

Claude VOGEL

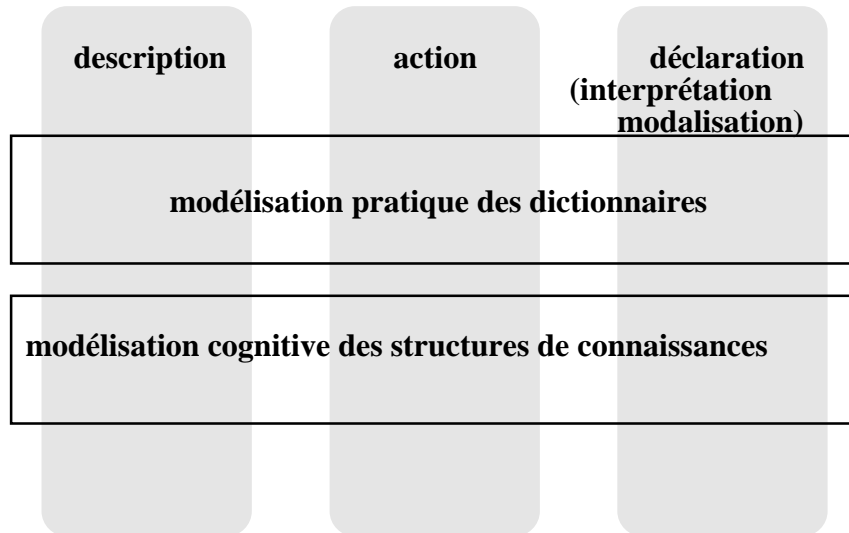
EXPRESSION LANGAGIÈRE DE L'EXPERTISE : PROBLÈMES D'ANALYSE

On a longtemps répété que les problèmes rencontrés lors de l'acquisition des connaissances transformaient cette phase en "goulet d'étranglement" des systèmes experts. Des méthodes telles que KADS (Knowledge Acquisition and Design Support), ou K.O.D. (Knowledge Oriented Design) permettent désormais de maîtriser cette phase critique du cycle de vie en proposant un ensemble de concepts méthodologiques pour encadrer la spécification d'un système à base de connaissances à partir d'une transcription d'entretiens oraux. Après une rapide présentation de ces concepts, on tentera de dresser un bilan des problèmes restés en suspens, l'ensemble de ces problèmes constituant le champ dans lequel devront nécessairement converger génie logiciel et génie linguistique.

I. Formalisation du recueil

La méthode K.O.D. (Vogel, 1988b, 1990) propose d'enchaîner deux phases de modélisation, pratique et cognitive ; la première permettant d'identifier des dictionnaires à partir d'une analyse textuelle des entretiens, la seconde de structurer les entités découpées au premier niveau d'analyse pour produire l'ossature de la base de connaissances. À chacun de ces niveaux, l'analyse considère tour à tour trois paradigmes de représentation :

description statique, action dynamique, et interprétation / modalisation déclarative.



La modélisation pratique s'appuie sur les transcriptions des entretiens libres. Le but de cette tâche est d'identifier à l'intérieur des entretiens toutes les entités qui sont de même nature sémantique. Pour la rapidité de l'exposé, on ne considérera dans les deux points suivants que la sémantique des données statiques ¹.

Dans le domaine des “données”, c'est-à-dire des entités qui servent à énoncer et à décrire les composants de l'univers physique considéré, la modélisation pratique consiste à relever les données citées ou utilisées, et à recenser les traits qui permettent de les caractériser. Pour chaque donnée nouvelle, on crée une entrée du dictionnaire des données.

¹ Cette présentation pratique de la méthode est pour une part le résultat d'une réflexion méthodologique conduite pour le Département Génie Logiciel et Méthodes du Service National Informatique des Télécoms.

1. Définitions

Une donnée est soit une caractérisation d'un objet du monde réel (un objet “photographiable”), soit un concept manipulé comme une entité physique. Les données sont identifiées par des formes descriptives du type : “être dans un état” ou “avoir une propriété”.

