

Mesure d'ergonomie des interfaces d'ALAO et connaissances sur la relation apprenant-tâche-outil à l'ordinateur

Marie-Josée Hamel
Université d'Ottawa, marie-josee.hamel@uottawa.ca

Résumé

Dans le cadre d'une recherche-développement portant sur la conception de ressources en ligne pour l'apprentissage des langues assisté par ordinateur (ALAO), nous avons élaboré une méthode d'analyse de données empiriques adaptée du domaine de l'ergonomie des interfaces Web qui permet une mesure d'utilisabilité, c'est-à-dire de la qualité de la relation « apprenant-tâche-outil » dans un contexte spécifique d'apprentissage. La collecte se fait par le biais d'une capture vidéoécran du processus de tâche à l'ordinateur. L'analyse est construite sur la base de paramètres « observables » à l'écran, lesquels correspondent aux actions, aux comportements des apprenants, en particulier en relation avec les ressources consultées durant ce processus. Le contexte spécifique qui nous concerne est celui de la relation qu'ont les apprenants de langue avec le dictionnaire électronique en situation de tâche de production écrite. Les résultats de tests d'utilisabilité que nous avons menés nous ont permis d'améliorer un dictionnaire que nous développons et, en parallèle, de développer des connaissances sur les stratégies de consultation du dictionnaire par les apprenants.

Mots clés

Apprentissage des langues assisté par ordinateur, mesure d'utilisabilité, analyse du processus de tâche

1. Introduction

On peut définir l'ergonomie comme étant la qualité de la relation entre un utilisateur, les tâches qu'il doit accomplir et l'outil qu'il utilise pour mener à bien ces tâches (Nielsen, 1993). Plusieurs types d'analyses permettent de mesurer cette relation. Certaines sont typiquement menées en amont de la conception d'un outil (les analyses du contexte, des profils d'utilisateurs, des besoins, de la tâche, de la concurrence) tandis que d'autres, une fois l'outil prototypé (les analyses par tri de cartes, les « *walk-through* », les tests d'utilisabilité, etc.) (Kuniavski, 2003; Nogier, 2008). Toutes sont centrées sur l'utilisateur et les tâches pour lesquelles l'outil en question sera ou a été conçu. Dans cet article, nous nous intéresserons aux tests d'utilisabilité (Baccino, Bellino, & Colombi, 2005). Il s'agit de tests qui portent sur l'analyse du processus ainsi que du résultat de l'interaction « utilisateur-tâche-outil ». Ils sont fréquents dans l'industrie de la conception d'interfaces Web (cf. par exemple : <http://yucentrik.ca/>) et typiquement conduits avec recours à une technologie de capture vidéo écran (et d'oculométrie). Cette technologie permet une collecte de données dynamiques, en temps réel, d'informations enregistrées à l'écran de l'ordinateur.

En apprentissage des langues assisté par ordinateur (ALAO), la technologie de capture vidéoécran commence à donner lieu à toute une recherche centrée sur le processus d'accomplissement de la tâche langagière (cf. Fisher, 2007; Park & Kinginger, 2010; Smith, 2008). Il existe cependant à ce jour peu de méthodes pour en guider l'analyse et nous croyons que les techniques employées dans l'industrie de la conception d'interfaces Web peuvent servir de repères. Le domaine de la recherche documentaire (*Web searches; information retrieval*) fournit quant à lui aussi des modèles pour l'analyse de l'activité de recherche, notamment des parcours de navigation (cf. Boubée & Tricot, 2010; Kovacs, Gaunet, & Briffault, 2004).

Inspirée par ces pratiques, et dans le cadre d'une recherche-développement (Guichon, 2007; Rabi, 2005) ancrée dans le domaine de la conception d'interfaces d'ALAO (Colpaert, 2004; Depover, 2000; Hémar, 2003), nous avons élaboré une méthode d'analyse de données empiriques pour tester l'utilisabilité d'un prototype de dictionnaire en ligne pour apprenants de français langue seconde que nous développons (Hamel, 2010), et ce, dans avec pour but de l'améliorer. Dans cet article, nous décrivons cette méthode et présentons son contexte d'application ainsi qu'une synthèse du type de résultats qu'elle a permis d'obtenir.

