

Laurent KARSENTY*

Une définition psychologique de l'explication

Ce texte propose une définition psychologique de l'explication. Cette définition s'appuie sur une série de travaux émanant de différents champs des sciences cognitives: la pragmatique linguistique, la psychologie des processus de compréhension, la psychologie des processus d'apprentissage et l'étude des mécanismes d'attribution causale. L'idée centrale qui ressort de cette analyse est que l'explication vise à ajuster le contexte cognitif du sujet de façon à rendre une information nouvelle compréhensible. Cet ajustement contextuel implique différents processus cognitifs: généralisation, particularisation, raisonnement contrastif, justification. En outre, nous décrivons comment ces processus interagissent dans la construction de l'explication. Enfin, l'intérêt de la définition proposée est discuté dans le cadre de l'étude des dialogues explicatifs : on montre comment cette définition rend compte de l'une des principales caractéristiques de l'explication dans le dialogue, à savoir son caractère interactif.

Mots-clés : explication, compréhension, contexte, dialogue explicatif.

A psychological definition of explanation. This text proposes a psychological definition of explanation. This definition relies on a synthesis of several previous studies coming from different fields of Cognitive Science : Pragmatics, Psychology of understanding, Psychology of learning processes and the study of causal attribution mechanisms. The main point emerging from this analysis is that explanations aim at adjusting the cognitive context of an individual in order to make new information understandable. This contextual adjustment implies different cognitive processes : generalization, particularization, contrastive reasoning and justification. In addition, we describe how these different processes interact in the construction of an explanation. Finally, the relevance of the definition proposed is discussed. In particular, we argue that this definition accounts

* ARAMIHS, Groupe de Recherche en Ingénierie Cognitive, 31, rue des Cosmonautes, 31077 Toulouse Cedex, France. Tél.: +33 (0)5 2 19 77 63, Fax ; +33 (0)5 62 19 58 53. e-mail: karsenty@anubis.tls.mms.fr.

for an important feature of explanations in a dialogue, their interactive dimension.

Key Words : Explanation, Understanding, Context, Explanatory Dialogue.

INTRODUCTION

Le concept d'explication, ou l'acte d'expliquer, ont plusieurs acceptions dans le langage commun. L'une d'elles nous intéressera particulièrement dans le cadre de ce texte : expliquer, c'est "faire comprendre ou faire connaître en détail par un développement oral ou écrit" (Petit Larousse, 1995). L'objet premier de ce texte est de donner un contenu psychologique à une telle définition. Nous tentons donc de répondre à la question : Qu'est-ce que *faire comprendre* ou *faire connaître en détail* signifient en termes de processus cognitifs ?

Ce texte apporte une réponse à cette question sous la forme d'une définition psychologique de l'explication, en se basant sur un certain nombre de travaux sur l'explication menés dans le champ des sciences cognitives. Ces travaux touchent notamment l'étude linguistique de l'explication, l'étude des processus de compréhension créative, des processus d'apprentissage et des mécanismes d'attribution causale. Précisons que nous n'avons pas cherché à être exhaustif dans la présentation des travaux traitant de l'explication. Notre souci a plutôt été de regrouper un ensemble de travaux appartenant à des champs disparates dont le rapprochement a permis de dégager une conception *psychologique* de l'explication. Autrement dit, les quelques travaux empruntés aux disciplines précédemment citées ont été choisis car ils permettent, chacun de leur point de vue, d'argumenter la définition qui sera proposée.

Deux raisons principales ont motivé cette réflexion. La première est qu'il n'existe pas, en psychologie cognitive, une conception théorique de l'explication indépendante d'un champ de recherche particulier. Comme on pourra le constater dans la suite, les définitions (rares) que l'on rencontre dans la littérature scientifique sont très liées au cadre de travail de leurs auteurs. Nous pensons que la définition proposée dans ce texte a un caractère plus général, et permet donc de comprendre plus globalement le processus d'explication. L'autre raison est liée à des objectifs appliqués de notre champ de recherche, l'ergonomie cognitive des dialogues homme-système. L'évolution des technologies place aujourd'hui l'homme face à des systèmes techniques de plus en plus difficiles à comprendre, donc à utiliser. Un certain nombre de solutions techniques a été avancé par des ingénieurs pour répondre à ce problème. Par exemple, un système à base de connaissances peut produire une trace de son raisonnement¹ pour expliquer la solution qu'il

¹ Une trace du raisonnement reprend l'enchaînement des règles utilisées par un système expert ou système à base de connaissances pour résoudre un problème.

a produit. On peut ainsi répondre à des questions de type "Pourquoi as-tu besoin de telle information?" ou "Comment as-tu déterminé cette solution?" (voir à ce sujet l'ouvrage de Buchanan & Shortliffe, 1984). Il s'est avéré cependant que ce type de solution ne correspondait que très partiellement, voire pas du tout aux besoins des utilisateurs de ces systèmes (Karsenty & Brézillon, 1995b). On a ainsi montré que lors de dialogues de validation de schémas conceptuels de bases de données, réunissant un concepteur et des futurs utilisateurs, aucune explication apportée par le concepteur (l'expert) ne portait sur son raisonnement à l'origine de la structure de base de données proposée. Ses explications, qui pour la grande majorité étaient données spontanément, portaient plutôt sur (i) la signification des éléments composant les schémas conceptuels, (ii) leur justification en terme de besoins utilisateurs, (iii) leur mode d'utilisation dans le travail quotidien des utilisateurs. Ce n'est donc pas la "trace du raisonnement" de cet expert qui est nécessaire pour rendre ses propositions plus acceptables, mais la prise en compte du point de vue des utilisateurs (c'est à dire leurs activités préalables, leurs objectifs, leurs contraintes, etc.) sur la future base de données.

Une voie possible pour résoudre ce problème serait de chercher à identifier le plus exhaustivement possible les types d'explication dont l'utilisateur a besoin, et de mettre en machine les connaissances nécessaires pour y répondre. C'est la voie que beaucoup ont poursuivi jusqu'ici, notamment en intelligence artificielle (voir par exemple Swartout, 1983, Belkin, 1988, Cawsey, 1993).

