



**HAL**  
open science

## Quelques critères d'utilisation d'un système d'évaluation automatique de synthèses de cours à distance

Damien Dupré, Hussein T. A. Salem, Mathieu Loiseau, Philippe Dessus,  
Stephane Simonian

### ► To cite this version:

Damien Dupré, Hussein T. A. Salem, Mathieu Loiseau, Philippe Dessus, Stephane Simonian. Quelques critères d'utilisation d'un système d'évaluation automatique de synthèses de cours à distance. Recherches & éducations, Société Binet Simon, 2012, 6, pp.143-159. hal-00843616

**HAL Id: hal-00843616**

**<https://hal.archives-ouvertes.fr/hal-00843616>**

Submitted on 11 Jul 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dupré, D., Salem, H. T. A., Loiseau, M., Dessus, P., & Simonian, S. (2012). Quelques critères d'utilisation d'un système d'évaluation automatique de synthèses de cours à distance. *Recherches & Educations*, 6, 143–159.

## Quelques critères d'utilisation d'un système d'évaluation automatique de synthèses de cours à distance

### Résumé

Dans le cadre du projet européen *Language Technologies for Lifelong Learning*, nous avons développé *Pensum*, un système informatique d'aide à la rédaction de synthèse de textes pour les apprentissages à distance. L'étude de validation de ce logiciel auprès d'étudiants du CNED nous a permis d'examiner les modalités d'appropriation du système dans une utilisation à distance. L'étude était composée d'une période de test associée à un questionnaire et une série d'entretiens. Les différents indicateurs utilisés (ex : efficacité perçue, charge cognitive) nous permettent de faire des propositions plus générales quant à l'usage et au non-usage des TIC.

### Mots clés

Enseignement à distance, usages, feedback, LSA, synthèse

### Abstract

In the framework of the *Language Technologies for Lifelong Learning* European project, we have developed *Pensum*, a system for automatic assessment of text summaries for distance-learning. The validation study of this widget with CNED students has allowed us to study the modalities of appropriation of the system. The study was constituted of a test phase followed by a questionnaire and a series of interviews. The various indi-

## Recherches & Educations

cators used (e.g. perceived lack of efficiency or excessive cognitive load) allow us to express more general propositions about use of ICT artefacts for learning (or the lack thereof).

### Key-words

Distance learning, uses, feedback, LSA, synthesis

### Introduction

Si les outils de correction orthographique et grammaticale sont largement disponibles et intégrés aux logiciels (traitement de texte, client mail, applications web, *etc.*), ce n'est pas encore le cas pour des outils de plus haut niveau, comme ceux analysant des aspects sémantiques. Cela tient en grande partie à la complexité des traitements en jeu et des activités qu'ils autorisent. Toutefois, des exemples de l'utilisation de tels outils existent dans le champ des technologies éducatives. Des techniques de Traitement Automatique des Langues (TAL) permettent l'implémentation d'outils expérimentaux utilisant une analyse sémantique en vue d'évaluer automatiquement certains aspects des productions d'étudiants (p. ex., *Intelligent Essay Assessor* de Landauer, Laham & Foltz, 2000). Il devient donc possible d'utiliser de tels systèmes dans des contextes d'enseignement à distance, dans lesquels les enseignants et tuteurs n'ont souvent que peu de temps pour interagir avec les étudiants et ont besoin d'une telle assistance.

L'usage productif d'un outil informatique investissant des traitements complexes est tributaire de la compréhension de leurs limites (p. ex., Murray & Barnes, 1998, sur les outils d'apprentissage des langues assisté par ordinateur), comme le souligne le problème de l'usage de correcteurs orthographique et grammatical dans l'enseignement des langues (Durel, 2006). L'analyse sémantique de la production écrite des apprenants ouvre donc de nouvelles perspectives, mais pose également des questions majeures quant à leur intégration dans les usages.

Les TICE (technologies de l'information et de la communication pour l'éducation) sont depuis longtemps suspectées voire accusées d'être sur-vendues, sous-utilisées, et de n'avoir que peu d'effets sur l'apprentissage (Clark, 2001 ; Cuban, Kirkpatrick & Peck, 2001). L'une des questions qui

### **Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance**

se pose donc est la suivante : selon quels facteurs et critères un outil logiciel procurant à des étudiants des rétroactions significatives dans un réel contexte d'enseignement à distance peut-il être efficacement utilisé ? Pour fournir des éléments de réponse à cette question, nous présenterons les résultats du test d'une application, *Pensum*, en contexte d'enseignement à distance écologique. Celle-ci utilise des techniques statistiques d'inférence du sens pour procurer aux apprenants des rétroactions automatiques sur leur compréhension d'un corpus de cours. Plus précisément, nous avons observé l'usage et la compréhension des rétroactions de *Pensum* : un comparatif a été initié pour connaître l'effet des rétroactions, ainsi qu'une étude qualitative pour connaître les usages réels comme leurs limites, ce qui permet d'identifier des incitateurs favorisant les usages et la compréhension des non-usages.

La première section de cet article présente brièvement *Pensum*. La seconde détaille le contexte et le matériel de l'étude et la troisième décrit les résultats. Pour interpréter ces derniers, la présente étude aura recours à certains concepts issus de l'approche instrumentale de Rabardel (1995), et s'appuie, du fait de la nature du projet, sur différents modèles d'adoption de la technologie.

