



**HAL**  
open science

## Évaluer l'utilisation d'hypermédias : intérêts et limites des variables de performance

Erica de Vries, André Tricot

► **To cite this version:**

Erica de Vries, André Tricot. Évaluer l'utilisation d'hypermédias : intérêts et limites des variables de performance. Hypertextes et Hypermédias, 1998, pp.175-190. edutice-00000214v1

**HAL Id: edutice-00000214**

**<https://edutice.hal.science/edutice-00000214v1>**

Submitted on 17 Nov 2003 (v1), last revised 27 Jul 2023 (v2)

**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.

## **Évaluer l'utilisation d'hypermédias : intérêts et limites des variables de performance**

### **Introduction**

Aujourd'hui, les hypermédias sont utilisés à des fins variées dans diverses situations. Mais, quelles que soient ces situations, ils ne peuvent qu'être des moyens pour l'utilisateur de réaliser une tâche. Le terme de tâche recouvre ici deux sens. Premièrement, un sens général que l'on peut appeler le contexte de l'activité. Le contexte fait référence à la finalité d'un ensemble d'activités, dont l'utilisation d'un hypermédia fait partie. Ce contexte peut être, par exemple, d'apprendre des concepts de physique, de concevoir une aire de jeu pour enfants ou encore de s'informer sur les études de médecine. Le terme de tâche a également un sens opérationnel que l'on peut appeler le but de l'activité. Le but réfère aux informations particulières recherchées à un moment donné. Le but de l'activité se traduit par la recherche d'un ou plusieurs nœuds dans l'hypermédia et des liens qu'il y a entre eux. Ces deux notions, contexte et but, ont leur pendant dans l'évaluation de l'utilisation d'un hypermédia. La réussite d'une utilisation peut s'exprimer en termes de contexte de l'activité (apprentissage, conception) ou en termes du but opérationnel (avoir trouvé les nœuds pertinents).

Dans le domaine de recherche d'information, de nombreuses variables sont employées pour mesurer la performance des systèmes d'informations. Leur application dans le domaine des hypermédias exige une évaluation de leur pertinence pour interpréter l'activité des utilisateurs. Plus précisément, il est nécessaire d'évaluer si la performance en terme de ces variables dépend des caractéristiques du contexte et de la nature du but de l'activité. Une meilleure connaissance des effets de contexte et du but sur l'utilisation d'hypermédias sera utile à deux titres. Du point de vue théorique, elle servira de base pour les recherches visant à étudier l'utilité au sens large des hypermédias pour l'apprentissage, pour la conception, et pour la recherche d'information. Du point de vue pratique, elle permet d'envisager précisément, lors de la conception d'un système hypermédia, les caractéristiques spécifiques aux tâches envisagées, que ce soit au niveau de l'organisation de l'information, des fonctionnalités ou de l'interface du système. Ainsi, cet article s'intéresse aux moyens de caractériser l'utilisation d'hypermédias. En particulier, la question de la sensibilité des variables de performance aux aspects du contexte et aux variations dans la nature du but est abordée en examinant des résultats d'études. La dernière section présente quelques conclusions sur ces sujets.

### *Analyse de la tâche*

Si l'on se place du point de vue de l'utilisateur, la tâche doit être définie à deux niveaux : d'une part l'utilisateur doit interpréter une consigne (interne ou externe), c'est-à-dire se faire une

représentation générale du contexte, d'autre part il doit former une représentation opérationnelle du but précis à atteindre. Cette dernière comporte une représentation des informations qu'il doit trouver dans le système pour atteindre le but qui lui est fixé (où qu'il s'est fixé lui-même), une représentation des moyens à mettre en oeuvre pour atteindre ce but, ainsi qu'une représentation de l'environnement (le système) dans lequel se déroule l'activité.

Ces deux niveaux de la représentation de la tâche, qui correspondent aux deux niveaux de description de l'activité, sont, on le voit, étroitement liés. Le niveau de la représentation opérationnelle du but, dont on suppose en psychologie cognitive qu'elle a pour rôle de diriger l'activité du sujet, peut être traduit *a priori*, de façon formelle, par l'observateur en termes d'ensemble de nœuds ou de séquence de nœuds du système que l'utilisateur doit ouvrir. Ces nœuds pertinents sont appelés des cibles, et l'ensemble des cibles la structure de but. Cette description peut aussi prendre en compte le degré de pertinence des cibles (degré de pertinence préalablement établi par l'observateur). Par exemple, telle tâche peut être décrite comme impliquant l'ouverture de seulement deux nœuds, tandis que telle autre implique l'ouverture d'un nombre important de nœuds. Une autre tâche impliquant l'ouverture de deux nœuds pertinents, peut impliquer en fait l'ouverture de  $n$  nœuds intermédiaires. Le traitement que l'utilisateur va effectuer une fois qu'il aura trouvé les nœuds, concerne des opérations telles que trier, copier, comprendre le lien avec un autre nœuds, et peut orienter la suite de sa recherche d'information. En même temps, ce traitement prendra son sens dans le contexte de l'activité.