Ce type d'approche, qui, précisons-le, ne fait pas l'unanimité même parmi les chercheurs en Intelligence Artificielle (voir par exemple Moore & Paris, 1992) repose cependant sur deux postulats : (1) l'explication peut être une simple réponse à une question, généralement du type Pourquoi ou Comment, (2) l'explication peut être déterminée par l'expliqueur seul. Or, l'étude de dialogues humains montre qu'une telle définition de l'explication présente des limites : on observe en effet que l'explication humaine est souvent *co-construite*, c'est à dire résulte d'une série d'ajustements sur plusieurs échanges entre l'expliqueur et l'expliqué (Baker, 1992, Karsenty & Brézillon, 1995). Ces cas traduisent des situations où l'explication ne peut être produite sans une aide active de l'expliqué.

Cette opposition de forme entre l'explication-machine et l'explication humaine conduit à se demander si l'insatisfaction des utilisateurs vis-à-vis des explications actuellement produites par une machine ne reposerait pas plus sur une fausse conception de ce qu'est

l'explication, plutôt que sur un manque de connaissances explicatives. Parce que la définition de l'explication adoptée par des concepteurs de systèmes complexes conditionne la forme et le contenu d'un dialogue explicatif instaurée entre l'homme et la machine, il est important de rechercher une définition de l'explication qui prenne un autre point de vue, le point de vue du sujet cognitif.

1. L'ETUDE LINGUISTIQUE DEL'EXPLICATION

Cette section rapporte les résultats du travail d'un groupe ayant mené une analyse linguistico-pragmatique des marques de l'explication (Groupe λ -1, 1975)². Ce travail a permis de définir des contraintes sur la réalisation d'un acte d'explication et de cibler le type de connaissances transmis par l'acte d'explication. Les membres du Groupe λ -1 considèrent une explication comme un énoncé constitué de trois composants : [l'information-à-expliquer p] [la relation explicative] [le contenu explicatif q]. Un énoncé de la forme p parce que q ³ représente pour ces auteurs un exemple typique de l'acte explicatif.

Pour définir la spécificité de ce type d'énoncé, ils envisagent alors les contraintes (1) sur p , (2) sur q , et (3) sur le lien entre p et q :

1.1. Contrainte sur p (l'information-à-expliquer)

La vérité de p doit toujours être reconnue. C'est d'ailleurs ce qui permet de distinguer sémantiquement les deux tournures suivantes : " p parce que q " et " q ; aussi p " comme le montrent les deux énoncés suivants :

- (1) Je suis sorti parce qu'il faisait beau
- (2) Il faisait beau ; aussi suis-je sorti.

Dans les deux cas, q est donné pour l'origine de p . Mais lorsqu'on emploie "parce que", p est considéré comme un fait déjà connu, que l'on

² Le groupe λ -1 ("Logique et Langage") rassemblait différents professeurs et chercheurs traitant des rapports possibles entre l'enseignement de la langue et celui de la pensée logique. L'article dont sont extraites les idées qui suivent a été rédigé par un sous-groupe comprenant M.C. Barbault, O. Ducrot, J. Dufour, J. Espagnon, C. Israel, D. Manesse.

³ Dans l'étude citée, le groupe λ -1 compare l'utilisation des conjonctions "parce que", "puisque" et "car", ce qui les amène notamment à faire la distinction suivante : "parce que" traduirait un acte explicatif, tandis que "puisque" et "car" traduiraient des actes de justification ou d'inférence.

cherche à expliquer, "cette admission préalable (de p) étant cela même qui permet de parler d'un acte d'explication" (Groupe λ -1, 1975, p. 262).

1.2. Contraintes sur q (le contenu explicatif)

Aucune contrainte particulière ne semble s'appliquer sur q . "Le locuteur ne présente pas nécessairement q comme déjà connu de l'auditeur, ni d'avantage comme inconnu". Mais il faut ajouter que "si l'auditeur refuse d'admettre la vérité de q , il ne pourra pas, a fortiori, tenir q pour cause de p (autrement dit, un énoncé de type : p parce que q "implique", parmi ses multiples conséquences, la vérité de q)" (ibid., p. 261)

1.3. Contraintes sur le lien entre p et q

Le contenu qui est asserté dans une explication est la relation causale entre p et q . C'est ce lien qui est présenté comme un apport d'information nouvelle. Ceci conduit d'ailleurs les auteurs à poser que la nécessité de croire q "découle logiquement, une fois p considéré comme incontestable, de la causalité affirmée entre p et q " (ibid., p.262)

Ces travaux, en se penchant sur l'utilisation de la conjonction "parce que", amènent donc à dire que la spécificité de l'acte d'explication repose sur deux aspects :

1. l'admission préalable de l'information-à-expliquer
2. l'énonciation d'une relation causale entre une information-à-expliquer et un ou plusieurs contenus explicatifs.

Si cette définition semble convenir pour l'assertion d'une relation causale, on peut constater qu'elle n'est pas tout à fait juste pour rendre compte de formes plus complexes d'explication. L'exemple présenté dans la suite (tiré de Schank, 1986) va illustrer cette limite :

Demande d'explication : Pourquoi le magasin fait-il son inventaire en pleine saison ?

Explication (possible) : Parce qu'il appartient à une chaîne qui en a décidé ainsi.

Tout d'abord, reconnaissons que l'explication satisfait bien les deux critères donnés par le Groupe λ -1: elle ne remet pas en cause le fait que "le magasin fait son inventaire en pleine saison", et elle consiste à mettre en relation causale ce fait avec un autre fait ("le magasin appartient à une chaîne qui en a décidé ainsi"). Toutefois, remarquons que le contenu explicatif est en fait décomposable en deux contenus c_1 et c_2 :

- c_1 : "le magasin appartient à une chaîne"

- c2 : "la chaîne de magasins a décidé de la date d'inventaire".

Seul le contenu c2 détermine causalement le fait à expliquer. Mais ce lien causal repose sur la connaissance de c1, qui constitue une information nouvelle essentielle pour faire comprendre la fermeture du magasin sans être pour autant en relation causale avec le fait à expliquer. Bref, si pour expliquer il s'agit effectivement de *mettre en relation un fait admis avec un ensemble d'autres faits*, nous devons considérer que cette mise en relation ne repose pas que sur des liens causaux.

