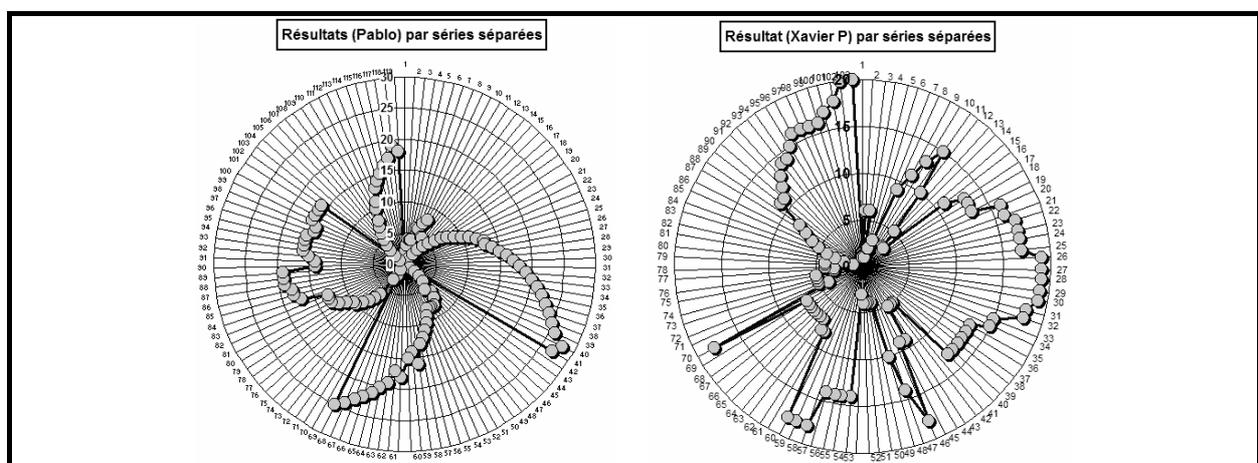
Farid

Quelques points d’approche pour déterminer certains “profils apprenants” récurrents

Pour le chercheur qui veut prendre plus de temps afin d’étudier ces parcours apprenants et d’essayer d’en faire apparaître quelques-uns plus récurrents que d’autres (alors que l’enseignant n’a évidemment pas le temps de faire cela), il a semblé utile de mettre au point quelques procédures permettant de visualiser les aires de déplacements et les effets d’insistance des apprenants dans les parcours prévus par chaque module.

Les deux graphes suivants se lisent toujours en partant de 12h, puis en suivant le sens des aiguilles d’une montre. Pour simplifier la lecture de tels graphes dans cet article, seuls quatre modules ont été retenus pour faire l’objet d’une analyse comparative.

Dans le premier cas, les courbes obtenues en suivant le parcours de Pablo (fig.19) sont nettement plus régulières que celles de Xavier; elles partent du centre et s’en écartent sans à-coups: cela indique que Pablo part, dans chaque module du premier item proposé, le traite, puis passe au suivant, et ainsi de suite, avec un peu moins de régularité toutefois dans le troisième quartier. Le parcours de Xavier suit visiblement beaucoup moins le parcours par défaut prévu par le module, ce qui est son droit le plus complet. Cela témoigne d’un type de comportement différent. Il préfère aller ici, puis là, revenir en arrière, sauter quelques items, etc. Pourtant, la dernière section de son parcours (juste avant de rejoindre le “point midi” du graphe) montre qu’il préfère, à ce moment de sa progression, suivre plus docilement le parcours proposé. En soi, cette comparaison n’a pas de sens à elle toute seule, ce n’est qu’en rapprochant ces éléments de nombre d’autres qu’on peut, au bout du compte, tirer quelques conclusions. En l’occurrence, on peut considérer que Pablo a un profil plus “algorithmique” et “systématique” que Xavier, dont le tempérament l’amène apparemment à “papillonner”, à butiner de façon plus aventureuse et libre que Pablo.



Pablo

Figure 21

Xavier

Ces deux profils n'ont pas été choisis au hasard. Ce sont, en effet, deux types de profils les plus saillants dans les parcours rencontrés.