2. Méthode

2.1 Conception de tests d'utilisabilité

La conception de tests d'utilisabilité repose en amont sur la connaissance de l'utilisateur : son profil, ses besoins, ses outils, son environnement de travail et, surtout, de ses tâches que l'on cherche à modéliser. L'élaboration de tâches qui fourniront des conditions à la fois authentiques et idéales d'observation de l'interaction « utilisateur-tâche-outil » à l'ordinateur est donc clé. Par authentiques, on entend des tâches qui ressemblent de très près, voire qui sont, celles de l'utilisateur ciblé tandis que par conditions idéales, on entend des conditions qui maximiseront, voire forceront le recours à l'outil testé. Ces tâches pourront solliciter des fonctionnalités en particulier de l'outil que l'on souhaite tester.

2.1.1 Contexte

L'utilisateur visé par la présente recherche est un jeune adulte, apprenant intermédiaire-avancé de français langue seconde (niveau B2). Étant donné son profil et ses besoins (Hamel & Milicevic, 2007), nous savons qu'il devra faire appel à l'outil pour accomplir avec succès (certains items de) la tâche.

L'outil testé est le prototype dictionnaire électronique *Dire autrement* (<http://web5.uottawa.ca/direautrement/>). Ce dictionnaire comporte un certain nombre de fonctionnalités permettant la recherche et la consultation d'informations lexicales qui portent, notamment, sur la synonymie et la cooccurrence des mots (Hamel, 2010). C'est un dictionnaire orienté sur la production, et en particulier la reformulation du texte (Hamel & Milicevic, 2007).

La tâche que nous avons conçue pour tester l'utilisabilité du prototype est une (micro-) tâche de production écrite orientée sur la connaissance de la synonymie et de la cooccurrence lexicale des apprenants visés. Elle comporte des items lexicaux (semi-contrôlés par le contexte phrastique) à remplacer ou à trouver dans des phrases ou dans un texte, ainsi conçue de sorte à maximiser l'accès au prototype (exemple d'item : *avoir de l'admiration* à *éprouver; ressentir*). Nous avons cherché à mesurer combien utilisé et combien utile s'est avéré le recours au prototype par des apprenants de langue dans ce contexte spécifique de tâche.

2.2. Définition de paramètres d'analyse

La conception de tests d'utilisabilité repose en aval sur la définition de paramètres qui serviront à l'analyse de l'interaction « utilisateur-tâche-outil » à l'ordinateur. Ces paramètres permettront une mesure (essentiellement quantitative, mais aussi qualitative) des données empiriques collectionnées sous forme de vidéos. Ils pourront être établis en fonction d'une part, des fonctionnalités spécifiques de l'outil que l'on cherche à mesurer et d'autres, des comportements de l'utilisateur. Certains de ces comportements sont prévisibles, alors que d'autres émergeront de l'analyse des données empiriques. Les comportements de l'utilisateur font référence à des (traces d') actions « observables » (Kovacs et al., 2004) visibles à l'écran l'ordinateur. Mises à la suite les unes des autres, elles formeront des parcours que l'on cherchera à mesurer. Puisque la mesure d'utilisabilité se rapporte tant au processus qu'au résultat de l'interaction « utilisateur-tâche-outil » à l'ordinateur, les paramètres d'analyse seront de deux types : efforts/durée pour la mesure du processus et qualité/quantité pour celle du résultat. Grosso modo, l'équation moins d'efforts/temps, plus de qualité/quantité équivaudra à un parcours plus ergonomique et à l'inverse, celle de plus d'efforts/temps, moins de qualité/quantité à un parcours moins ergonomique.

2.2.1 Mesure de l'interaction apprenant-tâche-dictionnaire

Les paramètres que nous avons définis pour la mesure orientée sur le processus de l'interaction « apprenant-tâche-dictionnaire » à l'ordinateur sont des paramètres qui visent les fonctions de recherche (recherche par mots clés ou par champs sémantiques, par exemple) et de consultation du prototype de dictionnaire (consultation des collocatifs ou de la définition, par exemple). Ils concernent aussi les comportements anticipés de l'apprenant dans ce contexte spécifique de tâche d'encodage du texte : production/révision d'input, va-et-vient entre la tâche et l'outil, pauses, etc. Les paramètres d'efficience ont été comptabilisés en termes de types et nombre d'efforts d'accès aux fonctions de recherche et de consultation du dictionnaire; ceux d'efficacité, en termes d'items lexicaux réussis, passables, non réussis et nil (non répondus).

