

Aspects psychologiques lies a l'interrogation d'une base de donnees

Y. Corson

► **To cite this version:**

Y. Corson. Aspects psychologiques lies a l'interrogation d'une base de donnees. RR-0126, INRIA. 1982. inria-00076434

HAL Id: inria-00076434

<https://hal.inria.fr/inria-00076434>

Submitted on 24 May 2006

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.

IRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél. 954 90 20

Rapports de Recherche

N° 126

**ASPECTS PSYCHOLOGIQUES
LIÉS À L'INTERROGATION
D'UNE BASE DE DONNÉES**

Yves CORSON

Avril 1982

IRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.: 954 90 20

Rapports de Recherche

N° 126

ASPECTS PSYCHOLOGIQUES LIÉS À L'INTERROGATION D'UNE BASE DE DONNÉES

Yves CORSON

Avril 1982

ASPECTS PSYCHOLOGIQUES LIES A
L'INTERROGATION D'UNE BASE DE DONNEES

YVES CORSON

Rapport.

BUR. 82.03. R 08

Cette étude a été menée grâce à une convention de recherche n° 81081 avec l'Agence de l'Informatique dans le cadre du projet KAYAK.

RESUME

Le développement des études concernant les aspects psychologiques liés à l'interrogation de bases de données est récent.

Sans prétendre à l'exhaustivité, nous en abordons ici quelques points sous forme d'une revue de littérature centrée principalement sur les langages artificiels d'interrogation et l'incidence sur la recherche de la structuration et de la représentation des informations.

L'apport de la méthodologie expérimentale propre aux recherches psychologiques permet de quantifier la "facilité d'utilisation" des divers langages existants, la syntaxe servant de variable dépendante.

D'autre part, des recherches plus ambitieuses conduisent à des études spécifiques des fonctions syntaxiques ou à des tentatives de modélisation des comportements, toutes deux nécessaires à notre compréhension des processus cognitifs en jeu dans l'établissement des requêtes.

Une seconde partie montre l'importance de la vision conceptuelle que peuvent avoir les opérateurs de l'organisation de la base de données.

La mise en évidence de structures représentatives a priori ou des liens entre stratégies de recherche d'information et restructuration opératoire de ces informations en fonction des objectifs ou des types de tâche peut être d'un grand intérêt ergonomique dans les recherches futures.

SUMMARY

The development of studies on psychological aspects of data base retrieval is quite recent.

In this paper some aspects of those studies, are reviewed and two issues discussed : the artificial query languages and the influence of both data structure and representation on retrieval.

The experimental methods used in psychology allow the measurement of the "ease of use" for some existing query languages. The variable most oftenly dealt with is the syntax.

Besides, more ambitious research concerned specific studies on syntactical functions or attempts to model the user's behaviour. These two topics are both necessary to a better knowledge of cognitive processes involved in queries formulation.

A second part stresses the importance of the user's conceptual perception of the data base organization. The identification of existing conceptual structures or of links between retrieval strategies and operative re-organization of the information as a function of the objectives or of the tasks can be of prime interest for future human factors research.

SOMMAIRE

ASPECTS PSYCHOLOGIQUES LIES A L'INTERROGATION D'UNE BASE DE DONNEES.

| | pages |
|---|-------|
| 1- Les utilisateurs | 8 |
| 2- A quels types de tâches peuvent être confrontés ces utilisateurs | 9 |
| 3- Apport des données psychologiques dans l'étude des syntaxes des langages | 10 |
| 4- Apport des données psychologiques dans l'étude de l'incidence de la structuration des information | 11 |

LES LANGAGES ARTIFICIELS

| | |
|---|----|
| I- Introduction | 13 |
| 1) Définition et position du problème | 13 |
| 2) Le modèle relationnel | 14 |
| II- <u>Présentation descriptive des langages</u> | 15 |
| 1) Aspects syntaxiques, mots-clés ou positionnel ?..... | 15 |
| 2) La procéduralité | 18 |
| 3) Etude affinée de quelques fonctions syntaxiques | 19 |
| a- Le mapping | 20 |
| b- Selection | 20 |
| c- Projection | 21 |
| d- Fonctions booléennes | 22 |
| e- Set operations (opérations sur les ensembles) | 23 |
| f- Built-in functions | 24 |
| g- Regroupement. (grouping)..... | 24 |
| h- Corrélation variable | 25 |
| III- <u>Aperçus méthodologiques</u> | 27 |
| 1) <u>Les variables indépendantes</u> | 27 |
| a- Les sujets | 27 |
| b- Les objectifs et hypothèses | 28 |
| c- Procédures et tâches | 31 |
| 1. Apprentissage | 31 |
| 2. Matériel | 32 |
| 3. Les problèmes | 32 |
| 4. Les tâches et tests | 34 |
| 2) <u>Les variables dépendantes</u> | 36 |
| a- Estimation subjective | 36 |
| b- Quantification objective | 36 |

En parallèle avec l'évolution respective des coûts du hardware et du software largement invoquée par ailleurs, l'apparition d'une nouvelle population d'utilisateurs est un des premiers déterminants des orientations psychologiques et cognitives que prennent les études actuelles sur l'interaction homme-ordinateur.

1- Les utilisateurs

Au vu de ces développements récents, Shneiderman (1978) a été amené à en distinguer 3 types :

- a)- L'utilisateur intermittent, non entraîné, qui n'a accès à une base de données que relativement rarement.
Il a peu de connaissances tant sur la syntaxe d'un langage interactif que sur l'organisation d'une base de données. Bien qu'il puisse comprendre les utilisations possibles appliquées à des domaines comme les réservations aériennes, par exemple, la recherche interactive suppose un intermédiaire humain ou un système informatisé rigide et dirigiste.
- b)- L'utilisateur fréquent, au fait des différentes contraintes syntaxiques, qui interroge quotidiennement la base de données. Quoiqu'il soit plus intéressé par son propre travail que par l'aspect spécifiquement informatique, il apprend une syntaxe simple et peut interagir directement, après quelques heures d'apprentissage avec la base de données.
Cette catégorie inclue la secrétaire, l'avocat, l'ingénieur, le médecin...
- c)- Enfin, l'utilisateur classique qu'est le professionnel de l'informatique.

L'utilisateur naïf tel que le définissent les études psychologiques correspond au second cas décrit par Shneiderman. Il s'agit donc de personnes sans formation informatique, mais capables d'apprendre un langage spécifiquement conçu pour l'interrogation de bases de données.

2- A quels types de tâches peuvent être confrontés ces utilisateurs ?

Les principales opérations réalisables sur une base de données concernent :

- a)- L'insertion d'un ou de plusieurs items.
- b)- L'effacement d'un ou de plusieurs items qui peut, par contre-coup, déclencher d'autres changements dans la base.
- c)- La recherche, l'extraction d'une information. Dans ce cas, l'utilisateur formule une question, une requête, dans le langage d'interrogation et la base renvoie l'information requise. C'est de cette dernière opération que nous traiterons.

Il est important de bien distinguer 2 aspects dans l'étude des processus de recherche d'information suivant que l'on s'y intéresse des points de vue informatique ou psychologique.

- c1)- D'un point de vue informatique, il existe de nombreux systèmes de recherche d'information.

Certaines études ont suggéré une méthode d'indexation probabiliste (Maron et Khuns 1960) ou des méthodes indirectes de recherche basées sur la dépendance entre documents (Goffman 1968, Doyle 1961, Mansur 1960).

D'autres recherches s'intéressent à des structures utiles pour les interrogations floues (Salton 1976) ou à des comparaisons entre systèmes du point de vue du coût (Croft et van Rijsbergen 1976).

- c2)- D'un point de vue psychologique, qui nous intéresse ici, c'est-à-dire des processus que l'utilisateur doit mettre en œuvre pour interagir efficacement avec la base, plusieurs domaines ont été inégalement abordés.

Des recherches de type psycho-moteur ou perceptif n'ont généralement pas été explicitement entreprises, bien que certains des travaux réalisés tant sur les entrées (GREV 1978) que sur les sorties (Miller et Thomas 1977) ont pu s'en inspirer.

Thomas (1977) mentionne, d'autre part, les données intéressantes que pourrait apporter l'importante littérature psychologique concernant la mémoire, la résolution de problèmes la motivation ou la structuration et l'organisation des informations.

L'apport des données psychologiques, d'un point de vue expérimental, est le plus important en ce qui concerne la syntaxe des langages et l'incidence de l'organisation des informations dans la base de données.

Nous avons donc choisi de privilégier ces deux aspects afin, d'une part, de rendre compte des travaux réalisés dans ces deux domaines, et, d'autre part, de souligner l'intérêt tant thématique que méthodologique de ces recherches dans le développement des interactions avec une base de données.

3- Apport des données psychologiques dans l'étude des syntaxes des langages.

La syntaxe des langages est certainement l'aspect qui permet le mieux d'évaluer leur "facilité d'utilisation".

Les divers langages testés jusqu'à aujourd'hui reposent en grande majorité sur des syntaxes présentant des structures classiques, qu'elles soient naturelles ou artificielles.

Signalons que certaines tentatives se sont démarquées de ces approches plus courantes. Nous ne ferons, pour les illustrer, que mentionner une technique intéressante d'extraction de l'information par positionnement dans un espace graphique.

Le lecteur intéressé pourra se reporter aux travaux de Herot sur la gestion spatiale des données (1979).

En ce qui concerne les langages plus classiques, d'interrogation de bases de données, Thomas et Gould (1975) proposent de distinguer les différents langages sur la base de leur ressemblance avec l'anglais.

Ceci permet de différencier trois approches selon qu'il s'agit d'une syntaxe naturelle, d'une syntaxe formelle mais utilisant un vocabulaire et une grammaire ressemblant à l'anglais (IQF, SQL...) ou d'un langage artificiel construit sans souci de ressemblance avec quelque langage naturel que ce soit (QBE).

Si l'utilisation d'un langage naturel élimine la nécessité astreignante d'un apprentissage, certaines critiques (Shneiderman 1978, Montgomery 1972) par contre, mettent l'accent sur de nombreuses difficultés.

Sans préjuger des possibilités offertes par les langages naturels, (dont une revue : sigart newsletter n° 61 février 1977 dénombre 52 projets de développement), il semble actuellement difficile de les rendre opérationnelles sans la lourdeur et la lenteur que leur impose la nécessité de feedbacks continuels entre l'utilisateur et la base.

Les recherches psychologiques se sont donc intéressées en priorité aux développements des langages artificiels et Nickerson et al. (1968) sont parmi les premiers à poser explicitement les questions qui seront débattues par la suite :

- quels sont les principes psychologiques qui doivent guider la conception des langages ?
- quels critères utiliser pour évaluer les langages existants et pour choisir entre plusieurs alternatives ?
- connaissons-nous assez les caractéristiques générales de la résolution de problème pour spécifier les déterminants d'un langage spécifiquement conçu pour la résolution de problème ?

Dans la première partie de cette revue de littérature, nous nous intéresserons donc aux langages artificiels.

Après une présentation descriptive des langages et des exemples de fonctions syntaxiques propres aux langages artificiels certains aspects méthodologiques seront abordés.

C'est sur ces bases nécessaires à la compréhension des recherches effectuées que nous présenterons les principaux résultats évaluatifs et comparatifs sur les langages en privilégiant des approches plus fines (Reisner 1976, Thomas 1976) qui illustrent bien l'apport de certaines connaissances ou méthodes psychologiques.

4- Apport des données psychologiques dans l'étude de l'incidence de la structuration des informations.

Dans la seconde partie, nous aborderons les problèmes, importants pour l'interaction efficace homme-ordinateur de la représentation et de la

structuration des informations.

Seront développées les recherches réalisées, entre autres, sur l'importance des relations sémantiques entre items de la base de données, sur l'incidence du choix de l'organisation de ces items et sur l'existence de certaines structures représentatives (matrices , hiérarchies, réseaux...).

Enfin, nous indiquerons, en guise de conclusion, les différentes directions de recherche envisagées.

LES LANGAGES ARTIFICIELS

I- INTRODUCTION

1- Définition et position du problème.

Comme nous l'avons précédemment mentionné, les systèmes existants qui permettent d'interroger une base de données sont nombreux. De l'ensemble important de tous les langages formalisés, nous ne traiterons que des plus performants parmi ceux ayant fait l'objet de quelques études psychologiques.

Reisner (1981) définit un langage d'interrogation comme un langage spécialement conçu afin de composer des questions ayant pour but de retrouver, d'extraire de l'information à partir d'une base de données stockée dans un ordinateur.

La motivation essentielle des recherches psychologiques entreprises est issue de la volonté de mesurer, de quantifier les facilités relatives d'utilisation des divers langages dans le but d'en améliorer les performances. Les objectifs des expériences sont multiples : en général, le chercheur s'attache à déterminer les conditions d'existence optimales des modalités d'un langage favorisant l'interaction entre la base de données et une population d'utilisateurs naïfs telle que nous l'avons définie auparavant.

Il paraît également important de découvrir, avant que ne soit définitivement développé le langage, les types d'erreurs les plus usuels et les caractéristiques fonctionnelles du langage qui posent problème. Une fois détectées, les principales difficultés peuvent alors être éliminées ou, à tout le moins, réduites en modifiant la fonction syntaxique en cause, en procurant des aides interactives (guidage, systèmes de questions-réponses, feedback...) ou en perfectionnant l'apprentissage.

