

Interpréter les liens entre utilisabilité et utilité des documents électroniques

André Tricot

*CERFI, IUFM de Toulouse, 56 avenue de l'URSS, 31 078 Toulouse cedex 4
& Laboratoire Travail et Cognition, UMR CNRS et Université de Toulouse 2*

andre.tricot@toulouse.iufm.fr

Résumé :

On admet généralement (*e.g.* Grudin, 1992) qu'un bon logiciel (par extension un bon document électronique) est à la fois utile et utilisable. Un document électronique très facile à utiliser, mais dans lequel on ne trouve aucune information n'a pas d'intérêt. Pas plus qu'un document électronique inutilisable, dans lequel se trouvent des informations très intéressantes pour l'utilisateur. Ceci étant, l'interprétation des liens entre l'utilité et l'utilisabilité est assez difficile. Par exemple, Nielsen (2000) rapporte que dans la majorité des cas, les utilisateurs ne trouvent pas ce qu'ils cherchent sur le Web. Il interprète ce résultat comme lié à un problème d'utilisabilité du Web, mais n'apporte aucune preuve en faveur de son interprétation. Peut-être, au moins pour une part, que ce résultat est lié à un problème d'utilité. Dans ce tutorial, je propose un cadre formel pour l'interprétation de quinze types de relations entre les variables d'utilité et d'utilisabilité, ainsi qu'une définition des mesures d'utilisabilité et d'utilité par les critères d'efficience, d'efficacité, de degré d'atteinte du but ou de degré d'utilisation. Je montre les liens entre ces mesures et : (a) les mesures d'efficacité spécifiques au domaine des documents électroniques (ou des bases de données : mesure de rappel, de précision, d'économie) ; (b) les mesures classiques d'utilisabilité (notamment par les critères d'apprenabilité, de prévention des erreurs, de satisfaction). Ainsi, ce tutorial tente de définir sous quelles conditions on peut espérer qu'une amélioration de l'utilisabilité d'un document électronique s'accompagne d'une amélioration de son utilité, et finalement, de sa qualité. Pour l'essentiel, ce tutorial reprend une communication de Tricot et Tricot (2000).

Mots clés : évaluation, utilisabilité, utilité, document électronique, ergonomie

1. Introduction

Lorsqu'on évalue un document électronique (et plus généralement un objet finalisé¹), on doit évaluer sa mise en œuvre (son utilisation) et ses résultats (les buts qu'il permet d'atteindre, leur conformité aux buts attendus). La mise en œuvre et les résultats sont décrits par des variables : respectivement, les variables d'utilisabilité et d'utilité. Grudin (1992) a discuté de ces deux notions dans le domaine du logiciel, décrit les champs de recherche et les pratiques qu'elles recouvrent et souligné la distance, difficile à franchir, qui les sépare. Il appelle de ses vœux un rapprochement des champs de recherche et espère une possible prise en compte des deux dimensions dans les pratiques de conception. Il me semble que, du côté de l'évaluation des documents électroniques, les progrès à accomplir concernent d'abord la définition des deux variables et des relations qu'elle entretiennent. Dans ce tutorial, je propose donc un cadre formel pour l'interprétation des liens entre ces deux variables. À ma connaissance en effet, une définition rigoureuse des relations entre utilisabilité et utilité n'existe pas, et une telle définition manque. La proposition de Le Coadic (1997) à ce propos me semble correcte mais trop peu précise. Pour cet auteur en effet, l'utilisabilité « peut servir d'antécédent causal pour l'utilité » (p. 56), ce qui pour lui, semble vouloir dire non-utilisabilité implique non-utilité et donc utilité implique utilisabilité. Si le principe de cette proposition me semble convenable, je regrette que l'auteur n'indique pas comment on exploite des résultats concernant l'utilité et l'utilisabilité (ce n'est d'ailleurs pas son but). Le Coadic n'envisage pas non plus les cas intermédiaires, où par exemple un système d'information est moyennement utile et plutôt utilisable. En bref, plutôt qu'une relation d'implication entre deux variables, j'envisage la description logique de l'ensemble des relations possibles entre ces deux variables, pour quelques valeurs standards de ces variables.

Selon moi, il manque aussi une définition rigoureuse des mesures d'utilisabilité et d'utilité par les critères d'efficacité, d'efficacités, de degré d'atteinte du but ou de degré d'utilisation. Je propose dans ce tutorial une définition de chacun de ces termes et un cadre formel pour interpréter les relations entre utilisabilité et utilité. Le cadre que je propose s'applique, sous certaines conditions que je définis, à tout objet finalisé que l'on voudrait évaluer et à toute comparaison entre objets finalisés.

