

Quelques éléments sur la dynamique de la prise d'information dans une tâche d'évaluation de connaissances par expert

FRANÇOIS GUERCIN
Université de Poitiers

JEAN-PAUL CAVERNI
Université de Provence

RESUMEN.—Las experiencias presentadas en este artículo han sido concebidas con vistas a precisar una problemática concerniente a las características de la elaboración de una representación cognitiva del universo de la evaluación entre expertos, en una tarea concreta. Los expertos son aquí docentes, y su tarea consiste en evaluar el nivel de dominio de la lengua materna de alumnos de último curso del Ciclo Medio.

Los soportes de evaluación son redacciones de francés, de las cuales se presentan varias versiones, haciendo variar la localización de la información concerniente al uso correcto o no de ciertas reglas de escritura. Del mismo modo, se hace variar la información dada *a priori* al experto en lo que se refiere al nivel del alumno. El seguimiento, en tiempo real, del proceso evaluativo durante la lectura se ve asegurado gracias a un paradigma experimental de ventana móvil. Los resultados permiten, en primer lugar, evidenciar los aspectos secuenciales de la presa de información, forzosamente limitada por el aspecto lineal del objeto, y validar parcialmente un modelo de organización de las representaciones entre expertos. En segundo lugar, encontramos que las heurísticas de juicios (y los sesgos asociados a ellas), puestas de manifiesto en situaciones de laboratorio, están igualmente presentes en una situación más próxima de la realidad cotidiana de una tarea de evaluación.

PROBLÉMATIQUE¹

On peut spécifier la psychologie contemporaine comme une perspective dans laquelle on tente de rendre compte de l'élaboration de comportements dans une situation

1. Une partie de ce travail a été effectué à Genève, et nous remercions la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Genève, qui a accueilli François Guercin dans le cadre d'une bourse de la Communauté de Travail des Alpes Occidentales. Nous remercions également vivement les instituteurs et institutrices du Canton de Genève et de la Ville de Poitiers ayant bien voulu participer à l'expérience.

ayant ses caractéristiques propres, si possible proche d'une tâche du monde courant (voir e.g. Logie & Bruce, 1990; Davies & Logie, 1991). Si certains champs de la psychologie du jugement comme les études sur la prise de décision ont plutôt été avant-coureurs d'une approche cognitive, cette démarche s'étend également maintenant à une psychologie du jugement évaluatif (voir Berkeley & Humphreys, 1982; Kleinmuntz, 1985; Pitz & Sachs, 1984), et on s'intéressera plus particulièrement dans cet article au domaine de l'évaluation de connaissances par expert (voir Caverni, 1987a).

Deux questions seront abordées. Premier point, on étudiera la validité de l'heuristique d'ancrage-ajustement dans une situation courante (proche de l'activité quotidienne des sujets experts). Deuxième point, on étudiera l'élaboration d'une représentation de l'objet évalué, à travers la dynamique de la saisie d'indices d'évaluation. Enfin on caractérisera l'influence de l'état initial de la représentation de l'évaluation (en fonction de l'ancrage) sur l'aspect séquentiel de son élaboration.

Originellement mis en évidence par Tversky & Kahneman (1974), l'heuristique d'ancrage-ajustement a été souvent critiquée pour son aspect artificiel, ou trop lié à des situations had-hoc ou de laboratoire (Berkeley & Humphrey, 1982; Evans, 1983; Polar & Evans, 1983).

Cependant, Northcraft & Neale (1987) ont pu révéler son utilisation dans le cas de situations plus proches du monde courant. Dans une situation de jugement par experts, Caverni & Peris (1990) précisent que si l'effet d'assimilation à une ancre est en conflit avec un effet de contraste dû au contexte, seul l'effet d'assimilation joue.

S'il semble bien établi que des critères fonctionnent comme opérateurs cognitifs lors de la saisie d'indices d'évaluation (Duter, 1986), il reste à comprendre comment s'élabore un jugement évaluatif à partir de ces indices.

En effet, dans le cadre conceptuel défini par Caverni & Gonzalez (1986), l'évaluation a pour fonction d'inférer les savoirs qui sont maîtrisés par l'auteur de la production jugée, et de les situer par rapport à des normes de savoirs. C'est à dire que se pose le problème de la constitution d'une représentation de l'objet jugé et de ses liens avec le savoir de l'évaluateur.

Les experts évaluateurs considérés, sont ici des enseignants. Pour évaluer une production l'enseignant fait intervenir d'une part des savoirs spécifiques à l'enseignement en question (règles du corps de savoirs concerné), d'autre part des savoirs spécifiques à la pratique évaluative, liés à l'exercice de son activité professionnelle. Ainsi, puisqu'il dispose de ces deux types de savoir, on peut considérer l'enseignant comme un expert.

Premier point, l'expert se réfère à un corps de savoirs constitués et organisés (espace des savoirs évalués). Deuxième point, au cours du processus d'évaluation un aspect particulier de l'épreuve jugée se trouve qualifié. Cette qualification est une unité de représentation, et une suite de qualifications constitue une représentation globale de l'épreuve.

Ainsi, le modèle de Caverni & Gonzalez pose que l'évaluateur, en tant qu'expert, dispose d'une *représentation mentale du corps de connaissances* à évaluer. Ce corps de connaissances est modélisé en termes d'*éléments de connaissances*, structurés par une *relation* dite de *recouvrement*². Si la maîtrise d'un élément de connaissance α implique

2. Considérons une épreuve de mathématiques visant notamment à faire calculer la surface d'un rectangle. Considérons une production dans laquelle ce calcul a été opéré par une addition de la longueur et de la largeur. Dans ce cas l'élément de connaissance «surface du rectangle» fera l'objet d'un marquage négatif. Ce marquage diffusera à tous les éléments de connaissance pour la

la maîtrise d'un élément de connaissance β , on dit que α recouvre β , ou que β est recouvert par α .

Lorsque l'évaluateur lit une production, il y recherche la mise en jeu des éléments de connaissance sur l'acquisition desquels il doit faire porter son évaluation. Deux états sont possibles lorsqu'un élément de connaissance est mis en jeu dans une production: ou bien la mise en jeu est conforme à (*confirme*) une norme, ou bien la mise en jeu n'est pas conforme à (*infirme*) une norme. Toute mise en jeu *conforme* dans la production suscite, dans la représentation du corps de connaissances qu'a l'évaluateur, un *marquage positif* de l'élément correspondant. Du fait de la relation de recouvrement, ce marquage positif *diffuse* à tous les éléments *recouverts* par l'élément concerné, lesquels sont conséquemment *marqués positivement* aussi. Toute mise en jeu *non conforme* dans la production suscite, dans la représentation qu'a l'évaluateur, un *marquage négatif* de l'élément correspondant. Du fait de la relation de recouvrement, ce marquage *diffuse* à tous les éléments *qui recouvrent* l'élément concerné, lesquels *sont donc marqués négativement* aussi. L'élaboration de l'évaluation s'opère donc par marquages successifs du corps des connaissances à évaluer.