<i>donnée physique</i>	<i>appartient à</i>	<i>une classe</i>
	<i>est composée de</i>	<i>données</i>
	<i>est décrit par</i>	<i>propriété</i>
	<i>etc.</i>	

Le modèle servant de référence à cette description est “orienté vers les objets”, et se distingue en cela des modèles “entité-relation”. Ces derniers sont construits en plaçant au premier plan le prédicat relationnel, tandis qu'un “objet” est décrit en prenant le point de vue de l'un des arguments. Chaque argument dupliquant la description de la relation, la notion de référence croisée, reprise ci-dessous, joue un rôle clé dans la gestion de la complétude de la base de connaissances. Si l'on considère le graphe totalisant les relations, un réseau sémantique propage les relations élémentaires, tandis qu'une structure objet structure et hiérarchise des objets-corporations encapsulant des propriétés et des services.

2. Identification d'une donnée dans le texte

Il est généralement facile d'identifier une donnée : elle correspond de façon commune à un objet physique, et elle se présente syntaxiquement sous la forme d'un nom

- sujet d'un prédicat statique : attribution d'une propriété à la donnée
- sujet d'un prédicat dynamique : sujet, destinataire ou ressource d'une action
- déclencheur d'une inférence.

Le premier cas (attribution) recouvre les propriétés primitives classiques d'inclusion, de composition et d'attribution unaire.

Les deux derniers cas alimentent de façon “indirecte” le modèle pratique des données : un énoncé centré de façon évidente sur une action ou une interprétation est d'abord traité comme tel, puis exploité dans un deuxième temps pour épuiser les “références croisées” qui vont permettre de caractériser les données utilisées pour manipuler l'action ou l'inférence.

Les références croisées font partie des procédures formelles de validation du modèle pratique. Au cours de la modélisation pratique des données, il vaut mieux centrer l'analyse sur l'identification des énoncés relevant du premier cas cité ci-dessus : la donnée est sujet d'un prédicat statique.

S'il est relativement trivial d'identifier comme tel un énoncé associant donnée et caractérisation statique de la donnée, il est souvent plus difficile de rejeter les énoncés qui, sous les apparences trompeuses d'un énoncé statique, véhiculent en fait une sémantique dynamique ou déclarative. Considérons les énoncés suivants :

- (1) cette pompe est hors d'état*
- (2) une fréquence d'encoche donne des raies supérieures à 1000 Hz*
- (3) après le terrassement, on édifie un mur de soutènement*
- (4) il faut être habilité pour pénétrer dans le centre*
- (5) si je constate un creux de tension, je pense à un court-circuit*

L'énoncé 1 constitue un exemple type de donnée, la pompe, caractérisée par une propriété, l'état. La pompe est clairement un objet physique et cet énoncé ne présente aucune ambiguïté. On marquera “cette” pompe en lui attribuant un nom.

pompe N
est une pompe
état hors d'état

L'énoncé 2 caractérise la fréquence d'encoche à partir d'une autre donnée, une raie. Comme dans le cas précédent, on notera cette propriété de la fréquence d'encoche. De plus, on créera une entrée pour ces raies, qui porteront provisoirement le nom de *raies de fréquence d'encoche*.

<i>fréquence d'encoche</i>	
<i>est une</i>	<i>fréquence</i>
<i>donne</i>	<i>raie de fréquence d'encoche</i>
<i>raie de fréquence d'encoche</i>	
<i>est une</i>	<i>raie</i>
<i>???</i>	<i>supérieur à 1000hz</i>

On remarquera également que l'énoncé 2 véhicule également une sémantique déclarative en fournissant un modèle de reconnaissance des fréquences d'encoche : la raie supérieure à 1000 hz est une sorte de "signature" de la fréquence d'encoche et cette information pourra être utilisée dans ce sens par le système d'identification ou de diagnostic final.

L'énoncé 3 représente un piège typique : le terrassement et le mur de soutènement, dans le contexte de cet énoncé, ne sont pas des entités de même nature. Si le mur de soutènement est un objet physique, il n'en va pas de même pour le terrassement qui nomme l'action de terrasser. D'une manière générale, on se méfiera des mots en "-tion" (qui nomment une action en cours) et des mots en "-ment" (qui nomment une action achevée).

<i>action : terrasser</i>	
<i>action de</i>	<i>???</i>
<i>décomposé par</i>	<i>???</i>
<i>suivie de</i>	<i>édifier un mur de soutènement</i>
<i>sujet</i>	<i>???</i>
<i>destinataire</i>	<i>???</i>
<i>état antérieur</i>	<i>???</i>
<i>état postérieur</i>	<i>???</i>
<i>ressource</i>	<i>???</i>
<i>contraintes</i>	<i>???</i>

action : édifier
action de ???
décomposé par ???
suivie de ???
sujet *on*
destinataire *mur de soutènement*
état antérieur ???
état postérieur ???
ressource ???
contraintes ???