### ***Pensum*, un système d'évaluation automatique de synthèses de cours**

*Pensum* (Loiseau, Dupré & Dessus, 2011) est l'un des prototypes développés au sein du projet européen *Language Technologies for Life-long Learning* (LTfLL), ayant pour but de proposer des outils pour faciliter l'apprentissage tout au long de la vie<sup>1</sup>. Ce système vise à améliorer la production de synthèses de cours dans un contexte d'apprentissage à distance. Son fonctionnement repose sur l'analyse sémantique latente (Laudauer & Dumais, 1997) du contenu de textes tout-venants (du point de vue des contenus, d'un point de vue technique seul le format texte est utilisable à l'heure actuelle). L'étudiant peut lire le cours dans la partie supérieure de la fenêtre et en rédiger la synthèse dans la partie inférieure (voir *Figure 1*).

---

<sup>1</sup> Plus d'informations à <http://webu2.upmf-grenoble.fr/sciedu/ltfll-pensum/>

## Recherches & Educations

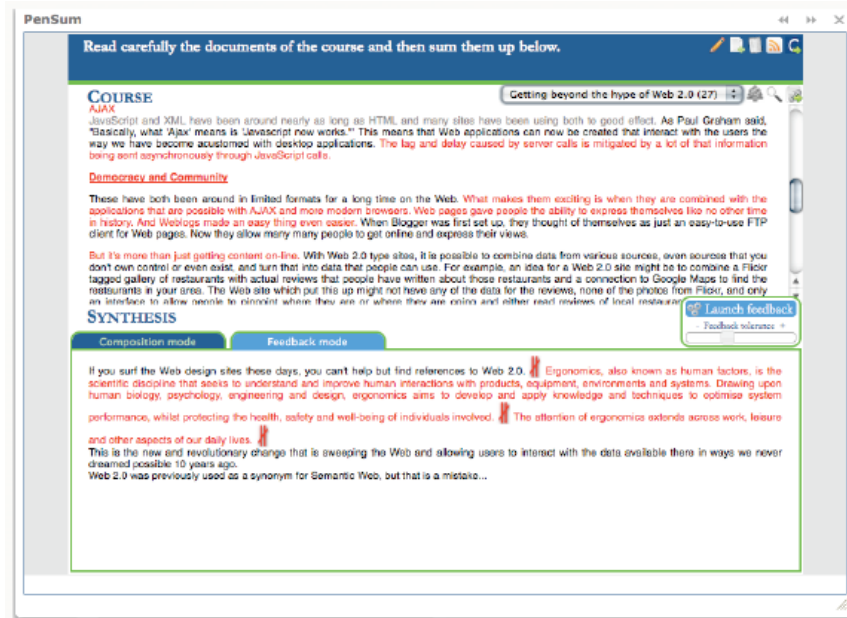


Figure 1. Interface du logiciel Pensum

*Pensum* donne à l'étudiant la possibilité de recevoir des rétroactions sur son travail quand il le souhaite et de manière automatique. Ces retours font suite à une analyse sémantique de la synthèse et du texte-source et sont de trois types : la mise en évidence des sauts de cohérence entre les phrases de la synthèse réalisée, des phrases de la synthèse sans lien avec le texte source et des phrases du texte source non reprises dans la synthèse. Ainsi, l'étudiant peut choisir de modifier le contenu de sa synthèse en accord avec le feedback fourni par *Pensum* ou bien de le remettre en question par un clic droit s'il estime que sa production satisfait ses intentions (Loiseau *et al.*, 2011).

### Comment des apprenants s'approprient-ils *Pensum* ? Une étude de validation

Afin d'évaluer l'usage et la possible appropriation de *Pensum*, nous avons mené une étude auprès de 36 étudiants au CNED (Centre National de l'Enseignement à Distance) de Lyon et de Rouen en Licence 3 de sciences de l'éducation. La méthodologie de cette étude consistait à faire

### **Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance**

utiliser le logiciel dans des conditions les plus proches d'un usage réel sur une durée suffisamment longue pour avoir des informations utilisables concernant son appropriation. Les étudiants devaient réaliser à distance deux synthèses de leur cours portant sur les TICE : l'une sans utiliser *Pensum* (une interface identique à celle de *Pensum* a été construite, mais qui ne procurait aucune évaluation), l'autre en l'utilisant. Parallèlement, ils étaient suivis par quatre tutrices formées au préalable à l'utilisation du logiciel. À la fin de la rédaction de chaque synthèse, les participants étaient invités à remplir un questionnaire portant sur leur expérience d'utilisation de *Pensum* afin d'évaluer leur propension à se l'approprier et l'utiliser dans le futur. Certains d'entre eux ont, ensuite, été interrogés par téléphone sur leur usage du logiciel, de manière plus ouverte et qualitative.

### **Participants**

Afin de recruter un nombre suffisant d'étudiants, un premier appel à participation a été déposé sur la plate-forme d'enseignement à distance dédiée à leur cours (Campus FORSE de Rouen et Lyon). Une présentation de l'étude a été effectuée en face à face aux étudiants présents sur le site de Lyon lors de leur réunion semestrielle. 7 étudiants de Rouen et 85 de Lyon ont accepté de participer à la validation de *Pensum*. Néanmoins, après avoir pris connaissance des consignes précises de l'expérience et après s'être investis dans la tâche, seuls 36 participants (4 de Rouen et 32 de Lyon) ont été retenus pour nos analyses. Cette mortalité expérimentale peut s'expliquer par la défection de participants (certains n'ont pas commencé la tâche, d'autres ne l'ont pas terminée), en raison de l'effort et du temps qu'elle demande. La participation de chaque étudiant était récompensée d'une clé USB. Afin de motiver les étudiants à réaliser l'ensemble des tâches, deux participants ayant réalisé toutes les phases de la validation ont été tirés au sort et se sont vu attribuer un disque dur externe.