### *L'évaluation de l'utilisation d'hypermédias*

Canter, Rivers et Storrs (1985) ont caractérisé l'utilisation d'hypermédias et le chemin parcouru par l'utilisateur en termes de chemins, d'anneaux, de boucles et de pointes, ainsi qu'en degré d'exploration et degré de redondance. L'exploration (nombre de nœuds ouverts / nombre de nœuds dans l'hypermédia) est une variable générale représentant le degré de couverture de l'hypermédia. La redondance ( $1 - \text{nombre de nœuds ouverts} / \text{nombre d'ouvertures de nœuds}$ ) mesure le nombre de retours aux mêmes nœuds. La question qui se pose est de savoir si ces mesures permettent de faire des inférences sur l'état mental de l'utilisateur. Notamment, un débat a lieu depuis quelques années : des indicateurs comme la redondance (parcours non-économique) ou les loopings (passer plus de trois fois par un même nœud) s'interprètent-ils comme le fait que le sujet est perdu, n'atteint pas son but, ne comprend pas les informations qu'il traite ? Après avoir répondu oui (Edwards & Hardman, 1989 ; Foss, 1988 ; Rouet, 1990) à cette question, les auteurs ont eu tendance à répondre non (Bernstein et al., 1992) , argumentant de façon très générale (Bernstein, 1993) que cela dépend du type d'activité, des objectifs, ou du système étudié.

Dans certaines situations, une plus grande redondance peut être favorable à la réussite d'une tâche. Par exemple, Tricot et Coste (1995b) ont mené une étude sur "Correl...", un ensemble de tutoriels hypermédia d'enseignement de la physique (vibrations et propagations, mécanique). Dans cet hypermédia volumineux (1350 nœuds), les parcours étaient très peu économiques en ce sens que seulement 36% des nœuds sélectionnés étaient pertinents et que ces nœuds étaient en moyenne sélectionnés 5 fois chacun. Mais surtout, répondre correctement à la question posée impliquait que le sujet passe plus de 4 fois par le ou les nœuds pertinents. En revanche, le fait de passer plus de 4 fois par un nœud pertinent ne garantissait pas une réponse correcte à la question correspondante. La confrontation entre cette observation et les travaux de Edwards et Hardman (1989), Foss (1988) ou Rouet (1990) montre qu'il existe bien une relation entre les caractéristiques du parcours des utilisateurs et leur performance sur la tâche principale. Cependant, la nature de cette relation dépendrait de la situation précise, et notamment d'autres variables comme la structure du but à atteindre.

D'autres variables souvent employées sont l'efficacité (bonnes réponses) et la vitesse lors de tâches de type question-réponse. Chen et Rada (1996), au moyen d'une méta-analyse d'études

empiriques, ont étudié l'effet de la complexité de la tâche et de la structure de l'hypermédia. Les résultats sur l'efficacité et la vitesse varient selon les caractéristiques de la situation étudiée. Les tâches ouvertes (sans objectif clair) sont exécutées plus efficacement, mais aussi plus lentement que les tâches fermées. Par contre, la performance est tantôt plus tantôt moins efficace, mais en tout cas plus lente dans les structures hypertextuelles que dans les structures non-hypertextuelles. L'analyse de Chen et Rada montre que ces variables posent problème quant à la comparaison de la performance de l'utilisation d'hypertextes dans différents contextes.

Les notions de cible et de structure de but peuvent également jouer un rôle dans l'évaluation d'une activité avec un hypermédia. La réussite d'une activité peut être mesurée en termes de proximité entre l'ensemble de nœuds réellement ouverts et la structure de but définie *a priori* par l'observateur. Dans le domaine de la recherche d'information, on définit classiquement (Sweets, 1969 ; Salton & McGill, 1983) une recherche efficace comme celle où un maximum de cibles sont sélectionnées (indice de rappel) et où un minimum de non-cibles sont sélectionnées (indice de précision). On pondère les deux variables par le nombre maximum de référence : par le nombre de cibles dans la base pour le rappel, et par le nombre de non-cibles dans la base pour la précision. Ainsi, la précision tient compte de la sélectivité de la tâche. Un sujet ouvrant peu de nœuds non-pertinents aura un score de précision élevé dans une base à pourcentage de cibles modeste, et un score peu élevé dans une base à pourcentage de cibles important. Enfin, pour calculer un score d'économie, le nombre de cibles différentes ouvertes peut aussi être pondéré par le nombre d'ouvertures de nœuds. L'économie tient compte, non pas de la sélectivité de la tâche, mais du niveau d'activité manifestée dans l'utilisation de l'hypermédia.

- *Rappel*. Nombre de cibles différentes ouvertes / nombre total de cibles.
- *Précision*. Nombre de non-cibles non-ouvertes / nombre total de non-cibles.
- *Économie*. Nombre de cibles différentes ouvertes / nombre d'ouvertures.

Ces trois variables de performance ont une valeur qui varie entre 0 et 1. L'économie est un indice plus spécifique des hypertextes : il mesure le degré de pertinence des nœuds ouverts.