Le formalisme utilisé pour décrire l'action combine une notation casuelle et une notation fonctionnelle. La première organise autour de l'action les différents pôles usuels de l'actance. La seconde développe la transformation sous-jacente. Les pointeurs de décomposition et de séquence permettent de hiérarchiser des graphes d'activités co-planaires. Le modèle linguistique est ainsi associé à un modèle transactionnel exécutable. Le même *format* est utilisé pour toutes les actions énoncées : expression des processus qualitatifs de fonctionnement des systèmes physiques, transformations expertes du monde physique ou mécanismes de traitement cognitif, bien que le *statut* accordé à ces caractérisations dans le cycle de vie soit différent (ci-dessous § 5). La formalisation de ces deux actions entraîne la création indirecte de l'entrée "mur de soutènement" :

mur de soutènement
est un *mur*
destinataire de *édifier*

Dans un autre contexte, on pourrait rencontrer le terrassement en temps que chose : support physique de la construction, qui deviendrait une entrée du dictionnaire des données.

L'énoncé 4 met en relation une modalité : *il faut être habilité*, avec une action : *pénétrer dans le centre*.

action : pénétrer dans le centre
action de ???
décomposé par ???
suivie de ???
sujet ???
destinataire ???
état antérieur ???
état postérieur ???
ressource ???
contraintes habilitation
contrainte : habilitation
contraint pénétrer dans le centre
thème ???
contenu ??? / habilitation / oui
poids ???
palliatif ???

Comme dans le cas précédent, la contrainte entraîne la création indirecte d'une entrée :

???
 habilitation

L'énoncé 5 met en relation un signe constaté : le creux de tension, avec une explication : le court-circuit. Le creux de tension appartient au monde physique tandis que l'explication fournie lui échappe. Le court-circuit est une abstraction utilisée par l'analyste pour rendre compte du fait observé :

règle : court-circuit
SI court-circuit
ALORS creux de tension
FAIRE ???

Le signe devient indirectement une entrée :

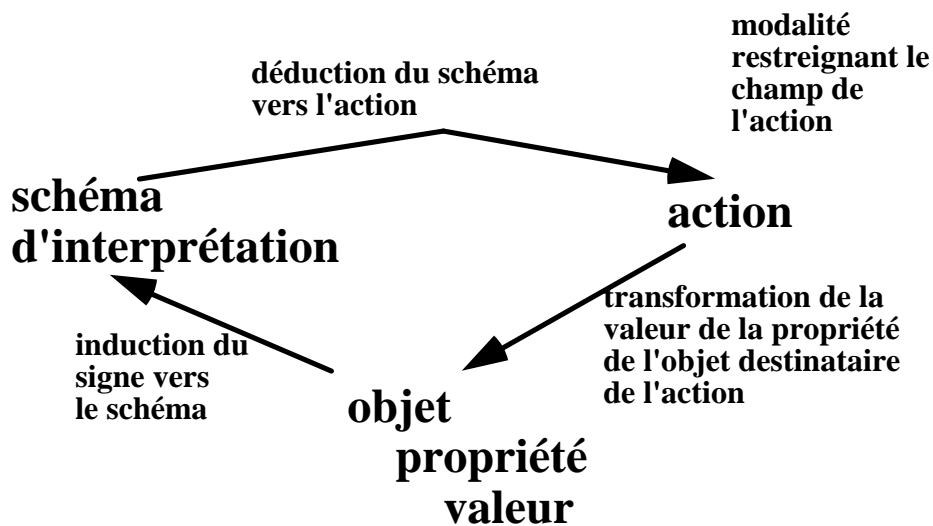
creux de tension
???

Il est clair que toutes les données identifiées, directement ou indirectement, appellent un complément d'information. Ce complément sera obtenu au cours des séances de validation.

3. Modélisation cognitive

Les données listées dans le modèle pratique sont ensuite intégrées dans les classifications du modèle cognitif sous forme de taxinomies. De la même façon, les actions, les contraintes et les inférences relevées dans la modélisation pratique sont structurées en actinomies, en schémas de conduite et en schémas d'interprétation.