Etudier les aspects psychologiques impliqués dans l'utilisation d'un langage d'interrogation suppose donc de savoir ce que font les sujets quand ils écrivent des questions ou des requêtes, pourquoi ils font tel ou tel type d'erreur, ou encore ce qui rend telle ou telle fonction syntaxique difficile à apprendre ou à manipuler.

2- Le modèle relationnel

Sous-jacents aux langages, plusieurs types de modèles d'organisation des données peuvent être utilisés.

Malgré le nombre très restreint d'expérimentations sur ce thème, il semble que la structuration sous forme de relations (ou tables), de hiérarchies ou de réseaux, ait un impact significatif sur la facilité de manipulation des données et de représentation de leurs relations.

Nous n'aborderons que par la suite l'étude des expériences qui rendent compte de comparaisons entre modèles.

Toutefois, précisons dès à présent que les langages que nous mentionnerons reposent tous sur le modèle relationnel. Dans ce cas, les données de la base sont représentées sous forme de tables qui ont un nom et un nombre variable de colonnes et de lignes.

Ces représentations matricielles sont définies par les propriétés suivantes (ZLOOF 1975) :

- a) toutes les lignes ou rangées sont distinctes
- b) l'ordre des rangées est non pertinent
- c) l'ordre des colonnes est également non pertinent et chacune d'elles est désignée par un nom distinctif.

Les figures suivantes présentent 2 exemples de table décrivant les employés et points de vente d'une entreprise de distribution de livres.

Dans la première relation, chaque rangée de la table EMPLOYES relie le nom d'un employé à son lieu de travail, à son salaire et au numéro du service auquel il appartient.

| EMPLOYES | NOMS | LOCALISATION | SALAIRE | N° SERVICE |
|----------|--------|--------------|---------|------------|
| | DUPONT | PARIS | 12000 | 12 |
| | MARTIN | LYON | 8000 | 50 |
| | ROGER | PARIS | 9500 | 44 |
| | CARIER | BORDEAUX | 10100 | 36 |
| | ⋮ | ⋮ | ⋮ | ⋮ |

| PTS DE VENTE | LOCALISATION | GENRE | CHIFFRE D'AFFAIRE |
|--------------|--------------|-----------------|-------------------|
| | PARIS | SCIENCE FICTION | 355000 |
| | MARSEILLE | HISTOIRE | 275000 |
| | STRABOURG | ECONOMIE | 300000 |
| | ⋮ | ⋮ | ⋮ |

II- PRESENTATION DESCRIPTIVE DES LANGAGES

Avant d'aborder véritablement le propos de cette étude, nous présentons, dans les quelques pages suivantes, des données essentiellement descriptives nécessaires à une meilleure compréhension des problématiques sous-jacentes aux recherches psychologiques entreprises.

1- Aspects syntaxiques : mots-clés ou positionnel ?

La syntaxe est un critère évident de différenciation entre langages.

Certains projets de langage d'interrogation comme QUEL, SQL sont à base de mots-clés, leur structure reposant sur une conception traditionnelle de la programmation.

D'autres projets ont introduit une notation bi-dimensionnelle dans laquelle la position des items est pertinente : SQUARE, CUPID, QBE, FORAL, LP. Dans ce cas, l'utilisation de support graphique fourni par le système évite presque totalement le maniement de mots-clés. Ainsi, en ce qui concerne le QBE, l'opérateur a à sa disposition un schéma structuré sous forme de tableau et les questions sont composées en remplissant les colonnes.

A ce propos, on peut noter que les partisans respectifs de ces 2 optiques basent leur choix sur les 2 facettes du même argument : si, d'un côté, certains auteurs pensent que les mots-clés aident à l'apprentissage et à la composition des questions en associant l'aspect sémantique de la requête à la terminologie familière (exemple du SQL : SELECT, FROM WHERE), de l'autre, une notation positionnelle ou des figurations spatiales sont justement supposées éviter des confusions éventuelles nées de cette association.

En tout état de cause, les aspects psychologiques fondamentaux liés à l'utilisation de mots-clés ou de notations plus figuratives ne sont pas étudiés.

Il est clair qu'en ce qui concerne ce problème, il ne s'agit plus de travailler au niveau d'une simple caractéristique d'un langage, mais au niveau d'une conception globale de la syntaxe et que les processus psychologiques sous-jacents sont aussi complexes que variés.

Les langages SQL (Structured English Query Language) et QBE (Query By Example) étant d'une part les plus performants et, d'autre part, représentatifs des langages par mots-clés et notation positionnelle sont ceux dont nous approfondirons certains des aspects.

D'une manière générale, SQL utilise donc un système de mots-clés à connotations sémantiques similaires à celles du langage naturel. Dans l'exemple utilisé, la fonction appelée "mapping" donne à l'utilisateur une liste de tous les employés (SELECT NAME) correspondant au critère défini par le mot WHERE (WHERE DEPTNO = 50), la relation entre ces deux valeurs se faisant à l'intérieur de la table EMPLOYEE (FROM EMP).

Avec le QBE, l'utilisateur formule sa requête en remplissant les lignes appropriées de la table au moyen d'un exemple de réponse possible. Pour les questions les plus simples, comme le mapping, l'utilisateur a seulement besoin de distinguer entre les 2 critères suivants :

- a) l'élément exemple (variable) qui doit être souligné :
DUPONT
- b) l'élément constant correspondant au critère de sélection des noms : 50 dans la colonne DEPTNO.

La lettre p. signifie "imprimer" et spécifie la colonne dont l'utilisateur veut voir sortir les éléments.

Ainsi, l'entrée p.DUPONT dans la colonne NAME correspond au SELECT NAME du SQL (le mot DUPONT est facultatif et peut être omis).

les langages SQUARE (specifying Queries As Relational Expressions) et TABLET présentent d'autres exemples de la même requête.

Le premier utilise une notation positionnelle, mais différente du QBE, tandis que le second est différent de par sa structure et sa conception dans la mesure où entre en jeu la notation de procéduralité.

2- La procéduralité

La procéduralité ou non-procéduralité d'un langage d'interrogation est à l'instar de la syntaxe une caractéristique importante distinguant les divers langages.

Date (1977) définit un langage non-procédural --que Shneiderman (1978) nomme "specification language"-- comme établissant simplement ce que le résultat de la question doit être, sans spécifier les modalités d'obtention de ce résultat.

Un langage procédural décrit par contre les processus, les différentes étapes qui sont nécessaires pour atteindre le but désiré.

Cette classification n'est cependant pas discrète. Divers langages peuvent, en effet, se situer différemment sur le continuum entre ces 2 extrêmes.

Les langages SQL et QBE sont considérés comme non-procéduraux puisque seules les caractéristiques d'un ensemble doivent être spécifiées pour le retrouver.

Par contre, le langage TABLET présente une structure nettement procédurale. En effet, les 3 mots-clés du langage (voir tableau) définissent les étapes de la recherche :

```
FORM   crée la table DEPTFIFTY à partir des colonnes NAME et
        DEPTNO de la table EMP
KEEP   extrait les lignes de la table créée où le DEPTNO est
        50
PRINT  permet d'obtenir le résultat final.
```

Welty et Stemple (1981) ont réalisé une expérience dans laquelle la plus grande facilité d'utilisation du langage TABLET sur le langage SQL pour les questions les plus complexes les conduit à postuler que les différences observées sont dues aux différences de procéduralité des modèles sous-jacents.

Toutefois, les fondements psychologiques de ces résultats sont, en fait, peu étudiés. L'existence éventuelle d'une variation inter-sujet ou inter-tâche n'est pas abordée. De même, les effets possibles d'une interaction entre les modalités de ce facteur et les fonctions syntaxiques ou les différentes étapes de traduction d'une requête n'ont fait l'objet d'aucune expérimentation.

Les questions sont nombreuses : y-a-t-il des variables de type cognitif qui puissent indiquer quelle méthode est préférable pour certains sujets ou pour certaines tâches ? Est-ce que la composition serait plus facile avec un langage procédurale et la compréhension plus aisée avec un "specification language" ?

Autant de questions auxquelles des réponses sont nécessaires avant de statuer sur cette distinction entre langages procéduraux et non-procéduraux.

Syntaxe et procéduralité sont les points de focalisation les plus prégnants lors d'une première approche des langages d'interrogation.

Cependant, pour répondre aux questions posées, l'étude globale, générale de ces 2 aspects ne suffit pas.

Il est, en effet, nécessaire d'approfondir certaines caractéristiques particulières par le biais de l'étude des différentes opérations qu'il est possible d'effectuer grâce aux fonctions spécifiques que fournit chacune des syntaxes.

3- Etude affinée de quelques fonctions syntaxiques

Les études psychologiques globales, qu'elles soient évaluatives ou comparatives, n'ont d'intérêt que si elles sont centrées sur les déterminants, les causes des effets observés.

S'il est intéressant de savoir que tel type de langage est plus approprié à tel type de tâche, les améliorations que peut apporter l'expérimentation psychologique suppose une approche plus fine que la simple mise en évidence d'une certaine "facilité d'utilisation".

D'autre part, toute étude expérimentale a pour but si ce n'est d'établir les caractéristiques complètes d'un bon langage, tout au moins d'essayer d'améliorer les points faibles.

Ceci suppose donc d'identifier les problèmes posés par l'étude spécifique des fonctions difficiles à traiter et d'agir au niveau des causes de ces difficultés.

Il existe, suivant les langages et leurs niveaux respectifs de développement, un nombre variable de fonctions qui permettent d'effectuer diverses manipulations sur la base de données.

En conservant comme fil conducteur les 2 langages SQL et QBE (ainsi que le SQUARE en tant qu'illustration secondaire), nous décrivons dans les pages qui suivent quelques unes de ces fonctions au moyen d'exemples. (1)

Il ne s'agit bien sûr pas d'une description exhaustive mais simplement d'une présentation partielle des principales possibilités offertes par les langages les plus courants.

Ceci devrait permettre d'une part de clarifier certaines notions, d'autre part, d'aborder des exemples de ce qu'une étude, plus complexe et détaillée, tenant compte des processus psychologiques sous-jacents à l'utilisation de ces fonctions, peut apporter quant à la mise au point des langages.

a) Le mapping

Le concept central dans chaque langage, est celui de mapping qui extrait les données associées à une autre donnée connue dans une colonne de la table choisie. Si les items sont le numéro de l'employé et son nom, le mapping entre eux est simple : chaque nom a un numéro d'employé qui lui est associé et vice-versa.

Exemple et explication du mapping sont donnés en page 16.

b) Selection

Permet d'obtenir toutes les données d'une ligne en recevant toutes les valeurs associées à une donnée spécifique. En ce qui concerne les langages SQL et SQUARE, les noms des colonnes qui doivent être sélectionnées sont omis de la question.

(1) Les exemples concernant le SQL et le SQUARE sont tirés de Reisner (1976). Les exemples concernant le QBE sont composés en fonction des requêtes illustrant chaque fonction des 2 premiers langages.

Exemple : "imprimer toutes les données de l'employé dont le nom est DUPONT".

SQL : SELECT EMP
 WHERE NOM = 'DUPONT'

SQUARE : EMP ('DUPONT')
 NOM

QBE :

| EMP | NOM | LOCALISATION | SALAIRE | N° SERVICE |
|-----|--------|--------------|---------|------------|
| p. | DUPONT | | | |

c) Projection

Cette fonction permet de faire sortir toutes les données d'une colonne.

La clause WHERE est alors omise.

Exemple : "donner la liste de tous les employés".

SQL : SELECT NOM
 FROM EMP

SQUARE : EMP
 NOM

QBE :

| EMP | NOM | LOCALISATION | SALAIRE | N° SERVICE |
|-----|------------------|--------------|---------|------------|
| | p. <u>DUPONT</u> | | | |

d) Fonctions booléennes

Il s'agit de fonctions qui permettent l'utilisation des connecteurs AND/OR/NOT.

Ces fonctions de conjonction et de disjonction s'utilisent à l'intérieur d'un mapping.

Exemple : "trouver les noms des employés qui travaillent sous la direction du manager Martin et qui ne sont pas dans le service 50".

SQL: SELECT NOM
 FROM EMP
 WHERE MGR = 'MARTIN'
 AND N° SERVICE ≠ 50

SQUARE : EMP ('MARTIN, ≠ '50')
 NOM MGR, N° SERVICE

QBE :

| EMP | NOM | N° SERVICE | MGR |
|-----|------------------|------------|--------|
| | p. <u>DUPONT</u> | ≠ 50 | MARTIN |

Autres exemples : "Trouver les noms des employés qui travaillent pour Martin et Durand".

QBE :

| EMP | NOM | N° SERVICE | MGR |
|-----|------------------|------------|--------|
| | p. <u>DUPONT</u> | | MARTIN |
| | <u>DUPONT</u> | | DURAND |

"trouver les noms des employés qui travaillent pour Martin ou Durand".