L'évaluation de l'utilisabilité des logiciels est un enjeu important depuis quelques années (depuis la publication en 1993 de l'ouvrage de Nielsen les références se comptent en milliers). Mais, plus encore que les logiciels, ce thème de recherche et d'ingénierie ergonomique concerne les documents électroniques, et en particulier les sites Web (Buckingham Shum & McKinght, 1997). Le succès de la notion d'utilisabilité est tel que tout se passe comme si l'utilisabilité devenait

¹ Je ne m'occupe pas ici d'instruments au sens de Rabardel (1995), qui définit l'instrument comme un objet technique, ou *artefact* + un schème d'utilisation. Je traite d'« objets finalisés », c'est-à-dire d'objets techniques conçus pour être utilisés par certains sujets humains (les utilisateurs) et pour leur permettre d'atteindre certains buts définis dans certains environnements définis. Ainsi, on considèrera qu'un document électronique est conçu pour être utilisé par certains utilisateurs afin qu'ils y trouvent certaines informations pertinentes dans certains environnements de travail, de loisirs, d'apprentissage, etc.

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

progressivement synonyme de qualité générale d'un document électronique. Par exemple, les cinq critères d'utilisabilité proposés par Nielsen (1993) intègrent : l'efficacité (c'est le fait d'atteindre sans perdre trop de temps le but que l'on s'est fixé) ; l'apprenabilité (c'est la facilité ou la rapidité avec laquelle l'utilisateur apprend à utiliser le système d'information) ; la mémorisation (c'est le fait que l'utilisateur parvienne à mémoriser « comment ça marche » et plus généralement « ce qu'il a fait ») ; la fiabilité (c'est la prévention ou la gestion des erreurs par le système) ; la satisfaction de l'utilisateur. On peut s'étonner que le critère d'efficacité (soit un critère lié au but, et pas seulement à l'utilisation) soit considéré comme un critère d'utilisabilité. Certains spécialistes (Meads & Stubbs, 2001) vont même jusqu'à intégrer le critère d'utilité parmi les critères d'utilisabilité. En bref, il me semble que non seulement l'utilisabilité fonctionne comme un critère d'évaluation globale de la qualité des documents électroniques (pour Sullivan (1997), « valeur », « qualité » et « utilisabilité » d'un site Web sont mêmes synonymes), mais aussi qu'une certaine confusion est liée à la notion d'utilisabilité et à ses liens avec l'utilité.

Les critères ergonomiques de Scapin et Bastien (1997) sont un bon exemple de définition rigoureuse de critères d'évaluation ergonomique des systèmes d'information qui n'utilise pas la notion fourre-tout d'utilisabilité. Parmi les critères de Scapin et Bastien, l'adaptabilité est assez proche d'une certaine acception de l'utilité. Ma proposition est cependant assez éloignée de celle de ces auteurs, car l'adaptabilité est strictement dépendante des buts de l'utilisateur, alors que l'utilité telle que je la conçois est l'adéquation entre les buts de l'utilisateur et la finalité du document électronique (ou du logiciel, d'ailleurs).

Dans ce tutorial, j'admets donc avec Grudin (1992) qu'un bon document électronique est à la fois utile et utilisable. Je ne peux pas suivre Nielsen (1999) quand il rapporte que dans la majorité des cas, les utilisateurs ne trouvent pas ce qu'ils cherchent sur le Web et qu'il interprète ce résultat comme lié à un problème d'utilisabilité. En effet, ce résultat est peut-être lié à un problème d'utilité : le Web n'est peut-être pas utile pour toute recherche d'information, de produits, de références, etc. Il me semble que la confusion liée au concept d'utilisabilité est assez générale, y compris dans les travaux les plus récents et publiés dans les meilleures revues. Par exemple, Park (2000) a réalisé une étude comparative de deux types de système d'information où il mesure l'utilisabilité, les préférences des utilisateurs, l'efficacité des systèmes ainsi que les comportements des utilisateurs. Pour évaluer l'utilisabilité, l'auteur utilise trois critères : l'opinion des utilisateurs quant à la facilité d'utilisation, leur degré de satisfaction quant aux résultats de leur recherche d'information et leur jugement sur le temps mis pour réaliser leur recherche (suffisant ou pas). Il distingue ces critères d'utilisabilité de la satisfaction des utilisateurs (sic), et de l'efficacité du système, qui est mesurée par le critère classique de « taux de rappel » (voir définition note 3). Enfin, il caractérise le comportement des utilisateurs par une série de treize critères, parmi lesquels on trouve... le temps passé pour réaliser la recherche d'information, soit une mesure non-pondérée (re-sic) de l'efficacité.