Ces différentes structures sont coordonnées entre elles dans un modèle de références croisées qui garantit leur complétude et qui assure dynamiquement l'exécution de leur spécification :



La charge sémantique véhiculée par ces modèles pose des questions complètement neuves au génie logiciel : comment prendre en compte la nouveauté sémantique de ces connaissances au cours des différentes phases composant un cycle de vie (spécification, choix d'architecture, conception, programmation, tests, généralisation), et en retour, comment appliquer sur le développement des systèmes experts les principes de base du génie logiciel ?

II. Le cycle de vie des bases de connaissances : clarifications

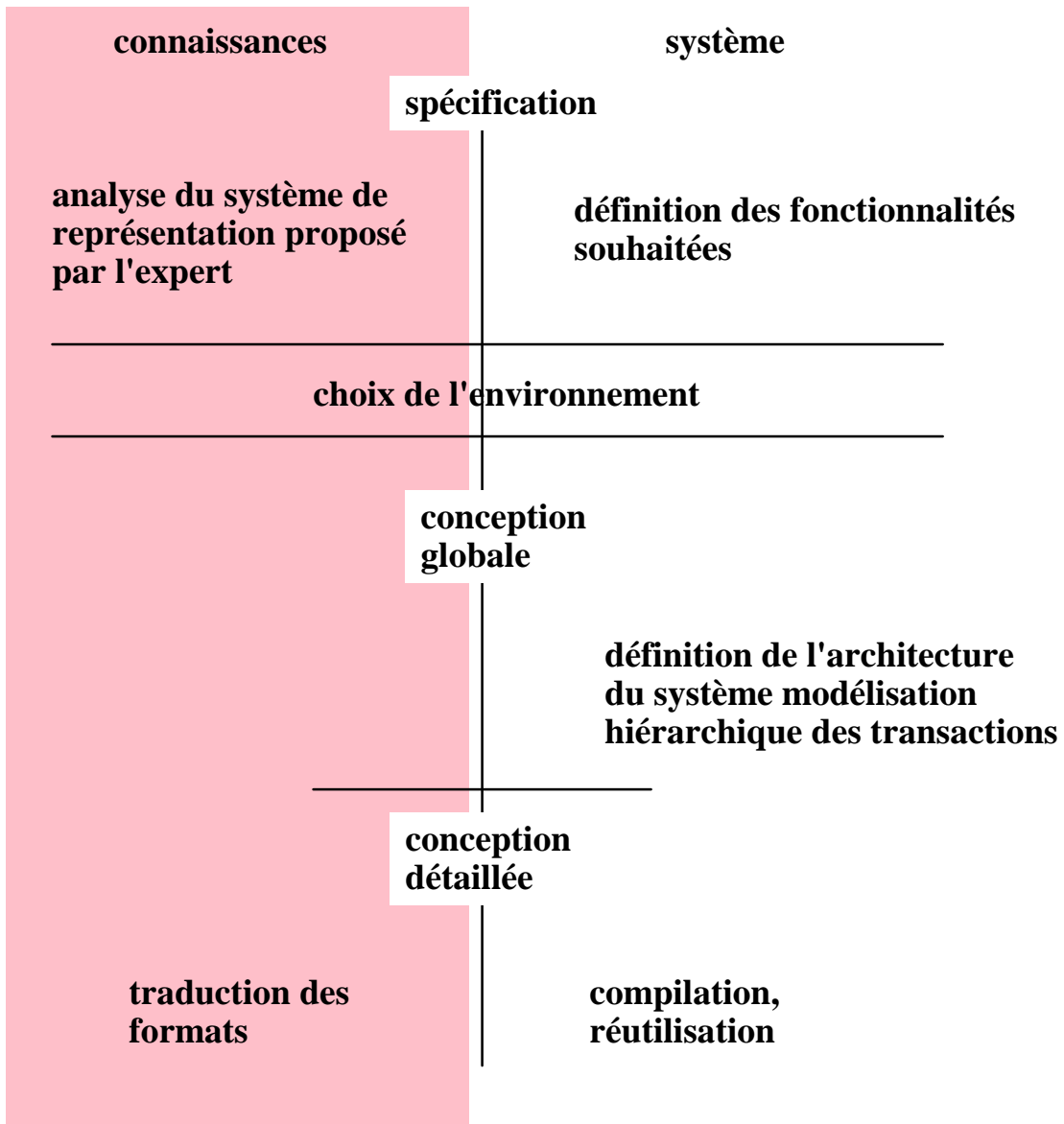
1. Spécification et conception

De façon classique, les méthodes de conduite de projet issues du génie logiciel enchaînent spécification et conception, et placent le choix de l'architecture du système à développer après la spécification. La spécification est définie comme la définition du *problème*, et la conception comme la définition de la *solution*. Cette distinction est introduite pour rendre la définition du problème indépendante de la recherche de la solution, et inversement pour arrêter la visibilité de la solution, aux yeux de l'utilisateur ou du décideur, à la fin de la spécification. Le “besoin” est quant à lui défini en amont des phases de spécification. A l'intérieur de ce cycle de vie, le choix de l'architecture du système et des supports de développement apparaît comme une phase-pivot permettant d'arrêter les stratégies de développement du système final, au terme de la spécification.

Les “objets” qui vont être modélisés dans la conception générale du système *ne* sont généralement *pas* les “objets” définis dans la spécification générale du système. La spécification générale du système a permis de définir — du point de vue de l'expert — les objets qui appartiennent à l'espace du problème (et qui deviendront des données de la solution) ; la conception générale du système va définir — du point de vue du concepteur — les objets qui appartiennent strictement à l'espace de la solution. Un objet appartient strictement à l'espace de la solution s'il n'a

aucune réalité dans l'espace du problème (il n'appartient pas au monde de l'expert), et s'il constitue un élément impératif de la réponse apportée au problème posé.

2. Objets du problème et objets de la solution



Un système à base de connaissances est défini par :

- sa forme : son ergonomie :
- son contenu : l'ensemble des connaissances manipulées ;
- les mécanismes de support qui assurent le bon fonctionnement de la forme et la bonne utilisation du contenu.

