

# Un cadre formel pour interpréter les liens entre utilisabilité et utilité des systèmes d'information (et généralisation à l'évaluation d'objets finalisés)

André Tricot (\*) et Marie Tricot (\*\*)

(\*) Centre de recherches en psychologie, cognition et communication, Université de Rennes 2 et IUFM de Bretagne (adresse actuelle : IUFM de Toulouse, 56 avenue de l'URSS, 31 078 Toulouse cedex 4 [andre.tricot@toulouse.iufm.fr](mailto:andre.tricot@toulouse.iufm.fr))

(\*\*) École Nationale de la Santé Publique, Avenue du Pr Léon Bernard, 35 043 Rennes Cedex [mtricot@ensp.fr](mailto:mtricot@ensp.fr)

## RESUME

Lorsqu'on évalue un système d'information (et plus généralement un objet finalisé), on doit évaluer sa mise en œuvre (son utilisation) et ses résultats (les buts qu'il permet d'atteindre, leur conformité aux buts visés). La mise en œuvre et les résultats sont décrits par des variables : respectivement, les variables d'utilisabilité et d'utilité. Dans cette communication, nous proposons un cadre formel pour l'interprétation de quinze types de relations entre ces deux variables ainsi qu'une définition 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. Le cadre que nous proposons, conçu à l'origine pour l'évaluation de la qualité des systèmes d'information, s'applique, sous certaines conditions que nous définissons, à tout objet finalisé que l'on voudrait évaluer et à toute comparaison entre objets finalisés.

**MOTS CLES** : évaluation, utilisabilité, utilité, systèmes d'information, ergonomie

## INTRODUCTION

Lorsqu'on évalue un système d'information (et plus généralement un objet finalisé<sup>1</sup>), 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 [2] a discuté ces deux notions, 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. Selon nous, du côté de l'évaluation, les progrès à accomplir concernent d'abord la définition des deux variables et

des relations qu'elle entretiennent. Dans ce papier, nous proposons donc un cadre formel pour l'interprétation des liens entre ces deux variables. À notre 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 [3] à ce propos nous 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 nous semble convenable, nous regrettons 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, nous envisageons la description logique de l'ensemble des relations possibles entre ces deux variables, pour quelques valeurs standards de ces variables.

Selon nous, 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. Nous proposons dans cette communication une définition de chacun de ces termes et un cadre formel pour interpréter les relations entre utilisabilité et utilité. Le cadre que nous proposons, conçu à l'origine pour l'évaluation de la qualité des systèmes d'information, s'applique, sous certaines conditions que nous définissons, à tout objet finalisé que l'on voudrait évaluer et à toute comparaison entre objets finalisés. Mais nous aborderons principalement ici l'évaluation des systèmes d'information.

L'évaluation de l'utilisabilité des systèmes d'information est un enjeu important depuis quelques années (depuis la publication en 1993 de l'ouvrage de Nielsen [9], les références se comptent en milliers). Ce thème de recherche et d'ingénierie ergonomique concerne notamment des systèmes d'information sur de nouveaux

<sup>1</sup> Nous ne nous occupons pas ici d'instruments au sens de Rabardel [13], qui définit l'instrument comme un objet technique, ou *artefact* + un schéma d'utilisation. Nous traitons 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.

supports électroniques comme les sites Web [1]. Le succès de la notion d'utilisabilité est tel que tout se passe comme si l'utilisabilité devenait progressivement synonyme de *qualité générale d'un système d'information*. Par exemple, les cinq critères d'utilisabilité proposés par Nielsen [9] 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 [5] vont même jusqu'à intégrer le critère d'utilité parmi les critères d'utilisabilité. En bref, il nous semble que non seulement l'utilisabilité fonctionne comme un critère d'évaluation globale de la qualité des systèmes d'information, 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 [14] 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. Notre 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 nous la concevons est l'adéquation entre les buts de l'utilisateur et la finalité du système d'information.

Dans cette communication, nous admettons donc avec Grudin [2] qu'un bon système d'information est à la fois utile et utilisable. Nous ne suivons pas Nielsen [10] 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 nous 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 [12] 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.