La caractérisation linguistique de l'explication faite par le groupe λ -1, qui met l'accent sur la mise en relation d'un contenu accepté avec des contenus nouveaux peut servir de base à une investigation psychologique de l'explication. L'étape suivante doit alors consister à définir la nature des *processus cognitifs* qui produisent cette mise en relation. C'est l'objet des deux sections suivantes.

2. L'EXPLICATION DANS LES MODELES DE LA COMPREHENSION

L'explication est vue par de nombreux auteurs comme un processus de compréhension prenant place quand l'information n'est pas immédiatement compréhensible (ex. : Schank, 1986, Abelson & Lalljee, 1988). L'explication consiste donc à *faire comprendre*. Pour préciser ce que cela signifie du point de vue de la psychologie cognitive, nous nous intéresserons dans un premier temps à faire ressortir quelques caractéristiques des processus de compréhension.

2.1. Deux étapes dans la compréhension

La plupart des modèles de la compréhension élaborés dans les sciences cognitives traite de la compréhension de texte. On distingue habituellement deux étapes dans la compréhension d'un énoncé :

Dans la première, une compréhension superficielle de l'énoncé donne lieu à une représentation propositionnelle, proche de la forme de surface des phrases [...] La seconde étape de la compréhension, qui est optionnelle, utilise entre autres mais non exclusivement les représentations propositionnelles et consiste à construire un modèle mental dont la structure est analogue à l'état de la situation décrit par le texte (Johnson-Laird, 1983, p. 244).

La seconde étape de la compréhension consiste pour une grande part à générer des inférences enrichissant la représentation propositionnelle de surface, et visant à obtenir une représentation cohérente du texte (Richard, 1990).

Type d'inférence	Description	Texte	Inférence
1. Référentiel	Mot ou phrase référentiellement lié à un précédent élément dans le texte	Le ballon roula jusqu'à lui. Il le prit.	Ballon est le référent de <i>le</i> .
2. Antécédent causal	L'inférence est une chaîne causale entre l'action, l'événement ou l'état décrit et le contexte d'un précédent passage.	L'homme s'arrêta. Il était épuisé.	L'homme n'a pas mangé depuis deux jours.
3. Conséquent causal	L'inférence est une chaîne causale anticipée, incluant des événements physiques et de nouveaux plans des agents.	Il provoqua un accident en grillant un feu rouge.	L'homme va perdre son bonus.
4. Moyen	L'inférence est un objet, ou une ressource utilisée quand un agent exécute une action intentionnelle.	Il serra fortement la vis.	L'homme utilise un tournevis.
5. But super-ordonné	L'inférence est un but qui motive l'action intentionnelle de l'agent.	Elle termina son café, prit son attaché-case, et courut pour attraper son train.	La femme ne veut pas arriver en retard au travail.
6. But	L'inférence est un but, un plan ou une action qui spécifie comment l'action d'un agent est réalisée.	Il referma la porte à double-tour.	L'homme a introduit la clé dans la porte et l'a tournée deux fois.

Tableau 1: Types d'inférences générées en lisant un texte (extrait et adapté de Graesser & Kreuz, 1993)

La génération d'inférences est basée sur des connaissances, et plus précisément sur des structures de connaissances. Ces structures de connaissances ont été formalisées sous forme de schémas (Rumelhart & Ortony, 1977), de scripts (Schank & Abelson, 1975) ou de frames (Minsky, 1975). Les inférences basées sur les connaissances sont héritées de ces structures de connaissances et conduisent à construire une représentation de la situation particulière décrite dans le texte. Le tableau 1 décrit quelques types d'inférence étudiés en Psychologie Cognitive.

2.2. Compréhension et contexte

Pour réaliser les inférences nécessaires à la compréhension d'un énoncé, des structures de connaissances ou des informations sur le texte antérieur doivent être présentes en mémoire de travail. Ces informations forment le *contexte représentationnel* de la compréhension :

Le contexte "représentationnel" [...] est composé de représentations qui sont, à un moment donné, en mémoire opérationnelle, c'est-à-dire activées. Par exemple, ce que je viens de lire se trouve dans un état d'activation transitoire au moment où je traite l'information subséquente. [...] Le concept de "schéma" ou de "script" [...] représente une excellente description hypothétique de ces représentations contextuelles (Tiberghien, 1985, p. 179).

Plus généralement, nous définirons le *contexte* comme l'ensemble des représentations mentales mises en jeu par un sujet dans le but de produire ou d'interpréter un message donné. Par exemple, l'inférence 3 dans le tableau 1 repose sur un contexte cognitif contenant notamment les règles : "si un conducteur provoque un accident alors qu'il est en tort, alors il perd tout ou partie de son bonus" et "griller un feu rouge place le conducteur en position de fautif".

Un certain nombre de travaux sur la compréhension ont montré par ailleurs que la génération d'inférences est dépendante de la finalité de l'activité du lecteur. Certains auteurs distinguent alors la compréhension à visée épistémique et la compréhension à visée pragmatique (Denhière & Richard, 1990). Dans la première situation, "la finalité est l'intégration de l'information à des structures mnésiques en vue d'une récupération ultérieure" (ibid., p.70). Dans la seconde situation, la finalité est "l'obtention d'un résultat par des actions appropriées" (ibid., p.70). Richard (1986) montre, dans le cas de la compréhension de consignes, que comprendre implique l'inférence des buts superordonnés des actions décrites, ainsi que toutes les conditions particulières et le détail de ces actions. Par contre, en compréhension de récit, les lecteurs infèrent essentiellement les buts superordonnés des acteurs du récit, et non le détail de leurs actions. Le but du lecteur semble ainsi impliquer la recherche de certaines inférences particulières.

Sur la base de ce qui vient d'être dit jusqu'ici, on retiendra que la compréhension nécessite d'intégrer une information nouvelle au contexte du sujet composé d'informations anciennes (cf. Haviland & Clark, 1974), ou de pouvoir exploiter l'information nouvelle dans le contexte des informations anciennes pour générer des inférences exigées par les buts du sujet.