Indépendamment de la situation particulière, on peut caractériser l'utilisation d'hypermédiats en termes de ces trois variables de performance. Toutefois, leur sensibilité aux aspects de la situation reste à déterminer. Pour évaluer la sensibilité, les résultats sur les variables rappel, précision, et économie dans deux études différentes sont présentés.

### **Les effets de la structure du but sur l'utilisation de la base**

Pour étudier si les caractéristiques d'utilisation dépendent de la structure du but (nombre de cibles), Tricot et ses collaborateurs ont conduit une expérience où 3 tâches impliquaient trois nombres différents de cibles dans une même base, donc une organisation de données identique (Tricot et al., in press). Les caractéristiques d'utilisation relevées sont le rappel, la précision et l'économie. Cette étude se situe dans un contexte d'information sur les formations et les métiers, et en particulier, il s'agissait pour les sujets de s'informer sur la profession de psychiatre.

#### *Méthode*

L'expérimentation a eu lieu sur une pile HyperCard™ décrivant la profession de psychiatre (Figure 1) contenant 60 cartes (dont 4 cartes outils : sommaire, index, accueil, rubriques). Chacune des 56 cartes de contenu était un texte descriptif de 80 mots en moyenne.

Une des trois tâches (T1, T2, T3) étaient proposées au sujets. Pour la tâche T1, les sujets

devaient répondre à trois questions : " Qu'est-ce que le métier de neurologue ? ", " Qu'est-ce que l'obsession? ", " Quelles sont les qualités morales du psychiatre ? ". Un nœud était pertinent pour répondre à chacune des questions, soit 3 cibles en tout. Pour la tâche T2, les sujets devaient répondre à la question suivante : " A l'aide de quelques exemples, peux-tu me dire la différence entre une névrose et une psychose ? ". Pour répondre à cette question, 9 nœuds étaient pertinents. Pour la tâche T3, on demandait aux sujets : " Imagine que tu aies à présenter le métier de psychiatre à des adolescents. Pour cela, tu disposes d'un logiciel comportant toutes les informations nécessaires ". Tous les nœuds de contenu (56) étaient pertinents.

Les sujets étaient étudiants à l'Université de Provence, en excluant ceux qui étudiaient la psychopathologie et l'informatique. Chaque condition expérimentale était passée par 10 sujets.

### Résultats

Le Tableau 1 montre les moyennes sur les caractéristiques d'utilisation en fonction de la structure du but (nombre de cibles). Concernant le rappel, on observe un effet significatif de la tâche ( $F(2, 27) = 6.49$  ;  $p < .0001$ ). Le rappel est le plus élevé pour la tâche 1, suivie par la tâche 3, et enfin le rappel est le plus faible pour la tâche 2. Outre l'effet global de la tâche, on observe des différences significatives dans les comparaisons deux à deux. Ces résultats montrent que la réussite de la tâche en termes de rappel ne dépend pas de façon linéaire du nombre de cibles. En termes de rappel, les sujets parviennent mieux à ouvrir 56 cibles que 9.

Tâche (cibles)	Caractéristique		
	Rappel	Précision	Économie
T1 (3)	1.00	.73	.12
T2 (9)	.60	.69	.23
T3 (56)	.78	-	.46

**Tableau 1.** Caractéristiques de l'utilisation en fonction du nombre de cibles. Note. Il n'y a pas de précision pour la tâche 3 puisqu'il n'y a pas de non-cibles

On ne note pas d'effet de la tâche sur la précision pour les tâches T1 et T2 ( $F(1,18) = 0.03$ , ns).

Il existe un effet de la tâche sur l'économie ( $F(2, 27) = 18.4$  ;  $p < .0001$ ). En effet, les sujets sont plus économes pour la tâche 3 que pour les deux autres tâches. Néanmoins, le score de .46 pour la tâche 3 indique que les sujets ne rencontrent en moyenne un nouveau nœud que toutes les deux ouvertures.

### Conclusions préliminaires

Les résultats obtenus font penser que l'approche rationnelle, qui décrit les parcours en termes de rappel, de précision et d'économie, ne constitue pas un bon cadre d'analyse. Le rappel ne montre pas une relation linéaire avec le nombre de cibles, ce qui aurait pu donner une indication de la difficulté de la tâche. En revanche, cette expérience semble indiquer qu'il y a un lien entre la proportion de cibles sur le nombre total de nœuds dans le système et la réussite de la tâche. Ce lien serait un effet en U : avec une proportion très faible ou très forte de cibles sur le nombre total de nœuds dans le système, la performance en rappel est meilleure qu'avec une proportion médiane.

Le fait qu'il n'y ait pas d'effet lié à la précision, et seulement une différence de la tâche 3 par rapport aux tâches 1 et 2 pour l'économie, semble confirmer l'inadéquation de l'approche

rationnelle.

Une hypothèse pour expliquer l'inadéquation de ce cadre d'analyse, est que le principe rationnel d'économie correspond à un cadre d'analyse dans lequel un algorithme de recherche ou un mode d'indexation doivent être efficaces : non seulement atteindre le but, mais aussi avec précision et dans un temps le plus bref possible. Or, de toute évidence, les utilisateurs ne sont pas contraints de se représenter la tâche en ces termes. Dans l'expérience, certains sujets ayant trouvé les nœuds pertinents exploraient la pile par curiosité. L'évaluation de l'utilité des descripteurs nécessite alors l'introduction d'une contrainte dans la consigne pour introduire la notion d'efficacité.