Deux aspects de ce modèle nous intéressent directement ici.

D'une part, on peut supposer que les marquages initiaux sont sensibles à l'influence de renseignements donnés avant la tâche d'évaluation elle-même. De plus cet état initial influe la recherche active d'information dans l'objet à évaluer. En effet on peut s'attendre à ce que l'évaluation soit sensible à un biais bien connu dans l'étude du diagnostic médical (Elstein, 1990), qui consiste à attribuer «sans raison» un signe nouveau à une hypothèse antérieure, plutôt que de poser une nouvelle hypothèse, ou de garder l'attribution en suspens.

D'autre part on peut supposer que l'organisation spatiale et temporelle des informations dans les productions va modifier les marquages. En effet on sait que les individus ont des stratégies dans les prises de décision (Payne, Bettman & Johnson, 1988), et que ces dernières sont en interaction avec les caractéristiques de la tâche, ce qui se manifeste par une certaine sensibilité à des effets d'ordre de présentation (Hogarth & Einhorn, 1992).

On peut également noter que l'un des objets de l'expérience est d'insister sur le caractère dynamique du comportement de jugement, dans le sens où ce qui nous intéresse ici n'est pas seulement les conditions dans lesquelles une réponse est donnée, mais plutôt comment cette réponse est élaborée (voir Vleck & Keren, 1993). Le pistage de cette élaboration est lié au développement de nouveaux modes d'investigation, d'analyse et de validation (Tiberghien, 1985; Gonzalez, 1988; Estes, 1990; Guercin, Roussey & Piolat, 1990).

confirmation desquels la connaissance du calcul de la surface du rectangle est requise (par exemple calcul de la surface du parallélogramme). En elle-même, l'infirmité de l'élément de connaissance «surface du rectangle» ne permet ici aucun marquage négatif par diffusion de l'élément de connaissance «multiplication». Par contre, si nous considérons une production qui explicite un calcul conforme et exact de la surface, l'élément de connaissance «surface du rectangle» fera l'objet d'un marquage positif. Ce marquage diffusera à tous les éléments de connaissance requis pour le calcul de la surface du rectangle, notamment la multiplication. Par contre, la confirmation de l'élément de connaissance «surface du rectangle» ne permet ici aucun marquage positif par diffusion de l'élément de connaissance «surface du parallélogramme».

En effet, lorsque les seuls observables de la situation pris en compte sont les données du jugement et la réponse finale, le processus étudié recouvre une période d'activité du sujet qui n'est marquée d'aucun observable. Aussi se développe un intérêt pour le pistage d'indices relatifs à l'activité de jugement pendant le déroulement même de l'élaboration de celui-ci (voir Payne, 1982; Russo & Doshier, 1983; Caverni, 1987). Cet intérêt conduit d'une part à développer un modèle intégrant la temporalité du processus de jugement, et d'autre part à développer de nouveaux outils d'investigation.

Ainsi, bien que la première variable recueillie soit classiquement l'expression du jugement à la fin du processus d'évaluation, les deux expériences présentées utilisent un paradigme expérimental permettant le recueil des données en temps réel, ainsi que l'examen de la succession des actes finalisés des sujets (Guercin, 1990³). La mise en perspective chronologique est effectuée en ne présentant au sujet que des informations partielles dont l'accès est limité à une seule à la fois, l'obligeant ainsi à mettre à plat sa stratégie de recherche d'information. Dans le cas d'informations visuelles en deux dimensions (lecture), on peut repérer la chronologie dans la succession de déplacements spatiaux d'une fenêtre ne laissant filtrer qu'une information partielle (partie de texte), et pouvant être déplacée au gré du sujet.

EXPÉRIENCE 1

Cette première expérience a été construite en vue d'un double but. Premièrement étudier l'influence de renseignements préliminaires ainsi que celle de l'ordre de la prise d'information sur le jugement final. Deuxièmement préciser les aspects les plus dynamiques du processus de prise de l'information, i.e. en essayant de clarifier l'interaction entre les caractéristiques des textes évalués et la stratégie de lecture employée par un évaluateur. A cette fin l'expérience est assistée par ordinateur et une méthode de recueil des données en temps réel est utilisée.

MÉTHODE

Sujets

Vingt-quatre enseignants de sept écoles du Canton de Genève (Suisse) ont accepté de collaborer gracieusement à l'expérience. Ces enseignants sont des instituteurs ou institutrices de dernière année du cycle primaire. Il leur était demandé d'évaluer des rédactions en langue française, présentées sur un écran d'ordinateur. L'évaluation est exprimée par une note sur 6, selon le système de notation en usage dans l'enseignement Helvétique.

3. Guercin, F. (1990, mai), *Le paradigme de la fenêtre baladeuse et l'étude des processus d'évaluation*, Congrès annuel de la S.F.P.: «L'évaluation: Processus, Méthodes et Pratiques». Bordeaux, 10-12 mai.

Matériel

Seize textes expérimentaux et un texte d'entraînement ont été construits pour les motifs de l'expérience. Il s'agit à l'origine de productions effectivement rédigées par des élèves du niveau de fin de cycle primaire. Ces productions ont été recomposées en fonction du montage expérimental.

Dans chaque texte, quatre éléments de trois classes de connaissances (correspondant à des critères d'évaluation usuellement déclarés) ont été manipulés: «correction de l'orthographe», «correction de la syntaxe», «correction du lexique»⁴.

Les seize textes expérimentaux sont composés de quatre productions modifiées selon quatre versions. Une version comprend toujours le même nombre d'éléments de connaissance (12). Mais la localisation dans les douze lieux du texte de leur mise en jeu correcte (+) ou incorrecte (-) varie (tableau 1). Pour le texte d'entraînement la localisation est répartie afin d'avoir des éléments présentant un aspect à la fois équilibré et de caractère aléatoire (Durup, 1967).

Les textes sont segmentés pour être présentés sur un écran selon une technique d'Auto. Présentation Segmentée (Pynte, 1974; Aaronson & Scarborough, 1976). La segmentation est choisie de sorte à optimiser les opérations de lecture (Pynte & Noizet, 1980). Ainsi l'enseignant peut se focaliser sur l'activité d'évaluation.