Commentaire :

Commentaire :

Soulignons que cette définition du contexte prend un point de vue qui n'est pas celui que l'on trouve partout. On désigne très souvent en effet avec cette notion la situation dans laquelle se trouve les participants au dialogue--le point de vue est alors objectif-- et non leur représentation de cette situation--le point de vue est alors subjectif. Dans la suite, nous emploierons donc l'expression de "contexte cognitif" ou tout simplement le terme de "contexte" exclusivement pour désigner des représentations mentales. Le terme de "situation" sera utilisé sinon.

Si l'on reprend la proposition énoncée en introduction de cette section, qui consiste à dire que l'explication est nécessaire face à quelque chose de non immédiatement compréhensible, on peut en conclure que le besoin d'explication doit naître d'une incompatibilité du contexte du sujet avec l'information nouvelle, et que l'explication doit pouvoir d'une certaine façon modifier ce contexte. Nous allons maintenant confronter cette reformulation à différentes théories de l'explication proposées dans le champ des Sciences Cognitives.

2.3. L'explication, et la généralisation d'événements inattendus

Pour Schank (1986), nous ne sommes pas toujours en train de chercher à tout expliquer. Nous ne cherchons à expliquer que les phénomènes perçus qui s'écartent de nos attentes ou qui ne sont pas pris en compte par ces attentes. Cette proposition repose sur l'idée que comprendre, c'est rattacher de nouvelles expériences à des expériences antérieures stockées en mémoire. Quand nous accédons directement à ces souvenirs, nous n'avons pas besoin d'explication. Quand nous n'accédons à aucun souvenir, ou quand nous avons des attentes déclenchées par nos connaissances et contredites par l'expérience nouvelle, nous recherchons une explication. Le modèle proposé par Schank considère donc la compréhension comme un processus fortement guidé par les connaissances. Les psychologues Abelson et Lalljee défendent une position similaire :

We view the process of explanation [...] as a commonsense problem of connecting the thing to be explained with some available conceptual pattern, appropriately modified to fit the circumstances. [...] Explanation, in other words, is a form of understanding." (Abelson & Lalljee, 1988, p. 175)...

Ces auteurs font alors une distinction entre les explications contrastives et les explications constructives. Les premières rendent compte des situations où un événement attendu ne se présente pas. Les secondes résolvent des situations où un événement surprenant ("anomalous") se présente sans qu'il y ait d'attentes précises.

L'explication, dans ces travaux, est considérée comme un *processus de généralisation* d'un événement inattendu (ex. : Clancey, 1983, Schank, 1986, Chi et coll., 1989). L'objet de ce processus est soit de rechercher en mémoire, soit de construire des connaissances qui permettraient d'attendre, et ainsi d'intégrer, l'événement inattendu⁴. Ces connaissances doivent avoir par essence un caractère plus général que

⁴ Cela rejoint la position du philosophe Gärdenfors (1980) qui considère qu'un phénomène est compris quand il peut être prévu.

Commentaire : Enfin, précisons que cette définition laisse encore une ambiguïté : Considère-t-on le contexte comme les représentations mentales présentes en mémoire de travail au moment où l'on reçoit le message ? Ou considère-t-on le contexte comme des représentations mentales construites pour aboutir à une certaine interprétation du message ? Les travaux en Psychologie Cognitive sur la compréhension de textes et la généralisation d'inférences se situent essentiellement dans le cadre de la première proposition, et considèrent donc le contexte comme quelque chose de donné (cf par exemple la citation de Tiberghien ci-dessus). D'autres travaux viennent destabiliser cette position [cf notamment Sperber & Wilson, 1989]. Ils se situent typiquement dans le cadre de la deuxième proposition, et amènent à définir le contexte comme un ensemble d'informations accédées ou construites pour donner du sens au message. Reste bien sûr à définir ce que l'on entend par "sens", ce qui sera l'une des préoccupations centrales du chapitre 3.

l'événement inattendu, puisque à la base elles rendent compte d'événements en partie différents de celui qui doit être expliqué.

2.4. L'explication et la particularisation des schémas de connaissances

Des travaux portant sur la relation entre compréhension, résolution de problème et apprentissage précisent la description du processus d'explication. Chi et coll. (1989) par exemple, analysent les "auto-explications" (*self-explanations*) que des "bons" et des "mauvais" élèves produisent alors qu'ils étudient des exemples de problème de mécanique corrigés. Ces auteurs cherchent dans le même temps à constater l'influence de ces exemples étudiés sur la résolution de nouveaux problèmes. Leurs analyses montrent une relation importante entre apprentissage et explication : les "bons" élèves extraient une connaissance indépendante des exemples étudiés grâce à une forte activité auto-explicative. L'explication est alors définie comme :

a mechanism of study that allows student to infer and explicate the conditions and consequences of each procedural step in the example, as well as to apply the principles and definitions of concepts to justify them (p.151).

D'après cette définition, l'explication explicite la connaissance implicite dans une procédure. Ce processus conduit à particulariser les connaissances du sujet sur une action, c'est à dire à spécifier les particularités de l'action dans une situation donnée. Il s'agit, pour Chi *et coll.*, d'une fonction importante de l'explication :

In order to learn with understanding, a student needs to overcome the incompleteness of an example by drawing conclusions and making inferences from the presented information [...]. To do so, a student needs to provide explanations for why a particular action is taken (p.149)

Nous pouvons donc dire que l'activité auto-explicative des bons étudiants les amène à élaborer le contexte particulier des actions au-delà de ce qui est donné dans l'exemple.

2.5. Lien entre généralisation et particularisation

Certains travaux mettent donc l'accent sur la fonction de généralisation de l'explication. D'autres travaux accentuent par contre la fonction d'explicitation de la situation remplie par l'explication. En reconsidérant plus précisément le modèle de la compréhension proposé par Schank (1986), on s'aperçoit que cette opposition n'est en fait qu'apparente.