Graphes portant sur l'écoute et la prononciation

Dans un autre travail en cours, il a semblé utile de faire apparaître de manière résumée dans un graphe spécifique les difficultés d'écoute que rencontrent les étudiants. Un module un peu général d'initiation à la phonétique et un petit outil auxiliaire (qui sert d'aide pour divers exercices proposés dans d'autres modules portant sur des sujets variés, mais où une place est presque toujours réservée pour les questions de prononciation⁵) permettent aux étudiants de travailler sur la reconnaissance des phonèmes de bases de la langue anglaise. Une sorte de clavier graphique permet également de les écrire dans des "cases réponses" ciblant des difficultés particulières d'écoute et de prononciation. Comme chaque activation de bouton correspondant à l'écoute de chaque phonème entraîne une écriture dans le fichier trace de chaque étudiant, on peut savoir quels sont les phonèmes, et même les enchaînements de phonèmes, qui font davantage que d'autres l'objet d'écoutes et de réécoutes, et donc qui semblent résulter d'une difficulté de discernement dans la phase d'écoute. Afin de pouvoir aider un peu plus finement l'apprenant, en répondant aux difficultés qu'il souligne d'ailleurs lui-même, il est intéressant de faire apparaître visuellement les points de difficultés les plus récurrents qu'il a rencontrés sur un graphique comme celui qui est représenté dans la figure suivante (fig.20). Pour ne pas alourdir ce type de graphiques, n'apparaissent ici que trois points – de grosseurs différentes suivant le nombre d'accès demandés aux écoutes correspondantes –, chacun correspondant aux trois phonèmes les plus travaillés, dans le cas présent: |i:], |I| et |e|. Cette procédure se fait en attribuant à chaque phonème une position (abscisses et ordonnées) spécifique par rapport à la figure de fond du graphique, et en lui donnant une taille correspondant au nombre d'accès demandés. Là encore l'absence de couleurs ne facilite pas la lecture du graphique.

⁵ On doit rappeler que ces exercices amenant les étudiants à travailler sur la prononciation et sur les questions de phonétique, loin de rebuter nos étudiants non-spécialistes, sont très appréciés. Ils sont tout particulièrement recherchés par ceux qui, ayant un niveau de langue un peu faible, préfèrent travailler sur les formes sonores de la langue avant d'être testés sur leur capacité de compréhension, inquiets qu'ils sont de ne pas réussir autant qu'ils le voudraient. La compréhension bénéficie toujours, il faut le noter, de ces "travaux d'approche" passant par le travail sur la strate sonore de la langue.

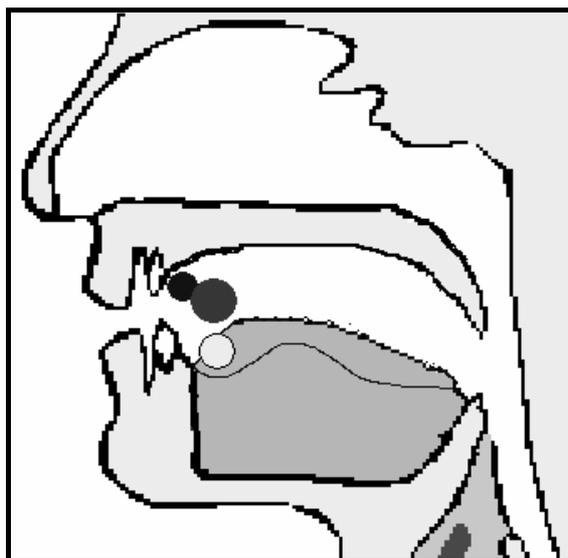


Figure 22 - Sons: |i:| - |I| - |e|

Il faudrait évidemment pouvoir faire de même, non pas pour l'écoute seulement mais pour la prononciation. Cependant les outils pour analyser la production orale (reconnaissance vocale, entre autres) ne semblent pas suffisamment probants, aujourd'hui encore, pour pouvoir faire un travail précis dans ce domaine. Dans certains cas, suivant le type d'exercice proposé, il est malgré tout déjà possible par déduction de tirer quelques conclusions quant aux difficultés de prononciation que rencontrent les apprenants concernés.