## DEFINITIONS, PRINCIPES, CADRE DE TRAVAIL

### 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 de l'information. 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. L'intérêt de cette synonymie en sciences de l'information réside dans la possibilité d'exploiter le travail immense qui a été consacré aux tentatives de définition de la notion de pertinence. Mizzaro [6] 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).

Nous retiendrons donc cette conception de la pertinence et l'étendrons, dans le cadre des systèmes d'information, à la notion d'utilité : *l'utilité d'un système d'information et plus généralement l'utilité d'un objet finalisé, est l'adéquation entre la finalité de l'objet et le but de l'utilisateur, pour un domaine, une exploitation et un environnement donnés.*

### Contextes d'usage des évaluations d'utilisabilité et d'utilité ; définition des conditions d'application de notre 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 l'objet finalisé par des utilisateurs, dans des environnements et pour des buts qui sont censés, tous les

trois (utilisateur, environnement, but<sup>2</sup>), représenter un contexte d'utilisation typique. Nous ne discuterons pas ici des conditions d'obtention de cette représentativité. Nous insistons 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 l'objet que l'on conçoit permet aux utilisateurs d'atteindre les buts pour lesquels on a conçu l'objet 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 objet finalisé : on compare alors les performances des utilisateurs avec les deux objets finalisés ;
- dans le cadre d'une situation de référence où un autre objet finalisé 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 de notre cadre sont les suivantes : *pour évaluer l'utilisabilité et l'utilité d'un objet finalisé, 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 (d'objets finalisés), il faut disposer d'un modèle a priori de l'utilisabilité et de l'utilité attendue de l'objet évalué.*

### Définitions

- L'utilisabilité d'un objet finalisé désigne la possibilité d'utilisation de cet objet.
- L'utilité d'un objet finalisé désigne la possibilité d'atteindre un but visé avec cet objet.

<sup>2</sup> Nous distinguons évidemment but, tâche et activité, de façon tout à fait classique [4]. À un niveau plus général nous définissons 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).

- L'utilisabilité et l'utilité peuvent être mesurées en termes absolus (possibilité vs impossibilité) ou relatifs (efficacité, efficacité, degré d'atteinte du but ou d'utilisation de l'objet finalisé).
- 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 (quelle proportion du but a-t-elle été atteinte?).
- Le degré d'utilisation est une quantification relative à l'objet (quelle part, quels éléments, quelles fonctions, de l'objet 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 gestes ou de mouvements, d'attention, etc.) de l'utilisation d'un objet finalisé et de l'atteinte du but recherché. Dans le domaine de l'utilisation des systèmes d'information, l'efficacité peut être mesurée par les critères de "taux de rappel", de "taux de précision" et d'économie<sup>3</sup> (voir [16] pour une présentation et une discussion de ces critères).
- L'efficacité mesure l'économie *en termes de temps* de l'utilisation d'un objet finalisé et de l'atteinte du but recherché.
- L'efficacité et l'efficacité, ainsi que toutes leurs mesures, n'ont pas de pertinence absolue. Cette pertinence est strictement relative à l'objet finalisé et à l'environnement dans lequel l'objet finalisé est utilisé. Cette pertinence doit être définie avant l'évaluation de l'objet.
- La qualité d'un objet finalisé est mesurée par son utilisabilité et par son utilité.

### Formalisme

Soit une tâche à réaliser par un sujet, composée d'un but B, d'un objet finalisé O, d'un environnement de travail E.

À l'état initial, l'environnement est dans un état zéro, noté  $E_0$ . L'objet finalisé est dans un état zéro, noté  $O_0$ . L'objet finalisé va changer d'état. Il existe au moins un état  $O_B$ , correspondant au but. L'action sur l'objet finalisé 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  $\Omega = \{O_0, O_\alpha, O_B, \dots, O_B\}$  des états de O et un ensemble  $\Sigma = \{E_0, E_\alpha, E_B, \dots, E_B\}$  des états de E. Le but B peut donc être décrit par le couple  $(O_B, E_B)$ .  $\Omega$  et  $\Sigma$  peuvent être des échelles ordinales ou non. Il existe un sous ensemble  $\Omega' \subset \Omega$ , ordonné ou non, correspondant à la façon la plus efficace d'atteindre