Schank propose que l'explication construite par le sujet, i.e., la généralisation de l'événement inattendu, dépende en partie de la

Commentaire : Cette distinction est précisée dans O'Malley (1987, cité dans i.Draper, 1987;), qui interprète le conflit cognitif en terme de "sentiment de surprise" (voir aussi i.Dessalles, 1992;), et en liste trois sources :

- conflit avec une prédiction active : Étant donné ses connaissances, le sujet s'attend à voir un événement se produire, lequel ne se produit pas.
- conflit avec une hypothèse passive : voir un chien vert, alors que les hypothèses par défaut sur les chiens n'incluent pas cette possibilité.
- conflit avec des attentes a-posteriori : e.g., les conséquences d'une action, non prévues, peuvent ne pas être expliquées immédiatement.

description de l'événement retenue. La description de l'événement n'est en effet pas donnée ; il y a toujours plusieurs façons de décrire la même réalité, donc de la comprendre. Les énoncés (A1) et (A2) présentent deux descriptions possibles du même événement perçu (A). Les énoncés (A1') et (A2') présentent deux explications de (A).

(A) Un jeune chat mort

(A1) Un jeune chat mort est étendu au bord de la route

(A2) Un jeune chat mort a le crâne fracassé.

(A1') Une voiture a écrasé le chat

(A2') M. Dupont s'est encore énervé sur un chat qui devait miauler toute la nuit.

Schank insiste sur le lien qui existe entre les propriétés de la situation que l'on retient dans la description et l'explication que l'on recherche. Il formalise ce lien de la façon suivante :

- la description de la situation construite par l'individu retient des éléments qui vont jouer le rôle d'index en mémoire : ils doivent permettre d'accéder à des patterns explicatifs (*explanation patterns*) contenant cet index. La première description de l'événement (A) fournit ainsi l'index "jeune chat mort" (ou "mort d'un jeune chat") ;

- le sujet récupère alors le ou les pattern(s) explicatif(s) contenant les index ;

- chaque pattern sélectionné définit des conditions de validité qui doivent être vérifiées dans la situation pour qu'il soit applicable. La validation de l'explication candidate conduit alors à rechercher certaines particularités de la situation (i.e. particulariser la représentation de la situation). L'explication (A1') a conduit à ajouter l'élément "au bord de la route" à la description (A), cet élément renforçant l'hypothèse explicative (A1'). L'explication (A2') a conduit à ajouter l'élément "crâne fracassé", cet élément renforçant l'hypothèse explicative (A2').

Retenons au passage une idée importante : le besoin de s'expliquer un fait nouveau conduit le sujet à *s'approprier* ce fait, c'est à dire à lui donner un sens très dépendant de ses connaissances propres. Pour l'un, le fait "un jeune chat mort" sera interprété comme "le chat a été écrasé par une voiture". Pour l'autre, ce même fait sera interprété comme "le chat a été tué par M. Dupont". Voir l'explication comme l'expression d'une appropriation de l'information nouvelle signifie aussi que toute

explication ne sera pas identiquement acceptable par le sujet : les connaissances du sujet conditionnent l'acceptabilité d'une explication.

La description partielle du modèle de Schank donnée ci-dessus permet donc de préciser le lien entre particularisation et généralisation : d'un côté, la description de la situation doit donner les indices qui permettront d'accéder à un pattern explicatif valide, donc de généraliser ; de l'autre, un pattern explicatif sélectionné peut conduire à développer la représentation de la situation, donc à particulariser.

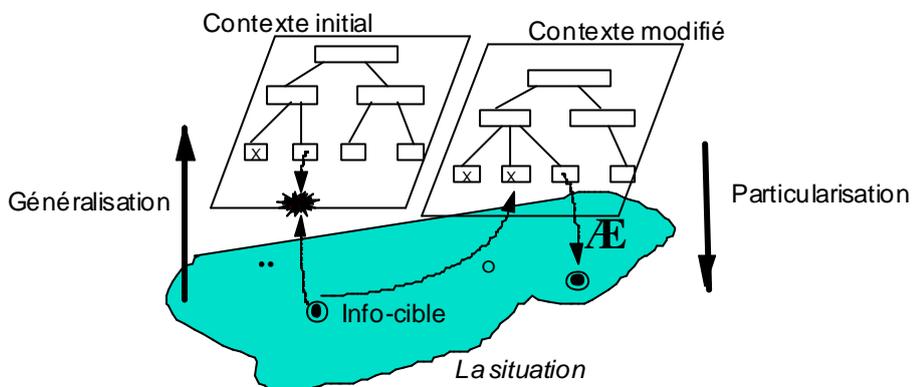


Figure 1 : Les processus de généralisation et de particularisation dans la recherche d'une explication

La figure 1 illustre graphiquement ce double mouvement, et permet de visualiser le travail qui se réalise au niveau cognitif. Le sujet cherche à comprendre une information-cible prélevée dans la situation. Il est alors dans un certain contexte, que nous représentons de façon simplifiée (et quelque peu simpliste) sous forme d'un schéma composé de variables reliées entre elles. Dans un premier temps, le sujet cherche à rendre compte de l'information prélevée avec ce schéma (mouvement de généralisation -1-). Cette tentative n'aboutit pas, ce qui conduit le sujet à tenter une autre généralisation en modifiant tout ou partie du schéma précédemment utilisé (-2-). Dans le même temps, et pour s'assurer que le changement contextuel reste en accord avec la réalité externe, il peut être conduit à vérifier un certain nombre de caractéristiques de la situation (particularisation du nouveau schéma -3-). Si certaines caractéristiques de la situation ne sont pas vérifiées, le sujet continuera à adapter son contexte de la sorte (donc modifiera l'explication de l'information-cible) jusqu'à obtenir un certain degré d'accord entre sa représentation de la situation et la réalité externe.

Ce double mouvement généralisation-particularisation, que l'on peut voir comme un changement de point de vue sur la situation, conduit le

sujet à ajuster peu à peu son contexte initial par rapport à la situation externe. La modification, ou la remise en cause du contexte cognitif, est donc au centre du processus explicatif. D'une certaine façon, une question telle que "Pourquoi X ?" signifierait du point de vue de l'activité mentale "Qu'est-ce qui dans mes représentations contextuelles ne correspond pas à la réalité ?"