La forme est spécifiée à partir des besoins exprimés par les utilisateurs. Le contenu est décrit dans la spécification générale du système. Les mécanismes de support appartiennent par contre directement à l'espace de la solution : ni l'utilisateur ni l'expert ne peuvent nous aider à définir les flux d'information internes au système (i.e. en étendant la notion de transaction, les transactions entre moteur d'inférences, règles et événements), et les liens de dépendance qu'ils établissent entre les différents modules.

3. Logique des modèles et pragmatique

L'orientation des deux phases de spécification et de conception est donc radicalement différente. Curieusement, la dimension pragmatique de la communication, que l'on avait tenu à écarter durant la phase de spécification, revient en force dans la phase de conception.

D'une manière générale, les connexions entre modules compliquent une architecture : un module est plus difficile à appréhender et à contrôler s'il est étroitement interconnecté à d'autres modules. Les trois relations de clientèle (à travers les transactions), d'assemblage et de classification ne sont pas équivalentes de ce point de vue.

Les relations de classification (qu'il s'agisse de classes ou de prototypes) sont fondées sur une sémantique unique, portent avant tout sur des catégories, et sont régulièrement hiérarchiques. Elles dominent la phase de spécification.

Les relations d'assemblage qui constituent le cœur des mécanismes sont fondées sur des sémantiques multiples, sont issues de- ou sont rendues

intelligibles par- la pragmatique des mécanismes, et les graphes construits sont irréguliers.

Les relations de clientèle sont toujours pragmatiques. Ce sont elles qui dessinent de façon décisive le système des relations entre les éléments de la solution.

III. Problèmes de communication

1. L'indexicalité

L'indexicalité recouvre l'ensemble des déterminations qui viennent lier un énoncé à ses conditions de production :

Cela désigne donc l'incomplétude naturelle des mots, qui ne prennent leur sens "complet" que dans leur contexte de production, que s'ils sont "indexés" à une situation d'échange linguistique (A. Coulon, 1987 : 29).

En changeant de domaine de compétence, d'expert ou même de situation d'entretien, on introduit une variation dans l'indexation de l'information livrée. La prise en compte de l'indexicalité ne dépassera pourtant pas l'horizon du domaine. La prise en compte explicite de changements d'experts ne sera assurée que dans le cas de changements de domaines, ou de points de vue typés (ci-dessous).

La négation de l'indexicalité sera accentuée par l'anonymat des dépouillements, l'analyste étant coupé des interactions de l'entretien et bien incapable de présumer quoi que ce soit de la biographie ou des intentions du locuteur.

2. La "documentation" des entretiens

Au fil de l'entretien, l'expert présuppose un ensemble de connaissances de sens commun détenu par l'analyste. Il réévalue éventuellement

(généralement à la baisse) ce corps de connaissances au cours de l'entretien, en tenant compte de la pertinence des questions posées.