<sup>3</sup> taux de rappel = nombre de cibles atteintes / nombre de cibles existantes ; taux de précision = nombre de cibles atteintes / nombre d'items ouverts ; économie = 1 - (nombre d'items différents ouverts / nombre total d'items ouverts).

l'état  $O_B$ . Il peut être pertinent, dans les environnements "à temps contraint", de pondérer  $O_B$  et le cas échéant  $\Omega'$  par le temps passé à atteindre le but. On mesure alors l'efficacité. La représentation de la tâche peut être composée d'une représentation de l'ensemble des états de l'objet finalisé  $R(\Omega)$ , de la façon efficace ou efficiente d'atteindre le but  $R(\Omega')$  et d'une représentation de l'ensemble des états de l'environnement  $R(\Sigma)$ .

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

$VO$  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  $\{O_0, O_\alpha, \dots\}$  et l'ensemble  $\Omega'$ .

### INTERPRETATION DES LIENS ENTRE LES VARIABLES D'UTILISABILITE ET D'UTILITE

Les liens entre les variables d'utilisabilité et d'utilité sont décrits par la fréquence  $f$  des co-occurrences des états de  $VO$  et  $VE$ . La somme de ces fréquences est évidemment égale à 1.

#### Cas des variables binaires

	VO (noté $VO_1$ )	$\neg VO$ (noté $VO_0$ )	
VE (noté $VE_1$ )	$f_{VO_1 VE_1}$	$f_{VO_0 VE_1}$	$f_{VE_1}$
$\neg VE$ (noté $VE_0$ )	$f_{VO_1 VE_0}$	$f_{VO_0 VE_0}$	$f_{VE_0}$
	$f_{VO_1}$	$f_{VO_0}$	

Table de contingence

	$VO_1$	$VO_0$
$VE_1$	$f=1$	$f=0$
$VE_0$	$f=0$	$f=0$

Objet finalisé parfait : utilisable et utile

Conjonction :  $VO \dot{\cup} VE$

	$VO_1$	$VO_0$
$VE_1$	0	0
$VE_0$	1	0

Objet utilisable mais inutile

Implication :  $VO \dot{\supset} \neg VE$

	$VO_1$	$VO_0$
$VE_1$	0	0
$VE_0$	0	1

Objet mauvais ou inadéquat : inutile et inutilisable

NOR (non ou) ou "ni ...ni ...":  $VO \bar{\cup} VE$

	$VO_1$	$VO_0$
$VE_1$	0	1
$VE_0$	0	0

Objet paradoxal

Implication :  $\neg VO \dot{\supset} VE$

	$VO_1$	$VO_0$
$VE_1$	0,5	0
$VE_0$	0	0,5

Objet "spécifique" (nécessaire et suffisant)

Équivalence :  $VO \dot{\cup} VE$

	$VO_1$	$VO_0$
$VE_1$	0	0,5
$VE_0$	0,5	0

Objet nuisible

Ou exclusif :  $VE \dot{\cup} VO$

	$VO_1$	$VO_0$
$VE_1$	0,5	0
$VE_0$	0,5	0

Objet utilisable mais moyennement utile

Indépendance :  $VO$  est vrai, "  $VE$

	$VO_1$	$VO_0$
$VE_1$	0	0,5
$VE_0$	0	0,5

Objet inutilisable

Indépendance :  $VO$  est faux, "  $VE$

	$VO_1$	$VO_0$
$VE_1$	0,5	0,5
$VE_0$	0	0

Objet placebo

Indépendance :  $VE$  est vrai, "  $VO$

	$VO_1$	$VO_0$
$VE_1$	0	0
$VE_0$	0,5	0,5

Objet inutile, bien que moyennement utilisable

Indépendance :  $VE$  est faux, "  $VO$

	$VO_1$	$VO_0$
$VE_1$	0,33	0,33
$VE_0$	0	0,33

Objet suffisant mais non nécessaire.