3. L'EXPLICATION DANS LES MODELES D'ATTRIBUTION CAUSALE

Précisons tout d'abord que dans le champ de recherches sur l'attribution causale, mais cela est aussi vrai dans certaines recherches issues de la philosophie (ex. : Garfinkel, 1981 ;, Achinstein, 1983), l'explication est synonyme de "cause". Ce champ de recherche a pour objectif essentiel de comprendre comment un sujet humain attribue une certaine cause à un phénomène. Le problème abordé par ces études est que pour la plupart des événements de la vie quotidienne, il est toujours possible d'invoquer un grand nombre de causes.

Notons que malgré l'intitulé différent, l'objectif de ces modèles est aussi de décrire des processus de compréhension, mais exclusivement centrés sur la compréhension causale. C'est pourquoi nous tentons de rapprocher ces deux champs d'étude, dans le but d'aboutir à une définition de l'explication plus générale.

3.1. Raisonnement contrastif et génération d'hypothèses causales

Les modèles classiques de l'attribution causale font appel à un critère de co-variation pour déterminer la cause. Ainsi la définition de "cause" donnée par Kelley & Michela, (1980) est "la condition qui est présente quand l'effet est présent et absente quand l'effet est absent".

Cependant, des déviations systématiques par rapport aux prédictions de ces modèles indiquent que les sujets emploient d'autres critères (Hilton & Slugoski, 1986). Un modèle alternatif, élaboré dans le champ de la philosophie du langage et de la psychologie sociale, postule que les sujets utilisent des critères contrafactuel et contrastif, réunis dans la notion de "condition anormale" (Hilton, 1988). Les conditions anormales sont identifiées comme causes, parce qu'elles sont les conditions nécessaires pour l'apparition d'un événement-cible, qui contrastent avec les conditions obtenues dans un cas comparable dans lequel l'événement-cible n'apparaît pas (Hilton & Slugoski, 1986, p. 77).

Expliquer repose donc sur une *comparaison* entre des situations proches dans lesquelles l'événement-cible peut ou non apparaître. L'implication essentielle de cette conception de l'explication causale

Commentaire :

Des travaux issus de la philosophie présentent une vision assez similaire de l'explication :

Explanation consists in subsuming the situation to be explained under a law. [É] the law is supposed to capture the idea of a connection between A and B, the causal connection whose existence means that A and B are not merely accidentally co-occurring [i.Garfinkel, 1981, p.21]

est que ce qu'un sujet est prêt à reconnaître comme une cause dépend avant tout de ses connaissances. Un exemple proposé par Hesslow (1988) va permettre d'illustrer cette théorie de l'explication causale.

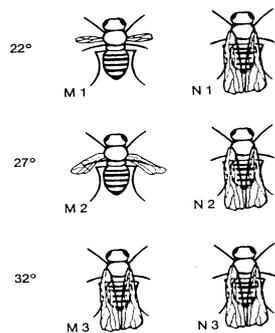


Figure 2 : Familles de mouches et conditions d'élevage (tiré de Hesslow, 1988)

La figure 2 présente différents types de mouches, toutes issues d'une même souche. En ligne, on distingue des mouches élevées dans des conditions environnementales différentes, en colonne, des mouches ayant un patrimoine génétique différent. Comme on peut le constater, les mouches M1 et N1, élevées dans des conditions environnementales identiques, n'ont pas des ailes de même longueur. Si l'on pose la question "Pourquoi M1 a de si petites ailes ?", la réponse naturelle sera de dire que la cause est génétique. Maintenant, supposons que l'on n'ait jamais vu les mouches normales, et que l'on ait observé M1 en même temps que M2 et M3. Ces mouches sont génétiquement identiques à M1, mais ont été élevées dans des conditions de température différentes. Si l'on pose de nouveau la question "Pourquoi M1 a de si petites ailes ?", la réponse naturelle sera de dire que la cause est la variation de température. La même question obtient deux réponses valides différentes !

Cette contradiction est résolue si l'on considère que ce n'est pas la même question qui a été posée. Dans le premier cas, on n'explique pas pourquoi M1 a de si petites ailes, mais pourquoi M1 a des ailes plus petites que N1. Et dans le second cas, on explique pourquoi M1 a des ailes plus petites que M2 et M3. Dans les deux cas, la comparaison est implicite.

L'objet à expliquer n'est donc pas un objet ayant une certaine propriété, mais un objet de comparaison, une différence. Cette position se retrouve dans le travail de Garfinkel (1981) qui affirme qu'un

événement, en étant expliqué, est toujours distingué d'un ensemble d'alternatives (ensemble qu'il appelle "contrast class").

Pour faire le lien avec ce qui a été dit précédemment, nous pouvons réinterpréter cette approche théorique de l'explication en considérant que la classe d'alternatives dont parle Garfinkel, ou l'objet de comparaison invoqué par Hesslow, est inclus dans le contexte cognitif du sujet. C'est même sa présence qui provoque une difficulté à intégrer l'information nouvelle et déclenche une recherche d'explication. L'explication apparaît alors comme l'information qui va contraster le contexte initial du contexte-cible (cas d'*explication contrastive*, selon la terminologie d'Abelson & Lalljee, 1988).

3.2. La confirmation d'une hypothèse causale

Le modèle de l'attribution causale, tel qu'il est présenté jusqu'ici, ne rend pas vraiment compte de l'explication d'un phénomène, et donc de sa compréhension, mais plutôt de la génération d'hypothèses explicatives. Le contraste entre deux situations fait apparaître des différences pouvant déterminer causalement un fait donné. Une étape supplémentaire dans le raisonnement explicatif consiste à confirmer l'une des hypothèses.

Hesslow (1988) propose que les connaissances du sujet, entendues comme structures génériques, soient invoquées pour justifier une hypothèse d'explication donnée. Il illustre cette proposition en soulignant la différence entre dire "La mort de Smith a été causée par une attaque cardiaque" et "les attaques cardiaques provoquent la mort". Le premier énoncé a un caractère particulier, le second, un caractère général. Ces deux types de propositions seraient articulés dans le raisonnement explicatif de telle façon que si Smith a eu un infarctus et est mort, nous utiliserions notre connaissance générale pour appuyer la croyance que cette mort a bien été causée par l'infarctus. Le raisonnement explicatif aboutirait en réussissant à justifier l'interprétation d'un événement perçu avec les connaissances antérieures du sujet.