#### *L'approche rationnelle : introduction d'une contrainte dans la consigne*

Tricot et ses collaborateurs (in press) ont conduit une seconde expérience, dans laquelle ils demandaient aux sujets d'être efficaces. Les auteurs pensaient pouvoir évaluer ainsi si l'inadéquation du cadre rationnel provient seulement de cette différence de représentation quant à un objectif d'efficacité ou d'économie. Il est classique en effet que le sujet se représente ce qu'il doit faire d'une façon différente de l'expérimentateur : l'inadéquation des critères d'efficacité ou d'économie dans la précédente expérience vient peut être du simple fait que les sujets ne se représentaient leur tâche comme devant être réalisée de façon efficace ou économique.

Dans une seconde expérience, conduite sur le même matériel dans les mêmes conditions, ils ont introduit une contrainte lors de la présentation de la consigne. Ils demandaient aux sujets de réaliser la tâche en faisant le moins d'ouvertures de nœuds possible. Seules les deux tâches les plus simples (T1 et T2) étaient utilisées. Les résultats présentés sur cette condition contrainte (voir Tableau 2) sont comparés à l'expérience précédente que l'on peut qualifier de libre (voir Tableau 1). Les sujets étaient issus de la même population que dans la première expérience et il y avait 10 sujets dans chacune des deux conditions expérimentales T1 et T2.

#### *Résultats*

Le Tableau 2 présente les résultats de cette expérience. La différence de score de rappel entre la condition libre et la condition contrainte n'est significative ni avec la tâche 2 ( $F(1, 18) = 3.06$ , ns) ni avec la tâche 1 (puisqu'elle est réussie par 10/10 sujets dans la condition libre et 9/10 dans la condition contrainte).

Tâche (Cibles)	Caractéristique		
	Rappel	Précision	Économie
T1 (3)	.97	.95	.17
T2 (9)	.37	.86	.14

**Tableau 2.** Caractéristiques de l'utilisation dans la situation contrainte

On note un effet significatif de la consigne sur la précision ( $F(3, 36) = 3.877$  ;  $p < .05$ ). Les sujets sont plus précis en condition contrainte qu'en condition libre.

Concernant l'économie, il y a un effet significatif de l'interaction tâche x consigne ( $F(3,36) = 4.750$  ;  $p < .04$ ). En condition libre les sujets sont plus économes en T2 qu'en T1 ; en condition contrainte cette différence disparaît. Dans la condition contrainte, l'économie pour la tâche 2 se situe au même niveau que pour la tâche 1, c'est-à-dire une économie plus basse.

Il y a aussi un effet significatif de la consigne sur l'utilisation des outils ( $F(3, 36) = 4.757$  ;  $p < .04$ ). Les sujets utilisent proportionnellement plus de nœuds outils pour la consigne contrainte ( $M_{T1} =$

.60,  $M_{T2} = .62$ ) que pour la consigne libre ( $M_{T1} = .47$ ,  $M_{T2} = .46$ ).

### *Discussion*

La structure du but semble avoir un effet sur la performance des sujets. Dans l'expérience de Tricot et collaborateurs, cet effet concernait non pas directement le nombre de cibles mais la proportion de cibles dans le système. Cet effet ne serait pas linéaire mais en U.

L'introduction d'une contrainte montre qu'en condition libre, certains sujets se font une représentation de la tâche qui n'obéit pas aux critères d'économie : les sujets ne se représentent pas leur parcours comme devant être efficace, ou, en tous cas, sélectif. C'est ce que Waterworth et Chignell (1991) ont appelé l'orientation de l'utilisateur. Elle représente une dimension qui varie entre un état cognitif visant à trouver une information précise, et un état axé sur la découverte de l'information dans la base.

En condition contrainte, on observe une économie plus basse pour la tâche comportant un nombre moyen de cibles. Ainsi, l'utilisation de l'hypermédia en situation contrainte devient plus ciblée et précise (diminue l'ouverture des non-cibles), mais en même temps rend moins favorable la proportion de cibles sur le nombre d'ouvertures (économie). Cette baisse d'économie en condition contrainte s'explique par une plus grande utilisation des nœuds "outils" et des réouvertures des mêmes nœuds cibles.

On peut donc penser que l'effet de la nature de la tâche sur l'utilisation du système se manifeste en fonction du nombre de cibles, de la proportion de cibles dans le système et de la représentation que l'utilisateur se fait des contraintes (de temps, d'efficacité) qui pèsent sur lui et selon que sa représentation concerne plus les traitements à effectuer ou les cibles à sélectionner. Ces effets ont des liens entre eux et se manifestent différemment selon les situations d'utilisation. La section suivante étudie les effets de certains de ces aspects de la situation ayant un effet potentiel sur l'utilisation.