Tableau 1. *Les quatre séquences expérimentales définissant les versions.*

ei: ième élément de connaissance

Vi: version i

	V1	V2	V3	V4
e1	-	+	-	+
e2	-	+	+	-
e3	-	+	-	+
e4	-	+	+	-
e5	-	+	-	+
e6	-	+	+	-
e7	+	-	-	+
e8	+	-	+	-
e9	+	-	-	+
e10	+	-	+	-
e11	+	-	-	+
e12	+	-	+	-

4. Voici quelques exemples de chacun des types d'incorrections (on a imprimé en caractère gras souligné l'incorrection, et porté entre parenthèses et en italique la forme correcte): j'ai (*je suis*) sorti me promener / il faut que des mesures *sont* (*soient*) prises / nuire _ (*à*) la santé de quelqu'un / lorsque mon père acheta une voiture notre vie _ (*en*) *était* (*fut*) facilitée certains *estimerons* (*estimeront*) / s'est (c'est) pourquoi / *merveilleux* (*merveilleux*) / il est *crevant* (*fatiguant*) de courir.

Procédure

La tâche demandée aux enseignants est d'évaluer cinq rédactions en langue française portant sur le même thème, présentées selon la technique de la fenêtre baladeuse: le sujet est assis en face d'un écran vidéo; sur cet écran un texte est affiché de façon partielle par l'intermédiaire d'une fenêtre; le sujet contrôle le déplacement de la sur le texte par les ordres qu'il donne au clavier; ainsi il fait apparaître successivement les segments de texte dans lesquels il cherche une information. A la fin de la lecture d'un texte, l'enseignant donne sa note puis passe au texte suivant. Il peut réajuster les notes à la fin de la lecture des cinq rédactions.

Organisation des facteurs

Pour les mêmes éléments de connaissance, quatre versions se distinguent par les différentes séquences de «confirmation-infirmité» de leur mise en oeuvre (cfr. tableau 1). Les versions 1 & 2 s'opposent à 3 & 4 selon une dimension de regroupement ou dispersion des incorrections. La version 1 s'oppose à la version 2 selon la localisation des incorrections (première ou deuxième moitié du texte). Par conséquent, les versions 1 & 2 sont à la fois homogènes et contrastées, alors que les versions 3 & 4 sont hétérogènes et uniformes.

Notons que d'autres aspects que ceux manipulés peuvent être envisagés: l'organisation des idées, leur pertinence à l'égard du thème, leur originalité, etc. C'est pourquoi un plan de mélange répartit les versions sur le croisement des quatre productions et des sujets: un sujet évalue quatre textes expérimentaux, chacun avec une version différente (donc l'enseignant évalue les quatre versions). Le facteur «version» est croisé avec le facteur sujet⁵.

En revanche le facteur «ancrage» est emboîté avec le facteur sujet: trois groupes peuvent être distingués selon les instructions qui sont données avant la lecture des textes au cours de la présentation de la consigne.

Pour C+, une ancre positive est donnée sous cette forme: «Ces élèves font partie d'un groupe de bons élèves dont la moyenne en Français est habituellement de 5/6» (N.B.: le niveau moyen pour un élève correspond habituellement à la note 4).

Pour C-, une ancre négative est donnée sous cette forme: «Ces élèves font partie d'un groupe d'élèves en difficulté dont la moyenne en Français est habituellement de 3/6».

Pour Co aucune indication n'est donnée sur le niveau supposé de l'élève.

La première variable dépendante retenue est l'expression du jugement à la fin de la lecture, i.e. la note sur 6 délivrée à chaque version.

La seconde variables est l'ordre de lecture des différents segments. A partir de cette variable brute, on construira une variable concernant le nombre de fois que les segments expérimentaux sont relus.

5. Le plan de mélange définit quatre associations entre un texte et une version, qui sont distribuées aléatoirement dans l'ensemble des sujets.

Hypothèses

Notes

- Les notes différeront selon l'ancre délivrée: elles devraient être plus élevés pour l'ancre positive que pour l'ancre neutre, et on devrait trouver les notes les plus basses pour l'ancre négative.
- Les notes seront d'autant plus faibles que les incorrections auront été rencontrées plus tôt dans la lecture de la version. On attend en particulier que la moyenne la plus faible soit observée pour V1 et que la moyenne la plus élevée soit observée pour V2.

Retours sur le texte

- Si la stratégie de prise d'information est adaptée à la dispersion des incorrections, les versions homogènes (V1 & V2) permettant une confirmation du marquage des éléments donneront lieu à moins de retours, et les versions hétérogènes (V3 & V4) le plus.
- De même, si l'espace évaluatif se construit à partir des marquages initiaux, les versions avec un début contraire au marquage initial donneront plus de retours.
- La consigne avec ancre neutre (Co) devrait donner lieu à la recherche d'information la plus active donc à plus de retours.

RÉSULTATS

Les tests statistiques sont effectués selon une procédure classique d'analyse de variance⁶ correspondant à la planification décrite ci-avant.

En ce qui concerne le résultat de l'évaluation (figure 1) on retrouve l'effet classique (Noizet & Caverni, 1978) de l'information a priori sur l'élève ($F_{(2,21)} = 3,98$; $p = .034$): puisque la consigne à ancre positive donne la meilleure note (4,66), la consigne à ancre neutre une note moyenne (4,43) et enfin la consigne à ancre négative la note la plus basse (4,09). Toutefois il faut noter qu'en opposant deux à deux les consignes, seules les deux ancres extrêmes s'opposent significativement ($F_{(1,21)} = 7,86$; $p = .010$).

Cette opposition entre l'ancre positive et négative, n'est confirmée que pour les versions homogènes (dans V1, $F_{(1,14)} = 4,24$; $-p = .058$; dans V2, $F_{(1,14)} = 4,95$; $p = .043$). Pour leur part, les versions hétérogènes ne permettent pas une stratégie bien définie de marquage de l'espace évaluatif: pas de test significatif pour l'effet des consignes.

(dans V3, $F_{(1,14)} = 1,94$; $p = .18$;

dans V4, $F_{(1,14)} = 1,87$; $p = .19$).

Ainsi que nous nous y attendions, nous ne pouvons pas conclure à l'effet global des différentes versions ($F_{(3,63)} = 1,82$; $p = .15$). Par contre les versions les plus contrastées s'opposent ($F_{(1,21)} = 5,034$; $p = 0.35$): la version avec les incorrections en début de texte (V1) donne la note la plus mauvaise (4,125); la version avec les incorrections à la fin (V2) donne la meilleure note (4,57).