2. Définitions, principes, cadre de travail

2.1. Utilité et pertinence

On aura compris dans l'introduction que le concept d'utilité présente une analogie forte avec le concept de pertinence, concept lui-même au centre des sciences du document (ou « sciences de l'information » au sens des pays anglo-saxons). Dans cette discipline, les deux concepts ont quasiment la même valeur, ils peuvent être des synonymes : une information utile est une information pertinente et vice-versa. Un document électronique est un support contenant ou non des informations pertinentes / utiles. Par extension, un document contenant une information pertinente / utile est un document pertinent / utile. L'intérêt de cette synonymie en sciences du document réside dans la possibilité d'exploiter le travail immense qui a été consacré aux tentatives de définition de la notion de pertinence. Mizzaro (1998) a consacré une revue de littérature recensant près de 160 articles proposant une définition de la notion de pertinence. Pour cet auteur, la pertinence est une relation entre deux groupes : dans un groupe il y a un document, des descripteurs de ce document et de l'information (ce qui est reçu par l'utilisateur); dans l'autre groupe, il y a un problème d'un utilisateur, son besoin d'information (sa représentation du problème en termes informatifs), sa requête « naturelle » et sa requête « formalisée ». Chacune de ces entités peut être décomposée en trois registres : le domaine de contenu, l'exploitation que l'utilisateur va faire avec l'information trouvée et l'environnement (de travail, d'apprentissage, de recherche, etc). La pertinence est l'adéquation entre chacune des entités des deux groupes (le système d'information d'un côté, l'utilisateur de l'autre), pour les trois composants (domaine, exploitation, environnement).

Je retiens donc cette conception de la pertinence et l'applique à la notion d'utilité : l'utilité d'un document électronique est l'adéquation entre la finalité du document électronique et le but de l'utilisateur, pour un domaine, une exploitation et un environnement donnés.

2.2. Contextes d'usage des évaluations d'utilisabilité et d'utilité ; définition des conditions d'application du cadre interprétatif

On évalue l'utilisabilité et l'utilité dans deux contextes principaux : la conception et l'expérimentation. Dans ces deux contextes, l'évaluation consiste à faire utiliser le document électronique par des utilisateurs, dans des environnements et pour des buts qui sont censés, tous les trois (utilisateur, environnement, but²), représenter un

² Je distingue évidemment but, tâche et activité, de façon tout à fait classique (Leplat & Hoc, 1983). À un niveau plus général je défini un « projet » comme intégrant le problème général de l'utilisateur (pourquoi a-t-il ce but?) et l'exploitation du résultat obtenu (comment va-t-il exploiter l'état obtenu de l'environnement, l'information trouvée?). Si l'activité du sujet est affectée par son projet, la description rationnelle de la tâche ne l'est pas (*i.e.* une même description rationnelle peut concerner deux projets différents).

contexte d'utilisation typique. Je ne discute pas ici des conditions d'obtention de cette représentativité. J'insiste seulement sur le fait que si l'un des trois paramètres (utilisateur, environnement, but) n'est pas représentatif, alors l'évaluation n'a que peu de sens. Plus généralement, la signification de l'évaluation est strictement relative aux modalités « utilisateur, environnement, but » choisies.

Dans le contexte de la conception, l'évaluation est une phase au cours de laquelle on vérifie si le document que l'on conçoit permet aux utilisateurs d'atteindre les buts pour lesquels on a conçu le document électronique en question (utilité). On vérifie aussi que les sujets atteignent le but dans de bonnes conditions (utilisabilité). Dans le contexte de l'expérimentation, on peut être conduit à vouloir décrire l'activité de sujets pour telle tâche. Pour décrire cette activité, on peut naturellement être conduit à utiliser des variables d'utilité (performance des sujets en termes d'atteinte du but) ou des variables d'utilisabilité (par exemple l'apprenabilité : le temps d'apprentissage). La conception et l'expérimentation peuvent, en gros, être conduites :

- dans le cadre d'une comparaison avec un autre document électronique : on compare alors les performances des utilisateurs avec les deux objets finalisés ;
- dans le cadre d'une situation de référence où un autre document électronique n'existe pas : ce type de démarche est très délicate, car il faut avoir un modèle *a priori* de l'utilisabilité recherchée, voire de l'apprenabilité, de l'efficacité, du but à atteindre, etc.

En résumé donc, les conditions d'application du cadre proposé sont les suivantes : pour évaluer l'utilisabilité et l'utilité d'un document électronique, il faut que lors de l'évaluation, l'ensemble des propriétés de la tâche et si possible du projet (voir note 2), ainsi que l'ensemble des caractéristiques des utilisateurs - testeurs soient comparables à celles des tâches et des utilisateurs potentiels. En outre, si l'on n'est pas dans le cadre d'une comparaison (de documents électroniques), il faut disposer d'un modèle *a priori* de l'utilisabilité et de l'utilité attendue du document électronique.