QBE :

| EMP | NOM | N° SERVICE | MGR |
|-----|------------------|------------|--------|
| | p. <u>DUPONT</u> | | MARTIN |
| | p. <u>ROGER</u> | | DURAND |

L'utilisation des connecteurs booléens montrent ici une différence importante entre SQL et QBE : les principaux résultats obtenus indiquent que l'emploi pour le SQL de mots-clés (AND) identiques au langage naturel provoque des erreurs ; en effet, une formulation du type "mettre les pions rouges et verts dans la boîte" correspond, dans un langage de programmation à "mettre les pions (verts ou rouges) dans la boîte". Le QBE, introduisant une différence fondamentale de structure dans la syntaxe entre AND et OR évite les erreurs de ce type, fréquentes dans les autres langages.

e) Set operations (opérations sur les ensembles).

Ces fonctions permettent de réaliser des unions, des intersections ou toute opération sur des ensembles entre mapping.

Exemple : "trouver les numéros des services localisés à Bordeaux qui ont des employés gagnant moins de 5000 F".

```
SQL :      SELECT      N° SERVICE
           FROM        Pts DE VENTE
           WHERE       LOC = 'BORDEAUX'
           ∩
           SELECT      N° SERVICE
           FROM        EMP
           WHERE       SAL < 5000 F.
```

```
SQUARE :      PtsVENTE ('BORDEAUX') ∩      EMP (< '5000')
              N° SERVICE      LOC      N°SERVICE SAL
```

QBE :

| PTS DE VENTE | N° SERVICE | LOCALISATION |
|--------------|-------------|--------------|
| | P: <u>X</u> | BORDEAUX |

| EMP | N° SERVICE | SALAIRE |
|-----|------------|---------|
| | <u>X</u> | < 5000 |

f) Built-in functions

Permettent de réaliser certaines opérations sur les valeurs de la base.

Les principaux langages prévoient l'utilisation d'opérateurs du type MAXIMUM, MINIMUM, AVERAGE, COUNT, SUM...

Exemple : "donner le salaire moyen dans le service 50".

SQL : SELECT AVG (SAL)
 FROM EMP
 WHERE N°SERVICE = 50

SQUARE : AVG (EMP ('50'))
 SAL N°SERVICE

QBE :

| EMP | NOM | SALAIRE | MGR | N°SERVICE |
|-----|-----|-------------------|-----|-----------|
| | | p.AVE <u>5000</u> | | 50 |

g) Regroupement (Grouping)

Cette fonction permet de réaliser un regroupement d'items ayant une valeur commune.

Exemple : "donner les noms des services dont le salaire moyen est supérieur à 5000 F."

SQUARE introduit une notation appelée "variable libre" : l'utilisateur invente une variable (X) qui représente une ligne, une rangée, puis fait un test qui définit les lignes qu'il veut sélectionner.

X E EMP : AVG (EMP (X)) > 5000
 N°SERVICE SAL N°SERVICE N°SERVICE

Le langage SQL évite l'utilisation de ce type de variable en employant une opération appelée GROUP BY qui permet

de grouper les employés par service avant que ne soit réalisé le moyennage.

```
SELECT    N°SERVICE
FROM      EMP    GROUP BY N°SERVICE
WHERE     AVG    (SAL) > 5000
```

QBE :

| EMP | NOM | N° SERVICE | SALAIRE | MGR |
|-----|-----|--------------------|------------|-----|
| | | p. <u>ECONOMIE</u> | AVE > 5000 | |

h) Correlation variable

Certaines questions qui utilisent les variables libres dans le SQUARE utilisent les "correlations variables" dans le SQL.

Exemple : "trouver les noms des employés dont le salaire est plus élevé que celui de leur manager".

SQL :

```
B1 : SELECT    NOM
FROM          EMP
WHERE         SAL >
              SELECT    SAL
              FROM      EMP
              WHERE     NOM = B1.MGR
```

Dans cette question le mapping étiqueté B1 extrait le nom de toutes les rangées qui valident le test suivant : le salaire de la rangée B1 doit être plus grand que le salaire de la rangée dont le nom est le même que le manager de la rangée B1.

QBE :

| EMP | NOM | SALAIRE | MANAGER | N°SERVICE |
|-----|------------------|------------------|---------------|-----------|
| | <u>p. MARTIN</u> | <u>> 10 K</u> | <u>PIERRE</u> | |
| | <u>PIERRE</u> | <u>10 K</u> | | |

Si PIERRE est un exemple de manager gagnant 10K, alors MARTIN est un exemple d'un employé qui gagne plus que 10 K et donc plus que son manager.

Ces quelques exemples ne prétendent pas à l'exhaustivité. Il ne s'agit que d'illustrations descriptives permettant une meilleure compréhension des problématiques sous-jacentes aux études psychologiques entreprises.

Le lecteur intéressé pourra se reporter à Thomas et Gould 1975, Zloof 1975 et Reisner 1974.

III- APERCUS METHODOLOGIQUES

L'utilisation de la méthodologie expérimentale est très récente en ce qui concerne l'interrogation de bases de données.

Barret (1968) note que les premières études ne faisaient que comparer les quelques systèmes existants sur la base de questionnaires tendant à rendre compte de leur "facilité d'utilisation". La performance du système est alors estimée au moyen d'un indice de satisfaction de l'utilisateur dont les variations s'expliquent en premier lieu par les niveaux d'apprentissage ou de compétence.

En fait, cette méthode dépend surtout de la capacité individuelle du sujet à identifier les habiletés cognitives requises.

D'autre part, ces jugements subjectifs sur la facilité d'utilisation du mode interactif ou d'une caractéristique du langage ne sont pas toujours corrélés avec les mesures effectives de performances.

Il arrive fréquemment, en effet que les sujets trouvent la tâche facile alors qu'ils n'obtiennent que des résultats médiocres.

Les travaux dont nous rendons compte ont tous basé leurs expérimentations sur les principes de méthodologie expérimentale. Toutefois, certains préalables méthodologiques sont nécessaires avant toute discussion des résultats.

Nous nous intéresserons donc, dans un premier temps à une étude rapide des divers plans d'expérience avant d'aborder, dans un second temps, l'analyse des principales données, tant évaluatives que comparatives, obtenues sur les langages.

1- Les variables indépendantes.

a) Les sujets

Ce sont généralement des étudiants ou employés de provenances très diverses. Les différents langages testés devant être manipulés par des non-professionnels de l'informatique, il s'agit le plus souvent de sujets naïfs. Cependant certaines caractéristiques relevant de l'arrière-plan scientifique et éducatif ont servi de variables indépendantes.

Ainsi, le taux d'expérience de programmation (Reisner 1976), le niveau de mathématique et le nombre d'années scolaires (Greeblatt et Waxman 1978), l'âge ou le sexe ont parfois déterminé certaines conditions expérimentales.

b) Objectifs et hypothèses

L'objectif principal des études évaluatives est de fournir une estimation quantifiée de la facilité d'utilisation.

Nickerson et al. (1968) définissent la facilité d'utilisation comme permettant à un individu d'interagir effectivement avec le système après une série d'instruction minimum. Le problème revient donc à concevoir un système tel qu'un sujet débutant puisse l'utiliser effectivement après quelques minutes d'apprentissage, et puisse apprendre à l'utiliser efficacement à partir du feedback fourni par le système lui-même.

Cet objectif général explique que les procédures expérimentales mises en œuvre soient essentiellement exploratoires. Les variations systématiques réalisées sur les variables indépendantes concernent principalement les utilisateurs, le type de tâches et de tests proposé ou la structure de la base de données elle-même.

Le langage est alors considéré comme un ensemble dont on évalue l'efficacité et, à part quelques tentatives plus fines pour comprendre et modéliser les comportements de traduction observés (Reisner 1981), les conclusions sont essentiellement normatives et se contentent de localiser les difficultés sur telle ou telle fonction sans apporter de solutions alternatives.

La plupart des expérimentations tentent de répondre à des questions du type :

- 1- Les langages peuvent-ils être bien appris ? Y-a-t-il des différences dans les degrés d'apprentissage des divers langages ?

- 2- Y-a-t-il des différences inter ou intra-langages en ce qui concerne certaines fonctions ou opérations spécifiques? Y-a-t-il certaines fonctions particulièrement difficiles à apprendre ?

L'originalité de la procédure réside dans le fait que le sujet devait effectuer cette tâche en 3 étapes :

- 1- Pour chaque question, il doit retranscrire à sa manière ce que la requête signifie. Le temps enregistré donne un temps de formulation.
- 2- Le sujet doit ensuite organiser, planifier sa question.
- 3- Enfin, l'étape de codage consiste en la traduction dans le langage.

D'une manière générale, la distinction entre ces 3 étapes semble pertinente : si le temps de formulation dépend plutôt de la richesse de l'expression de la question, les temps de planification ou de codage dépendent des types de tâche à réaliser sur les éléments récupérés de la base.

- Thomas (1975) s'intéresse plus particulièrement, mais de manière incidente, aux processus de traduction des questions impliquant l'utilisation et le traitement des quantificateurs universels et existentiels.
- Greenblatt et Waxman (1978), pour éviter des comparaisons entre langages à partir de résultats obtenus sur la base de plans différents, testent 3 langages sur les mêmes critères d'apprentissage et font varier les facteurs dans les mêmes conditions.
- Reisner, Boyce et Chamberlain (1974) et Reisner (1976) étudient le SQL de telle manière qu'une analyse fine de certaines fonctions leur permet de tenter une modélisation du comportement des sujets par la prise en compte du nombre d'opérations nécessaires à la transformation d'une question du langage naturel dans le langage artificiel.
Ce modèle traitant des difficultés relatives de certaines opérations, peut être utilisé comme prédicteur d'erreurs dans la composition des requêtes.

Il est de ce fait assez rare que le choix des variables indépendantes repose sur des hypothèses précises pour lesquelles l'expérimentation apporterait confirmation ou infirmation.

Cependant, certains auteurs (Reisner 1976, Thomas et Gould 1975, Gould et Ascher 1975) pensent que le seul moyen d'améliorer véritablement un langage est de développer notre compréhension des processus impliqués dans la mise au point des requêtes, processus existants à l'arrière-plan des comportements observés.

Ceci suppose donc d'essayer :

- 3- D'identifier, pour les fonctions les plus difficiles, les sources de difficultés.
- 4- D'apporter des aides ou de trouver des solutions pour diminuer le taux des erreurs les plus fréquentes.
- 5- De modéliser le comportement de traduction à partir des différentes stratégies utilisées par les sujets.

Répondre à ses questions implique de concevoir des plans d'expériences rigoureux testant de façon sélective des hypothèses précises soit sur la totalité d'un ou de plusieurs langages, soit sur certaines caractéristiques spécifiques.

Ainsi :

- Gould et Ascher (1975) ont tenté, en étudiant le langage IQF (Interactive Query facility) d'analyser séparément la formulation, la planification et le codage des questions.

Le but principal de cette étude était de repérer les caractéristiques, les fonctions générales des langages qui sont les plus difficiles à traiter pour les non-programmeurs. La méthode consistait, après apprentissage à demander au sujet de traduire 15 questions de l'anglais dans l'IQF.

- Thomas et Gould (1975) ne font pas seulement des distinctions au niveau des fonctions, c'est-à-dire au niveau des opérations effectuées sur les éléments de la base, mais font varier également le nombre de relations et de connecteurs impliqués par la question.
- Welty et Stemple (1979) font varier les niveaux de complexité des questions pour tester l'existence éventuelle d'une interaction entre ce facteur et la procéduralité inhérente à certains langages.

c) Procédures et tâches

1- Apprentissage

Toute expérimentation, qu'elle soit évaluative ou comparative, suppose l'apprentissage de la syntaxe. C'est une priorité nécessaire et commune à tous les langages d'interrogation.

Cet apprentissage peut-être individuel (Gould et Ascher 1975 sur le langage QBE), ou collectif (Thomas et Gould 1975 sur le langage QBE), distribué (12 périodes de 50 minutes : Reisner 1976) ou massé (QBE). D'autre part, à l'issue d'une phase d'instruction à partir d'une plaquette mise au point par l'expérimentateur, celui-ci peut-être amené à donner des explications individualisées.

Enfin, certains auteurs profitent de la période d'apprentissage pour utiliser des bases de données différentes (monopoly, school administration, computer dating, car sales, airport administration, department store...).

2- Matériel

Peu d'expériences sont effectivement réalisées on-line. La grande majorité des études traitant exclusivement de l'interaction par l'intermédiaire du langage, les expérimentations sont réalisées avec papier/crayon. La consigne n'implique donc généralement pas la recherche réelle d'une information dans une base de données, mais la traduction ponctuelle d'une requête, c'est-à-dire la seule manipulation de la syntaxe.

Les tâches requises par la recherche on-line peuvent être considérées à plusieurs niveaux.

Un premier niveau est défini par le degré de difficulté de résolution de la situation posée.

Le second rend compte des différents types de tâches impliqués dans la mise au point d'une question spécifique.

Le premier niveau faisant référence à une catégorisation, plus générale, nous préférons employer le terme de problème.

3- Les problèmes.

T.Martin (1980) distingue 4 types de problèmes (qu'il nomme "information retrieval tasks").

Le plus simple est du type "quel est le numéro de téléphone de tel restaurant?".

Dans la mesure où la recherche à entreprendre ne requiert pas de mises en relation complexe entre descripteurs, les possibilités d'erreurs sont minimales.