### **Tutrices**

Pour tester *Pensum* dans un contexte le plus proche possible des modalités d'apprentissage à distance, quatre tutrices ont été recrutées pour gérer l'ensemble des requêtes et résoudre les éventuels problèmes des participants. Ces tutrices sont étudiantes en Master 2 de sciences de l'éducation de l'université Grenoble-2. Il est à noter que ces étudiantes avaient déjà participé à la validation d'une version précédente de *Pensum*

## Recherches & Educations

(Loiseau *et al.*, 2011). Après une formation sur l'utilisation du logiciel, chaque tutrice s'est vu attribuer un nombre équivalent d'étudiants et a été capable de répondre à toute question concernant le logiciel et l'organisation de l'étude.

### Matériel

Le choix du cours travaillé par les participants s'est opéré selon deux critères : 1) son adéquation avec le thème de l'expérience ; 2) le fait que chaque chapitre se termine par différentes activités à réaliser, dont la rédaction d'une synthèse. Le support de cours a été divisé en six chapitres importés au préalable dans *Pensum*. L'étude de validation portait spécifiquement sur les deux premières parties du cours, mais les élèves pouvaient utiliser *Pensum* librement sur les autres.

Dans le but de mesurer les déterminants de l'usage de *Pensum*, nous avons construit un questionnaire visant à étudier les pratiques et usages de *Pensum* (Armitt *et al.*, 2011). Les critères évalués par ce questionnaire se basent sur le modèle *Unified Theory of Acceptance and Use of Technology* (UTAUT, Venkatesh, Morris, Davis & Davis, 2003). Le questionnaire du groupe expérimental était composé de 30 items et celui du groupe-contrôle 25 (les 5 items supprimés traitaient de types de rétroactions différents et étaient par conséquent non applicables au groupe-contrôle). Les items étaient proposés sous forme d'affirmation, le participant devait alors indiquer son degré d'accord avec celle-ci. La réponse était mesurée sur une échelle de Likert en 5 points allant de « *pas du tout d'accord* » à « *tout à fait d'accord* ». Huit critères pouvant influencer l'usage de *Pensum* et son appropriation ont été déterminés (*cf.* Tableau 1 et Armitt *et al.*, 2011).

### Design expérimental

Un groupe-contrôle et un groupe expérimental ont été mis en place afin d'évaluer l'influence des rétroactions sur l'usage de *Pensum*. Le groupe expérimental pouvait utiliser *Pensum* tel quel, alors que le groupe-contrôle ne pouvait utiliser *Pensum* que dans une version sans rétroaction (les zones de lecture et d'écriture de synthèses étant identiques à la version expérimentale). Après la confirmation de leur participation, les étudiants ont été répartis au hasard, soit dans le groupe-contrôle soit dans le groupe expérimental pour la première synthèse. Pour la rédaction de la seconde synthèse, les groupes-contrôle et expérimental échangeaient

### **Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance**

leurs rôles, à des fins de contre-balancement. Néanmoins, nous ne présenterons dans cette étude que les résultats relatifs au premier groupe de participants, c'est-à-dire ceux ayant réalisé la première synthèse en tant que groupe-contrôle et la seconde en tant que groupe expérimental. Cette limitation vise à supprimer l'effet de « contamination » liée à l'utilisation de toutes les fonctionnalités de *Pensum* lors de la rédaction de la première synthèse sur le jugement de la version-contrôle de *Pensum* lors de la seconde synthèse.

### **Description de la tâche**

Les participants potentiels recevaient un courriel expliquant les conditions de l'expérimentation : première description des tâches à effectuer et leur planification. Les participants décidant de s'inscrire fournissaient des informations de contact et recevaient ensuite un second courriel avec leurs identifiants et mot de passe pour se connecter à *Pensum*. Il est à noter que toute l'expérience était réalisée à distance, et que toutes les interactions avec les participants (par exemple, entre apprenants et tutrices) étaient réalisées par courriel. Leur tâche était d'utiliser *Pensum* librement à distance afin d'écrire deux synthèses de deux premières parties de leur cours. Aucune contrainte de longueur n'a été donnée et dix jours ont été laissés aux étudiants pour écrire les deux synthèses (cinq jours par synthèse).

### **Résultats : Quelques déterminants de l'usage de *Pensum***

Les réponses aux questionnaires ont été étudiées selon la comparaison entre la moyenne de l'échantillon (expérimental ou contrôle) et l'indice central de l'échelle de Likert (*i.e.*, 3 sur l'échelle en 5 points) sur l'ensemble des questions qui composent un critère. Nous avons considéré ces données comme des données d'intervalle puisqu'il y a une symétrie entre les réponses positives et négatives et que les participants pouvaient clairement percevoir un égal espacement entre les différentes opinions exprimées. Un test *t* bilatéral de Student (contre la valeur médiane, 3) a été réalisé afin de mettre en évidence les critères pour lesquels l'opinion moyenne est significativement positive.



## Recherches & Educations

### L'opinion des apprenants sur l'utilisation de *Pensum*

Après la première synthèse, 20 participants du groupe-contrôle ont rempli le premier questionnaire. Après la seconde synthèse, 16 participants du groupe expérimental ont répondu au second questionnaire. Selon les tutrices, les contraintes liées à la charge de travail (disponibilité / effort mental) seraient à l'origine de ces quatre abandons. Tous avaient le français pour première langue. Le Tableau 1 détaille pour chaque dimension l'opinion des participants sur leur utilisation de *Pensum*.

La comparaison des résultats avec la moyenne révèle que les réponses des enquêtés selon les critères *Utilisabilité* et *Conditions facilitatrices* sont significativement supérieures à l'opinion médiane. *A contrario*, il s'avère que les critères *Effizienz perçue* et *Charge cognitive* apparaissent significativement inférieurs à l'opinion médiane. Il est à noter que ces deux critères apparaissaient déjà significativement inférieurs à l'opinion médiane dans le groupe-contrôle (première synthèse).