2.3. Définitions

- L'utilisabilité d'un document électronique désigne la possibilité d'utilisation de ce document.
- L'utilité d'un document électronique désigne la possibilité d'atteindre un but visé avec ce document, *i.e.* trouver de l'information pertinente.
- L'utilisabilité et l'utilité peuvent être mesurées en termes absolus (possibilité vs impossibilité) ou relatifs (efficacité, efficacité, degré de pertinence de l'information trouvée ou d'utilisation du document électronique).
- Le degré d'atteinte du but est une quantification relative à un groupe d'utilisateurs (quelle proportion du groupe a-t-elle atteint le but ?) voire au but lui-même (quel est le degré de pertinence de l'information trouvée ? *i.e.* quelle proportion du but a-t-elle été atteinte ?).

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

- Le degré d'utilisation est une quantification relative au document (quelle part, quels éléments, quelles fonctions, du document ont-ils été utilisés, pendant combien de temps, etc. ?).
- L'efficacité mesure l'économie en termes généraux (*e.g.* économie d'énergie, d'efforts, d'argent, de nombre de clics sur la souris, d'attention, etc.) de l'utilisation d'un document électronique et de la pertinence de l'information trouvée. L'efficacité peut être mesurée par les critères de « taux de rappel », de « taux de précision » et d'économie³ (voir Tricot et al., 1999, pour une présentation et une discussion de ces critères dans le domaine des hypertextes).
- L'efficience mesure l'économie en termes de temps de l'utilisation d'un document électronique et de la pertinence de l'information trouvée.
- L'efficacité et l'efficience, ainsi que toutes leurs mesures, n'ont pas de valeur absolue. Cette valeur est strictement relative au document électronique et à l'environnement dans lequel le document électronique est utilisé. Cette valeur doit être définie avant l'évaluation de l'objet.
- La qualité d'un document électronique est mesurée par son utilisabilité et par son utilité.

2.4. Formalisme

Soit une tâche à réaliser par un sujet, composée d'un but B, d'un document électronique D, d'un environnement de travail E.

À l'état initial, l'environnement est dans un état zéro, noté E_0 . Le document électronique est dans un état zéro, noté D_0 . Le document électronique va changer d'état. Il existe au moins un état D_B , correspondant au but. L'action sur le document et/ou sur l'environnement va avoir un effet sur l'environnement. Il existe au moins un état E_B , correspondant au but.

Il existe donc un ensemble $\mathcal{D} = \{D_0, D_a, D_b, \dots, D_B\}$ des états de D et un ensemble $\mathcal{E} = \{E_0, E_a, E_b, \dots, E_B\}$ des états de E. Le but B peut donc être décrit par le couple (D_B, E_B) . \mathcal{D} et \mathcal{E} peuvent être des échelles ordinales ou non. Il existe un sous ensemble $\mathcal{D}' \subset \mathcal{D}$, ordonné ou non, correspondant à la façon la plus efficace d'atteindre l'état D_B . Il peut être pertinent, dans les environnements «à temps contraint», de pondérer \mathcal{D}' et le cas échéant \mathcal{E}' par le temps passé à atteindre le but. On mesure alors l'efficience. La représentation de la tâche peut être composée d'une représentation de l'ensemble des états du document électronique $R(\mathcal{D})$, de la façon efficace ou efficiente d'atteindre le but $R(\mathcal{D}')$ et d'une représentation de l'ensemble des états de l'environnement $R(\mathcal{E})$.

³ Les mesures de rappel et de précisions sont les plus classiques : taux de rappel = nombre de cibles atteintes / nombre de cibles existantes ; taux de précision = nombre de cibles atteintes / nombre d'items ouverts. Je rajoute la mesure d'économie, qui peut être particulièrement pertinente dans le domaine des documents électroniques, en particulier celui des hypertextes, dont on sait qu'ils peuvent entraîner des parcours très redondants : économie = 1 - (nombre d'items différents ouverts / nombre total d'items ouverts).

Une variable VD sert à décrire les états de D, avec $D_B = 1$ ou bien, si '?' est défini, '?' = 1; tout autre élément D_i de l'ensemble ? : $0 \leq D_i < 1$. Une variable VE sert à décrire les états de E, avec $E_B = 1$ et tout autre élément E_i de l'ensemble ? : $0 \leq E_i < 1$. VD et VE peuvent être des variables binaires, dans ce cas, $\forall D_i \neq D_B, D_i = 0$. Et $\forall E_i \neq E_B, E_i = 0$.

VD et VE peuvent être des variables non-binaires, discrètes ou continues. On peut alors décrire non seulement la distance entre $\forall O_i$ et O_B ou celle qui existe entre $\forall E_i$ et E_B , mais aussi décrire la distance entre tout ensemble { D_0, D_a, \dots } et l'ensemble '?'.