Le second degré de complexité concerne des questions du type "quels sont les documents qui traitent de l'enseignement secondaire en Bretagne et en Vendée?".

Le problème posé spécifie un certain nombre de mots caractérisant des concepts que l'utilisateur doit mettre en relation.

Ces relations peuvent être variables : intersection (enseignement et secondaire) ou union (Bretagne et Vendée) par exemple.

Martin (1980) nomme le troisième degré "aide à la décision" bien que formellement il n'y ait pas de différences avec le degré précédent.

Comme pour le second degré, le problème est bien structuré, les catégories d'informations sont connues et l'utilisateur doit exprimer ces catégories et les relations qu'il veut voir apparaître en langage formel.

Les exigences formelles sont en fait les mêmes pour ces deuxième et troisième degrés bien qu'il ne s'agisse plus d'une simple recherche d'un document mais de recherche de données, éventuellement nombreuses, pour aider à une décision quelconque.

Enfin, le niveau le plus difficile concerne la résolution de problèmes. Un problème du type "quelle route dois-je prendre pour me rendre à tel endroit de la ville ?" suppose la gestion d'une somme d'informations beaucoup plus importante et surtout une recherche dans le cadre d'une situation moins bien définie ou structurée que précédemment.

L'ensemble de la littérature traitant de la recherche d'informations à partir d'une base de données ne fait pas de distinction explicite entre ces différents niveaux bien que les exigences requises modifient de façon importante la complexité de la résolution.

Dans la discussion qui suit sur les langages nous ne tiendrons compte que des contraintes impliquées par les degrés 2 et 3, c'est-à-dire que nous ne traiterons que des performances réalisées entre l'établissement précis de l'information ponctuelles recherchée et la mise au point de la requête en langage formel.

Le quatrième degré sera éventuellement abordé par la suite lorsque nous discuterons de l'importance de la structuration des données dans la recherche et le traitement de l'information.

4- Les tâches et tests

Afin de comparer ou d'évaluer les divers langages, les expérimentateurs ont mis au point un certain nombre de tâches différentes.

- Dans la plupart des cas, la tâche principale est une tâche de traduction et d'écriture de questions : le sujet doit traduire une question formulée en langage naturel dans le langage d'interrogation testé.

Ces questions peuvent tester des fonctions de base du langage ou les capacités des sujets à combiner les fonctions d'une manière qu'ils n'ont pas vue précédemment.

D'autre part, les requêtes sont construites de telle manière qu'elles permettent les comparaisons de fonctions spécifiques comme AND/OR (Reisner 1976) ou SUM/COUNT (Thomas et Gould 1975) par exemple.

- Dans le cas de lecture ou de compréhension, les tâches sont inversées : on propose alors au sujet une question dans le langage formel et il doit :

- donner un équivalent en Anglais
- trouver la réponse parmi un ensemble de tables servant de base de données.

- Enfin, sont également utilisées des tâches de mémorisation dans lesquelles le sujet doit mémoriser et reproduire une base de données ou plus rarement de résolution de problèmes dans lesquelles le sujet doit générer lui-même les questions permettant de découvrir la solution.

Pour une même tâche, entrent en jeu également plusieurs niveaux de complexité des questions en fonction du nombre de variables établissant des relations qui peuvent être inter ou intra-tables.

D'autre part, différentes questions peuvent varier en fonction du nombre de contraintes conjonctives, du

nombre d'ensembles disjonctifs ou du type d'opérateurs impliqués.

L'exemple suivant illustre une question avec 2 ensembles disjonctifs requérant chacun 4 contraintes conjonctives :

On veut savoir combien d'hommes mariés de moins de 30 ans travaillent dans le département "science fiction" et combien d'employés célibataires travaillent depuis plus de 5 ans à Bordeaux dans le département "économie".

Ces tâches peuvent être utilisées dans des buts et à des moments différents suivant ce que l'expérimentateur veut tester :

- Le test le plus fréquent consiste en un examen final, qui a lieu après apprentissage, de l'ensemble des fonctions du langage.
- Par ailleurs, certaines épreuves peuvent être utilisées pour tester la compréhension immédiate d'une fonction au cours de l'apprentissage, le sujet connaît alors la fonction à utiliser, ou de l'ensemble des fonctions apprises précédemment, le sujet devant alors déterminer laquelle des fonctions est appropriée au problème posé.
- Enfin, certains expérimentateurs utilisent également des tests de rétention rendant compte de l'impact de l'apprentissage à long terme.
- L'importante variété des tâches et tests employés dans les diverses recherches rend difficiles, voir dangereuses les comparaisons entre langages d'une expérience à l'autre. Les buts, souvent explicites, sont pourtant bien d'estimer les possibilités d'un langage et par extension de les confronter à d'autres systèmes.

Reisner (1981) montre, dans une revue de question intéressante, que la diversité des paramètres expérimentaux n'autorise pas de conclusions comparatives ou seulement sur des résultats très ponctuels.

2- Les variables dépendantes.

Plusieurs types de variables dépendantes sont utilisées. Une distinction peut-être faite selon qu'il s'agit de variables permettant une quantification objective ou de variables rendant compte d'une estimation subjective de la part des utilisateurs.

a) Estimation subjective.

Dans ce cas, l'expérimentateur peut utiliser différents moyens, comme, par exemple :

- Un questionnaire informel dans lequel on demande au sujet son opinion sur le langage et l'apprentissage, de suggérer des améliorations de décrire leurs niveaux de motivation (Reisner, Boyce et Chamberlain 1974).
- Les sujets donnent leurs niveaux de confiance dans l'exactitude de chaque question qu'ils écrivent (Thomas et Gould 1975, Greenblatt et Waxman 1978).
- Les sujets doivent ranger les fonctions qu'il ont apprises et utilisées en ordre de difficulté.

b) Quantification objective

Plusieurs types de variables dépendantes quantifiables servent à définir les degrés de performance. :

- 1- L'aspect le plus fin et surtout le plus informatif concerne l'analyse des erreurs.

Cependant, il est rare qu'une réponse soit complètement correcte ou incorrecte ce qui a conduit certains auteurs à créer des degrés dans l'échelle des erreurs. Par exemple, oublier une parenthèse dans la syntaxe d'une question est moins un indice d'incompréhension du langage que de faire des erreurs systématiques sur certaines fonctions spécifiques. Reisner, Boyce et Chamberlain (1974) distinguent 3 types d'erreurs dont la gravité va croissant :

- Erreur minime de données ou de langage, par exemple, oubli de guillemets, erreurs d'orthographe utilisation d'une partie d'un nom (Jones) au lieu du nom complet (John Jones) ...etc.

Ce type d'erreurs n'est pas imputable au langage lui-même.

- Erreur de substance qui concerne les formes valides au niveau syntaxique mais produisant une réponse erronée. Dans ce cas, la formalisation de la requête est bonne, mais ne répond pas aux objectifs désirés.
- Erreur de forme qui concerne spécifiquement la syntaxe.

2- D'autres critères de performance sont utilisés.

De fait, la manière dont la performance est évaluée dépend de la nature et de la complexité du problème posé.

Ainsi, la recherche ponctuelle d'un élément dont on peut aisément découvrir ou estimer la localisation dans la base de données ne pose pas de difficultés importantes.

Dans ce cas, le temps nécessaire à l'établissement de la requête sera déterminant pour quantifier la performance. En ce qui concerne les recherches plus complexes, qui exigent une formalisation de la requête mettant en relation divers ensembles au moyen d'un certain nombre de connecteurs, plusieurs variables dépendantes peuvent être utilisés simultanément.

Le temps moyen nécessaire à l'apprentissage, à l'écriture des questions, le pourcentage des questions correctement écrites sont autant d'indicateurs permettant des comparaisons entre langages dans la mesure où les plans sont identiques. (Greenblatt et Waxman 1978)

IV- RESULTATS EVALUATIFS ET COMPARATIFS GLOBAUX

L'examen des résultats obtenus à partir des expérimentations doit être entrepris avec la prudence qu'imposent les impératifs méthodologiques déjà mentionnés.

1- Etudes évaluatives

Les études évaluatives du SQL montrent que tant en ce qui concerne l'apprentissage que l'utilisation effective du langage, les programmeurs (c'est-à-dire, dans les expérimentations réalisées, souvent des sujets ayant quelques notions de programmation !!), réalisent de meilleures performances que les non-programmeurs (Reisner 1976).

D'autre part, les caractéristiques du SQL permettent un haut niveau de rétention dans la mesure où les scores observés après apprentissage présentent peu de différences par rapport à ceux observés lors d'un test de mémorisation.

Un autre résultat intéressant indique que les sujets savent aussi bien utiliser les caractéristiques ou fonctions apprises que combiner celles-ci en de nouvelles formes. Reisner, Boyce et Chamberlain (1974) en concluent que, plutôt que de sélectionner certains patterns à partir d'un ensemble connu, les sujets expriment des idées dans un nouveau langage.

Enfin, certaines fonctions sont beaucoup plus difficiles que d'autres à apprendre et à utiliser. Reisner (1976) propose alors que le langage soit enseigné en 3 étapes, selon 3 degrés de complexité dont rendraient compte les difficultés des sujets à réaliser certaines opérations sur la base de données.

Tandis qu'avec le langage IQF (Gould et Ascher 1975), les sujets n'écrivent correctement que 35% des questions, le QBE (Thomas et Gould 1975), par contre, avec 67% de requêtes exactes, permet un résultat plus important. Par ailleurs, Thomas et Gould ont établi l'existence d'une corrélation entre d'une part la réussite, et d'autre part, le niveau de confiance ($r = 0.86$; $p < 0.001$).

Une donnée intéressante concerne également la possibilité de prédire l'exactitude d'une requête en fonction de 4 paramètres : le nombre d'opérateurs, le nombre de lignes, le nombre de colonnes et le nombre de relations inter-tables (Thomas et Gould 1975).

Comme pour l'étude menée sur l'IQF, il existe pour le QBE une augmentation de la difficulté avec l'accroissement des mesures objectives de complexité de la question.

2- Etudes comparatives

Les études comparatives semblent indiquer que les performances du SQL sont supérieures à celles du SQUARE pour les non-programmeurs (Reisner 1976).

L'expérience de Greenblatt et Waxman (1978) comparant le QBE et le SQL dans les mêmes conditions d'apprentissage, de procédure et de test montre que le QBE :

- a) requiert moins de temps d'apprentissage
- b) nécessite moins de temps pour les tests
- c) permet une plus grande exactitude dans l'établissement des requêtes
- d) et provoque un meilleur niveau de confiance.

Plusieurs raisons ont été invoquées, rendant compte de la supériorité apparente du QBE. Il convient cependant de garder à l'esprit qu'il n'est possible de statuer sur cette supériorité que de manière très globale et que les divers langages n'en sont encore qu'au stade de la conception.

En ce sens ces raisons ne sont en fait que des hypothèses qu'aucune expérimentation n'a confirmé ou infirmé.

Parmi celles-ci :

a) La représentation tabulaire ou relationnelle fournie par le système semble faciliter la tâche du sujet au moins pour 3 raisons :

- Celui-ci n'a pas à générer sa question indépendamment de toute structure. On sait, en effet, l'importance d'un schéma structuré dans la manipulation et l'organisation des informations.

- Cette représentation présente d'autre part une structure identique à la vision que le sujet aurait de la base de données.
 - Enfin, il semble que la possibilité d'entrer les éléments d'une question dans n'importe quel ordre est important.
- b) Le système n'utilise pratiquement pas de mots. Ceci élimine les possibilités de confusion avec le langage naturel (exemple du AND/OR dans le SQL).
- c) Enfin le système est assez facile à apprendre pour que la motivation reste importante.

Nous ne rapporterons pas ici les autres résultats comparatifs, qui, quoique présentant des données allant généralement dans le même sens, sont beaucoup plus sujets à caution du fait des variations des plans expérimentaux.

A défaut de pouvoir véritablement comparer ces différents langages, il est possible de constater l'existence de certaines similitudes au moins à 2 niveaux :

- 1- Le type de formulation utilisé en langage naturel peut provoquer des erreurs.

Thomas (1977) donne l'exemple d'une difficulté, que les sujets sont amenés à rencontrer quel que soit le langage et qui est dû à une possible discordance entre la formulation de la requête et la manière dont les données sont stockées.

Par exemple, un utilisateur peut demander la liste des "gens qui travaillent depuis plus de 5 ans".

Si les diverses tables de la base de données ont une colonne appelée "année d'engagement", la solution consiste, en traduisant, à remplacer la question "plus de 5 ans" par inférieur à 1976".

Les sujets capables de transcrire la formulation originale en "ceux qui ont été engagés avant 1976" s'exposent beaucoup moins à l'erreur.

- 2- Suivant les contraintes syntaxiques, il existe des niveaux de difficulté distincts dans le traitement et l'utilisation des fonctions.

En effet, selon les langages, les avantages ou inconvénients liés à la syntaxe ne sont pas les mêmes.

- Ainsi, le SQL provoque plus d'erreurs que le QBE dans le cas de l'utilisation des connecteurs logiques AND et OR.

Dans l'exemple suivant :

"list the names of the psysicians and children under 5 years old", le AND fait référence à une disjonction.