### **L'influence du type de question et de l'organisation des données**

Les résultats présentés montrent que les variables de performance posent des problèmes d'interprétation dans la caractérisation de l'utilisation d'hypermédia, et ne sont pas très sensibles à la structure du but en soi. Le nombre de cibles pourrait jouer un rôle en combinaison avec d'autres facteurs, tels que la difficulté des questions et l'organisation de la base.

En effet, il est possible de concevoir une activité, par exemple de formuler des questions, en faisant varier indépendamment difficulté et nombre de cibles. On peut faire varier la difficulté en jouant sur le niveau d'abstraction nécessaire pour comprendre ce qui est demandé. Dans ce qui suit, deux types de questions sont distingués, des questions superficielles et des questions conceptuelles.

Un deuxième aspect ayant un effet potentiel est l'organisation des données dans la base. Dans nombre de situations, l'organisation en réseau oblige l'ouverture de nœuds non-pertinents pour atteindre éventuellement les nœuds pertinents. Autrement dit, l'organisation de la base entraîne d'elle-même un niveau de précision et d'économie (Canter, Rivers & Storrs, 1985). Pour pouvoir estimer l'importance de cet effet, une organisation en hypertexte doit être comparée avec une organisation de type index. Dans ce dernier type d'organisation, tout nœud peut être ouvert sans passer par d'autres nœuds.

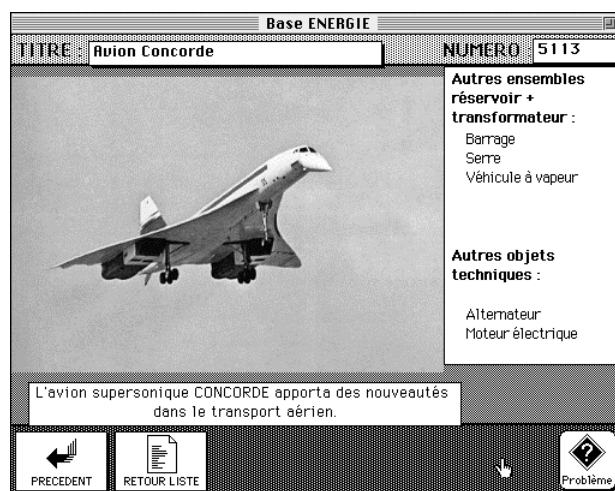
Enfin, par rapport à l'approche rationnelle, les notions de but et de structure de but permettent également de différencier d'une part la quantité d'informations pertinentes présente dans le système (le nombre de cibles), et d'autre part la quantité d'informations exigée par l'utilisateur. Dans certains cas, il n'est pas nécessaire pour l'utilisateur de trouver toutes les informations pertinentes, mais seulement d'en trouver suffisamment. Ce seuil peut être défini en fonction du

contexte dans lequel la recherche d'informations a lieu. Il provient du fait qu'un certain nombre d'informations sont soit demandées, par l'utilisateur lui-même, soit imposées par son environnement. Dans d'autres situations, l'effort à fournir pour continuer l'activité de recherche peut devenir trop important par rapport au gain estimé en termes d'information trouvée.

### *Description de la situation de recueil de données*

Cette section présente une analyse de données provenant d'une étude menée par de Vries (1995) sur le rôle d'hypermédiâs dans des situations d'apprentissage. Ces données permettent d'étudier l'influence de la structure du but sur l'utilisation en fonction de deux facteurs : le type de question (conceptuelle versus superficielle) et l'organisation des informations dans l'hypermédiâ (réseau versus index). Le contexte est celui de l'apprentissage du concept d'énergie en physique en classe de seconde. L'activité avec l'hypermédiâ faisait partie d'une séquence d'enseignement sur l'énergie. Elle avait pour but de familiariser les élèves avec les notions associées au concept d'énergie en physique, et de les préparer à une activité de résolution de problèmes mettant en jeu les notions abordées dans l'hypermédiâ.

La base utilisée dans cette étude consiste en une pile HyperCard™ de 54 cartes. Les cartes présentaient des informations au sujet de l'énergie à partir de textes, images, sons et vidéos. Chaque carte (voir Figure 2) est composée d'un titre, d'un élément (une photo, une vidéo, etc.) et d'une légende. Des exemples de titres sont Avion Concorde, Barrage, Moteur essence, Puissance vapeur. L'organisation de réseau est réalisée à partir de liens établis entre les cartes. Ces liens désignaient, du point de vue l'énergie, une similarité de deux cartes. Une version index du même matériel est obtenue en établissant une liste des titres de toutes les cartes disponibles. La navigation était très simple : un bouton Retour Liste permet de retourner à la liste de titres (index) ou à la première carte (réseau), puis un bouton Précédent permet de retourner à la carte ouverte juste avant.