3. Interprétation des liens entre les variables d'utilisabilité et d'utilité

Les liens entre les variables d'utilisabilité et d'utilité sont décrits par la fréquence f des co-occurrences des états de VD et VE. La somme de ces fréquences est évidemment égale à 1. L'analyse implicative (Bernard & Charron, 1996) permet, à partir de la distribution des fréquences dans cette table de contingence, d'inférer la relation logique entre ces deux variables. Il s'agit simplement donc de considérer une table de contingence 2x2 comme une ligne de la table de vérité : les fréquences = 0 correspondant aux états «faux» (disons impossibles), les fréquences > 0 correspondant à des états « vrais » (disons possibles). On considère que la somme = 1 des fréquences > 0 correspond à une équirépartition de ces fréquences. Voici quelques exemples d'interprétation des liens entre ces deux variables pour quelques valeurs standards des fréquences.

3.1. Cas des variables binaires

	VD (noté VD_1)	\neg VD (noté VD_0)	
VE (noté VE_1)	$f_{VD_1 VE_1}$	$f_{VD_0 VE_1}$	f_{VE_1}
\neg VE (noté VE_0)	$f_{VD_1 VE_0}$	$f_{VD_0 VE_0}$	f_{VE_0}
	f_{VD_1}	f_{VD_0}	

Table de contingence

	VD_1	VD_0
VE_1	$f=1$	$f=0$
VE_0	$f=0$	$f=0$

Document électronique parfait : utilisable et utile

Conjonction : $VD \wedge VE$

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

	VD ₁	VD ₀
VE ₁	0	0
VE ₀	1	0

Document électronique utilisable mais inutile

Implication : $VD \Rightarrow \neg VE$

	VD ₁	VD ₀
VE ₁	0	0
VE ₀	0	1

Document électronique mauvais ou inadéquat : inutile et inutilisable

NOR (non ou) ou " ni ...ni ..." : $VD \downarrow VE$

	VD ₁	VD ₀
VE ₁	0	1
VE ₀	0	0

Document électronique paradoxal

Implication : $\neg VD \Rightarrow VE$

	VD ₁	VD ₀
VE ₁	0,5	0
VE ₀	0	0,5

Document électronique « spécifique » (nécessaire et suffisant)

Équivalence : $VD \Leftrightarrow VE$

	VD ₁	VD ₀
VE ₁	0	0,5
VE ₀	0,5	0

Document électronique nuisible

Ou exclusif : $VE \omega VD$

	VD ₁	VD ₀
VE ₁	0,5	0
VE ₀	0,5	0

Document électronique utilisable mais moyennement utile

Indépendance : VD est vrai, $\forall VE$

	VD ₁	VD ₀
VE ₁	0	0,5
VE ₀	0	0,5

Document électronique inutilisable

Indépendance : VD est faux, $\forall VE$

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

	VD ₁	VD ₀
VE ₁	0,5	0,5
VE ₀	0	0

Document électronique placebo

Indépendance : VE est vrai, \forall VD

	VD ₁	VD ₀
VE ₁	0	0
VE ₀	0,5	0,5

Document électronique inutile, bien que moyennement utilisable

Indépendance : VE est faux, \forall VD

	VD ₁	VD ₀
VE ₁	0,33	0,33
VE ₀	0	0,33

Document électronique suffisant mais non nécessaire.

Implication : VD \Rightarrow VE

	VD ₁	VD ₀
VE ₁	0,33	0
VE ₀	0,33	0,33

Document électronique nécessaire mais pas suffisant

Implication : VE \Rightarrow VD

	VD ₁	VD ₀
VE ₁	0,33	0,33
VE ₀	0,33	0

Document électronique peu spécifique (ni nécessaire ni suffisant)

Ou inclusif : VE \vee VD

	VD ₁	VD ₀
VE ₁	0	0,33
VE ₀	0,33	0,33

Document électronique nuisible : la réussite à l'utilisation implique l'échec au but

Fonction NAND VE / VD : incompatibilité

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

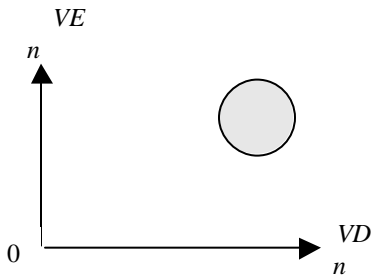
	VD ₁	VD ₀
VE ₁	0,25	0,25
VE ₀	0,25	0,25

Document électronique « moyen »

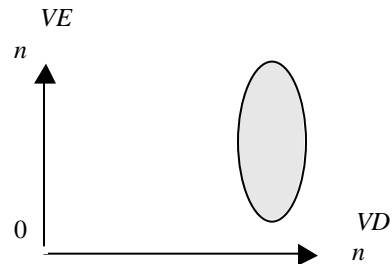
3.2. Cas des variables non binaires, discrètes ou continues

Il s'agit des variables où l'on quantifie la différence entre chaque élément E_i de ? avec E_B ainsi que la différence de chaque élément D_i de ? avec D_B. Ces différences d_E ou d_D sont telles que $0 \leq d_E < 1$ et $0 \leq d_D < 1$.