Le modèle proposé par Hesslow est en fait très proche du modèle présenté par Schank (1986), ce qu'illustre la citation suivante :

Because explanations themselves are such strong indices, other memories, especially one in similar contexts [...] become activated whenever an explanation is concocted. [...] They allow us to check the validity of the explanations themselves (p.21)

On peut toutefois relever une différence : Hesslow parle de connaissances générales pour appuyer des hypothèses causales, Schank considère qu'un souvenir peut jouer ce rôle. Par conséquent, et pour généraliser les deux positions, on peut dire que la réussite du raisonnement explicatif suppose de pouvoir justifier l'interprétation d'un événement perçu avec d'autres interprétations présentes en mémoire, qu'elles soient génériques ou épisodiques.

4. CONCLUSION : L'EXPLICATION COMME PROCESSUS D'AJUSTEMENT CONTEXTUEL

Les modèles des processus de compréhension ou les modèles de l'attribution causale donnent une place prépondérante au contexte cognitif du sujet. Dans les premiers, il correspond essentiellement aux connaissances activées en mémoire de travail, nécessaires pour intégrer l'information nouvelle et réaliser certaines inférences. Dans les seconds, il correspond à une situation de référence à laquelle est comparée l'événement nouveau, ainsi qu'à des connaissances confirmant une hypothèse explicative. Dans les deux cas, un individu a besoin d'explication si son contexte ne permet pas d'intégrer une information nouvelle. L'explication, pour être satisfaisante, doit donc permettre de différencier le contexte initial dans lequel l'information nouvelle n'était pas attendue et le contexte dans lequel elle devient prévisible.

Ces modèles nous conduisent à proposer la définition suivante de l'explication :

L'explication vise à ajuster ou à construire les représentations contextuelles adéquates, quand celles qui sont activées en mémoire de travail ne permettent pas de reconnaître l'information ancienne dans l'information nouvelle, ou ne permettent pas de générer les inférences recherchées à partir de cette information nouvelle.

Faire comprendre, c'est donc modifier certaines représentations contextuelles qui entretiennent un lien avec l'information nouvelle pour lui donner un certain sens.

Notons que cette définition peut être vue comme une traduction en termes psychologiques de la définition de l'explication donnée par les linguistes du Groupe λ -1. Ceux-ci mettent en effet l'accent sur la fonction de mise en relation de l'explication entre un fait admis et d'autres faits. Le processus qui réalise cette mise en relation n'est autre qu'un *processus de contextualisation* d'un point de vue cognitif.

L'explication, ou plus exactement le processus explicatif, traduit un ajustement contextuel nécessaire à l'intégration d'une information

nouvelle. Cet ajustement contextuel repose sur l'interaction entre une phase de génération d'hypothèses explicatives, passant par des processus de généralisation ou un raisonnement contrastif, et une phase de confirmation des hypothèses générées, pouvant passer par un processus de particularisation ou la recherche d'une justification par les connaissances du sujet. Ces phases sont toutes deux très dépendantes des connaissances du sujet. C'est pourquoi nous pouvons considérer l'explication comme l'expression d'une *appropriation* de l'information nouvelle.

Une autre conclusion que l'on doit tirer est que l'origine d'un besoin d'explication n'est pas attachée à l'information nouvelle en soi, mais naît de l'incompatibilité d'une information nouvelle avec les représentations contextuelles du sujet. Cette conclusion conduit à briser l'un des présupposés implicites présents dans un grand nombre de travaux sur la génération automatique d'explications, à savoir qu'il est possible de répondre à un besoin d'explication de type "Pourquoi X?" juste en identifiant le type d'information auquel appartient "X" (voir par exemple Gilbert, 1987). La réponse à un besoin d'explication doit en effet modifier les représentations contextuelles du sujet, lesquelles sont par nature situées, c'est à dire dépendantes de la situation actuelle et particulière du sujet. Ceci signifie que l'explication doit non seulement partir d'une catégorisation sémantique de "X", mais aussi d'une connaissance du contexte cognitif dans lequel est envisagé "X".

On comprend à présent pourquoi l'explication dans le dialogue apparaît moins comme un acte unilatéral, de celui qui explique à celui qui demande l'explication, que comme un acte coopératif, reposant sur les actions des deux interlocuteurs (e.g., Goguen, Weiner & Linde, 1983 ; Moore, 1989 ; Gilbert, 1989 ; Gaulmyn, 1991 ; Baker, 1992 ; Karsenty & Brézillon, 1995 a, b). Parce que le contexte cognitif sous-tendant la demande d'explication est par nature inobservable, l'explicateur ne peut y avoir accès directement. En fait, il dispose pour cela de deux possibilités : (i) il peut avoir accès au contexte de son interlocuteur grâce à un modèle mental qu'il a pu s'en construire (voir par exemple Cahour & Falzon, 1992), mais il faut alors noter que ce modèle peut être incomplet voire incorrect, et de ce fait insuffisant pour produire des explications satisfaisantes "du premier coup" (Moore & Paris, 1992) ; (ii) il peut avoir accès au contexte de son interlocuteur grâce à des explicitations produites par celui-ci au cours d'une série d'interactions. Ces explicitations peuvent être de nature différente (Karsenty & Brézillon, 1995b), par exemple : reformulation de l'explication apportée, explication de la demande d'explication, énoncé d'hypothèses d'explication avec la demande d'explication. Tous ces

comportements traduisent la nécessité pour le demandeur d'explication de participer à la construction de l'explication. Nous pensons que la définition de l'explication proposée dans ce texte rend compte de cette nécessité : l'expliqueur ne peut pas être sûr d'avoir réalisé l'ajustement contextuel nécessaire à la compréhension d'une information nouvelle sans la participation du demandeur d'explication, car il s'agit avant tout de permettre à celui-ci de s'approprier l'information nouvelle.