**Figure 2.** Exemple de carte de la pile sur l'énergie

La consigne donnée aux élèves est de chercher et de sélectionner trois cartes avec des appareils ou des installations qui utilisent de la lumière (4 cibles), qui utilisent le mouvement pour produire de l'électricité (6 cibles), qui produisent de la chaleur (23 cibles). Ces questions emploient des termes relatifs à l'énergie. Une version superficielle des questions demandait aux élèves de chercher et de sélectionner trois cartes avec des appareils ou des installations qui utilisent des produits pétroliers pour des moyens de transport (5 cibles), qui produisent de la vapeur d'eau (9 cibles), qui utilisent de l'eau (14 cibles). Pour chacune de ces questions, le nombre de cibles possible est différent, mais l'élève n'est pas obligé de toutes les repérer. Le seuil introduit par la consigne est de trois cartes.

36 élèves de seconde ont participé, repartis aux quatre conditions organisation - tâche de la



façon suivante : réseau - conceptuelle 10, réseau - superficielle 10, index - conceptuelle 9, et index - superficielle 7 élèves. Les élèves dans la condition réseau - conceptuelle étaient supposés bénéficier le plus de leur activité avec l'hypermédia dans la mesure où ils traitaient davantage les concepts relatifs à l'énergie.

### Résultats

Le Tableau 3 montre les moyennes sur les variables de performance. Deux types de comparaisons sont faits. Premièrement, des moyennes sur les trois tâches pour les variables rappel, précision et économie, sont comparées en utilisant le test de Wilcoxon (3x3 comparaisons). Deuxièmement, les effets de l'organisation et du type de question pour chaque variable et pour chaque tâche (2x3x3 comparaisons) sont testés en utilisant le test de Mann-Whitney. Pour limiter la probabilité de trouver un effet significatif par chance, alpha est divisé par le nombre de tests ( $\alpha = .05/27 = .002$ ).

Les résultats montrent un rappel décroissant avec un nombre croissant de cibles ( $w_{T1,T2} = -3.09$  ;  $p < .002$  ;  $w_{T1,T3} = -4.83$  ;  $p < .002$  ;  $w_{T2,T3} = -3.65$  ;  $p < .002$ ). Par ailleurs, le rappel ne descend pas au niveau du seuil introduit par la consigne. Les sujets ont donc trouvé plus de cibles qu'il n'en fallait. Ceci semble particulièrement le cas quand le nombre de cibles constitue une proportion importante des nœuds de la base. Les moyennes en précision ne montrent pas d'effet du nombre de cibles. En revanche, il y a une différence en économie entre les deux tâches extrêmes ( $w_{T1,T3} = -4.9$  ;  $p < .002$ ). Un plus grand nombre de cibles produit une plus grande économie.

		Structure						
Tâche (Cibles)	Seuil	Réseau		Index		Réseau		
		Rappel	Précision	Économie	Rappel	Précision	Économie	
Conceptuelle								
T1 (4)		.75	.83	.75	.69	.95	.12	.40
T2 (6)		.50	.53	.60	.69	.88	.18	.35
T3 (23)		.13	.38	.21	.81	.95	.45	.67
Superficielle								
T1 (5)		.60	.62	.58	.63	.92	.09	.34
T2 (9)		.30	.57	.47	.76	.94	.28	.51
T3 (14)		.21	.44	.27	.70	.96	.25	.54

**Tableau 3.** Caractéristiques de l'utilisation en fonction de l'organisation de la base et du type de question (seuil introduit par la consigne dans la 1ère colonne)

Le réseau permet un rappel plus élevé que l'index lorsque le nombre de cibles est important ( $u = 4.5$  ;  $p < .002$ ). Les utilisateurs de l'index s'arrêtent lorsqu'ils ont atteint le seuil du nombre de cibles, alors que les utilisateurs du réseau en trouvent davantage en navigant. En ce qui

concerne la précision et l'économie, elles sont plus importantes dans l'index que dans le réseau quel que soit le nombre de cibles (les cartes d'entrée du réseau et de l'index ne sont pas prises en compte dans le calcul puisque leur fonction est différente dans les deux structures). L'index permet donc à la fois de viser les cibles et d'éviter les non-cibles. Des effets du type de question (conceptuelle versus superficielle) sur le rappel, la précision, et l'économie ne sont pas détectés.

### Conclusions préliminaires

Ces résultats montrent que les caractéristiques d'utilisation ne sont pas très sensibles au nombre de cibles isolément. L'organisation de la base a une influence, surtout dans le cas d'un grand nombre de cibles. Le rappel est alors plus élevé dans le réseau que dans l'index, et cet effet montre qu'une organisation de type réseau est mieux adaptée à des questions de découverte. Par contre, ce résultat est atteint avec une plus faible précision et une plus faible économie.

Le type de question n'a pas influencé l'utilisation. Une analyse des cartes effectivement marquées comme pertinentes par les élèves dans cette même étude (de Vries, 1995) montre que les élèves de seconde n'ont pas de difficultés à répondre à l'un ou l'autre type de questions. On peut se demander si le niveau scolaire des élèves a un lien avec ces effets observés. Des élèves de CM2 participant à l'étude avaient davantage de difficultés à répondre aux questions conceptuelles qu'aux questions superficielles. C'est pourquoi la section suivante présente les mêmes analyses sur les données de ces élèves.