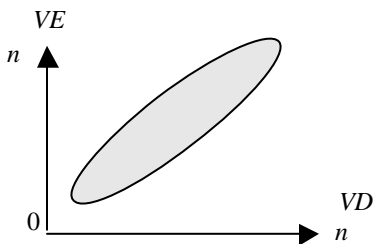
On construit ensuite un nuage de points, qui est interprété comme les tables de contingences ci-dessus.



Document électronique parfait



Document électronique utilisable



Document électronique « spécifique »

Etc.

4. Traitements

J e propose d'aborder très simplement le traitement des données.

Du côté de la conception, la raison de cette simplicité est d'abord pratique : en évaluation d'utilisabilité, Nielsen et Landauer (1994) ont montré que plus de 80%

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

des problèmes étaient identifiés avec cinq utilisateurs – testeurs, et 100% des problèmes avec quinze utilisateurs - testeurs. L'évaluation est de plus en plus abordée de façon itérative au cours de la conception, dans des modèles opportunistes de conception (Nanard & Nanard, 1998). Pour Nielsen (2000), il vaut mieux évaluer trois fois avec cinq utilisateurs – testeurs, en ayant à chaque fois essayé de remédier aux problèmes d'utilisabilité identifiés la fois précédente, que d'évaluer une fois avec quinze utilisateurs – testeurs.

Du côté de l'expérimentation, le caractère rudimentaire de ma proposition tient simplement à la nouveauté du thème. La plupart des expérimentations conduites dans le domaine de l'analyse de l'interaction utilisateur – document électronique ne traitent pas les relations entre variables d'utilisation et variables d'atteintes du but (voir par exemple la recension de Tricot et al., 1998). Je considère donc ma proposition comme un « premier pas ».

4.1. Cas des variables binaires

Je propose de contrôler d'abord avec un calcul de χ^2 si la distribution des valeurs dans la table de contingence est significativement différente de l'équipartition (table correspondant au « document électronique moyen »). Si la distribution observée est significativement différente de l'équipartition, alors la procédure consiste à repérer les deux ou trois distributions du modèle les plus proches de la distribution observée. Puis, pour chaque effectif attendu $E_{f_{\text{attendu}}}$ (*i.e.* la fréquence définie par le modèle multipliée par l'effectif total de l'échantillon), calculer la somme des différences à l'effectif observé $E_{f_{\text{obs}}}$.

Cette différence est donc la mesure de la distance entre les différentes distributions testées et la distribution observée. Il suffit de comparer chacune des valeurs d . La plus petite valeur de d correspond à la distribution théorique de référence.

On peut, par exemple, vouloir évaluer un document électronique de type « site Web » où le but est la « réservation d'un titre de transport » et où l'utilisabilité est évaluée par un seul critère : « temps d'utilisation inférieur à 5 minutes » (imaginons par exemple que ces 5 minutes représentent le temps moyen d'atteinte du but avec le « moyen de référence » *i.e.* le guichet⁴). L'évaluation va consister à demander à un groupe représentatif d'utilisateurs de ce moyen de transport (disons 50 individus) de réserver un titre de transport pour une destination A, B, C ou D (chacune de ces destinations étant représentative de l'ensemble des destinations de ce moyen de transport). VE est mesurée en vérifiant que l'utilisateur a effectivement réservé son titre de transport pour la destination prévue (VE=1 s'il a réussi, VE=0 s'il a échoué). VD est mesurée en temps supérieur (VD=0) ou inférieur (VD=1) à 5 minutes. L'évaluation globale consiste simplement à indiquer le nombre d'utilisateur parmi 50 dans chaque case de la table de contingence. Par exemple :

⁴ Notons que l'on pourrait ici accorder un peu plus de temps, compte tenu du déplacement à effectuer pour se rendre au guichet. Ou bien considérer que le moyen de référence est le Minitel. Cet exemple illustre bien l'importance que j'indiquais au paragraphe 2.3 d'une définition a priori des critères d'utilisabilité choisis.