Implication :  $VO \dot{\supset} VE$

	$VO_1$	$VO_0$
$VE_1$	0,33	0
$VE_0$	0,33	0,33

Objet nécessaire mais pas suffisant

Implication :  $VE \dot{\supset} VO$

	VO <sub>1</sub>	VO <sub>0</sub>
VE <sub>1</sub>	0,33	0,33
VE <sub>0</sub>	0,33	0

Objet peu spécifique (ni nécessaire ni suffisant)

Ou inclusif:  $VE \dot{\cup} VO$

	VO <sub>1</sub>	VO <sub>0</sub>
VE <sub>1</sub>	0	0,33
VE <sub>0</sub>	0,33	0,33

Objet nuisible : la réussite à l'utilisation implique l'échec au but *Fonction NAND VE / VO : incompatibilité*

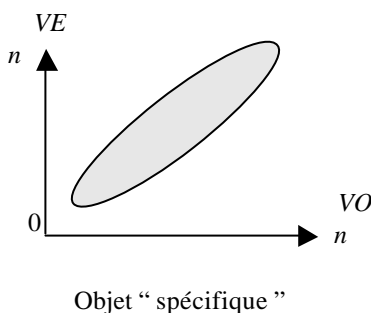
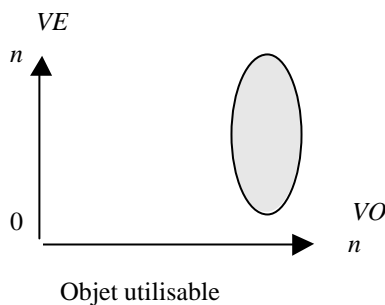
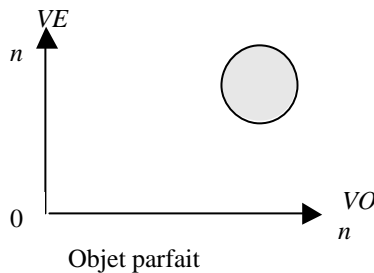
	VO <sub>1</sub>	VO <sub>0</sub>
VE <sub>1</sub>	0,25	0,25
VE <sub>0</sub>	0,25	0,25

Objet finalisé "moyen"

### 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  $\Sigma$  avec  $E_B$  ainsi que la différence de chaque élément  $Q_i$  de  $\Omega$  avec  $O_B$ . Ces différences  $d_E$  ou  $d_O$  sont telles que  $0 \leq d_E < 1$  et  $0 \leq d_O < 1$ .

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



Etc.

### TRAITEMENTS

Nous proposons 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 [8] ont montré que plus de 80% 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 [7]. Pour Nielsen [11], 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 notre proposition tient simplement à la nouveauté du thème. La plupart des expérimentations conduites dans le domaine de l'analyse de l'interaction utilisateur – système d'information ne traitent pas les relations entre variables d'utilisation et variables d'atteintes du but (voir par exemple la recension [15]). Nous considérons donc notre proposition comme un "premier pas".

Dans cette communication, nous n'aborderons pas la question de l'inférence statistique. Un article ultérieur devrait y être entièrement consacré.

### Cas des variables binaires

Notre proposition est de contrôler d'abord avec un calcul de  $\chi^2$  si la distribution des valeurs dans la table de contingence est significativement différente de l'équirépartition (table correspondant à l'"objet finalisé moyen"). Si la distribution observée est significativement différente de l'équirépartition, 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  $Ef_{attendu}$  (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  $Ef_{obs}$ .

$$d = \sum |Ef_{attendu} - Ef_{obs}|$$

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 système d'information de type "borne interactive" 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

guichet). 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).  $VO$  est mesurée en temps supérieur ( $VO=0$ ) ou inférieur ( $VO=1$ ) 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 :

	$VO_1$	$VO_0$
$VE_1$	25	3
$VE_0$	10	12

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

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 l'objet est nécessaire mais pas suffisant. On s'inquiètera en particulier du fait 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 la borne (ils ont abandonné plus rapidement que le temps moyen d'attente au guichet).

Le principal intérêt de notre cadre 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 système d'information présentant un cours sur le thème étudié ". On peut décrire cette tâche comme un ensemble A de 10 sous-tâches :  $A = \{ (question_1 ; cible_1), (question_1 ; cible_2), \dots (question_{10} ; cible_{10}) \}$ . On peut alors, pour chacune de ces sous-tâches, construire une table de contingence, en mesurant l'utilisabilité par le fait de l'utilisateur atteigne la cible  $i$  ( $VO_i = 1$  s'il a atteint la cible  $i$  ;  $VO_i = 0$  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  $VO_i$  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 éventuellement choisir de contrôler le niveau initial de connaissances déclaratives ou procédurales.