Or, le langage d'interrogation SQL utilise respectivement les opérateurs AND et OR pour la conjonction et la disjonction. De ce fait, un sujet utilisant une stratégie de traduction mot à mot, ce qui semble être fréquemment le cas, établira une requête qui tentera d'extraire la conjonction, l'intersection de l'ensemble des médecins et de l'ensemble des enfants de moins de 5 ans, qui bien sûr sera vide.

- La syntaxe du QBE, par contre, facilite la manipulation des connecteurs AND et OR en exprimant ces notions non pas à l'aide de mots-clés mais par des différences fondamentales de structure. (voir exemple page 22).

D'un autre côté, 25% des erreurs, en ce qui concerne le QBE, sont dues à l'utilisation non maîtrisée des quantificateurs. Ainsi, Thomas (1977) a-t-il remarqué de manière incidente que lorsque les sujets génèrent eux-mêmes leur requête en langage naturel, ils utilisent plusieurs stratégies qui leur permettent d'éviter une formulation nécessitant la manipulation effective de quantificateurs.

Ces dernières précisions sur les aspects évaluatifs et comparatifs des divers langages montrent clairement que l'amélioration et le développement d'un langage passe plutôt par l'étude précise de certaines fonctions spécifiques que par une tentative d'estimation globale de la facilité d'utilisation.

Très peu d'études ont été entreprises en ce sens.

Nous avons choisi de détailler, à titre d'exemple, 2 approches plus fines illustrant ce que pourrait être une recherche explicative et approfondie sur certaines fonctions parmi celles qui occasionnent les taux d'erreurs les plus élevés.

L'étude évaluative globale telle qu'elle est généralement entreprise ne permet, comme le remarque Reisner (1981), que de situer où se trouvent les difficultés, mais certainement pas de comprendre des raisons et la genèse de ces difficultés.

Reisner (1976) et Thomas (1976) ont tous deux tenté, de manière différente, d'apporter des réponses plus nuancées et plus opérationnelles.

3- Les travaux de Reisner

Reisner aborde cet aspect des langages d'interrogation d'une part, par le biais de l'étude des stratégies mises en œuvre par les sujets dans l'utilisation d'une fonction spécifique et d'autre part, au moyen d'une tentative de modélisation du comportement de traduction du langage naturel dans le langage artificiel.

a) Exemple d'une approche spécifique : la fonction GROUP BY

Pour une fonction bien précise, Reisner (1976) se pose les questions suivantes :

- Pourquoi est-elle difficile à apprendre ou à utiliser et que faire pour y remédier ?

- Où les changements doivent-ils être réalisés : au niveau de la syntaxe ou de l'apprentissage ?
- Quels doivent être ces changements ?

les résultats obtenus au niveau de la fonction GROUP BY montrent que les sujets savent comment l'utiliser (son rôle syntaxique dans le cadre d'une question bien définie est donc compris) mais ne savent pas quand l'utiliser. En effet, 89% des questions impliquant la fonction GROUP BY sont incorrectes du fait de la simple omission de cette fonction.

L'approche subjective (voir les variables dépendantes page 36) montre au moyen de questionnaires que les sujets ne reconnaissent pas l'utilité ni la pertinence de cette fonction. Cependant, les raisons de ces difficultés ne peuvent ressortir de ce simple constat.

C'est donc là le rôle de l'expérimentateur de suggérer une explication et si possible de la tester en tant qu'hypothèse.

Reisner suppose donc que les sujets développent différentes stratégies, dans l'écriture des requêtes, basées sur l'ordre dans lequel le langage SQL impose l'utilisation des clauses SELECT, FROM et WHERE.

L'auteur distingue 2 types de stratégies :

- Une stratégie de transfert, de translation dans laquelle les sujets recherchent des mots dans la formulation naturelle et tentent de les insérer dans la formulation en SQL.

D'après Reisner, ceci serait favorisé par le fait que la première clause est SELECT.

- Une stratégie "d'opérations sur tables".

En utilisant cette stratégie induite par le fait que la clause FROM intervient en premier dans certaines formulations de la requête, le sujet demanderait à l'ordinateur de se positionner sur la table appropriée et de réaliser les opérations nécessaires.

Reisner met donc en évidence un certain décalage entre le type de stratégie induit par le SQL (translation) et le fait que la fonction GROUP BY requiert que les sujets doivent réaliser des partitions en sous-groupes sur les tables (la stratégie d'opération sur tables serait alors plus appropriée).

Cette hypothèse n'a pas été testée, mais a au moins le mérite de mettre l'accent sur ces aspects importants que sont les stratégies utilisées par les sujets dans la mise au point des requêtes.

b) Tentative de modélisation du comportement.

Moran (1981) insiste sur l'importance de l'approche du comportement par une tentative de modélisation qui identifie, étape par étape, les opérations mentales et cognitives que le sujet doit mettre en œuvre pour effectuer les tâches demandées.

C'est ainsi qu'un modèle peut fournir un cadre permettant d'analyser la tâche par le biais des processus psychologiques mis en jeu.

Ce type de modélisation repose sur l'étude de la manière dont les sujets pensent que fonctionne le langage et peut permettre d'améliorer l'apprentissage ou d'aider les sujets à mieux comprendre les contraintes de la syntaxe formelle.

Reisner (1981) donne un exemple de ce que pourrait être un modèle de ce type pour le langage SQL en détaillant ce que suggère chez les sujets l'action de chacune des clauses SELECT, FROM et WHERE sur les étapes de recherche entreprises par l'ordinateur.

Un deuxième type de modèle s'intéresse plus spécifiquement aux stratégies utilisées dans la traduction des requêtes du langage naturel dans le langage artificiel.

Reisner fait alors l'hypothèse que le sujet crée un patron, une forme type, puis transforme les mots de la formulation naturelle en termes du langage d'interrogation pour finalement insérer ces termes dans le patron.

L'auteur a ainsi identifié plusieurs transformations possibles :

- Augmentation : cette transformation requiert l'insertion dans la question en SQL de mots absents de la formulation initiale.

Dans l'exemple suivant : "how many 727's can be dispatched from Laguardia", les mots "plane" et "airport" sont omis alors qu'ils doivent être intégrés dans la requête.

- Suppression : cette opération est inverse de la précédente. Un mot présent dans le langage naturel n'est pas nécessaire à la formulation en SQL.

Par exemple, en utilisant la fonction "projection" (voir page 14), pour trouver les noms de tous les employés, le mot "tous" doit être supprimé.

- Remplacement : un mot de la formulation naturelle doit être remplacé par un autre mot dans la formulation artificielle.

La difficulté d'utilisation des connecteurs AND/OR en est un bon exemple :

"donner la liste des étudiants de 1^{er} et de 2^{ème} année" doit être traduit de la façon suivante :

```
SELECT    NAME
FROM      STUDENTS
WHERE     YEAR = 1
OR      YEAR = 2
```

- L'auteur a pu mettre en évidence l'existence effective de telles transformations à partir de l'analyse des erreurs, et a tenté également de mettre au point un indicateur de complexité transformationnelle.

Ceci permet donc de faire l'hypothèse qu'une formulation en langage naturel qui requiert le plus de transformations serait la plus difficile à traiter comparée à une formulation différente de la même question.

Comme il a été signalé plus tôt, la difficulté de traduction sera de ce fait influencée par les formulations originales de la requête .

Ces tentatives n'en sont qu'à leur début mais ont le mérite d'exprimer le souci d'une compréhension plus profonde et plus opérationnelle que le simple constat d'une "facilité d'utilisation" qui prévalait jusqu'à ces travaux.

4- Les travaux de Thomas

De l'étude menée par Thomas et Gould (1975) sur le QBE, il ressort que l'ensemble des différentes erreurs se distribue de la manière suivante :

| type d'erreurs | probabilité d'erreurs |
|-------------------------|-----------------------|
| Quantification | 0.25 |
| Count/sum/compact-count | 0.25 |
| < > ≤ ≥ | 0.03 |
| Soulignage | 0.01 |
| Et/ou | 0.003 |
| Erreurs de valeur | 0.001 |

Nous n'avons pas rapporté d'exemples de la fonction quantificatrice dans la partie descriptive précédente. Cette fonction est en effet particulière sous cette forme au langage QBE.

Zloof (1975) en décrit les différents aspects au moyen des exemples suivants :

-n°1 "trouver les départements qui vendent tous les items fournis par le fournisseur Parker".

| | | |
|-------|---------------------|------------------------|
| sales | dept | item |
| | p. <u>household</u> | [all <u>pen</u>] • |

| | | |
|--------|----------------|----------|
| supply | item | supplier |
| | all <u>pen</u> | Parker |

all pen représente l'ensemble de tous les items fournis par Parker. Le point (.) sous all pen indique que le(s) département(s) en question peut(vent) vendre autre chose que tous les items fournis par Parker

- n°2 : "Trouver les départements dont tous les items sont fournis par Parker".

| sales | dept | item |
|-------|---------------------|----------------|
| | p. <u>household</u> | all <u>pen</u> |

| supply | item | supplier |
|--------|-------------------|----------|
| | {all <u>pen</u> } | Parker |

Cette fois-ci, le point est dans la colonne item de la table supply. Cela signifie que le fournisseur Parker peut fournir plus que tous les items vendus par le(s) département(s) en question.

- n°3 : "trouver les départements qui vendent seulement tous les items fournis par Parker".

| sales | dept | item |
|-------|---------------------|----------------|
| | p. <u>household</u> | all <u>pen</u> |

| supply | item | supplier |
|--------|----------------|----------|
| | all <u>pen</u> | Parker |

Il n'y a pas de point dans la mesure où les 2 ensembles doivent être égaux.

La fonction quantificatrice rend donc compte d'une part importante des erreurs observées lors de l'utilisation du langage QBE.

Thomas et Carroll (1981) remarquent que presque tous les systèmes d'interrogation de base de données utilisent les quantificateurs. De ce fait, l'étude de la compréhension, par les sujets, des relations entre ensembles et des concepts sous-jacents aux fonctions quantificatrices est de première importance.

Les travaux de ce type directement liés à l'interaction homme-machine sont inexistantes. Les études menées en ce sens par Thomas (1976) reposent sur des observations incidentes et ne permettent pas de faire des inférences précises dans la mesure où l'utilisation de quantificateurs n'était qu'un facteur secondaire des plans d'expérience.

- Sur cette base, Thomas pose la question de la manière dont les sujets utilisent spontanément les quantificateurs. Des sujets sans apprentissage spécifique en logique doivent décrire des relations représentées par des diagrammes de Venn. L'étude montre qu'une grande variété d'expression est utilisée pour décrire chaque relation tandis que certaines sont décrites plus précisément que d'autres. Ainsi, les ensembles disjoints sont toujours décrits sans ambiguïté ("no A are B and vice-versa", "A are not B and vice-versa") mais les relations entre ensembles se recouvrant partiellement posent le plus de problèmes. Toutefois, d'une manière générale, aucune description n'est totalement en contradiction avec les diagrammes de Venn. L'utilisation de ceux-ci procure une bonne représentation à des sujets non-programmeurs et les erreurs consistent surtout en des oublis quant à certaines caractéristiques qu'implique la relation plutôt qu'en une description totalement fausse.

- Une seconde approche consiste non plus à savoir comment des sujets expriment des relations en langage naturel, mais comment ils interprètent des formulations quantifiées. La tâche, dans ce cas, est donc inverse puisque les sujets doivent dessiner les diagrammes de Venn rendant compte des interprétations possibles de phrases, ou doivent juger de l'équivalence de plusieurs formulations. Les résultats confirment les données précédentes dans la mesure où les relations d'ensembles disjoints sont les mieux comprises, tandis que le recouvrement partiel ou les formulations admettant plusieurs représentations distinctes (quelques A sont des B) sont les plus difficiles à traiter.

En fait, ces observations incidentes semblent indiquer que la difficulté d'utilisation des quantificateurs n'est pas due à la syntaxe du langage d'interrogation mais à une incompréhension plus fondamentale.

Il apparaît donc qu'une direction de recherche importante soit d'approfondir les aspects psychologiques liés à l'utilisation des quantificateurs tant universels qu'existentiels

du fait que la syntaxe des divers langages n'est pas en cause, que le 1/4 des erreurs leur sont imputables et qu'enfin aucune étude spécifique dans le cadre de l'interrogation de bases de données n'a été entreprise.

V- LANGAGES ET MODELES DE STRUCTURE DE DONNEES

On ne peut approfondir l'étude des langages d'interrogation sans aborder le problème des modèles de structure de données qui leur sont sous-jacents.

Tester l'incidence de ces modèles sur les performances des langages suppose de s'intéresser à la vision conceptuelle qu'ont les sujets de la manière dont ces données sont stockées et non pas à la manière dont elles sont effectivement stockées.

Les 3 modèles de structuration des données décrits par Date (1975) sont les plus utilisés ; il s'agit

- du modèle relationnel où les données sont stockées sous forme de tables, de relations.
- du modèle hiérarchique où les données sont stockées sous forme de structure d'arbre.
- du modèle de réseau, en treillis (network) où les données sont stockées sous forme de structure de graphe.

De nombreuses recherches informatiques se sont penchées sur chacun de ces modèles. Parmi l'importante littérature sur ce thème, le lecteur intéressé pourra se reporter à Earnest (1974), Martin (1975), Borkin (1980), Sandberg (1981).