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

	VD ₁	VD ₀
VE ₁	25	3
VE ₀	10	12

On vérifie que la distribution est différente de l'équipartition ($\chi^2 = 20,2$; $p < 0,01$)

On se demande si cette distribution correspond plutôt à :

	VO ₁	VO ₀	Soit :	VO ₁	VO ₀
VE ₁	0,5	0		25	0
VE ₀	0	0,5		0	25

Objet « spécifique » (nécessaire et suffisant)

ou à :

	VO ₁	VO ₀	Soit :	VO ₁	VO ₀
VE ₁	0,33	0		16,5	0
VE ₀	0,33	0,33		16,5	16,5

Objet nécessaire mais pas suffisant

Dans le premier cas, $d = 0+3+10+13 = 26$

Dans le second cas, $d = 8,5+3+6,5+4,5 = 22,5$

On peut donc conclure que le document est nécessaire mais pas suffisant. On s'inquiètera en particulier du fait que 22 utilisateurs parmi 50 n'ont pas réussi à réserver leur titre de transport, et surtout du fait que 10 utilisateurs parmi ces 22 sont restés moins de 5 minutes sur le site Web (ils ont abandonné plus rapidement que le temps moyen d'attente au guichet).

Le principal intérêt du cadre proposé réside sans doute dans la possibilité d'interpréter une série de traitements locaux sur des « sous-tâches ». Imaginons par exemple que l'on fasse une recherche d'information dans le cadre d'un « projet » d'apprentissage, la tâche étant de « répondre à une série de 10 questions sur le thème étudié à l'aide d'un document électronique présentant un cours sur le thème étudié ». On peut décrire cette tâche comme un ensemble A de 10 sous-tâches : $A = \{ (\text{question1} ; \text{cible1}), (\text{question1} ; \text{cible2}), \dots (\text{question10} ; \text{cible10}) \}$. On peut alors, pour chacune de ces sous-tâches, construire une table de contingence, en mesurant l'utilisabilité par le fait que l'utilisateur atteigne la cible i ($VD_i = 1$ s'il a atteint la cible i; $VD_i = 0$ s'il ne l'a pas atteint), et en mesurant l'utilité par le fait que l'utilisateur réponde correctement à la question i ($VE_i = 1$ s'il a correctement répondu à la question i; $VE_i = 0$ s'il n'a pas correctement répondu). On peut alors procéder à une analyse locale de la qualité du système : par exemple, « pour la question 1, le système est utilisable mais moyennement utile, pour la question 2, l'outil est spécifique, etc. ». On diagnostique ainsi un ensemble de points locaux à améliorer ainsi que la nature de l'amélioration (améliorer l'utilisabilité, améliorer l'utilité). Le lecteur aura remarqué que la moyenne des valeurs de VD_i n'est autre que le taux global de rappel et que la somme des valeurs de VE_i est la performance globale au questionnaire.

On peut utiliser ce principe des traitements locaux pour s'intéresser aux différences de performances entre les utilisateurs - testeurs, dont on pourra

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

éventuellement choisir de contrôler le niveau initial de connaissances déclaratives ou procédurales.

4.2. Traitement des variables continues

Si une table de contingence est toujours facile à interpréter avec le calcul de d , il est souvent difficile d'interpréter aussi directement un nuage de points. L'analyse en composantes principales et d'autres techniques d'analyses des données peuvent produire des « résumés » (des droites, des courbes) ne rentrant pas dans notre cadre d'interprétation.

5. Discussion

Dans ce tutorial, j'ai proposé un cadre formel pour l'interprétation de quinze relations possibles entre deux variables : l'utilisabilité et l'utilité. Ce cadre permet d'interpréter des résultats lors d'évaluations de documents électroniques et plus largement d'outils finalisés, sous réserve que les tests mesurent indépendamment l'utilisabilité et l'utilité. Il permet de souligner un aspect important de l'ergonomie des documents électroniques : seuls les cas correspondants à un document « nécessaire et suffisant » permettent d'envisager à coup sûr qu'une amélioration de l'utilisabilité du document entraînera une amélioration de son utilité et donc de sa qualité (ou valeur) globale.

Ce cadre permet en outre, dans les recherches expérimentales sur l'utilisation des documents électroniques, de décrire l'activité des sujets en intégrant des variables d'utilisation et des variables d'atteinte du but.

Les conditions d'application du cadre sont contraignantes. Elles supposent notamment une description univoque du but. Si le document est détourné de sa finalité il faut pouvoir décrire ce détournement *a priori*, car l'évaluation ne concernera alors que cette nouvelle finalité (détournée). Dans la mesure du possible il faut pouvoir décrire l'ensemble des états du document électronique conduisant au but (l'ensemble noté ?').