6. Programme d'analyse «VAR3». Rouanet & Lépine (1977).

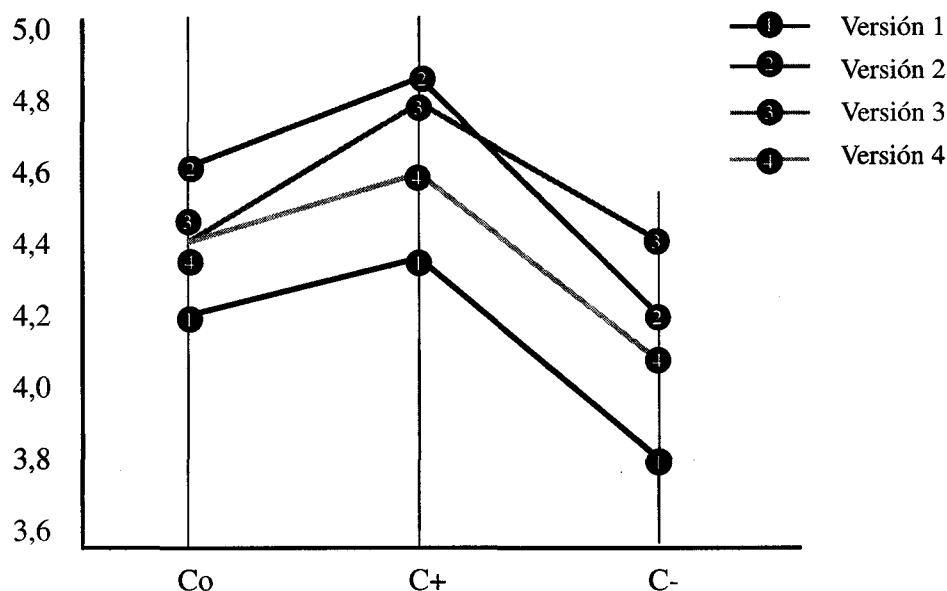


Figure 1. *Moyenne des notes attribuées aux copies, en fonction des différentes versions et des différentes consignes (Co: ancre neutre; C+: ancre positive; C-: ancre négative).*

En ce qui concerne les retours en arrière sur le texte (figure 2), en premier lieu, il faut noter l'importance de la contrainte linéaire de lecture du texte. Beaucoup de sujets effectuent une lecture dans l'ordre d'apparition des segments, sans relecture. Ainsi nous avons globalement un nombre assez faible de retours en arrière, et une variance importante liée à cette variable.

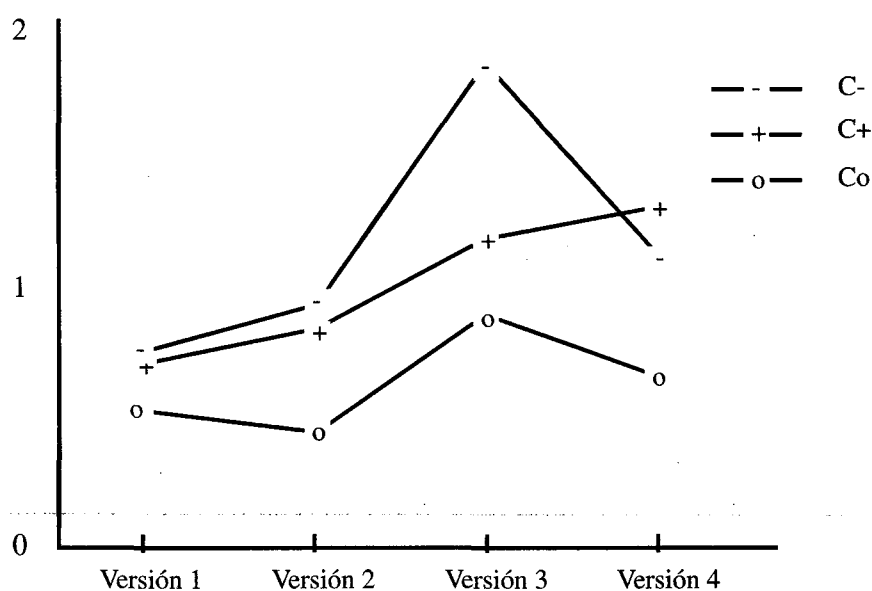


Figure 2. *Moyenne du nombre de retours en arrière sur le texte, en fonction des différentes versions et des différentes consignes (Co: ancre neutre; C+: ancre positive; C-: ancre négative).*

Contrairement à ce que nous attendions, la consigne «neutre» donne lieu au nombre de retours le plus faible ($F_{(1,21)} = 5,88$; $p = .024$). Ainsi la prise d'information sans marquage initial donne lieu à une lecture plus linéaire du texte.

On trouve moins de retours dans les versions homogènes (V1 & V2) que dans les autres ($F_{(1,63)} = 30,73$; $p < .001$). Dans la version V1 les deux consignes opposées C+ et C- donnent lieu au même nombre de retours, et dans la version 2, on n'obtient aucun test significatif concernant l'effet des consignes ($F_{(2,21)} < 1$). Par contre, dans les versions 3 et 4 le nombre de retours est plus important pour la consigne donnant un effet d'ancrage opposé à la mise en jeu du premier élément de connaissance rencontré (dans V3, $F_{(2,21)} = 3$; $p = 0.71$; dans V4, $F_{(2,21)} = 3.34$; $p = .055$).

COMMENTAIRES

Il est manifeste que l'évaluateur ne s'en tient pas à la simple prise en compte *hic & nunc* des éléments de connaissance à évaluer.

Tout d'abord, ils doivent être présentés de façon consistante pour obtenir un effet: les mêmes éléments n'ont pas d'influence notable lorsqu'ils sont dispersés, alors que la notation y est sensible lorsqu'ils sont regroupés. De même, les retours en arrière sont plus nombreux dans le cas de versions hétérogènes, comme si l'évaluateur avait du mal à se faire une opinion précise.

Ensuite, alors que les textes comportent le même nombre et le même type d'éléments à évaluer, la notation est sensible à l'effet des versions. Ainsi le seul effet de l'ordre dans lequel sont présentés les éléments de connaissances pris en compte suffit à influencer le jugement final: la mise en jeu incorrecte au début donne une plus mauvaise note, alors que la mise en jeu incorrecte à la fin donne une meilleure note.

Ce qui signifie également que le début de la production a un rôle déterminant puisque les deux versions les plus contrastées s'opposent. En ce qui concerne les retours en arrière sur le texte, les versions 3 & 4 qui ont un nombre de retours plus important, ont d'autant plus de retours que le début du texte est contraire aux renseignements donnés dans la consigne. Ainsi l'effet d'ancrage crée une attente, une structuration des marquages initiaux qui induit une recherche d'information adaptée: si la mise en jeu des éléments de connaissance du début est conforme au marquage créé par l'ancre, la prise d'information est linéaire, dans le cas contraire la prise d'information nécessite des retours en arrière sur le texte.