Les études plus psychologiques de l'effet des modèles de structure sur les performances ont été réalisées de 2 manières suivant que les comparaisons de ces modèles s'effectuaient sur la base ou non de l'utilisation d'un langage d'interrogation.

1- Comparaisons avec langages d'interrogation

Lochovski et Tschritzis (1977) et Lochovski (1978) ont réalisé 2 expériences permettant de comparer ces 3 modèles au moyen de 3 langages conçus pour refléter les approches relationnelles, hiérarchiques et en réseaux.

D'une manière générale, qu'il s'agisse du temps ou de la précision de codage, de la compréhension ou de la composition des requêtes, le système relationnel provoque les meilleures performances.

Toutefois, ces résultats sont hypothéqués par le fait que les effets observés ne peuvent être, de manière certaine, imputables exclusivement aux différences de modèles dans la mesure où le plan d'expérience utilisé ne permettait pas de contrôler l'incidence du langage.

Ceci a conduit certains auteurs à tester l'effet des modèles de structure de données indépendamment des langages d'interrogation.

2- Comparaisons sans langage d'interrogation

Brosey et Shneiderman (1978) ont mené des études comparatives, en proposant des diagrammes, reproduisant les modèles hiérarchique et relationnel, qui permettent d'éliminer la variable "langage d'interrogation".

La tâche de chacun des 38 sujets (étudiants en informatique) consistait à répondre à certaines questions au vu des diagrammes et des relations entre éléments de ces diagrammes.

L'expérience impliquait des tests de compréhension de plusieurs niveaux de difficulté, des situations de résolution de problèmes (quel ensemble de questions poseriez-vous qui permettraient de résoudre tel problème grâce à l'information contenue dans le diagramme ?) et des tests de mémorisation.

Les résultats obtenus montrent que le modèle hiérarchique est plus facile à utiliser, mais seulement pour les sujets sans expérience de programmation. Toutefois, les auteurs remarquent que la base de données utilisée présente une structure arborescente naturelle.

Enfin, certaines variables expérimentales limitent la portée de ces résultats. En effet, l'influence éventuelle non étudiée d'un apprentissage, le type particulier de base de données et de sujets, le fait qu'il s'agit d'une expérience papier/crayon nécessite d'approfondir ces études avant de statuer quant à l'incidence d'un modèle de structure.

3- Caractéristiques d'un modèle de structure des données

D'un point de vue psychologique, il existe encore beaucoup trop peu d'éléments pour se permettre quelque certitude que ce soit.

Si Mac Gee (1976) propose un ensemble de critères permettant d'évaluer ces modèles (simplicité, élégance, figurabilité ...), les interactions entre les performances de chacun d'eux et les tâches à effectuer ou les relations sématiques entre éléments ne sont pas encore bien établies.

D'une manière générale, un modèle en structure arborescente semble être assez lourd sauf quand les données sont perçues comme ayant une structure naturelle d'arbre.

Le modèle relationnel est élégant, mais est parfois considéré comme trop syntaxique et les modèles qui peuvent représenter de l'information plus sémantique sont parfois préférés.

Enfin, le modèle en réseau permet des structures sophistiquées, mais la complexité concomittante et l'utilisation de pointeurs mobiles augmentent la difficulté d'utilisation.

VI- CONCLUSION

Les résultats tels qu'ils apparaissent dans la littérature semble souvent ponctuels et se prêter difficilement à une vision globale.

Cependant, il est possible d'en extraire certaines constantes quel que soit le langage.

En tout premier lieu, on peut constater qu'il est possible pour les non-programmeurs d'apprendre un langage d'interrogation relativement puissant en un temps assez court.

D'autre part, la mise en évidence de points de difficulté et surtout les approches potentielles basées sur l'étude des processus cognitifs en jeu lors de l'utilisation de certaines fonctions syntaxiques permettent d'améliorer progressivement les performances d'un langage.

Ceci peut se réaliser :

- en modifiant les aspects syntaxiques difficiles sur la base d'une meilleure compréhension des processus de traitement qui sont sous-jacents.
- en développant des apprentissages spécifiquement tournés vers ces points de difficulté.
- en découpant l'apprentissage de la syntaxe sous forme de strates dont les difficultés vont croissant.
- en prévoyant que le système puisse anticiper les erreurs possibles à partir d'indicateurs de complexité d'une requête liés aux niveaux d'expérience ou de confiance de l'utilisateur.

Les diverses études réalisées sur les langages d'interrogation ne s'intéressent le plus souvent qu'aux modalités de traduction d'une requête du langage naturel dans le langage artificiel.

Cependant, l'interaction avec une base de données doit être envisagée dans un contexte plus large où la maîtrise d'un langage n'implique pas nécessairement qu'un sujet puisse utiliser une base de données facilement.

L'utilisation efficace d'une base de données ressort de la résolution de problèmes. En effet, le problème que se pose un utilisateur requiert souvent que soient posées plusieurs questions, et donc qu'une stratégie de recherche soit mise en œuvre.

De ce fait, le succès que rencontrera le sujet ne dépendra plus seulement de son habileté à transcrire une requête, mais également des stratégies utilisées, des manipulations qu'il aura à effectuer sur les données et donc des représentations qu'il aura de la manière dont sont structurées ces données.

Nous nous intéresserons donc, dans les pages qui suivent, aux aspects psychologiques liés à la représentation et à la structuration des informations. La littérature psychologique sur ce thème est importante mais nous nous limiterons aux recherches permettant d'élargir la discussion précédente sur les modèles de structure de données.

REPRESENTATION ET STRUCTURATION DES INFORMATIONS

I- INTRODUCTION

Lorsqu'un utilisateur veut récupérer des éléments d'information en questionnant une base de données informatisée, ses stratégies de recherche reposent sur des hypothèses concernant l'organisation de cette base.

Les connaissances actuelles en ce domaine laissent supposer que les difficultés éventuelles d'utilisation du système seraient en grande partie dues à une discordance trop importante entre les représentations des informations proposées et l'organisation de ces informations telle que la conçoit l'opérateur (voir, entre autre, l'exemple de Thomas page 40).

Si l'expérience montre qu'une représentation unique de la structure de la base de données -au moins au niveau du software- ne peut être envisageable indépendamment du type d'information, des tâches à réaliser, ou des opérateurs, la question se pose des principes devant guider les composantes de la représentation.

L'incidence des modalités de présentation des données a généralement été évaluée dans le cadre de l'étude des situations de résolution de problèmes ou des processus de récupération et de rappel de l'information.

En effet, toute résolution de problèmes suppose l'utilisation d'informations.

Celles-ci peuvent être directement disponibles, mais le plus souvent, elles doivent être récupérées, retrouvées en mémoire, quelle que soit cette mémoire.

Les performances réalisées dans cette récupération dépendent de la manière dont l'information est structurée et des stratégies utilisées dans la recherche et le traitement des données.

Il est communément admis que l'utilisation par le sujet d'un moyen graphique de présentation accroît la performance réalisée dans des tâches déductives.

Cependant, cette assertion de bon sens se doit d'être nuancée ou du moins précisée.

En effet, une étude de l'incidence directe de la présentation des informations sur les performances ne peut conduire qu'à une évaluation descriptive de son action.

Par contre, une compréhension réelle de son importance ne peut être réalisée que par l'analyse de la représentation de la structure du problème qu'elle provoque.

Or, cette représentation qu'aura le sujet dépend de plusieurs facteurs tels l'aspect sémantique des données, leur support contextuel et leur organisation implicite ainsi que l'existence éventuelle de structures représentatives a priori.

II- ASPECTS SEMANTIQUES

1- Incidence de la signification des informations

Pour étudier l'effet de la signification du matériel sur la recherche des données et la résolution de problèmes, Kieras et Greeno (1975) ont réalisé une expérience où l'information consistait en un échantillon fini de concepts. A l'inverse des expériences classiques où l'ensemble des connaissances du sujet sert de base de données, les auteurs pouvaient donc espérer que, la structure de l'information étant bien définie, les inférences sur la récupération seraient moins ambiguës.

Les résultats obtenus, confirmant les études précédentes, montrent qu'il est plus facile de récupérer de l'information chargée sémantiquement que de l'information non significative.

Cependant, dans la mesure où le choix d'un échantillon fini et spécifique comme base de données permettait de présenter l'information significative et non significative d'une manière identique, des différences observées ne peuvent être attribuées aux différences de structure formelle du matériel.

Les auteurs font donc l'hypothèse que l'action de la signification repose sur des facteurs organisationnels qui ne sont pas contenus dans le matériel spécifique mais qui dépendent des connaissances générales du sujet.

De ce fait, les stratégies de recherche seraient différentes pour les 2 types d'information, mais Kieras et Greeno échouent à déterminer la nature exacte des différences de processus dans la mesure où le plan expérimental utilisé ne permet pas de les inférer à partir des comportements des sujets.

Par la suite, Mayer et Greeno (1975) confirment que des structurations qualitativement différentes sont élaborées et utilisées pour les 2 types de matériel.

En effet, malgré la supériorité de l'information significative, l'organisation et la présentation en elles-mêmes des données sont mieux retenues pour l'information non significative.

Corrélativement avec l'hypothèse de Kieras et Greeno selon laquelle le sujet relierait le matériel significatif à ce qu'il connaît déjà, les auteurs pensent qu'il se produirait une restructuration et une intégration de la nouvelle information en fonction de l'arrière-plan conceptuel du sujet.

2- Incidence de l'arrière-plan conceptuel du sujet

Le résultat précédent est d'autre part corroboré par l'étude de Malin (1973) qui trouve des différences dans le choix des stratégies pour plusieurs matériels, ces différences étant principalement reliées à des caractéristiques physiques des stimuli qui facilitent la constitution de chunks plutôt que directement à la signification des stimuli.

De ces travaux, il ressort que l'aspect significatif du matériel n'a pas d'influence directe en soi mais augmente les performances par le biais de l'organisation et de la structuration qu'il provoque et des stratégies de recherche qu'il favorise.

L'importance de cette connexion entre l'information manipulée et les connaissances du sujet se retrouve dans l'étude de Broadbent (1978) sur l'appellation des catégories rendant compte des informations stockées dans une base de données.

En effet, la recherche est plus efficace lorsque les sujets utilisent leur propres descripteurs plutôt qu'une classification réalisée par un tiers.

Il apparaît donc qu'une direction de recherche intéressante ne serait pas tant d'examiner l'incidence de l'aspect sémantique de l'information que d'essayer de définir les déterminants de la restructuration réalisée à partir de la représentation qu'a le sujet des données.

Un dispositif informatique qui serait capable d'envisager l'ensemble possible des transformations structurelles réalisables par l'opérateur sur le matériel pourrait de ce fait apporter une aide efficace à l'utilisation des stratégies de récupération de l'information et de résolution.

III- COMPREHENSION ET STRUCTURATION DES INFORMATIONS BRUTES

La manière dont le sujet va restructurer l'information fournie sera influencée par et influencera la compréhension qu'il aura du problème. On peut en effet constater qu'il existe une sorte de feedback entre les modalités de présentation des données et la compréhension des relations entre éléments pertinents du problème.

D'autre part, des erreurs de raisonnement peuvent naître d'inférences correctes basées sur des représentations internes que des organisations inadéquates ou des restructurations erronées, ont faussées. Ces observations attestent que résoudre un problème suppose, avant que tout processus inférentiel ou déductif soit mis en jeu, que soient bien comprises et intégrées les informations brutes.

Mayer (1976) utilise différentes manières (sous formes verbale, d'organigramme ou d'exemple) de présenter la même information dans une tâche consistant simplement à intégrer cette information et ne nécessitant pas la mise en œuvre de processus de transformation des données pour découvrir la solution. L'auteur teste donc l'incidence de la présentation des informations et donc de la structuration qu'elle détermine avant que soient entreprises les diverses opérations généralement réalisées pour résoudre un problème. Il fait l'hypothèse que si une présentation mieux structurée aide à la compréhension en intégrant d'avantage d'information dans des chunks plus larges alors devrait être observée une performance supérieure seulement pour des problèmes plus longs et plus complexes.

Parallèlement à cette action supposée d'une représentation structurée, Moreau (1973) remarque que les processus d'organisation transforment les éléments d'information en unités plus larges et donc moins nombreuses.

Les résultats montrent que des diagrammes provoquent une meilleure performance s'ils remplacent des présentations verbales particulièrement complexes. Mayer attribue cet effet à une prise en compte plus rapide et plus précise des données. Par contre les diagrammes semblent moins performants si les représentations verbales sont simples et structurées. Enfin, aucune des données obtenues ne laisse supposer l'existence de différences qualitatives dans le processus de traitement entre les 2 types de modalité de présentation.

De l'hypothèse de l'auteur, il apparaît donc que les diverses manières de structurer la représentation d'un problème influencent, entre autre,

le taux d'information qu'un sujet est capable de traiter pendant la résolution.

IV- ORGANISATION MATRICIELLE ...

Schwartz (1971) et Polish et Schwartz (1974) ont également étudié l'incidence de la quantité d'information sur la représentation et la performance. Cependant, ils abordent leur étude de manière différente en utilisant par ailleurs des variations connues pour être pertinentes dans des tâches de formation de concepts ; l'information peut-être conjonctive, disjonctive ou conditionnelle suivant le connecteur logique utilisé, elle peut-être positive ou négative ou varier suivant le nombre de dimensions définissant un objet...