La “méthode documentaire d'interprétation” de Garfinkel (1967) cherche à identifier les “patterns” qui sous-tendent les échanges apparents, à clarifier l'interaction entre questions et réponses, à reconstruire et comprendre la façon dont les réponses (ici les réponses de l'expert) tentent de prolonger la signification accordée aux questions posées. Toutes les présuppositions dynamiques des interlocuteurs “documentent” l'échange, et c'est l'ensemble des paroles effectivement dites et des documentations de l'échange qui constitue finalement la communication. La méthode documentaire est souvent renforcée sur le terrain par une approche interactive (enregistrement vidéo des entretiens, projection des enregistrements aux experts, enregistrement des commentaires faits au cours de ces projections, cf. e.g. Pinsky et Theureau, 1985).

Dans la perspective de la méthode documentaire, les questions sont des ressources utilisées par l'expert pour produire son discours, et la logique de l'échange est participative. Il n'est pas sûr que cette logique soit celle des premiers entretiens, ceux qui précisément fournissent, à travers leur transcription, l'essentiel du corpus. Au cours des premiers échanges, l'expert est invité à “vider son sac”, à partir d'une interrogation suffisamment générale de l'analyste pour qu'elle affiche sans ambiguïté une ignorance à peu près totale du sujet traité. Les propriétés de redondance et de clôture du discours, qui sont des traits fondamentaux de son fonctionnement (Greimas, 1966 : 143), font que le discours de l'expert tend très vite à se fermer sur lui-même. Le long monologue de l'expert n'est ponctué que par de courtes interventions de l'analyste qui ne remplissent qu'une fonction phatique :

(BNP 1990) : début de l'entretien hors enregistrement; l'analyste se présente et demande à l'expert de lui expliquer son activité ; début de la transcription :

expert : Disons sur le plan fonctionnel, je ne veux pas entrer dans les détails informatiques, ça change.

analyste : Oui, tout à fait, oui.

expert : donc la situation ...

[suivent 47 lignes de monologue]

analyste : Bien sûr.

expert : Donc la règle ...

[suivent 79 lignes de monologue]

expert : ...Donc, maintenant, est-ce que vous souhaiteriez qu'on parle des contrôles ?

analyste : Oui, tout à fait.

expert : Donc, la situation...

[suivent 68 lignes de monologue, etc.]

Dans ce type d'entretien, on ne retrouve rien du jeu suivi des paires adjacentes de questions-réponses qui constitue habituellement la trame d'un entretien participatif (A. Coulon, 1987 : 68). Même lorsque les interventions de l'analyste se font plus pressantes, elles semblent davantage représenter des obstacles à éliminer que des ressources à consommer pour relancer l'interaction. De même, lorsque deux experts entrent en compétition, l'expert qui arrive à conserver une position dominante ignore purement et simplement les interventions du second expert :

E1. : On va lui dire : d'accord, vous serez transporté dans le train que vous avez demandé, ou dans un train de plus ou moins 2 heures. On fait des propositions, on essaye....

E2. : Ou on crée des trains spéciaux. Quand on aura beaucoup de demandes, on peut créer des trains spéciaux directs.

E1. : Mais à ce niveau là, si vous voulez, on ne lui donne qu'une indication. On lui dit : vous serez transporté, soit dans le train que vous avez demandé, soit dans un train dans une solution très proche, mais on ne donne pas ce que l'on appelle des références de places, on ne lui dit pas : "Vous serez dans telle voiture, à telle place".

C'est la gare qui transcrit les références sur la fiche de voyage du client. Pourquoi on fait cela : c'est pour que les clients aillent bien chercher leur billet, car si on mettait les références des places, on peut supposer que dans certains cas, ils pourraient voyager sans billet. C'est un garde fou.

E2. : Je vais apporter un petit complément. Lorsque nous donnons au client l'engagement de transport, nous disons dans le train demandé ou dans un train de dédoublement. Avant de forcer un train d'une voiture ou de faire dédoubler un train, on va essayer de les placer dans un autre train. Notre rôle, c'est d'essayer d'orienter les groupes bien managés dont on connaît les responsables, vers des trains de dédoublements.

E1. : Ce que je vous ai donné tout à l'heure comme processus, c'est le processus en période normale, creuse. Il est certain qu'en période de pointe, et là nous allons en parler, un travail un peu différent.