En dernier lieu, on peut noter l'influence de la consigne sur l'évaluation, car un même élément de connaissance n'est pas traité de la même manière selon l'ancre qui est délivrée avant la lecture de la copie. En particulier le nombre de retours en arrière est moins important dans le cas de la consigne neutre. Comme si la construction sans *a priori* permettait une construction du jugement évaluatif plus harmonieuse.

EXPÉRIENCE 2

Cette deuxième expérience est construite afin d'examiner le processus de prise de l'information au cours même de la construction de l'espace évaluatif i.e. au cours de la

lecture de la copie d'un élève, et non plus seulement inférer cette construction à partir des seuls indices recueillis en fin d'évaluation. Elle permettra également d'étudier l'influence de renseignements susceptibles de modifier les marquages initiaux, sur la prise d'information.

A cette fin la méthode employée est très proche de celle de l'expérience précédente. En particulier le même paradigme de fenêtre baladeuse est employé, mais dans cette deuxième expérience, notre intérêt se focalisera sur les temps de lecture. Caverni (1987b) a déjà validé cette méthode d'étude pour le pistage de processus d'évaluation. En particulier cet auteur observe que les segments de phrase porteurs d'incorrection font l'objet d'un temps de lecture plus long, seulement si la tâche demandée est une tâche d'évaluation versus une tâche de simple lecture, et seulement pour des experts.

Le montage de l'expérience étant très proche de celui de l'expérience précédente, nous ne détaillerons que les quelques points de différence.

MÉTHODE

Sujets

Vingt-quatre enseignants de l'Académie de Poitiers ont accepté de collaborer gracieusement à l'expérience. Ces enseignants sont des instituteurs ou institutrices de dernière année du cycle primaire. Il leur était demandé d'évaluer des rédactions en langue française, présentées sur un écran d'ordinateur.

Matériel

Le matériel utilisé est composé de seize textes expérimentaux et d'un texte d'entraînement, construits selon les mêmes principes que l'expérience 1: douze éléments de connaissance (quatre éléments de trois classes) sont manipulés.

Procédure

Identique à l'expérience précédente.

Organisation des facteurs

Le premier facteur concerne la localisation spatiale des éléments de connaissance dans le texte. Afin de caractériser le début et la fin du texte, les quatre premiers et les quatre derniers segments expérimentaux (sur les douze) seront pris en compte dans l'analyse. Ce facteur à huit modalités est croisé avec les sujets.

Le deuxième facteur est également croisé avec les sujets, il concerne les différentes séquences de confirmation-infirmité des éléments de connaissance, et qui déterminent quatre versions, identiques à l'expérience précédente (tableau 1).

Le troisième facteur, emboîté avec les sujets, concerne les instructions qui sont données avant la lecture des textes au cours de la présentation de la consigne.

Pour C+, une ancre positive est donnée sous cette forme: «Ces élèves font partie d'un groupe de bons élèves dont la moyenne en Français est habituellement de 13/20».

Pour C-, une ancre négative est donnée sous cette forme: «Ces élèves font partie d'un groupe d'élèves en difficulté dont la moyenne en Français est habituellement de 07/20».

Pour Co aucune indication n'est donnée sur le niveau supposé de l'élève.

Deux variables dépendantes sont retenues:

- L'expression de l'évaluation, i.e. la note sur 20 délivrée à chaque version⁷.
- Le temps de lecture des segments de texte manipulés expérimentalement. Ce temps est rapporté au nombre de caractères du segment; ce qui constitue la variable «temps/caractère» (TC).

Hypothèses

Conformément aux résultats et aux interprétations antérieures:

- Les TC d'affichage seront différents selon les consignes si l'effet de l'ancre correspond à des traitements différenciés de l'information contenue par la production évaluée.
- Le premier élément de connaissance rencontré dans le texte fera l'objet d'un traitement plus long (TC plus important), surtout pour la condition avec ancre neutre.
- Pour le début du texte les TC les plus longs seront observés pour la version 1, et les plus courts pour la version 2.
- Si la consigne contribue à modifier les marquages initiaux, c'est à dire l'attente a priori des sujets, il y aura une interaction entre les consignes, et les versions 1 et 2, au début du texte.
- Seules les versions contrastées auront un effet sur la localisation spatiale: pour les versions 1 et 2 on attend une modification de l'effet entre début et fin du texte, mais pas pour les versions 3 et 4.

RÉSULTATS

(précisons qu'en ce qui concerne la localisation spatiale, les quatre premiers segments constituent le «début» du texte et les quatre derniers la «fin»).

Globalement (cfr. tableau 2), on n'obtient pas de test significatif en ce qui concerne les différentes consignes, ainsi que les différentes versions. En effet, le nombre de mise en jeu correcte ou incorrecte des éléments de connaissance étant le même, leur distribution se répartit en moyenne, puisque les deux facteurs interagissent avec la localisation des éléments ($F_{(3,63)} = 4,57$; $p = .005$; $F_{(2,21)} = 4,98$; $p = .017$). Nous repérerons donc les effets des deux facteurs en effectuant des analyses partielles.

7. N.B.: La variable dépendante «note» étant aussi une variable de l'expérience 1, et les résultats étant tout à fait similaires, ils ne seront pas détaillés.

Tableau 2. *Moyenne des temps/caractère en fonction des différentes versions et des différentes consignes (Co: ancre neutre; C+: ancre positive; C-: ancre négative).*

	Co	C+	C-	
Version 1	79,7	79,6	75,2	78,2
Version 2	79,4	67,8	75,2	74,1
Version 3	80,7	73,8	79,4	77,9
Version 4	76,4	72,3	72,7	73,8
	79	73,4	75,6	

Le premier segment porteur d'un élément de connaissance rencontré est lu plus lentement que tous les autres (82,31/75,11; $F_{(1,21)} = 5,92$; $p = .024$). Toutefois si l'on décompose cet effet selon les consignes (cfr. tableau 3), le test n'est significatif que pour l'ancre positive ($F_{(1,7)} = 5,47$; $p = .051$).