3. Résultats

À ce jour, nous avons conduit trois tests d'utilisabilité, un avec des dictionnaires électroniques usuels (Hamel & Caws, 2010) et deux avec le prototype de dictionnaire *Dire autrement* (Hamel, 2012, à paraître, soumis). La procédure de collecte de données pour les deux premiers tests s'est déroulée comme dans l'industrie, c'est-à-dire avec un petit nombre de participants ($n = 6$), un nombre suffisant pour saturer l'information recherchée (Nielsen, 1993). Ils sont venus tour à tour accomplir la tâche langagière à l'ordinateur. Le dernier test, plus récent (Hamel, à paraître), s'est déroulé dans des conditions plus « écologiques ». La tâche langagière accomplie à l'ordinateur en laboratoire de langue par les participants ($n = 17$) était partie intégrante d'un cours de vocabulaire. Pour les trois tests, les vidéoécrans du processus d'accomplissement de la tâche à l'ordinateur ont été annotées de paramètres observés et les résultats compilés. Leur analyse a permis dans le cas du premier test d'attester des décisions de conception et de contenu tandis que de faire ressortir, dans le cas du pilote surtout, les aspects forts et faibles du prototype. Ces tests, deux premiers en particulier, ont conduit à des recommandations qui nous ont menées à des améliorations du prototype (Hamel, 2011; Hamel & Caws, 2010). Nous avons par exemple renommé certains onglets, supprimé des informations complexes aux yeux des apprenants, développé des exercices qui exploitent les collocations et synonymes de chaque lexie du dictionnaire. Nous avons surtout attesté des décisions de conception.

Ces tests d'utilisabilité ont par ailleurs aussi révélé des connaissances empiriques sur la relation « apprenant-tâche-dictionnaire » à l'ordinateur, le dernier test en particulier (Hamel, à paraître, soumis). Ainsi se sont dégagés des profils d'apprenants utilisateurs de dictionnaire (Hamel, à paraître). On a observé des apprenants efficaces et efficaces qui accomplissent la tâche avec succès, en peu d'efforts et, par conséquent, en moins de temps. D'autres sont efficaces, en ce qu'ils réussissent aussi la tâche, mais moins efficaces, surtout en amont du processus de recherche de mots. Cependant, ils apprennent vite de sorte que rendus à la mi-tâche, ils ont mis en place des stratégies plus efficaces de recherche de mots dans le dictionnaire. Dans cette catégorie, on trouve par ailleurs aussi des apprenants qui consultent systématiquement le sens des mots, qui lisent davantage, en gros qui « prennent leur temps ». Un troisième profil est celui d'apprenants techniquement efficaces qui déploient un minimum d'efforts durant la tâche et qui sont peu efficaces, réussissant mal la tâche, tant en quantité qu'en qualité. Enfin, un quatrième type de profil est celui d'apprenants peu efficaces et peu efficaces. Ces apprenants éprouvent beaucoup de difficulté à compléter la tâche durant laquelle ils dépensent beaucoup d'efforts tant au niveau de la recherche et que de la consultation d'informations dans le dictionnaire. Une analyse plus fine (qualitative) des résultats a permis de mettre en lumière la nature des choix lexicaux que font les apprenants dans le dictionnaire (Hamel, soumis). Ainsi, on a pu observer une sélection lexicale prudente, de mots fréquents, souvent congénères de l'anglais, de mots situés à proximité du mot recherché au niveau des résultats proposés par le dictionnaire (Hamel, soumis).