### Le niveau scolaire des élèves

Dans le but d'étudier l'utilisation d'hypermédia par des élèves de différents niveaux scolaire, l'étude incluait un groupe de 40 élèves de CM2, répartis en 5 binômes par condition. Dans la base, l'information elle-même était bien adaptée à des élèves de ce niveau scolaire. En revanche, le réseau reposait sur des concepts non maîtrisés par ces élèves. De plus, les questions conceptuelles exigeaient de mettre en jeu ces concepts, et devaient être plus difficiles pour eux.

### Résultats

Les moyennes des caractéristiques de l'utilisation sont présentées dans le Tableau 4. Les tests Wilcoxon ne donnent pas d'effet du nombre de cibles sur le rappel pour ces élèves. Les moyennes montrent que le rappel ne décroît pas pour un nombre croissant de cibles. Par ailleurs, les moyennes pour les questions conceptuelles n'atteignent pas le seuil introduit par la consigne.

		Structure					
Tâche (Cibles)		Réseau	Index	Réseau	Index	Réseau	Index
Seuil		Rappel		Précision		Économie	
Conceptuelle							
T1 (4)	.75	.10	.50	.14	.57	.01	.09
T2 (6)	.50	.23	.14	.41	.32	.05	.04
T3 (23)	.13	.27	.07	.57	.95	.25	.50

Superficielle							
T1 (5)	.60	.60	.52	.64	.82	.10	.15
T2 (9)	.30	.44	.60	.84	.96	.28	.54
T3 (14)	.21	.40	.33	.76	.88	.24	.35

**Tableau 4.** Caractéristiques de l'utilisation pour les élèves de CM2 en fonction de l'organisation de la base et du type de question

La précision et l'économie sont plus importantes quand le nombre de cibles est élevé que quand il est réduit ( $w_{T1,T3} = -3.5$  ;  $p < .002$  ;  $w_{T1,T3} = -3.8$  ;  $p < .002$ , respectivement).

L'organisation de la base n'a pas d'effet sur le rappel pour des élèves de CM2. De plus, contrairement aux élèves de seconde, il n'y a un effet de l'organisation de la base sur la précision que dans le cas d'un nombre élevé de cibles. Dans ce cas, la précision est supérieure dans l'organisation index. Enfin, les élèves de CM2 obtiennent les mêmes moyennes en économie dans les deux structures.

En ce qui concerne le type de question, les moyennes sur le rappel, la précision et l'économie semblent plus importantes pour les questions superficielles que pour les questions conceptuelles. Néanmoins, un effet du type de question sur le rappel et l'économie a été trouvé que pour un nombre moyen de cibles ( $u = 9.0$  ;  $p < .002$  ;  $u = 0.0$  ;  $p < .002$ , respectivement). Les élèves de CM2 trouvent donc davantage de cibles lorsque les questions sont superficielles, et en faisant moins d'ouvertures de nœuds.

### Discussion

Bien que cette étude n'ait pas été conçue dans ce but, les résultats obtenus donnent une appréciation de l'utilité des variables de performance telles que le rappel, la précision et l'économie. On peut notamment observer que leurs valeurs varient selon l'organisation de la base, le type de questions, et le niveau scolaire des utilisateurs. Pour les élèves de seconde, le rappel décroît avec le nombre de cibles, l'index est particulièrement précis et économe, et il n'y a pas de différence selon le type de question. En revanche, pour les élèves de CM2, le rappel ne décroît pas avec un nombre croissant de cibles, il n'y a pas de différence selon l'organisation de la base, mais une différence existe selon le type de question. Les questions conceptuelles étaient effectivement plus difficiles que les questions superficielles pour ces élèves. L'utilisation de la base par des élèves de CM2 paraît ainsi obéir à d'autres règles. Alors que un souci d'efficacité est observé chez les élèves de seconde, notamment dans l'utilisation de l'index, les élèves de CM2 paraissent plus libres dans leur exploration de la base. Leur plus faible niveau d'abstraction ne se traduit que partiellement en une différence de rappel et d'économie entre les questions conceptuelles et superficielles. Bien que l'on puisse observer une répercussion de divers facteurs sur ces variables, les variations semblent refléter l'interprétation qu'a le sujet de la situation, plutôt qu'une mesure de l'efficacité de l'ensemble situation - tâche - hypermédia.

### Conclusions

Lors de la description de résultats, il est important de préciser les aspects pertinents de la situation étudiée. La description du contexte de l'activité peut être détaillée, avec des descripteurs classiques du domaine étudié (apprentissage, conception, recherche d'information).