Tableau 3. *Moyenne des temps/caractère en fonction des différentes localisations et des différentes consignes (Co: ancre neutre; C+: ancre positive; C-: ancre négative).*

Co	79,99	71,18	80,08	74,81	78,80	76,99	87,40	83,19
C+	86,01	73,05	69,69	75,23	70,99	67,74	71,11	73,03
C-	80,92	70,47	68,20	75,42	80,14	79,83	76,21	73,81
	Segment	Segment	Segment	Segment	Segment	Segment	Segment	Segment
	1	2	3	4	5	6	7	8
	82,31	71,57	72,66	75,15	76,64	74,85	78,24	76,68

En début de texte (figure 3) on trouve les TC les plus longs pour la version 1 (inco-rrections au début; $F_{(1,21)} = 5,55$; $p = .028$), et les plus courts pour la version 2 (inco-rrections à la fin; $F_{(1,21)} = 7,09$; $p = .014$). Mais l'effet de ces deux versions n'est pas le même selon les consignes:

- Pour la consigne à ancre neutre, les versions 1 et 2 s'opposent avec un écart de 16,8 ($F_{(1,7)} = 10,54$; $p = .014$).
- La consigne à ancre positive accentue cette opposition, puisque l'écart est de 26,4 ($F_{(1,7)} = 9,27$; $p = .018$).
- La consigne à ancre négative atténue cette opposition, l'écart étant de 6,4, et le test non-significatif ($F_{(1,7)} < 1$).

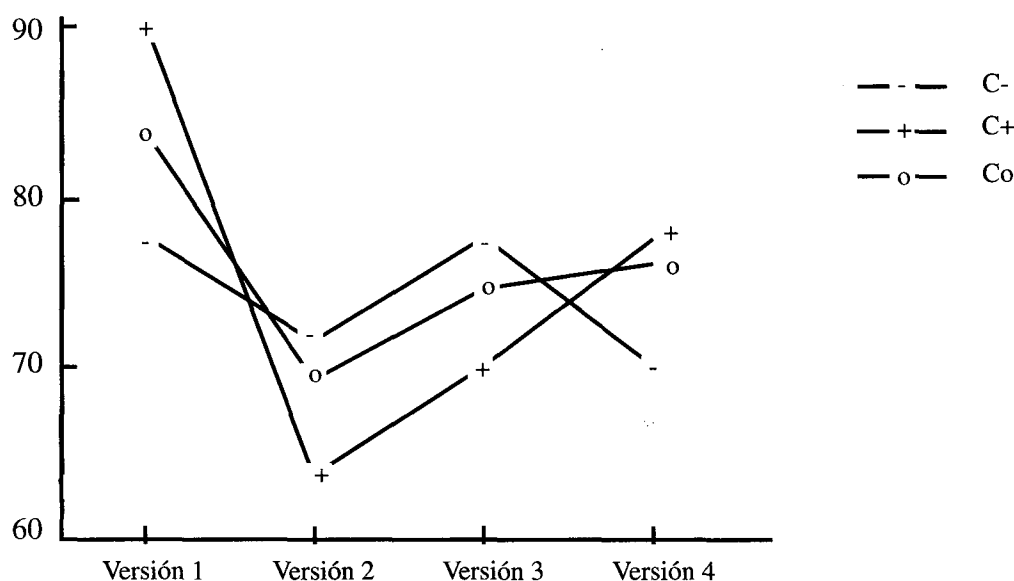


Figure 3. *Moyenne des temps / caractère pour le début du texte, en fonction des différentes versions et des différentes consignes (Co: ancre neutre; C+: ancre positive; C-: ancre négative).*

En fin de texte on ne trouve pas d'effet notable (tests non significatifs).

Entre le début et la fin du texte on peut observer une modification de l'effet des versions (figure 4).

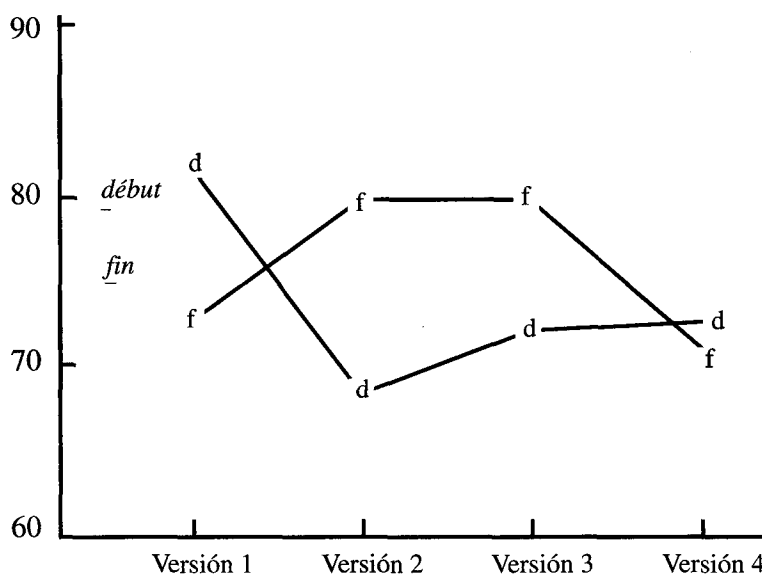


Figure 4. *Moyenne des temps / caractère en fonction des différentes versions et de la localisation en début ou fin de texte.*

En ce qui concerne les versions contrastées (V1 & V2) l'effet s'inverse entre le début et la fin du texte ($F_{(1,21)} = 8,84$; $p = .007$): pour la version 1, les TC les plus longs se trouvent au début ($F_{(1,21)} = 5,21$; $p = .033$), alors que pour la version 2, les TC les plus longs se trouvent à la fin ($F_{(1,21)} = 8,32$; $p = .008$). En ce qui concerne les versions hété-

rogènes on n'obtient aucun test significatif concernant l'effet de la localisation ($F_{(1,21)} = 1,37$; $p = .25$): pour V3 $F_{(1,21)} = 1,8$; $p = .19$; pour V4 $F_{(1,21)} < 1$.

COMMENTAIRE

Globalement on peut se rendre compte que la recherche d'information dans le texte afin de se construire une représentation de la maîtrise des éléments de connaissance pris en compte dans l'évaluation semble obéir à certaines contraintes.

Tout d'abord il faut noter que le début de la copie exerce une influence déterminante. D'une part le premier élément de connaissance rencontré fait l'objet d'un traitement plus long. D'autre part seul le début de la copie est sensible à l'effet de l'ancre et des versions. Ce qui signifie que l'ancre structure les marquages initiaux, et qu'en fonction de cette structure initiale se développe une certaine stratégie de prise d'information consistant à intégrer des éléments plus ou moins cohérents avec cette structure. Celle-ci n'est modifiée que dans le cas où un nombre trop important d'éléments contradictoires sont rencontrés.

Ensuite le traitement des éléments de connaissance dépend de l'accord entre ceux-ci et la représentation induite sous l'influence de l'ancre. On constate que les segments de texte porteurs d'une information cohérente avec une consigne sont lus plus rapidement, et qu'à l'inverse, les segments de texte porteurs d'une information contradictoire sont lus plus lentement.