Bibliographie

- ABELSON R.P. et LALLJEE M. (1988) Knowledge Structures and Causal Explanation, in HILTON D.J. (Ed.) *Contemporary Science and Natural Explanation. Commonsense Conceptions of Causality*. Harvester Press.
- ACHINSTEIN P. (1983) *The Nature of Explanation*, Oxford University Press.
- BAKER M. (1992) Le rôle de la collaboration dans la construction d'explications. *Actes des 2èmes Journées Explication du PRC-GDR-IA du CNRS*, Sophia-Antipolis, 17-19 Juin, pp. 25-40.
- BELKIN N.J. (1988) On the nature and function of explanation in intelligent information retrieval. *11th Conf. on Research and Development in Information Retrieval* (pp. 135-145), Grenoble.
- BUCHANAN B.G. & SHORTLIFFE E.H. (1984) *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Mass.
- CAHOUR B. et FALZON P. (1992) Assistance à l'opérateur et modélisation de sa compétence. *Intellectica*, 12, pp. 159-186.
- CAWSEY A. (1993) Planning interactive explanations. *International Journal of Man-Machine Studies*, 38, pp. 169-199.
- CHI M.T., BASSOK M., LEWIS M.W., REINMANN P. et GLASER R. (1989) Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 13, pp. 145-182.
- CLANCEY W.J. (1983) The Epistemology of a Rule-Based Expert System - A Framework for Explanation. *Artificial Intelligence*, 20, pp. 215-251.
- DENHIÈRE G. et RICHARD J.F. (1990) Compréhension et construction des représentations, in BONNET C., GHIGLIONE R. & RICHARD J.F. *Traité de Psychologie Cognitive*, vol. 2, ch. 2.2, Paris : Dunod.
- GÄRDENFORS P., (1980) A pragmatic approach to explanations, *Philosophy of Science*, 47, pp. 405-423.
- GARFINKEL A. (1981) *Forms of Explanation. Rethinking the Questions in Social Theory*, New Haven, CT: Yale University Press.
- GAULMYN M.M. (de) (1991) La question dans tous ses états. Les cinq types de questions de l'explication conversationnelle, in KERBRAT-ORECCHIONIC. (Ed.) *La Question*. Presse Universitaire de Lyon.

- GILBERT N. (1987) Question and Answer Types, in MORALEE D.S. (Ed.) *Research and Development in Expert Systems III*. Cambridge University Press.
- GILBERT N. (1989) Explanation and Dialogue. *The Knowledge Engineering Review*, 4(3), pp. 235-247.
- GOGUEN J.A., WEINER J.L. & LINDE C. (1983) Reasoning and natural explanation. *International Journal of Man-Machine Studies*, 19, pp. 521-559.
- GRAESSER A.C. & KREUZ R.J. (1993) A Theory of Inference Generation During Text Comprehension. *Discourse Processes*, 16, pp. 145-160.
- GROUPE λ-1 (1975) Car, parce que, puisque. *Revue Romane*, 10(2), pp. 248-280.
- HAVILAND S.E. & CLARK H.H. (1974) What's New ? Acquiring New Information as a Process in Comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, pp. 512-521.
- HESSLOW G. (1988) The Problem of Causal Selection, in HILTON D.J. (Ed.) *Contemporary Science and Natural Explanation. Commonsense Conceptions of Causality*. Harvester Press.
- HILTON D.J. & SLUGOSKI B.R. (1986) Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93(1), pp. 75-88.
- HILTON D.J. (1988) Logic and Causal Attribution, in HILTON D.J. (Ed.) *Contemporary Science and Natural Explanation. Commonsense Conceptions of Causality*, Harvester Press.
- JOHNSON-LAIRD P.N. (1983) *Mental Models*, Cambridge, Cambridge University Press.
- KARSENTY L. & BREZILLON P. (1995a) Cooperative problem solving and explanation. *Expert Systems with Applications*, 8(4), pp. 445-462.
- KARSENTY L. & BREZILLON P. (1995b) Coopération Homme-Machine et Explication. *Le Travail Humain*, 58(4), pp. 291-312.
- KELLEY H.H. & MICHELA J.L. (1980) Attribution theory and research. *Annual Review of Psychology*, 31, pp. 457-503.
- MINSKY M. (1975) A Framework for Representation Knowledge, in WINSTON P. (Ed.) *The Psychology of Computer Vision* (pp. 211-277), New York : Mc Graw Hill.
- MOORE J.D. & PARIS C.L. (1992) Exploiting User Feedback to Compensate for the Unreliability of User Models. *User Modeling and User-Adapted Interaction*, 2, pp. 287-330.
- MOORE J.D. (1989) *A Reactive Approach to Explanation*. Thèse de Doctorat en Informatique de l'Université de Californie, UCLA, Décembre 1989.

- RAM A. (1991) A theory of questions and question asking. *The Journal of the Learning Sciences*, 1(3&4), pp. 273-318.
- RICHARD J.F. (1986) *The semantics of action : its processing as a function of the task*. Rapport INRIA (Rocquencourt) n°542.
- RICHARD J.F. (1990) *Les Activités Mentales. Comprendre, Reasonner, Trouver des Solutions*, Paris: Armand Colin.
- RUMELHART D.E. & ORTONY A. (1977) The representation of knowledge in memory, in : ANDERSON R.C., SPIRO R.J. & MONTAGUE W.E. (Eds.) *Schooling and the acquisition of knowledge* (p. 99-136), Hillsdale, N.J. : Lawrence Erlbaum Associates.
- SCHANK R. & ABELSON R. (1975) Scripts, Plans and Knowledge. *Proc. of the 4th Joint Conf. on Artificial Intelligence (IJCAI'75)*
- SCHANK R.C. (1986) *Explanation Patterns. Understanding Mechanically and Creatively*, LEA, Hillsdale: NJ.
- SWARTOUT W.R. (1983) XPLAIN : a System for Creating and Explaining Expert Consulting Programs. *Artificial Intelligence*, 21(3), pp. 285-325.
- TIBERGHIE G. (1985) Mais où sont les stimulus d'antan ? *Psychologie Française*, 30(2), pp. 177-183.