Ainsi, si l'on compare les résultats des questionnaires après la première et la seconde synthèse, notre premier résultat indique que l'utilisation des rétroactions se révèle être ajustée aux attentes des participants (*Utilisabilité*). De plus, il semblerait que les participants aient toutes les capacités pour utiliser ces rétroactions dans une juste mesure (*Conditions facilitatrices*). Cependant, les participants ayant commencé par utiliser *Pensum* dans le groupe-contrôle ont un jugement négatif vis-à-vis de sa pertinence dans la rédaction de synthèses (*Effizienz perçue*) et du coût en termes de ressources (*Charge cognitive*) qu'entraîne son utilisation de manière générale (*Ergonomie de l'interface et fonctionnement de base*).

Même si les synthèses ont été produites avec minutie, les questionnaires d'opinion des participants ont soulevé des faiblesses du logiciel, dont le lien avec d'éventuelles barrières à l'usage quotidien de *Pensum* est à analyser plus précisément. Pour tenter d'expliquer les résultats obtenus, nous avons effectué des entretiens avec un échantillon des participants.

### Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance

Tableau 1 - Résultats du questionnaire d'opinion et résultats du test *t*. *M* : moyenne,  $\sigma$  : écart type. Lire (ligne 2) : les participants évaluent l'efficacité du logiciel de manière significativement (valeur de *p*) négative (signe de *t*), comparée à la valeur centrale, neutre.

Critères	Synthèse 1 (Contrôle)			Synthèse 2 (Expérimental)		
	M	$\sigma$	<i>t</i>	M	$\sigma$	<i>t</i>
Efficacité du logiciel	2,90	0,94	-0,47	2,68	0,72	-1,85
Efficience	2,40	0,90	-2,99*	2,69	0,59	-2,16*
Charge cognitive	2,20	1,11	-3,24**	2,47	1,13	-1,94*
Utilisabilité	3,25	1,20	0,93	3,46	0,89	2,12*
Satisfaction	3,17	1,01	0,74	2,75	0,89	-1,14
Conditions facilitatrices	3,45	1,08	1,87	3,71	0,96	3,04*
Auto-efficacité	3,32	1,32	1,10	3,27	0,95	1,19
Intention d'utilisation	3,75	1,29	2,60*	2,88	1,29	-0,38

\*  $p < .05$  ; \*\*  $p < .01$

### Comprendre le non-usage de *Pensum* : des entretiens qualitatifs

L'opinion de 12 des participants de l'étude a été sollicitée afin de mettre en lumière leur expérience d'utilisateur de *Pensum*. Ils étaient également distribués dans les deux groupes. L'entretien avec les participants s'est déroulé par téléphone en raison de la distance les séparant, parfois élevée. Une grille d'entretien semi-directive a été élaborée de manière à aborder tous les points d'intérêt : les raisons d'adoption et les barrières à l'utilisation, la généralisation de l'utilisation de *Pensum*, l'identification des changements désirés pour le système, les sources de motivation et enfin l'exploration des résultats aux questionnaires non attendus.

#### Un avis favorable...

D'une manière générale, les participants interrogés ont émis un avis favorable sur leur utilisation de *Pensum*, 6 parmi les 12 participants interrogés répondent « oui » à la question « *Voudriez vous utiliser Pensum à*

## Recherches & Educations

*nouveau ?* », en soulignant le gain pour l'apprentissage à distance et pour la méthode pédagogique. De même, 9 participants pourraient l'utiliser dans d'autres cours et le recommanderaient à un ami.

Concernant la question « *Pourquoi utiliserez-vous Pensum à nouveau ?* » les participants identifient deux raisons. Pour la majorité d'entre eux (9 participants), *Pensum* permet d'augmenter la qualité de la synthèse. Quelques arguments donnés par les participants révèlent que *Pensum* permet, notamment, d'identifier plus facilement des éléments qui n'ont pas été pris en compte lors de la rédaction de la synthèse, de mettre en évidence les parties du résumé qui sont en surplus, et surtout de guider le travail de rédaction. La seconde raison identifiée par d'autres étudiants (pour 2 participants) est que *Pensum* permet, selon eux, d'améliorer les habiletés en écriture par l'entraînement.

De plus, 7 participants ont indiqué que *Pensum* avait eu un effet positif sur leur manière d'apprendre (« *Est-ce que Pensum a eu un effet sur votre manière d'apprendre ?* »). Les retours fournis par *Pensum* sont perçus comme ayant eut un effet, par exemple, sur une meilleure identification des mots-clés du texte et sur la manière de rédiger une synthèse. Un second aspect de l'effet de l'usage de *Pensum* sur la manière d'apprendre se situe dans l'investissement cognitif demandé aux participants. En effet, comme l'avance un participant, « *C'est un outil qui interpelle, qui demande d'aller plus loin* ». Le travail de synthèse est guidé par *Pensum* mais il requiert une analyse approfondie du texte source par l'apprenant qui n'est pas forcément réalisée lors de la rédaction de synthèses libres.