Enfin, le traitement des éléments de connaissance dépend du lieu où ils se situent dans le texte, et de la densité avec laquelle ils se présentent. Pour les versions homogènes présentant une structure prégnante en ce qui concerne la recherche d'information, on trouve une inversion de la vitesse de lecture, selon la mise en jeu correcte ou incorrecte, en début ou en fin de texte. Par contre, pour les versions hétérogènes, les segments sont lus avec une certaine constance de vitesse, que les segments soient au début ou en fin de texte.

DISCUSSION

Par rapport à l'ensemble des résultats, on peut pointer quelques caractéristiques de la conduite de l'expert en train d'évaluer: ce dernier recherche l'information qui lui est utile, de façon différenciés selon les mécanismes de construction de l'espace évaluatif, selon l'organisation de cet espace, selon la séquence dans laquelle l'information est structurée.

Quand l'information présente une certaine consistance (versions contrastées), la construction peut se faire sur une prise d'indice homogène. On constate alors que les temps de traitement suivent la règle de l'allongement quand il y a intégration d'une mise en jeu incorrecte d'un élément de connaissance. Qui plus est, pour ces versions le début de la copie a une influence déterminante sur l'évaluation finale. Par contre, si l'information est distribuée de façon hétérogène, et que sans cesse l'évaluateur est confronté à des mises en jeu correcte et incorrecte, aucun élément ne peut être intégré correctement, et aucun mécanisme de construction de l'espace évaluatif ne peut être caractérisé. Ce résultat

tat peut se rapprocher de celui de Pennington & Hastie (1988): la cohérence de la présentation des arguments améliore la confiance dans la prononciation d'un verdict juridique. De même Denis & Cocude (1992) trouvent que la structure d'un texte descriptif affecte la structure de la représentation cognitive induite par la lecture, et de plus affecte les opérations mentales effectuées sur cette représentation. Ainsi pour notre expérience, dans le cas des versions homogènes («bien structurées») on trouve peu de retours en arrière sur le texte. De plus le nombre de retours en arrière est indifférencié selon l'ancre, car l'évaluateur développe une stratégie de confirmation d'une hypothèse. Dans ce cas, on peut dire qu'il n'y a pas une prise en compte très différente des éléments favorables ou contraires au marquage initial. Par conséquent on peut dissocier deux phénomènes: bien qu'au plan de l'intégration de l'information il y ait un traitement spécifique des mises en jeu incorrectes, l'espace des marquages, une fois structuré sous l'effet de l'ancre⁸, semble très stable.

Au total pour qu'un mécanisme de construction de l'espace évaluatif puisse se dégager il faut que le texte évalué présente une structure cohérente du point de vue de la mise en jeu des éléments de connaissance. Si c'est le cas, on peut pointer l'apparition d'un biais de confirmation puisque le nombre d'éléments de connaissances contraires à la structure de l'espace évaluatif doit être assez important avant qu'ils soient pris en compte.

On peut également souligner l'effet de l'information donnée avant la tâche d'évaluation elle-même. Tout d'abord la connaissance du niveau de l'élève influence le résultat de l'évaluation finale, confirmant ainsi l'objectivité de l'heuristique d'ancrage-ajustement dans une tâche d'évaluation dans une situation pratique. La valeur initiale donnée sert de point de référence (d'ancre) sur lequel s'ajuste la valeur évaluative. Mais le plus intéressant est que les différentes ancres ne jouent pas seulement à la fin de l'élaboration du jugement comme un «modérateur». En effet, les marquages initiaux de l'espace évaluatif, modifiés par l'effet de l'ancre, influencent activement la prise d'information. Selon que le sujet s'attend à trouver des mises en jeu incorrectes d'éléments de connaissance à évaluer (ancre négative), des mises en jeu correctes (ancre positive), ou n'a aucune attente *a priori*, le traitement de l'information rencontrée n'est pas le même. Cet effet joue essentiellement au début du texte, quand l'évaluateur n'a pas encore modifié son espace des marquages évaluatifs en fonction des éléments rencontrés. Par exemple, pour la version avec toutes les mises en jeu d'éléments de connaissance incorrectes au début, ces éléments sont traités le plus rapidement quand la consigne avec ancre négative est donnée, et le plus lentement pour la consigne avec ancre positive. Ce qui signifie que l'évaluateur intègre rapidement les informations correspondant à son attente, alors qu'il est plus lent dans le cas contraire.

D'une part, ces résultats sont cohérents avec le modèle de recouvrement structurant l'espace évaluatif. D'autre part, ils permettent de lever une contradiction pointée par Pennington & Hastie (1986) à propos des situations dans lesquelles le modèle d'ancrage-ajustement semble inopérant. En effet, si sous l'effet de l'ancre, les marquages initiaux se manifestent par des stratégies de prise d'information, l'heuristique d'ancrage ne se traduit pas forcément simplement par des lois additives ou multiplicatives dans des tâches complexes.

8. L'ancre est induite soit par la consigne, soit par construction en fonction du début du texte.

Enfin la localisation et la séquence de la mise en jeu des éléments de connaissance considérés a une influence sur la recherche et le traitement de ces éléments. Dans le domaine de l'évaluation, Tseptsov & Coirier (1992)⁹ trouvent que la séquence dans laquelle sont présentés les différents arguments dans un texte ont une influence sur le jugement de l'événement rapporté, et sur la stratégie de l'élaboration de la réponse de jugement. Dans notre expérience, si l'on considère les différences de temps de traitement des éléments entre le début et la fin du texte, on s'aperçoit que cette différence dépend des versions, i.e. des différentes séquences de présentation des éléments de connaissances à évaluer. De même, dans les cas où le texte est mal structuré¹⁰, si l'évaluateur rencontre un élément contraire à son attente, il a tendance à retourner prendre de l'information dans le texte déjà lu, alors qu'il a tendance à l'intégrer dans la structure actualisée de son espace évaluatif dans le cas de textes homogènes.

Ces résultats vont dans le sens d'une thèse selon laquelle la structure de présentation de l'information influence les opérations mentales en mémoire de travail (Kulhavy, Stock, Verdi, Rittschof & Savenye, 1993).