4. Conclusion

En conclusion, les techniques de mesure d'utilisabilité de l'industrie de la conception d'interfaces Web couplées aux modèles d'analyse issus du domaine de la recherche documentaire ont permis de guider non seulement notre analyse du processus et du produit de l'interaction « apprenant-tâche-dictionnaire », mais aussi de fournir un éclairage sur les comportements des apprenants dans ce contexte. Les résultats ont conduit à des recommandations pour l'amélioration de l'outil, mais en parallèle aussi à des recommandations pour l'optimisation des comportements des apprenants en processus de tâche, à savoir : réfléchir davantage au processus (se détacher du produit pour se concentrer sur les étapes de la tâche); déconstruire ce processus pour mieux le modéliser, notamment les stratégies exemplaires ayant trait à la recherche et la consultation d'informations lexicales dans les dictionnaires électroniques. Au final, l'ergonomie en ALAO vise la pleine autonomie des apprenants investis dans des tâches langagières à l'ordinateur avec des outils dont on aura mesuré l'utilisabilité et dont on connaîtra par conséquent les fonctionnalités, les forces et les faiblesses.

Références

- Baccino, T., Bellino, C., & Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces*. Paris, France : Hermès Science.
- Boubée, N., & Tricot, A. (2010). *Qu'est-ce que rechercher de l'information?* Villeurbanne, France : Presses de l'enssib.
- Colpaert, J. (2004). *Design of online interactive language courseware. Conceptualization, specification and prototyping: Research into the impact of linguistic-didactic functionality on software architecture* (Thèse de doctorat inédite). Université d'Anvers, Anvers, Belgique.

- Depover, C., Quintin, J.-J., & De Lièvre, B. (2000). La conception d'environnement d'apprentissage : de la théorie à la pratique/de la pratique à la théorie. *Alsic*, 3(1), 3-18.
- Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *CALICO Journal*, 20, 409-442.
- Guichon, N. (2007). Recherche-développement et didactique des langues. *Les Cahiers de l'Acedle*, 4, 37-54.
- Hamel, M.-J. (2010). Prototype d'un dictionnaire électronique de reformulation pour apprenants avancés de français langue seconde. *Cahiers de l'APLIUT*, XXIX(1), 73-82.
- Hamel, M.-J. (2012). Testing aspects of the usability of an online learner dictionary prototype: A product and process oriented study. *Computer Assisted Language Learning*, 25(4), 339-365.
- Hamel, M.-J. (à paraître). Analyse de l'activité de recherche d'apprenants dans un prototype de dictionnaire en ligne. *ALSIC*.
- Hamel, M.-J. (soumis). *Analyse de la sélection lexicale d'apprenants dans un prototype de dictionnaire électronique*.
- Hamel, M.-J., & Caws, C. (2010). Usability tests in CALL development: Pilot studies in the context of the Dire autrement and the FrancoToile projects. *CALICO Journal*, 27, 491-504.
- Hamel, M.-J., & Milicevic, J. (2007). Analyse d'erreurs lexicales d'apprenants du FLS : Démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage. *Revue canadienne de linguistique appliquée*, 10, 25-45.
- Hémard, D. (2003). Language learning online: Designing towards user acceptability. Dans U. Felix (dir.), *Language learning online: Towards best practice* (p. 21-46). Lisse, Pays-Bas : Swets & Zeitlinger.
- Kovacs, B., Gaunet, F., & Briffault, X. (2004). *Les techniques d'analyse de l'activité pour l'IHM*. Paris, France : Lavoisier.
- Kuniavski, M. (2003). *Observing the user experience. A practitioner's guide to user research*. San Francisco, CA : Morgan Kaufmann Publishers.
- Milicevic, J., & Hamel, M.-J. (2007). Un dictionnaire de reformulation pour apprenants avancés du français langue seconde. *Revue de l'Université Moncton*, numéro hors série, 145-167.
- Nogier, J.-F. (2008). *Ergonomie du logiciel et design Web : le manuel des interfaces utilisateur (4^e éd.)*. Paris, France : Dunod.
- Nielsen, J. (1993). *Usability engineering*. San Francisco, CA : Morgan Kaufmann.
- Park, K., & Kinginger, C. (2010). Writing/thinking in real time : Digital video and corpus query analysis. *Language Learning & Technology*, 14, 31-50.
- Rabi, F. (2005). A user-centered ergonomic approach to CALL research. Dans J. Egbert, & G. Petrie (dir.), *CALL research perspectives* (p. 179-190). Cambridge, MA : CUP.
- Smith, B. (2008). Methodological hurdles in capturing CMC data : The case of the missing self-repair. *Language Learning and Technology*, 12(1), 85-103.