Enfin, la moitié des participants indique que l'usage de *Pensum* leur a fait gagner du temps, soit sur la réalisation de leurs synthèses, soit sur leur activité d'apprentissage. D'une part, la correction des rétroactions dispensées par *Pensum* permet non seulement d'améliorer la synthèse des participants mais aussi d'acquérir des schèmes de rédaction. En effet, pour certains participants, la correction de leurs erreurs permet d'apprendre à anticiper les rétroactions pour travailler plus rapidement et plus efficacement :

*Après, avec du recul, oui il y a un gain quant à l'acquisition et la faculté de synthétiser un texte.*

*Le logiciel fait gagner du temps car il permet de se perfectionner en synthèse. En en faisant plusieurs pour travailler nos cours, on apprend à rédiger correctement les synthèses.*

### **Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance**

Cette acquisition nécessite néanmoins un usage répété du logiciel par l'étudiant :

*Une fois qu'on le maîtrise ça peut faire gagner du temps. Peut-être qu'à la troisième utilisation ça peut être un outil intéressant.*

D'autre part, l'utilisation de *Pensum*, à travers l'analyse approfondie du texte, est perçue comme ayant un effet immédiat sur l'apprentissage du cours synthétisé. La réflexion de l'étudiant et la rédaction de la synthèse guidées par *Pensum* font alors partie intégrante du processus d'apprentissage. Ce phénomène est illustré par différents propos d'étudiants lors des entretiens, par exemple :

*À force d'utilisation, il fait gagner du temps en rapidité car on analyse plus rapidement les données.  
Mon temps de synthèse était introduit dans le temps d'apprentissage donc oui, globalement ça m'a fait gagner du temps sur mon apprentissage car c'est immédiat.*

#### **... contrasté par de vives critiques.**

En revanche, les étudiants identifient plusieurs raisons expliquant pourquoi ils n'utiliseraient pas *Pensum* à nouveau. La première réponse concerne les caractéristiques ergonomiques et notamment la présentation des textes. Il peut s'agir du texte de la synthèse : dans la version utilisée pour l'étude, *Pensum* ne fournissait qu'un éditeur de texte (pas de possibilité de jouer sur les polices de caractères) mais, en outre, le traitement des séparateurs (retour chariot, tabulation, espace) interdisait toute indentation ou sauts de ligne (dans la version courante du système, le traitement des séparateurs a été amélioré). L'éditeur reste toutefois un simple éditeur de textes :

*La mise en page n'est pas sauvegardée et c'est insupportable.*

L'agencement global des zones de textes dans l'interface manque de clarté, d'après les étudiants :

*Donc si je ne l'utilisais plus ça serait pour ça, le manque de vision globale.*

Ces témoignages gagnent en clarté une fois mis en perspective avec les réponses à d'autres questions, notamment vis-à-vis des fonctionnalités à développer en priorité :

*Le surlignage, je pense, car j'ai dû tout imprimer pour avoir le support papier, pouvoir me faire mes propres repères, c'est la culture « papier stylo » mais c'est indispensable pour moi.*

## Recherches & Educations

Ces critiques sont issues de la comparaison de *Pensum* avec les outils habituellement utilisés par les étudiants pour ce type de tâche et posent la question de son intégration dans les usages, ou plutôt de la manière dont les usages sont affectés par *Pensum*. La notion de « champ instrumental pour un utilisateur » peut nous aider à caractériser cette idée :

*« L'ensemble des schèmes d'utilisation de l'artefact où il est insérable pour former un instrument; l'ensemble des objets sur lesquels ils permet d'agir, l'ensemble des transformations, changements d'états qu'il permet de réaliser » (Rabardel, 1995, p. 119)*

Les remarques des étudiants indiquent que, par rapport à leurs pratiques habituelles, l'adoption de *Pensum* peut s'avérer difficile. Le cas de la mise en forme de la synthèse peut s'exprimer selon la notion de champ instrumental, dans la mesure où il s'agit d'un changement d'état (synthèse texte brut vers synthèse mise en forme) qui n'est plus réalisable du fait de l'usage du système (puisque la synthèse rendue à l'enseignant est, dans l'état actuel des usages, celle écrite et stockée au format texte dans *Pensum*). Le cas de l'annotation des textes-sources s'y inscrit également : *Pensum* est intégrable dans des schèmes d'utilisation qui comprennent des artefacts permettant d'annoter le texte (qu'ils soient numériques ou non). Cependant, le champ instrumental de *Pensum* est limité car il interagit directement avec les mêmes objets que ces artefacts (les textes à synthétiser) sans pour autant intégrer ses propres informations. Plus les transformations élémentaires de ces objets en « textes annotés » sont fastidieuses, plus le système risque de dissuader les apprenants les plus dépendants de ces pratiques de l'utiliser. Le témoignage ci-dessus indique qu'à les rendre trop fastidieuses, les transformations élémentaires risquent de ne pas émerger comme schème d'utilisation. Et ce, même si elles sont possibles en passant par la multiplication des représentations des textes ou, dans une moindre mesure, par le refus de certaines rétroactions. La notion de champ instrumental est liée aux usages effectifs des participants. Dans ce cas précis, cette impasse sur les fonctionnalités basiques, inhérente à ce type de prototype, est fréquemment identifiée par les participants comme un frein à l'usage.

Enfin, les rétroactions sont, elles-mêmes, remises en question. Pour certains participants, les retours de *Pensum* font l'objet de doutes en termes de fiabilité et de validité (« *Les feedback ne sont pas clairs et je n'ai pas confiance aux feedback* » ou « *Il y a de gros problèmes avec les feedback, ils ne font que 3 remarques et ils sont pour la plupart du temps faux* »). Une remise en question est également citée dans les améliora-

## Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance

tions clés à porter au système. D'autres témoignages montrent, de plus, une incompréhension des buts des fonctionnalités du système :

*Il n'y a pas de proposition de feedback pour corriger et faire le lien entre les phrases donc quand on pense avoir tout mis et que le logiciel signale une erreur de feedback, que fait-on?*

*Je pense qu'il faudrait plus d'indices sur pourquoi ce n'est pas pertinent, car on reformule une fois puis deux puis trois et ça décourage. Au moins qu'il nous guide sur ce qu'il attend.*