D'autre part, les auteurs envisagent leur recherche à un stade de résolution plus avancé puisqu'ils utilisent comme variable dépendante les structurations effectuées par le sujet. L'utilisation de problèmes déductifs de type "who-done-it" permet d'obtenir des protocoles rendant observables sinon évidentes les organisations réalisées dans la mesure où les manipulations et transformations nécessaires sur les données sont trop importantes pour être exécutées sans notes.

Schwartz classe les protocoles obtenus en 5 modes de représentation :

- 1- représentation matricielle,
- 2- groupement informel,
- 3- représentation graphique,
- 4- réécriture des données,
- 5- une dernière catégorie regroupant le reste.

Les résultats montrent une supériorité de la matrice (74%) sur les autres types de présentation (25% à 50%) en terme de réussite à la résolution du problème bien que son utilisation augmente à peine avec la complexité de la tâche. Par ailleurs, à l'inverse des résultats classiques sur la formation de concepts, le nombre de dimensions ainsi que le connecteur logique utilisé ont peu d'influence sur la performance.

Enfin, le succès au problème de raisonnement déductif semble être dû, pour une grande part, à la capacité de représenter précisément les données sous une forme favorisant les manipulations ultérieures.

De ce fait, Schwartz suggère que la représentation matricielle, plutôt que de favoriser le codage, aide à générer un espace-solution, où les transformations déductives, les synthèses et le stockage des résultats peuvent être plus facilement réalisés.

En effet, elle définit clairement les informations nécessaires (cases, vides), induit des ordres d'opérations fructueux et surtout favorise des tests faciles pour les solutions intermédiaires obtenues.

Sauf si les informations sont présentées sous forme négative, ces résultats montrent effectivement la supériorité de la représentation matricielle. Cependant, il est à noter que dans ces expériences, les problèmes testés ainsi que les divers types de relations entre informations ne peuvent efficacement et complètement se décrire que par une telle représentation.

Schwartz et Fattelah (1972), conscients de ces restrictions, proposent pour l'avenir de rechercher pourquoi une organisation matricielle est si efficace mais également de découvrir les caractéristiques rendant d'éventuels types de représentation plus adaptés à d'autres tâches.

V- ... OU ORGANISATION HIERARCHIQUE ?

Plusieurs travaux ont montré qu'une organisation hiérarchique permet un meilleur rappel des informations. Les nombreux avantages de cette structuration ont même conduit certains auteurs (Collins et Quillian 1969) à suggérer qu'elle rendait compte préférentiellement de la mémoire sémantique humaine.

Toutefois Broadbent (1966) fait remarquer les limitations inhérentes à un système hiérarchique en ce qui concerne la récupération d'information : en effet un item ne peut être accessible que par une branche de la structure arborescente.

Par contre, une organisation matricielle permet d'assigner à chaque information un certain nombre de descripteurs non mutuellement exclusifs rendant possibles la recherche d'un item par plusieurs entrées.

De ce fait, plutôt que de rechercher les caractéristiques d'une seule structuration optimale pouvant rendre compte de l'activité mnémotique dans son ensemble, Broadbent (1978) préfère une autre direction en étudiant les avantages et désavantages relatifs à chaque structure. Pour ce faire, l'auteur utilise des listes de 16 mots organisés par groupes de 4 et présentés sous forme hiérarchique ou matricielle. La structure hiérarchique suppose l'existence de 7 mots-titres représentant les points de bifurcation de l'arborescence. L'organisation matricielle nécessite 4 mots-titres figurant les 4 entrées obtenues à partir des 2 subdivisions orthogonales réalisées sur la liste des 16 mots.

La tâche dévolue au sujet consiste à rappeler les mots des 2 listes sans qu'un ordre quelconque soit imposé.

Broadbent réalise 4 expériences à partir de ce paradigme de base et obtient plusieurs résultats intéressants :

Tout d'abord, le bénéfice principal d'une organisation quelle qu'elle soit repose sur le regroupement de mots apparentés plutôt que sur la présence visible des mots-titres. Ceux-ci sont toutefois les points de départ des processus de rappel dans la mesure où leur récupération favorise celle des items de la catégorie correspondante.

D'autre part, il ne semble pas qu'une des 2 structures présente quelque supériorité sur l'autre. Cependant, les types de performance sont différents : l'organisation hiérarchique permet une récupération inégale du matériel, certains groupes d'items montrant des différences de rappel importantes tandis que l'organisation matricielle provoque un niveau égal moyen pour tous les mots.

Ceci confirme la remarque de Broadbent concernant la possibilité en cas d'utilisation d'une hiérarchie de perdre totalement un groupe d'items dû à l'oubli d'un mot-titre alors que la structure matricielle évite cette occultation complète d'une partie des informations puisqu'elles peuvent être retrouvées par plusieurs entrées.

A l'inverse, la structure hiérarchique favorise une récupération par contagion : chaque item dépend du supérieur et le rappel d'un petit ensemble initial rend aisé le rappel d'un autre situé sur l'embranchement. Une telle stratégie est évidemment impossible à utiliser avec une structure matricielle puisqu'elle subdivise les informations en éléments orthogonaux et indépendants.

Ces données tempèrent l'ambition sous-jacente à nombre de travaux de trouver un principe structurel unique rendant compte de l'ensemble des processus de rappel ou de récupération de l'information. Par ailleurs, il apparaît que les performances dépendent moins des connotations sémantiques de l'item recherché que de son ajustement à la structure utilisée.

Un système informatique d'interrogation de base de données se devra donc de ne pas imposer une présentation structurelle unique des données. Son efficacité sera au contraire accrue s'il favorise une certaine souplesse quant aux modalités de présentation. Celle-ci devra varier en fonction des informations, de leurs relations, des liens entre matériels que l'opérateur voudra voir privilégiés ainsi que des manipulations ou transformations que celui-ci voudra effectuer sur les informations stockées.

Cette nécessité d'une adaptation du système à des impératifs ponctuels pose le problème de l'existence éventuelle de structures représentatives a priori chez l'opérateur. Les divers travaux réalisés (Collins et Quillian sur la structure hiérarchique, Anderson sur les réseaux, Schwartz sur les matrices) montrent que les sujets peuvent utiliser plusieurs modèles d'organisation suivant les conditions de la tâche ou les buts à atteindre.

VI- STRUCTURES REPRESENTATIVES A PRIORI

Durding, Becker et Gould (1977) ont réalisé 3 expériences tendant à montrer la disponibilité et les conditions d'utilisation des diverses formes de structure. La première enjoignait aux sujets d'organiser les données, en l'occurrence des mots, afin de savoir si l'organisation choisie reposerait sur des relations sémantiques prédéfinies des items ou sur toute autre structure indépendante de ces relations.

Les résultats indiquent que les sujets sont capables de découvrir les relations sémantiques entre les mots et ont, de ce fait, toutes les structures organisationnelles testées disponibles (hiérarchie, matrice réseau, liste).

D'autre part, l'absence de tout effet d'apprentissage conduit à postuler l'existence avant-expérience de ces structures représentatives.

Dans une deuxième expérience, l'utilisation d'ossatures schématiques, rendant compte des différentes structures et consistant en espaces blancs reliés par des lignes augmentait de manière significative les performances des sujets.

La troisième montrait la difficulté qu'ont les sujets à insérer des items dans des structures inappropriées.

Les auteurs tirent de cette étude 3 implications principales quant à un système de récupération de l'information.

- Tout d'abord, la structure de la base de données, ou tout au moins son expression, doit se conformer aux relations sémantiques entre les éléments.
- Ensuite, doit être disponible pour l'opérateur un ensemble important de relations structurelles dont les diverses caractéristiques puissent être exprimées par le langage utilisé.
- Enfin, le fait que l'opérateur ait à sa disposition des ossatures schématiques reproduisant les types de relations existant entre données favorisera certainement la performance dans la recherche de l'information.

VII- LE SUPPORT CONTEXTUEL

Structurer la présentation des informations en fonction de leurs connotations sémantiques ne suffit pas à contrôler la représentation qu'auront les sujets du problème.

Les performances réalisées sont également affectées par les formulations différentes qui sous-tendent un même problème. Thomas (1978) remarque que, bien que l'on sache que les processus impliqués dans la formulation d'un problème sont au moins aussi importants que ceux impliqués dans sa résolution une fois formulé, il y a très peu d'études expérimentales sur ce point.

A la suite de Reed et al. (1974) et de Simon et Hayes (1976), Carroll, Thomas et Malhotra (1980) ont étudié l'incidence de l'utilisation de supports contextuels différents dans 2 tâches de résolution de problèmes de conception dont les logiques sont isomorphes.

Le premier facteur de l'expérience oppose pour un même problème une présentation temporelle à une présentation spatiale des conditions de résolution.

Le deuxième prévoit 3 modalités de présentation des informations dont les divers degrés d'organisation explicitent de manière croissante les caractéristiques du but à atteindre.

Les résultats établissent que les sujets structurent plus aisément les informations et ont une représentation plus adéquate des données sous la condition spatiale. Par ailleurs, le fait de fournir aux sujets une représentation graphique aussi performante dans les deux conditions diminue sensiblement les différences observées.

Les auteurs en concluent donc que les différences de performance sont dues à l'utilisation d'une représentation adéquate des données plus disponible dans la condition spatiale et non pas à des divergences conceptuelles plus fondamentales.

Le deuxième facteur du plan expérimental était supposé manipuler le niveau auquel était rendue apparente la structure implicite du but à atteindre.

Les résultats ne montrent aucune différence entre les modalités de ce facteur. Les auteurs expliquent cet échec par le fait que la procédure utilisée oblige le sujet à analyser le problème entièrement à chaque fois qu'il utilise une nouvelle information. Ils font donc l'hypothèse que la structuration de l'information permettrait d'organiser l'ensemble du problème à résoudre en sous-problèmes, en buts intermédiaires.

De la même manière Schwartz avait déjà fait une hypothèse analogue selon laquelle l'avantage de la représentation graphique reposerait sur le fait qu'elle procure un moyen de mémorisation aidant à maintenir et à intégrer les solutions intermédiaires précédentes.

CONCLUSION

Malgré certains résultats ponctuels qui ressortent des études évaluatives ou comparatives, les recherches actuelles n'ont permis que la constitution d'un corpus de connaissances restreint sur le plan des applications.

Cependant, les directions prises ainsi que les méthodes utilisées ne peuvent que contribuer à une meilleure connaissance des activités psychologiques sous-jacentes à l'interaction avec une base de données.

Nous avons développé deux aspects parmi ceux ayant suscité le plus de travaux : les langages artificiels d'interrogation et la structuration et la représentation des données.

1- Les langages artificiels

1.1 Les recherches actuelles

Des études spécifiques ont permis des évaluations de syntaxe quant à leur facilité d'utilisation.

les résultats obtenus sont le plus souvent normatifs, et de ce fait, ne permettent pas la compréhension plus fondamentale des processus cognitifs en jeu nécessaire à une conception rationnelle de ces langages.

Thomas (1976) et Reisner (1976), conscients de ces limitations, ont montré des voies de recherche beaucoup plus prometteuses en tentant une analyse des syntaxes au niveau plus fin des fonctions qui les constituent et une modélisation des comportements de traduction du langage naturel au langage artificiel.

Nous avons de ce point de vue suggéré quelques expérimentations possibles concernant entre autres certaines fonctions du SQL ou la fonction quantificatrice dans le QBE.

Un certain nombre de recherches ont été menées concernant les aspects psychologiques sous-jacents à l'utilisation des quantificateurs "tout", "quelque" ... ect.

Des expérimentations basées sur ces résultats et liées directement à l'interrogation de base de données seraient certainement profitables du point de vue des applications (rappelons que 25% des erreurs dans le QBE sont dues à la fonction quantificatrice).

Toutefois, beaucoup de questions restent sans réponse et notamment, à d'autres niveaux, peu abordés, que celui de la simple traduction d'une requête :

1.2 Les recherches possibles

On peut en effet se demander si les différences observées entre les langages sont dues aux diverses conceptions qui sous-tendent les syntaxes : l'incidence spécifique de la dichotomie mot-clé/position (SQL et QBE) n'est pas étudiée. Les processus psychologiques en jeu peuvent être, en effet, fort différents.

L'aspect procédural ou non procédural des langages, bien que peu étudié, semble de première importance. La question se pose des types de tâche ou de stratégies que favoriseraient des syntaxes procédurales.

Une direction importante paraît également se situer dans le développement de la compréhension des processus d'écriture des requêtes indépendamment d'un langage spécifique.

En effet, l'ensemble des études réalisées jusqu'à présent est essentiellement centré sur la traduction ou la compréhension ponctuelle d'une question précise.

Or, l'interaction efficace avec une base de données suppose que l'opérateur génère et organise une série de questions. Se profile donc tout un ensemble de recherches possibles ayant pour but de savoir comment l'utilisateur organise sa tâche, quels sont les processus de génération des requêtes en langage naturel et de quel ordre sont les dépendances entre requêtes tendant à résoudre un problème.

2- L'organisation des données.

2.1 Organisation spatiale et sémantique

Les nombreuses études comparatives des types de présentation de l'information ont le plus souvent consisté à établir une correspondance directe et descriptive entre la disposition des données sur l'écran et les performances.