Conclusion

La seule présentation de quelques graphes ne peut suffire à convaincre de l'intérêt – et à mon sens, de la nécessité même – de travailler sur la mise au point et l'intégration de ce type d'"aide au diagnostic" dans les outils de suivi que nous serons amenés à utiliser de plus en plus souvent et de plus en plus intensément dans nos travaux d'accompagnement pédagogique et dans nos tâches de certification de compétences. Il est impossible de présenter via un support statique – comme l'est cet article – le bénéfice de ce type de développement. Ces aides sont notablement plus nécessaires dans un enseignement-apprentissage de langues vivantes étrangères que dans d'autres cadres pédagogiques, étant donné la nécessité de prendre en compte l'interaction orale entre les apprenants et l'interaction d'étayage pédagogique dans la relation spécifique que ce type de co-construction langagière demande entre les apprenants et l'enseignant-accompagnant.

La poursuite de ces travaux permettra dans de futures publications de présenter un certain nombre de conclusions, différentes, à l'évidence, suivant qu'on travaille en présentiel, en mode synchrone ou en mode asynchrone, à l'écrit ou à l'oral. Jusqu'à maintenant, il a semblé nécessaire d'insister sur

l'accompagnement en mode synchrone, écrit et oral, pour faciliter une pose de diagnostic en temps réel. Pour ce faire, il a fallu, comme on l'a vu, prendre en compte de façon combinée un choix particulier d'éléments: temps d'attente, intensité, répétitivité, insistance, etc., qui permettent surtout de savoir quand intervenir ou non, quand relancer la motivation et quand réorienter le travail quand celui-ci ralentit ou même se bloque.

Il apparaît que le mode à distance et asynchrone favorise les travaux de réflexion, se prête davantage à un effort de compréhension plus dense et approfondi, et, de ce fait, également à l'expression écrite. Le traitement de telles productions écrites devrait néanmoins pouvoir bénéficier d'utilitaires complémentaires d'annotation pour faciliter la mise en place d'un "retour pédagogique" (*feedback*) suffisamment fourni pour stimuler le récepteur de manière conséquente, sans devenir trop chronophage malgré tout pour l'accompagnement enseignant. Il ne faudra pas négliger ces pistes de recherche à l'avenir dans nos travaux sur le suivi-apprenant en langues vivantes.

Moodle et le suivi des apprenants en langues vivantes

Commentaires d'une présentation de l'utilisation de Moodle pour le suivi des apprenants en langues vivantes

Olivier Catteau, Enseignant à l'IUT A de l'Université Toulouse 3, Chargé de mission TICE (sites de Castres, Auch, Albi et Toulouse), doctorant de l'Institut de Recherche en Informatique de Toulouse (Problématique d'environnement informatique pour l'apprentissage humain).

Mise en forme assurée par *Christine Vaillant Sirdey*, Maître de Conférences en anglais, chercheure au laboratoire LAIRDIL, Université Toulouse 3.

La plateforme Moodle propose une approche modulaire de diffusion et de distribution de *ressources* ainsi qu'un choix important d'*activités*. Les outils de suivi s'appliquent à toutes les activités et une notation intégrée est automatiquement disponible dans les bases de données, les devoirs, les forums et les tests. On peut également ajouter des modules développés par des entreprises ou organismes tiers dont les notations sont reprises et intégrées automatiquement au suivi de Moodle. Moodle est développé dans le cadre du logiciel libre.

Le suivi permet de connaître instantanément les ressources mises à disposition d'un apprenant et les actions qu'il a effectuées. Plusieurs visualisations sont disponibles: résumé, détaillé ou par histogrammes.

Quelques exemples concrets et parlants pouvant faire l'objet d'un suivi

À l'écrit, Moodle permet de construire des tests sous forme de *QCM* ou d'*exercices à trous*. Si l'apprenant se trompe, le concepteur peut avoir autorisé des retours qui permettent de trouver la bonne réponse avec des retours d'information écrits ou sonores. Toute mauvaise réponse entraîne des pénalités (paramétrables) et le nombre d'essais est pris en compte dans la note finale.