En s'intéressant de plus près à ces commentaires, il apparaît qu'aussi bien les doutes émis sur la fiabilité et la validité des rétroactions que les attentes d'une proposition de réponse indiquent que ces participants ont perçu les rétroactions de *Pensum* comme l'expression d'une erreur dans leur synthèse. Ces témoignages traduisent donc des attentes qui outrepassent les capacités réelles du système. Celui-ci n'a pas vocation à proposer une correction automatique ou des conseils de rédaction, mais simplement de proposer à l'utilisateur des rétroactions comme autant de questions à se poser : cette phrase de ma synthèse trahit-elle réellement l'idée exprimée dans les textes ? Cette phrase du texte a-t-elle réellement été ignorée dans ma synthèse ? L'ai-je ignorée à bon escient ? Y a-t-il réellement un saut de cohérence entre ces deux cas ? La possibilité de remettre en question chaque rétroaction proposée par *Pensum* était censée favoriser ce type de raisonnement (cf. Figure 2).

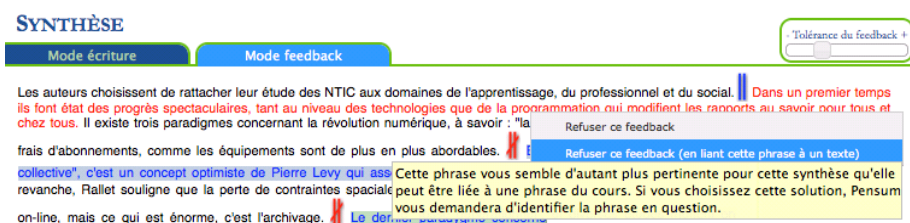


Figure 2. Rétroaction contestée par l'utilisateur, Pensum lui indique qu'il peut refuser le feedback et justifier son désaccord.

Pour préciser les modifications à apporter en priorité à *Pensum*, nous avons demandé aux participants de choisir librement quel serait le principal changement à effectuer. Il n'y a pas de consensus quant à l'amélioration à effectuer en priorité.

Outre les deux problèmes évoqués précédemment, certains étudiants ciblent des points liés au contexte d'enseignement à distance, notamment la non-réalisation de leurs attentes concernant les modalités de correction

## Recherches & Educations

de la synthèse, la communication avec les professeurs ou tuteurs et le type d'interaction avec un tiers au moyen de *Pensum*.

Le premier problème soulevé (mise en forme de la synthèse et outils d'annotation) appelle avant tout une solution technique qui, bien que chronophage, ne présente pas de difficulté particulière (des outils *open source* peuvent être intégrés au système). Afin de répondre à la question de la fiabilité des retours, une étude préliminaire a été menée afin de tenter d'estimer la précision de *Pensum* et d'identifier les leviers par lesquels rappel et précision peuvent être améliorés. Toutefois, les résultats n'étant pas encore disponibles, faire varier ces différents leviers et paramètres et les évaluer avec une nouvelle étude n'est encore qu'une perspective. En revanche, qu'il s'agisse des retours ou des interactions avec les tuteurs, des éclairages peuvent être offerts par une analyse sommaire des traces.

### Analyse des corrélations opinion-utilisation

Comme nous l'avons signalé plus haut, certains avis favorables indiquent avoir nécessité plusieurs utilisations pour réaliser l'intérêt du système. Certaines des critiques négatives témoignent d'attentes dépassant l'objet du système, mais également les possibilités des outils d'analyse sémantique. Nous nous sommes donc demandé si l'appréciation du système était liée à l'usage qui en avait été fait. En d'autres termes, ce n'est pas parce que les étudiants ont eu accès aux fonctionnalités avancées de la version expérimentale de *Pensum* qu'ils s'en sont effectivement servis. Nous avons donc réalisé une deuxième série d'analyses afin de mettre en relation leur opinion (dans le questionnaire global) avec leur utilisation effective.

Les résultats indiquent qu'il existe une corrélation modérée mais significative entre l'usage de *Pensum* et l'opinion des participants. Autrement dit, plus les participants ont reçu de rétroactions distinctes sur leur synthèse, plus ils avaient une opinion positive sur le fait que *Pensum* augmente leur curiosité à propos du contenu du cours ( $r = 0,44, p < .01$ ), que *Pensum* rend l'apprentissage plus intéressant ( $r = 0,40, p < .05$ ) et que *Pensum* les motive à explorer le cours ( $r = 0,38, p < .05$ ).

Les résultats révèlent également une corrélation modérée mais significative entre l'intention de réutiliser *Pensum* avec le nombre de rétroactions obtenues ( $r = 0,41, p < .05$ ) et avec le nombre de modifications faites sur la synthèse ( $r = 0,48, p < .01$ ). Enfin, il existe une corrélation modérée mais significative entre le nombre de rétroactions obtenues sur

### **Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance**

l'ensemble de la rédaction de la synthèse et la satisfaction de l'utilisation de *Pensum* ( $r = 0,54, p < .01$ ). Ces corrélations modérées ne nous empêchent pas de postuler une relation probable entre l'appréciation du système, la probabilité de réutilisation et l'intensité de son usage, nous mettant en présence d'un potentiel cercle « vertueux » : plus les apprenants se servent du système et plus ils risquent d'être amenés à s'en resservir d'eux-mêmes (cette étude ne nous permet évidemment pas d'en connaître l'éventuel point de rupture). Toutefois, ce potentiel cercle vertueux en révèle à notre avis un autre, sur lequel il est plus aisé d'agir : plus l'apprenant utilise le logiciel, plus il en comprend le fonctionnement et les limites, plus il est à même d'en tirer parti et de le réutiliser.