Le cadre utilisé distingue deux niveaux de description de la tâche. D'une part, le niveau du contexte associé à des situations d'apprentissage ou de résolution de problème, et d'autre part, le niveau du but opérationnel pour le sujet à un instant précis. Ce but opérationnel est matérialisé dans une base hypermédia comme la structure du but exprimée en nombre de cibles. La définition de la structure du but présente un avantage de comparabilité. On peut, dans le domaine des hypermédiats, conduire des expériences dans des contextes différents dont les résultats sont formellement comparables entre eux. Cette comparaison comporte des variables dépendantes calculées également à partir du nombre de cibles. Les variables utilisées sont le rappel, la précision et l'économie, directement issues de l'évaluation des systèmes d'information (évaluation d'un algorithme, d'une technique d'indexage, etc.). Les résultats de nos études montrent l'intérêt des variables formelles pour évaluer l'utilisation de bases hypermédia. Premièrement, elles permettent de décrire l'effet des caractéristiques de la tâche du point de vue de la base : le nombre ou la proportion de cibles. Ensuite, les variables formelles se sont montrées révélatrices d'effets de l'organisation de la base. Cette organisation elle-même se caractérise par un certain niveau de précision et d'économie qui peut se conjuguer selon le type d'utilisateurs. Enfin, les résultats ont montré qu'il existe des interactions entre le type de questions et le niveau de connaissance des utilisateurs.

A présent, il s'agit de déterminer la place du rappel, de la précision, et de l'économie dans le cadre plus large de l'étude de l'utilisation des hypermédiats en contexte. En première analyse, le domaine de pertinence des variables issues de l'approche rationnelle semble restreint aux tâches de recherche d'information précise. Leurs résultats ne permettent pas d'interprétation au niveau de la réussite de l'activité en termes d'exploration ou de découverte d'information non-recherchée mais pertinente. Plus particulièrement, l'orientation du sujet introduite par la consigne, sa propre interprétation de la situation, ne s'identifie pas à un but précis en nombre de cibles. L'effet de l'interprétation du sujet (le niveau contexte) sur le rappel, la précision, et l'économie se confond avec l'effet de la structure du but et l'organisation de la base (le niveau but). C'est pourquoi ces variables ne peuvent prendre leur sens qu'en s'intégrant dans une analyse qui prend en compte des variables de description d'une activité principale, telle que la compréhension, l'apprentissage ou la résolution de problèmes. Cela peut conduire à penser que l'efficacité n'est pas une mesure de la performance dans le domaine des hypermédiats. Mais dans un second temps, il existe aujourd'hui une large variété de documents qui supportent des objectifs très divers, des plus larges et flous aux plus précis. Ainsi, le délicat problème de la catégorisation des tâches de recherche d'information a son pendant dans le choix des mesures de réussite de la tâche (voir aussi de Vries, 1997 ). Par ailleurs, le degré de précision de l'objectif dépend autant du nombre et de la disposition des cibles dans le système que de la représentation que l'utilisateur s'en fait. On peut imaginer qu'il sera possible un jour d'inférer le but et la structure du but à partir de l'analyse du comportement de l'utilisateur. A partir de cela, l'analyse de l'activité de l'utilisateur ainsi que l'évaluation et la conception des systèmes multi-usages pourront être envisagées de façon plus satisfaisante.

## Références

Bernstein M., Joyce M., Levine D., " Contours of constructive hypertexts ", in D. Lucarella, J. Nanard, M. Nanard, P. Paolini (Eds.), *Proceedings of ECHT'92 Conference*, ACM Press, p. 161-170, 1992.

Bernstein M., " Enactment in informations farming ", *Proceedings of Hypertext'93 Conference*, ACM Press, p. 242-249, 1993.

Canter D., Rivers R., Storrs G., " Characterizing user navigation through complex data structures ", *Behaviour and Information Technology*, vol. 4, n° 2, p. 93-102, 1985.

Chen C., Rada R., " Interacting with hypertext : A meta-analysis of experimental studies ", *Human-Computer Interaction*, vol. 11, n° 1, p. 125-156, 1996.

Edwards D.M., Hardman L., " 'Lost in Hyperspace' : Cognitive mapping and navigation in a hypertext environment ", in R. McAleese (Ed.), *Hypertext : Theory into Practice*, Intellect, p. 105-125, 1989.

Foss C.L., " Effective browsing in hypertexts systems ", in *RAIO Conference : User-oriented content based text and image handling*, Cambridge, 1988.

Rouet J.-F., " Interactive text processing in inexperienced (hyper-) readers ", in A. Rizk, N. Streitz, J. André (Eds.), *Hypertexts : Concepts, systems and applications*, Cambridge University Press, p. 250-260, 1990.

Salton G., McGill M.J., *Introduction to modern information retrieval*, McGraw Hill, 1983.

Swets J., " Effectiveness of information retrieval methods ", *American Documentation*, vol. 20, p. 72-89, 1969.

Tricot A., Coste J.-P., " Evaluating complex learner-computer interaction : what criteria for what task ? ", *EARLI'95 Conference*, Nijmegen, 1995.

Tricot A., Puigserver E., Berdugo D., Diallo M., " The validity of rational criteria for the interpretation of user-hypertext interaction ", *Interacting with Computers*, in press.

Vries E., de *Hypermedia for physics learning : Activating prior knowledge ?*, (Intern report nr CR-12/95). COAST, CNRS-Université de Lyon-2, 1995.

Vries E. de, Jong T. de, " Using information systems while performing complex tasks : an example from architectural design ", *International Journal of Human-Computer Studies*, vol. 46, n° 1, p. 31-54, 1997.

Waterworth J.A., Chignell M.H., " A model for information exploration ", *Hypermedia*, vol. 3, n° 1, p. 35-58 1991.