Les résultats les plus importants insistent sur l'intérêt de privilégier une représentation spatiale des données ou de structurer les informations en fonction de leurs relations sémantiques.

D'autre part, la recherche des données, reposant en partie sur les aspects sémantiques reconnus par les sujets en fonction de leur arrière-plan conceptuel, il semble préférable de laisser les opérateurs organiser eux-mêmes la base au moyen de leurs propres descripteurs. (Voir Broadbent 1978).

2.2 Organisation, représentation et stratégie

Directement liées à ces aspects, des recherches concernant les modèles d'organisation des données (tout n'est pas forcément hiérarchique ou matriciel) et les visions conceptuelles et structurelles des bases de données qu'ont les utilisateurs, apporteraient des informations sur les méthodes optimales de stockage et d'extraction des données.

Une direction intéressante paraît se situer dans l'existence des structures représentatives propres aux utilisateurs. Certaines de ces structures sont manifestement plus efficaces (matrices, hiérarchies, organigrammes ...) mais leurs conditions d'utilisation en fonction de critères spécifiques restent à définir. D'autre part, s'il est reconnu que la nature de la tâche proposée ou sa complexité ou le but visé favorisent telle ou telle représentation, les recherches entreprises ne font souvent état que de résultats ponctuels basés sur une hypothèse a priori quant à la structuration de la mémoire.

En effet, Durdin, Becker et Gould (1977) remarquent que les modèles de structure conçoivent l'organisation de la mémoire en termes de variables structurelles.

Les nombreux travaux réalisés en ce sens montrent que semble vouée à l'échec la recherche d'une structure représentative absolue témoignant de la manière dont serait effectivement organisée la mémoire et ceci indépendamment des caractéristiques impliquées dans la tâche.

Le but visé, lorsqu'un opérateur questionne une base de données, est la recherche d'informations, mais les modalités d'obtention de ces informations suppose la mise en oeuvre de stratégies propres à la résolution de problèmes.

Une nouvelle approche consisterait donc à penser que la structuration de l'information dépendrait plutôt des stratégies utilisées que d'une organisation hypothétique de la mémoire. De ce point de vue, Durdin note que "les exigences de résolution de problèmes et de recherche de l'information seraient les premiers déterminants des structures organisationnelles observées".

Par ailleurs, Schwartz et Fattelah (1972) en remarquant que l'organisation matricielle favorisait les transformations déductives nécessaires à la résolution, considéraient déjà la structuration des informations comme une étape opératoire directement incluse dans les processus responsables de l'émergence des stratégies de recherche.

Plusieurs questions, chacune indiquant des directions de recherche complémentaires, se posent alors sur :

- Les critères de choix de telle ou telle représentation.
- La manière dont les sujets intègrent de nouvelles données dans des structures déjà existantes.
- La manière dont les sujets modifient des structures et donc leurs stratégies de recherche en fonction de nouvelles données.
- La possibilité du choix d'une représentation en fonction d'une stratégie de découpage par sous-buts intermédiaires, par étapes dans la résolution.
- La relation entre la restructuration des informations réalisées par les sujets et les stratégies mises en œuvre.

Ces aspects sont prometteurs mais supposent que la première étape réside dans la recherche des caractéristiques pertinentes rendant compte des stratégies utilisées.

Ceci pose donc le problème de déterminer tout d'abord quels sont les types de tâches existants dans la récupération de l'information.

Par la suite, mieux comprendre le lien existant entre structuration et stratégie sera déterminant.

Il est en effet certain que pour beaucoup de gens, les techniques optimales de résolution ne seront pas spontanément utilisées et que, de ce fait, des suggestions informatisées de stratégies déduites des structurations effectuées par le sujet augmenteraient grandement l'efficacité générale du système.

BIBLIOGRAPHIE

- ANDERSON (J.R.) - "A simulation of free recall" - In G.H. BOWER (ED)
"The Psychology of learning and motivation : Advance in
research and theory" - (vol 5)
New-York. Acedemic Press, 1972,
- BARRET (G.V.), THORNTON (C.L.), CABE (P.A.) - "Human factors evaluation of
a computer based information storage and retrieval system" -
Human factors, 1968, 10(4), 431-436.
- BORKIN (S.A.) - "Data models : a semantic approach for data base system" -
1980, M.I.T. Press, Cambridge, Mass.
- BROADBENT (D.E.) - " The well-ordered mind" -
American educational research journal, 1966, 3, 281-295
- BROADBENT (D.E.), COOPER (P.J.), BROADBENT (M.H.P.) - " A comparison of
hierarchical and matrix retrieval schemes in recall" -
Journal of experimental psychology ; Human learning and
behavior 1978, 4, 5, 486-497.
- BROADBENT (D.E.), BROADBENT (M.H.P.) - "The allocation of descriptor terms
by individuals in simulated retrieval system" -
Ergonomics, 1978, 21, 5, 343-354.
- BROSEY (M.), SHNEIDERMAN (B.) - "2 experimental comparisons of relational
and hierarchical database models" -
International journal of man-machine studies, 1978, 10,
625-637.
- CARROLL (J.M.), THOMAS (J.C.), MALHOTRA (A.) - "Presentation and repre-
sentation in design problem solving" -
British journal of psychology, 1980, 71, 143-153.

- COLLINS (A.M.), QUILLIAN (M.R.) - "Retrieval time from semantic memory" -
Journal of verbal learning and verbal behavior, 1969, 8,
240-247
- CROFT (W.B.), VAN RIJSBERGEN (C.J.) - "An evaluation of Goffman's indi-
rect retrieval method" -
Inform. proc. Management, 1976, 12, 327-331.
- DATE (C.J.) - "An introduction to database systems" -
Addison-Wesley, 1975.
- DOYLE (L.P.) - "Semantic road maps for literature searches" -
J.ACM, 1961, 8, 574-578.
- DURDING (B.M.), BECKER (C.A.), GOULD (J.D.) - "Data organisation" -
Human factors, 1977, 19(1), 1-14.
- EARNEST (C.) - "A comparison of the network and relational data structure
models" -
Computer sciences organisation, El Segundo California, 1974.
- ENGEL (S.), GRANDA (R.) - Guidelines for man/display interfaces " -
IBM technical report TR 002720, 1975.
- GOFFMAN (W.) - "An indirect method of information retrieval" -
Inform. Stor. Retr. 1968, 4, 4.
- GOULD (J.), ASCHER (R.) - "Use of a IQF-like query language by non-pro-
grammers" -
IBM research report, RC 5279, 1975.
- GREENBLATT (D.), WAXMAN (J.) - "A study of 3 database query languages" -
In Databases : Improving usability and responsiveness,
B. Shneiderman
Ed. Academic Press, New-York, 1978.

- Grev-Colloque international "vision travail". Groupe de recherche sur les écrans de visualisation.
Rodez-Toulouse, 23-25 nov. 1978.
- HEROT (C.F.) - "Spatial management of data" -
Computer corporation of America
Technical report CCA-79-25, June 30, 1979
- KIERAS (D.E.), GREENO (J.G.) - "Effects of meaningfulness on judgments of computability" -
Memory and cognition, 1975, 3, 4, 349-355.
- LANCASTER (D.W.), FAYEN (E.G.), - "Information retrieval on-line" -
Melville Publishing co., Los Angeles, Calif. 1973, 597 p.
- LOCHOVSKI (F.H.) - "Data base management system user performance" -
Ph.D dissertation Univer. Toronto, Canada, 1978.
- LOCHOVSKI (F.H.), TSICHRITZIS (D.C.) - "User performance considerations in DBMS selection" -
In Proc. ACM SIGMOD, 1977, 128-134.
- MAC GEE (W.X.) - "On user criteria for data model evaluation" -
ACM Trans. Database Syst. 1976, 1, 4, 370-387.
- MALIN (J.) - "Information processing load in problem solving by network search" -
Journal of experimental psychology ; Human perception and performance 1979, 5, 2, 379-390.
- MANSUR (O.) - "An associative search strategy for information retrieval" -
Infor. proces. and management. 1980, 16, 129-137
- MARON (M.E.), KUHN (J.L.) - "On relevance, probabilistic indexing and information retrieval" -
J.ACM 1960, 7, 3.

- MARTIN (J.) - "Design of man-computer dialog"-
Englewood cliffs, New Jersey : Prentice Hall 1973.
- MARTIN (J.) - "Computer database organisation" -
Prentice-hall, Englewood cliffs, New Jersey, 1975.
- MARTIN (T.) - "In Smith et Green "Human interaction with computers" -
Academic Press in 1980, London.
- MAYER (R.) - "Comprehension as affected by structure of problem representation" -
Memory and cognition 1976, 4, 3, 249-255.
- MAYER (R.), GREENO (J.G.) - "Effects of meaningfulness and organisation on problem solving and computability judgments" -
Memory and cognition, 1975, 3, 4, 356-362.
- MILLER (L.A.), THOMAS (J.C.) - "Behavioral issues in the use of interactive system" -
Intern. journal of man-machine studies, 1977, 9, 5, 509-536.
- MONTGOMERY (C.A.) - "Is natural language an unnatural query language?" -
Proc. ACM nat. conf. N.Y. 1972, p. 1975-1078.
- MORAN (T.P.) - "An applied psychology of the user" -
ACM 1981, 13, 1.
- MOREAU (A.) - "Le rôle du schéma dans l'apprentissage et l'évocation d'une tâche verbale" -
Année psychologique, 1973, 73, 521-533.
- NICKERSON (R.S.), ELKIND (J.I.), CARBONNEL (J.R.) - "Human factors and the design of time sharing computer systems" -
Human factors, 1968, 10(2), 127-134.
- PARSONS (H.M.) - "The scope of human factors in computer-based data processing systems" -
Human factors, 1970, 12(2), 165-175.

- POLISH (K.M.), SCHWARTZ (S.H.) - "The effect of problem size on representation in deductive problem solving" -
Memory and cognition, 1974, 2, 4, 683-686.
- RAMSEY (H.R.), ATWOOD (M.E.) - "Human factors in computer systems : a review of the literature" -
Science applications Inc. Englewood CO 80111, 1979.
- RAMSEY (H.R.), ATWOOD (M.R.) - "Man-computer interface design guidance : state of the art" -
25th annual meeting 1980, 85-89.
- REED (S.K.), ERNEST (G.W.), BANERJI (R.) - The role of analogy in transfer between similar problem states" -
cognitive psychology, 1974, 6, 436-450.
- REISNER (P.) - "Use of psychological experimentation as an aid to development of a query language" -
IBM technical report, RJ 1707, 1976.
- REISNER (P.) - "Human factors studies of database query language : a survey and assesment" -
ACM, 1981, 13, 10, 13-31.
- REISNER (P.), BOYCE (R.F), CHAMBERLAIN (D.D.) - "Human factors evaluation of 2 database query langages- SQUARE and SEQUEL" -
IBM technical report, RJ 1478, 1974.
- ROUSE (W.B.) - "Design of man-computer interfaces for on-line interactive systems" -
Proceedings of the IEEE, 62, 6, 1975, 847-856.
- SALTON (G.) - "Dynamic information and library processing" -
Prentice-hall, Englewood cliffs, New Jersey, 1976

- SANDBERG (G.) - "A primer on relational data base concepts" -
IBM system journal, 1981, 20, 1, 23-40.
- SCHWARTZ (S.H.) - "Modes of representation and problem solving : well
evolved is half solved" -
Journal of experimental psychology, 1971, 91, 2, 347-350.
- SCHWARTZ (S.H.), FATTELAH (D.L.) - "Representation in deductive problem
solving : the matrix" -
Journal of experimental psychology, 1972, 95, 2, 343-348.
- SHNEIDERMAN (B.) - "Improving the human factors aspects of data base
interactions" -
ACM trans. database system, 1978, 3, 4, 417-439.
- SIMON (H.A.), HAYES (J.R.) - "The understanding process : problem iso-
morph" -
Cognitive psychology, 1976, 8, 165-190.
- STEWART (T.F.M.) - "Ergonomics aspects of man-computer problem solving" -
Applied ergonomics, 1974, 5, 4, 209-212.
- THOMAS (J.C.) - "Quantifiers and questions-asking".
IBM research report, RC 6468, 1976.
- THOMAS (J.C.) - "Psychological issues in database management" -
Proc. 3th conf. on very large data bases, Tokyo, 1977,
1977, 169-185.
- THOMAS (J.C.), CARROLL (J.M.) - "Human factors in communication" -
IBM system journal, 1981, 20, 2, 237-263.
- THOMAS (J.C.), GOULD (J.D.) - "A psychological study of query by example" -
Nat. computer conf. Proc. afips press, 1975, 44, 439-445

WELTY (C.) - "A comparison of a procedural and a non-procedural query language : syntactic metrics and human factors" -
PH.D. dissertation Univ. Massachussets Amherst, May 1979.

WELTY (C.), STEMPLE (D.N.) - "Human factors comparison af a procedural and a non procedural query langage" -
ACM trans. Database Syt. 1981, 6, 4, 626-649.

ZLOOF (M.M.) - "Query by example" -
Proc. nat. computer Conf. AFIPS press, 1975, 44, 431-437.

32

33

34

35

36

37