Ces exemples ne remettent certainement pas en cause le bien-fondé des travaux sur l'interaction, mais ils en montrent probablement les limites dans le contexte d'un recueil d'expertise : la passivité de l'analyste, sa méconnaissance affichée du sujet, le caractère technique (non passionnel) du sujet abordé limitent l'interaction et donnent à l'échange une orientation univoque.

IV. Problèmes d'identification

1. Le corpus

Le corpus est constitué par l'ensemble des entretiens réalisés avec un ou plusieurs experts, dans un ou plusieurs domaines. Il s'agit d'un espace de travail dont les limites sont décidées par l'analyste.

La sobriété et la rigueur logique de la définition ne font, en somme, que masquer le caractère intuitif des décisions que le descripteur sera amené à prendre à cette étape

de l'analyse. Un certain nombre de précautions et de conseils pratiques doivent donc entourer ce choix, afin de réduire, autant que possible, la part de subjectivité qui s'y manifeste. On dira qu'un corpus, pour être bien constitué, doit satisfaire à trois conditions : être représentatif, exhaustif et homogène (Greimas, 1966 : 143).

Représentativité et homogénéité sont liées à des critères non-linguistiques qui déplacent le problème vers des phases de projet antérieures au recueil proprement dit : la représentativité du corpus est liée à la représentativité de l'expert, et l'homogénéité est fonction des objectifs assignés au projet complet.

L'exhaustivité est par contre liée aux techniques de constitution du corpus. L'exhaustivité érudite est incompatible avec les très courtes durées des périodes de recueil (au mieux quelques semaines), et on parlera plutôt d'exhaustivité d'un sous-ensemble représentatif. Le comportement du modèle construit à partir de ce segment sera validé dans les phases de recette du projet, et le contenu de la base de connaissances pourra être étendu dans des phases ultérieures de maintenance. L'expérience montre qu'en raison des propriétés de redondance et de clôture “naturelles” du discours déjà évoquées, le sous-ensemble des verbalisations recueillies au cours des entretiens est potentiellement représentatif :

En se prolongeant, le texte devient non seulement de plus en plus redondant et introduit de moins en moins d'information, mais, du fait de la redondance des structures préférentielles, il développe en même temps un sous-code autonome. Cette clôture du texte par l'épuisement de l'information lui confère son caractère idiolectal : en effet, les dénominations contenues dans le texte sont déterminées par les définitions qui y sont présentes et uniquement par elles, de sorte que le texte constitue un micro-univers sémantique fermé sur lui-même. Cette propriété sémantique du discours rend légitimes les descriptions partielles, en établissant une sorte d'équation entre les textes finis et les univers signifiants clos (Greimas, 1966 : 93).

Cette potentialité de la représentativité deviendra réalité après que des entretiens complémentaires de qualification de la base aient permis de

croiser systématiquement les entrées du dictionnaire et d'épuiser les contenus restés implicites dans la surface des premiers entretiens libres.

2. Les techniques de fragmentation

La première tâche de l'analyste consiste à fragmenter ce texte de surface pour le déstructurer et le structurer en accord avec les normes de représentation fixées par la syntaxe de la méthode. Les opérateurs élémentaires sont la permutation et la distribution. Le principe de l'analyse distributionnelle consiste à regrouper dans une seule classe d'équivalence les éléments qui ont des distributions semblables, c'est-à-dire les éléments qui se rencontrent dans des environnements équivalents (Harris, 1951). L'extraction de la base de connaissance à partir d'un ensemble d'énoncés relève des mêmes principes. La familiarité que l'analyste entretient avec son corpus lui permet de choisir les équivalences pertinentes (la dimension "émique" de Pike) et de réduire toutes les occurrences à un petit nombre de propriétés regroupées sur des objets. Implicitement, l'analyste considère qu'il lui est possible de normer le corpus en plaçant régulièrement les unités lexicales qui l'intéressent en position de sujet, et en regroupant dans ce contexte l'ensemble des prédicats rencontrés dans le texte, puis en réduisant les différentes formes d'un seul prédicat à une seule formulation.

On retrouve ces notions de permutation et de distribution dans l'analyse tagmémique de Pike (éd. 1967) et dans l'ethnologie du langage de Hymes (1964). Ces auteurs élargissent simplement le champ des fragmentations en se donnant le droit de segmenter et de permuter des unités de taille et de niveau linguistique quelconque, du phonème au texte.