En ce sens, les instructions données, les fonctionnalités de contestation des rétroactions et l'interface correspondante doivent être améliorées pour rendre les capacités du système plus transparentes. À défaut, il est probable que tous les apprenants ne soient pas en mesure de s'approprier le logiciel avant d'en abandonner l'utilisation. Une bonne compréhension des limites du système diminuerait probablement les critiques à l'encontre de la fiabilité des rétroactions (qui reste quoi qu'il en soit un problème à considérer).

Une augmentation des interactions entre tuteurs et apprenants pourrait également permettre d'améliorer la compréhension du système par ces derniers. On peut, par exemple, imaginer des discussions entre tuteur et apprenant portant sur des rétroactions remises en question par l'utilisateur. Ce type d'interactions permettrait de savoir si certains témoignages (p. ex : « Des corrections plus utiles que ce qu'il propose ») sont le fruit d'une compréhension insuffisante du système ou bien s'il remet en question la fiabilité du système ou le type de rétroaction proposé.

Bien que les données de nos entretiens soulignent l'intérêt global des participants quant à la possibilité d'avoir des retours automatiques sur leur travail en termes d'apprentissage et de qualité de leur production, il est primordial de s'assurer que les apprenants évoluent dans un contexte favorable pour l'appropriation du logiciel en améliorant l'adéquation entre le système et les attentes des utilisateurs. Cela peut être effectué, d'une part, en circonscrivant les attentes à des fonctionnalités réalistes (amélioration de l'aide et présence accrue des tuteurs durant les premières utilisations), aussi bien qu'en étendant le spectre d'utilisation de l'outil.



## Conclusion

Nous avons mené une étude permettant à des étudiants d'un cours à distance de tester *Pensum*, un logiciel leur procurant différentes rétroactions sur leurs synthèses de cours, à un niveau sémantique. Ayant recueilli leur opinion à propos de leur expérience du logiciel, via un questionnaire fermé puis un entretien ouvert, nous avons tenté de mettre ces données en perspective pour analyser leur usage du logiciel à la lumière des attentes des apprenants, de leur activité et de leurs opinions.

L'opinion des participants de notre étude révèle une certaine satisfaction à l'égard de l'usage de *Pensum*. Particulièrement, l'opinion des participants semble être liée à l'utilisation du système : plus les rétroactions sont nombreuses et variées, plus les participants ont une opinion positive et seraient prêts à utiliser *Pensum* à nouveau. Pourtant, leur avis met en évidence des domaines problématiques (*Efficienc e perçue* et *Charge cognitive*). Ces domaines font émerger des conditions d'adoption *a priori* triviales et attendues qui tiennent en particulier au statut de prototype du logiciel : amélioration de l'ergonomie du logiciel et implémentation de fonctionnalités additionnelles prenant en compte l'intégralité de la tâche synthèse dans les modalités d'interaction avec les textes mais également avec les tuteurs (extension du spectre d'utilisation pour qu'il satisfasse mieux les attentes).

Au-delà de la solution technique, ces diverses remarques concernant les fonctionnalités de *Pensum* incitent à réfléchir en termes d'usages et de champ instrumental : plus que de retrouver la familiarité de l'interface, il semble que ce soient des usages rendus possibles par les outils habituels des apprenants qui leur manquent le plus dans le cadre de la tâche de synthèse. *Pensum* ne semble pas couvrir, à l'heure actuelle, une proportion suffisamment élevée des usages des étudiants dans le cadre de leurs pratiques habituelles pour la synthèse (voir Figure 3) – en opposition à leurs utilisations des mêmes outils (p. ex : traitement de texte) dans d'autres contextes (p. ex : écriture de courrier officiel).

### Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance

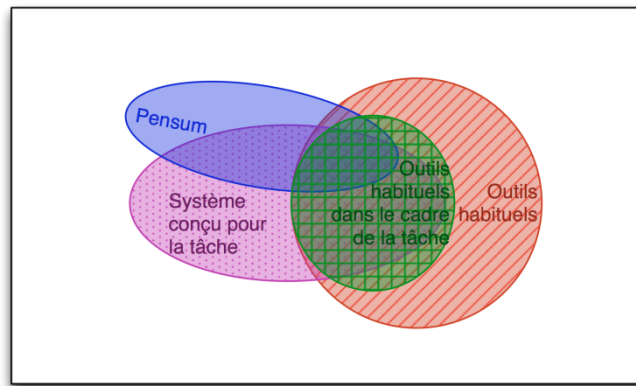


Figure 3. Champs instrumentaux respectifs d'outils dans le cadre de la réalisation d'une tâche d'apprentissage.

Dans la Figure 3, nous proposons d'exprimer les conclusions tirées pour *Pensum* en termes plus génériques. Ces conclusions peuvent constituer une hypothèse générale du non-usage d'un système d'aide à l'apprentissage, spécifiquement dans le cas où il ne s'agit pas de tenter de pointer la contre-productivité des pratiques actuelles des apprenants. Nous ne considérons pas dans ces réflexions le champ instrumental complet des artefacts que les apprenants utilisent dans le cadre de la tâche concernée par le système conçu (la synthèse de cours dans le cadre de *Pensum*). Nous nous intéressons ici à leur champ instrumental *pour les usages les plus courants des apprenants dans le cadre la tâche concernée* (ensemble quadrillé, par rapport à l'ensemble hachuré dans la Figure 3). Par exemple, dans le cadre de la synthèse de textes, on peut imaginer que la fonctionnalité de génération automatique de tables des figures dans un traitement de textes n'est pas utilisée par les apprenants et n'est, dans ce cas, pas concernée par nos analyses.