Je n'ai présenté qu'une façon rudimentaire de traiter des résultats. Dans la plupart des situations, notamment en conception, cela suffit amplement. L'analyse qualitative des tables de contingences me semble plus intéressante que le traitement quantitatif. Cependant, dans d'autres cas, il pourra être intéressant de pouvoir faire des inférences statistiques sur les résultats.

6. Bibliographie

(Bernard et Charron, 1996) Bernard, Jean-Marc, et Charron, Camilo, « L'analyse implicite bayésienne, une méthode pour l'étude de dépendances orientées I : données binaires », *Mathématiques, Informatique et Sciences Humaines*, vol. 134, 1996, p. 5-38.

In M. Mojahid & J. Virbel (Eds.), *Les documents électroniques, méthodes, démarches et techniques cognitives*. Paris : Europia. 2001

- (Buckingham Shum et McKight, 1997) Buckingham Shum, Simon, et McKight, Cliff. (Eds.), « Web usability », *International Journal of Human Computer Studies*, special issue, 1997, vol. 47, n°1, p.1-222. <http://ijhcs.open.ac.uk/>
- (Grudin, 1992) Grudin, Jonathan, « Utility and usability: research issues and development contexts », *Interacting with Computers*, 1992, vol. 4, n°2, p. 209-217.
- (Le Coadic, 1997) Le Coadic, Yves François, *Usages et usagers de l'information*, Paris, Nathan, 1997.
- (Leplat et Hoc, 1983) Leplat, Jacques et Hoc, Jean-Michel, « Tâche et activité dans l'analyse psychologique des situations », *Cahiers de Psychologie Cognitive*, 1983, vol. 3, n°1, p. 49-63.
- (Meads et Stubbs, 2001) Meads, Jon et Stubbs, David D. « Usability Architects », <http://www.usabilityarchitects.com>
- (Mizzaro, 1998) Mizzaro, Stephano. « Relevance, the whole history ». In T. Bellardo Hahn et M. Buckland (Eds.), *Historical studies in information science, Journal of the American Society for Information Science*, 1998, special issue, p. 221-243.
- (Nanard et Nanard, 1998) Nanard, Jocelyne et Nanard, Marc, « La conception d'hypermédias », In A. Tricot et J.-F. Rouet (Eds.), *Les hypermédias, approches cognitives et ergonomiques*. Paris, Hermès, 1998, p. 15-34.
- (Nielsen et Landauer, 1993) Nielsen, Jacob et Landauer, Thomas K. « A mathematical model of the finding of usability problems », *Proceedings of ACM InterCHI'93 Conference*, New York, ACM Press, 1993, p. 206-213.
- (Nielsen, 1993) Nielsen, Jacob, *Usability engineering*. Boston, Academic Press, 1993.
- (Nielsen, 1999) Nielsen, Jacob. « Voodoo usability », *Jakob Nielsen's Alterbox*, Dec 12, 1999 <http://www.useit.com>
- (Nielsen, 2000) Nielsen, Jacob, *Designing web usability: the practice of simplicity*, Indianapolis, New Riders, 2000.
- (Park, 2000) Park, S. « Usability, user preferences, effectiveness and user behaviors when searching individual and integrated full text databases: implications for digital libraries », *Journal of the American Society for Information Science*, 2000, vol.51, n°5, p.456-468.
- (Rabardel, 1995) Rabardel, Pierre, *Les hommes et les technologies*, Paris, Armand Colin, 1995.
- (Scapin et Bastien, 1997) Scapin, Dominique L, et Bastien, J.M Christian, « Ergonomic criteria for evaluating the ergonomic quality of interactive systems », *Behavior & Information Technology*, 1997, vol. 17, n° 4/5, p. 220-231.
- (Sullivan, 1997) Sullivan, Terry. « The Usable Web », *All Things Web*, 1997. <http://www.pantos.org/atw/usable.html>
- (Tricot et al., 1998) Tricot, André, Pierre-Demarcy, Corinne, et El Boussarghini, Rachid, « Un panorama des recherches consacrées à l'étude de l'activité mentale de l'utilisateur d'un hypermédia », In E. Bruillard et B. de La Passardière (Eds.), *Le livre électronique*, Paris, Hermès, 1998, p. 371-400.
- (Tricot et al., 1999) Tricot, André., Puigserver, E.mmanuelle, Berdugo, Dolly., and Diallo, Mariama, « The validity of rational criteria for the interpretation of user-hypertext interaction », *Interacting with Computers*, 1999, vol. 12, p. 23-36.
- (Tricot et Tricot 2000) Tricot, André et Tricot, Marie, « Un cadre formel pour interpréter les liens entre utilisabilité et utilité des systèmes d'information (et généralisation à l'évaluation d'objets finalisés) ». *Colloque Ergo-IHM 2000*, Biarritz, 3-6 octobre 2000.