En conclusion, cette expérience a permis non seulement de valider l'heuristique d'ancrage-ajustement dans une situation pratique, mais de replacer la problématique de l'élaboration d'une réponse évaluative dans une autre perspective. C'est en repérant certaines singularités de l'interaction entre les propriétés de l'objet à évaluer et les stratégies de l'évaluateur que l'on a pu caractériser certains processus cognitifs en jeu pendant l'élaboration d'une réponse évaluative. D'une part, cette réponse évaluative s'élabore non seulement à partir du cumul des différents indices, mais ces indices sont intégrés dans un espace évaluatif présentant des propriétés particulières. Ainsi, cette élaboration est sensible à un biais de confirmation qui se traduit par une intégration spéciale des informations contraires à une hypothèse actualisée. D'autre part, dans une tâche complexe comme celle d'évaluer des copies d'élèves, chez l'expert il y aurait un processus supérieur de gestion qui, bien que sensible à l'heuristique d'ancrage-ajustement, pourrait en utiliser d'autres, mieux adaptées à la tâche à exécuter.

RÉFÉRENCES BIBLIOGRAPHIQUES

- AARONSON, D. & SCABOROUGH, H. S. (1976). Performance theories for sentence coding: some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 56-70.
- BERKELEY, D. & HUMPHREYS, P. (1982). Structuring decision problems and the bias heuristic. *Acta Psychologica*, 50, 201-252.
- CAVERNI, J.-P. (1987a). Knowledge acquisition assessment by expert: effects and models of the cognitive functioning of evaluators. *European Journal of Psychology of Education*, 2, 119-131.

9. Tseptsov, V. & Coirier, P. (1992, mai), *Effet de séquence textuelle sur la vraisemblance d'une information événementielle*. Congrès de la Société Française de Psychologie. Poitiers, 13-15 mai.

10. Du point de vue de la répartition des éléments de connaissance à évaluer.

- CAVERNI, J.-P. (1987b). Self-Paced display time for process tracing in assessment of acquired knowledge. *European Bulletin of Cognitive Psychology*, 7, 633-651.
- CAVERNI, J.-P. & PÉRIS, J.L. (1990). The anchoring-adjustment heuristic in an «information rich, real world setting»: knowledge assessment by experts. In CAVERNI J.-P. et al. (Eds.). *Cognitive biases*. Amsterdam: North Holland.
- CAVERNI, J.-P. & GONZALEZ, M. (1986). Vers une modélisation des processus de jugement dans l'évaluation. *Bulletin de Psychologie*, 39, 375, 301-304.
- DAVIES, G. & LOGIE, R. H. (1991). Contemporary themes in cognitive psychology. *The psychologist: Bulletin of the British Psychological society*, 4, 291-292.
- DENIS, M. & COUDE, M. (1992). Structural properties of visual images constructed from poorly or well-structured verbal description. *Memory & Cognition*, 20, 497-506.
- DURUP, H. (1967). Graphes et plans d'expériences temporels, mots circulaires et plans toriques. *Mathématiques et Sciences Humaines*, 18, 1-31.
- DUTTER, V. (1986). Des critères fonctionnent-ils comme opérateurs cognitifs lors de la saisie d'indices d'évaluation? *Bulletin de Psychologie*, 39 (375), 311-318.
- ELSTEIN, A. S.; SHULMAN, L. S. & SPRAFKA, S. A. (1990). Medical problem solving: a ten-year retrospective. *Evaluation and the Health Professions*, 13 (1), 5-36.
- ESTES, W. K. (1990). Cognitive architecture from the standpoint of an experimental psychologist. *Annual review of psychology*, 42, 1-28.
- EVANS, J. St B.T. (1983). Selective processes in reasoning. In EVANS J. St B.T. (Ed.). *Thinking and reasoning: psychological approaches*. London: Routledge & Keagan Paul.
- GONZALEZ, M. (1988). A propos de la validité des modèles de simulation de processus. In CAVERNI, J.-P.; BASTIEN, C.; MENDELSON, P. & TIBERGHIE, G. (Eds.). *Psychologie cognitive: modèles et méthode* (pp. 443-464). Grenoble: Presses Universitaires de Grenoble.
- GUERCIN, F.; ROUSSEY, J.-Y. & PIOLAT, A. (1990). Time-series: a tool for studying complex cognitive activities. *European Bulletin of Cognitive Psychology / C.P.C.*, 10, pp. 79-110.
- HOGART, R. & EINHORN, H. (1992). Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology*, 24, 1-55.
- KULHAVY, R.; STOCK, W.; VERDI, M.; RITTSCHOF, K. & SAVENYE, W. (1993). Why maps improve memory for text: the influence of structural information on working memory operations. *European Journal of Cognitive Psychology*, 5 (4), 375-392.
- KLEINMUNTZ, D. N. (1985). Human decision processes: heuristics and task structure. In HANCOCK, P. A. (Ed.). *Human Factors Psychology*. Amsterdam: North Holland.
- LOGIE, R. H. & BRUCE, D. (1990). Developments and directions in applying cognitive psychology. *Applied Cognitive Psychology*, 4 (4), 349-358.
- NOIZET, G. & CAVERNI, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris: P.U.F.
- NORTHCRAFT, G. B. & NEALE, M. A. (1987). Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, 39, 84-97.
- PAYNE, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, 92, 382-402.
- PAYNE, J. W.; BETTMAN, J. R. & JOHNSON, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 534-552.

- PENNINGTON, N. & HASTIE, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51, 242-258.
- PENNINGTON, N. & HASTIE, R. (1988). Explanation-based decision making: effects of memory structure on judgement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 521-533.
- PITZ, G. F. & SACHS, N. J. (1984). Judgment and decision: theory and application. *Annual Review of Psychology*, 35, 139-163.
- POLLARD, P. & EVANS, J. St B.T. (1983). The role of 'representativeness' in statistical inference: a critical appraisal. In EVANS, J. St B.T. (Ed.). *Thinking and reasoning: psychological approaches*. London: Routledge & Keagan Paul.
- PYNTE, J. (1974). Une expérience automatisée en psycholinguistique. *Informatique et Sciences Humaines*, 23, 45-46.
- PYNTE, J. & NOIZET, G. (1980). Optimal segmentation for sentences displayed on a video screen. In KOLERS, P.; WROLSTAD, H. & BOUMA, H. (Eds.). *Processing of visible language, II* (pp. 375-386). New-York: Plenum Press.
- ROUANET, H. & LÉPINE, D. (1977). Introduction à l'analyse des comparaisons pour le traitement des données expérimentales. *Informatique et Sciences Humaines*, 33-34, 125p.
- RUSSO, J. E. & DOSHER, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 676-696.
- TIBERGHIE, G. (1985). Mais où sont les stimulus d'antan? *Psychologie Française*, 30 (2), 177-183.
- TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- VLECK, C. & KEREN, G. (1993). Behavioral decision theory and environmental risk management: assesment and resolution of four «survival» dilemmas. *Acta Psychologica*, 80, 249-278.