On peut aussi importer des tests développés avec un outil extérieur comme des mots-croisés créés sous Hot Potatoes™ (Half Baked Software Inc) dont les résultats seront intégrés automatiquement dans le suivi de Moodle.

L'activité "*devoir*" permet à l'étudiant de remettre son travail selon deux modalités: soit en répondant directement via un éditeur html simple à utiliser, soit par le dépôt d'un fichier que l'étudiant aura pris le temps de rédiger hors ligne et qu'il postera sur la plateforme dans les délais indiqués par l'enseignant. Ce dernier aura tout le récapitulatif de tous les devoirs et pourra apporter une notation et des commentaires.

Un autre exemple intéressant est celui de l'utilisation de la notation des *forums*. On lance un débat entre un petit groupe d'étudiants, l'enseignant analyse les échanges et analyse, commente et note les interventions. Là aussi les étudiants peuvent déposer un travail fait "hors ligne" avec des applications telles que Word, Excel, etc. L'enseignant peut analyser les échanges de manière asynchrone et donner une note sur les interventions des différents participants.

Un nouveau module est le module de *base de données* qui permet de renseigner des champs textuels, des images, etc. Par exemple, l'enseignant dépose toute une série

de visuels provenant de publicités et les rend accessibles à chaque étudiant. À chacun peut être associée une discussion, que l'enseignant notera suite à l'analyse des documents.

L'utilisation des plateformes par les étudiants est suivie via un outil de type SCORM. SCORM est un standard (en fait, un regroupement de différents standards) qui permet de déployer des ressources pédagogiques sur différentes plateformes. On peut ainsi réutiliser des ressources "efficaces". Evidemment, obtenir ces ressources demande plus de temps et plus de compétences. Mais c'est la manière la plus pertinente d'obtenir des ressources riches et mutualisables. Dans l'exemple donné, l'étudiant doit reconnaître en anglais différentes formes de graphismes. L'enseignant peut analyser et noter les explications.

Suivi de l'oral: test et QCM avec lecteur audio

L'étudiant écoute un enregistrement sonore et répond à des questions de compréhension posées par écrit ou par oral. L'IUT A de Toulouse 3 participe aujourd'hui à l'ajout de fonctionnalités multimédia telles que le magnétophone intégré dans Moodle. L'étudiant parle très peu dans un enseignement traditionnel des langues, donc il faut amener les étudiants à parler. L'outil fonctionne avec un serveur d'application multimédia. L'étudiant s'enregistre directement sur la plateforme. Ce qui est intéressant c'est que l'ensemble ne demande aucune installation de logiciel additionnel. Tout est directement intégré dans la plateforme. L'étudiant enregistre, valide et poste son enregistrement sur la plateforme.

Parmi les autres développements en cours on peut citer la possibilité d'organiser une vidéoconférence intégrée dans Moodle entre un enseignant et différents étudiants distants géographiquement les uns des autres, ou encore le *streaming* vidéo pour la mise en place duquel un serveur dédié est en cours d'acquisition au moment de la rédaction de ces notes.

En conclusion on peut dire que, hormis celui qui concerne la vidéoconférence, les outils de mesures proposés sont encore basiques. Ils permettent de savoir sous forme de textes, tableaux et graphiques, qu'à tel moment un étudiant a fait ceci ou cela, on peut très facilement "voir" son activité sur des périodes déterminées. Néanmoins on notera que toutes les données recueillies demeurent des mesures informatiques brutes.

La recherche universitaire menée actuellement en environnement informatique pour l'apprentissage humain vise à analyser ces résultats informatiques et à en donner des interprétations pédagogiques: pourquoi un étudiant a-t-il "réussi" une question mais passé dix minutes dessus? Il s'agit d'avoir des informations qualitatives sur l'apprentissage humain et être capable de repérer le degré de réussite ou d'échec.