Il nous semble que, dans ce contexte, il est pertinent de tenter de proposer un outil qui soit le plus proche possible de constituer un sur-ensemble du champ instrumental des outils habituels dans le cadre de la tâche concernée (ce qui signifie que les fonctionnalités doivent être accessibles aux utilisateurs novices). L'intégration aux outils usuels, qui par leur potentielle diversité est d'autant plus difficile à atteindre, ne semble pas primordiale. En revanche, le système conçu pour la tâche devrait tenter de limiter le moins possible les possibilités habituelles (pointillés dans la Figure 3) sans quoi la charge cognitive nécessaire à l'intégration

## Recherches & Educations

du système dans les schèmes d'utilisation des apprenants risque de les en dissuader.

C'est notamment en cela qu'a pêché *Pensum* et probablement d'autres systèmes avant lui. Cette constatation fait apparaître deux recommandations. Tout d'abord, la proposition d'outils visant à aider les apprenants dans une tâche donnée devrait s'appuyer sur une analyse des usages dans le cadre de la réalisation de la tâche afin d'identifier les schèmes d'utilisation-clés. Ensuite, la proposition de fonctionnalités supplémentaires ne devrait pas interdire ces usages éprouvés (à moins qu'une réflexion sur ces usages soit précisément l'un des objets de ces fonctionnalités). Si la première recommandation est inhérente au processus de recherche, la seconde peut augmenter drastiquement les coûts de développement. Une solution pour les limiter est de se tourner vers les logiciels libres, ce qui appelle le plus souvent de distribuer son travail sous licence libre en retour.

Cette étude fait également apparaître un second ensemble de critères, qui nous paraissent essentiels pour estimer l'adoption de ce type de logiciels dans un contexte d'enseignement à distance. Les étudiants ont des attentes précises quant aux possibilités de l'outil informatique, qui leur font supposer des interactions *très proches de celles qu'un enseignant aurait* (i.e. un retour orienté et correctif), alors que les objectifs des concepteurs du logiciel sont différents puisque dédiés à l'autonomisation et à l'implication de l'apprenant à distance dans l'enseignement. Les *scénarios d'utilisation et d'interaction avec les enseignants ou tuteurs* sont à concevoir avec la plus grande attention afin de maximiser la compréhension du système par les apprenants (Villiot-Leclercq *et al.*, à paraître). Sans un niveau de compréhension et d'investissement suffisants, l'apprenant ne pourra évaluer correctement les retours et encore moins leur utilité. Or, les attentes qu'il projette par rapport à ses buts sont déterminantes. Ce processus de mise à l'épreuve de l'usage peut alors se traduire par une insertion définitive des TICE dans les activités quotidiennes ou au contraire par des rejets, des abandons et, par là même, aboutir à son non-usage (Toussaint, 1993).

Ces résultats nous permettent de faire certaines préconisations quant à l'usage des TICE dans le cadre de l'enseignement à distance. Celles-ci doivent être pensées de manière à rendre autonome l'apprenant dans son enseignement tout en lui apportant une aide à l'acquisition de compétences techniques (ici savoir rédiger une synthèse de cours). Les objectifs

## Quelques critères d'utilisation d'un outil d'évaluation automatique de synthèses de cours à distance

de *Pensum* visent à satisfaire ces préconisations mais nos résultats révèlent que cette démarche dépend, en partie, des attentes des apprenants.

## Remerciements

Cet article a bénéficié du financement partiel du projet européen LTfLL, 7<sup>e</sup> PCRD. Nous remercions vivement les étudiants ayant participé à cette étude, ainsi que les responsables de la formation FORSE de nous avoir autorisé à mener cette étude. Nous tenons aussi à remercier nos relecteurs qui nous ont permis d'enrichir et d'orienter nos analyses.

## Bibliographie

- Armitt, G., Stoyanov, S., Hensgens, J., Smithies, A., Braidman, I., Mauerhofer, C., Osenova, P., Simov, K., *et al.* (2011). *D7.4 LTfLL Validation 4*. Heerlen : OUNL.
- Clark, R. E. (2001). *Learning from media*. Greenwich : I.A.P.
- Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal*, 38(4), 813–834.
- Durel, P. (2006). Utilisation de l'assistant grammatical Antidote dans le cadre d'activités de révision-Analyse exploratoire de protocoles d'observation. *ALSIC*, 9, 33–60.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2), 211–240.
- Landauer, T. K., Laham, D., & Foltz, P. W. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems*, 15(5), 27–31.
- Loiseau, M., Dupré, D., & Dessus, P. (2011). *Pensum*, un système d'aide à la compréhension de cours à distance. In M. Bétrancourt, C. Depover, V. Luengo, B. De Lièvre & G. Temperman (Eds.), *Conférence EIAH 2011* (pp. 287–299). Mons : Presses de l'université de Mons.
- Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: toward clarification of the construct and an integrative framework for research. *Information Systems Research*, 9(2), 126–163.

## Recherches & Educations

- Murray, L., & Barnes, A. (1998). Beyond the “wow” factor – Evaluating multimedia language learning software from a pedagogical viewpoint. *System*, 26(2), 249–259.
- Nielsen, J. (1994). Usability inspection methods. In C. Plaisant (Ed.) *Conference companion on Human factors in computing systems* (pp. 413–414). New York: ACM.
- Rabardel, P. (1995). *Les hommes et les technologies*. Paris : Colin.
- Toussaint, Y. (1993). Historique des usages de la télématique. In P. Chambat (Ed.), *Communication et lien social* (pp. 193–212). Paris : Editions Des-cartes/La cité des sciences et de l’industrie.
- Venkatesh, V., Morris, M., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 26(4), 425–478.
- Villiot-Leclercq, E., Dessus, P., Mandin, S., Zampa, V. & Loiseau, M. (à paraître). Une approche par scénarios pour la conception d'un dispositif d'accompagnement en ligne. *Recherche et Formation*.