### 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.

### DISCUSSION

Dans cette communication, nous avons 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 de tests de systèmes d'information et plus largement d'outils finalisés, sous réserve que les tests mesurent indépendamment l'utilisabilité et l'utilité. Il permet en outre, dans les recherches expérimentales sur l'utilisation des systèmes d'information, de décrire l'activité des sujets en intégrant des variables d'utilisation et des variables d'atteinte du but.

Les conditions d'application de notre cadre sont contraignantes. Elles supposent notamment une description univoque du but. Si l'objet 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 système d'information conduisant au but (l'ensemble noté  $\Omega$ ).

Nous n'avons 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 nous semble beaucoup plus intéressante que le traitement quantitatif. Cependant, dans d'autres cas, il pourra être intéressant de pouvoir faire des inférences statistiques sur les résultats. Nous aborderons ce point dans un

prochain article.

**Remerciements :** ce papier a largement bénéficié des commentaires de Béatrice Drot-Delange, Michel Deleau, Jean Julo, Jean-Claude Coulet, Pierre Merle, Luc Rodet et des *reviewers* anonymes du colloque. Christian Bastien et Jonathan Grudin nous ont fourni une documentation complémentaire à la première version qui s'est avérée très utile. Des discussions avec Gérard Sensevy, Annie Morin et Camilo Charron nous ont convaincu d'écrire un deuxième article qui abordera l'inférence statistique avec notre modèle.

## BIBLIOGRAPHIE

1. Buckingham Shum, S. and McKinght, C. (Eds.), (1997). Web usability. *International Journal of Human Computer Studies, special issue*, 47 (1), 1-222. <http://ijhcs.open.ac.uk/>
2. Grudin, J. Utility and usability: research issues and development contexts. *Interacting with Computers*, 1992, 4 (2), 209-217.
3. Le Coadic, Y.F. *Usages et usagers de l'information*. Paris, Nathan, 1997.
4. Leplat, J. and Hoc, J.-M. Tâche et activité dans l'analyse psychologique des situations. *Cahiers de Psychologie Cognitive*, 1983, 3 (1), 49-63.
5. Meads, J. and Stubbs, D.D. *Usability Architects*, <http://www.usabilityarchitects.com/>
6. Mizzaro, S. Relevance, the whole history. In T. Bellardo Hahn and M. Buckland (Eds.), Historical studies in information science, *Journal of the American Society for Information Science*, 1998, special issue, 221-243.
7. Nanard, J. and Nanard, M. La conception d'hypermédias. In A. Tricot and J.-F. Rouet (Eds.), *Les hypermédias, approches cognitives et ergonomiques*. Paris, Hermès, 1998, pp. 15-34.
8. Nielsen, J. and Landauer, T.K. A mathematical model of the finding of usability problems. *Proceedings of ACM InterCHI'93 Conference*, New York, ACM Press, 1993, pp. 206-213.
9. Nielsen, J. *Usability engineering*. Boston, Academic Press, 1993.
10. Nielsen, J. Voodoo usability. *Jakob Nielsen's Alterbox*, Dec 12, 1999 <http://www.useit.com>
11. Nielsen, J. *Designing web usability: the practice of simplicity*. Indianapolis, New Riders, 2000.
12. 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, 51 (5), 456-468.
13. Rabardel, P. *Les hommes et les technologies*. Paris, Armand Colin, 1995.
14. Scapin, D.L. and Bastien, J.M.C. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behavior & Information Technology*, 1997, 17 (4/5), 220-231.
15. Tricot, A., Pierre-Demarcy, C. and El Boussarghini, R. Un panorama des recherches consacrées à l'étude de l'activité mentale de l'utilisateur d'un hypermédia. In E. Bruillard and B. de La Passardière (Eds.), *Le livre électronique*, Paris, Hermès, 1998, pp. 371-400.
16. Tricot, A., Puigserver, E., Berdugo, D., and Diallo, M. The validity of rational criteria for the interpretation of user-hypertext interaction. *Interacting with Computers*, 1999, 12, 23-36.