Le modèle du tagmème, découpé par permutation et défini par l'ensemble de ses distributions, est le modèle linguistique sous-tendant le concept informatique d'hyper-texte (Nelson, 1966, 1980), bien que Nelson, et l'ensemble des outils d'aide à l'acquisition qu'il inspire actuellement (e.g. Anjewierden, 1987) semblent ignorer totalement les modèles linguistiques de fragmentation.

L'intérêt du modèle de l'hyper-texte est de proposer un ensemble de liens pratiques pouvant être établis par l'utilisateur entre des fragments qu'il découpe dans un texte de surface. L'absence d'un modèle linguistique interdit toutefois à ce type d'éditeur de dépasser le simple éditeur de liens textuels, et en particulier :

- de découper automatiquement des fragments ;
- d'indexer automatiquement des fragments sur d'autres fragments, des éléments d'un dictionnaire, ou des éléments d'une structure de représentation disponible dans un autre éditeur ;
- d'analyser la structure des fragments.

Nous avons souligné, lors du bref exposé des principes de l'organisation sémantique de la base de connaissances, l'imbrication transitive des différents types de références croisées. Ces références, de fragment à fragment, d'entité formelle à entité, d'entité à structure, et de structure à structure, permettent d'organiser la navigation entre les niveaux de constitution du modèle de spécification.

3. Les marqueurs

L'identification des règles constitue de loin le problème le plus épineux du modèle pratique. C'est pourtant paradoxalement celui qui semble le plus familier aux concepteurs de systèmes experts. En fait, et même s'il est possible de représenter sous une syntaxe de règles n'importe quel type d'énoncé, qu'il s'agisse d'un ensemble de données, d'un traitement ou d'une interprétation, il est beaucoup plus difficile de cerner le contenu proprement déclaratif d'un énoncé.

Dès que l'on passe de l'unité textuelle découpée dans l'entretien à sa description formelle, on se heurte au problème du découpage interne de ses composants et de l'identification des marqueurs qui en articulent le sens :

Problème qui ne serait pas difficile si nous travaillions sur un langage logique : un langage logique est en effet construit de façon telle que l'on peut, par simple inspection d'une formule, savoir, sans contestation possible, ce qui est relié par les connecteurs utilisés dans cette formule. D'une part, parce qu'un connecteur relie toujours deux segments de la formule où il apparaît ; d'autre part, parce que des règles explicites permettent de déterminer quels sont ces segments. Ainsi, dans la notation la plus habituelle du calcul propositionnel, on trouve toujours, immédiatement à droite d'un connecteur comme et, une et une seule suite de symboles qui constitue une proposition, et de même à gauche : ce sont ces deux suites, mécaniquement repérables, qui sont conjointes par le et. Or ni l'une ni l'autre de ces exigences ne sont satisfaites par les langues naturelles (Ducrot, 1980 : 15).

Malgré ces mises en garde, il semble tentant de prendre pour point de départ de l'identification des connecteurs logiques. La paire “si... alors” semble la plus évidente, et constitue pourtant un redoutable faux ami :

(1) S'il s'agit d'une pompe primaire, alors elle est raccordée au circuit primaire.

(2) Si le dossier est arrivé, alors il faut le traiter.

(3) Si le client envoie une lettre de réclamation, alors il faut réexaminer son dossier.

(4) Si des creux de tension sont constatés, alors il faut blinder les fils.

La première phrase énonce un critère d'appartenance à la classe des pompe primaires; cette propriété sera représentée dans le modèle pratique des données.

Les énoncés 2 et 4 semblent très proches : ils sont tous deux de la forme : si événement (un changement d'état relevé dans l'environnement) alors action. La formalisation devra prendre en compte le contexte implicite qui sous-tend l'action prescrite :

(2) événement et contexte sont identiques, le fait que le dossier soit arrivé implique directement le déclenchement de l'action (tout au moins cette implication peut directement s'énoncer comme un principe de sens commun), et il existe une sorte de continuité de l'action entre la réception

du dossier et son traitement; cet énoncé sera formalisé dans le modèle pratique des traitements. Dans ce cas, la relation si ... alors souligne la nécessité de la consécution des actions dans la chaîne de traitement.

(4) événement et contexte ne sont pas identiques, il manque le maillon de l'interprétation pour comprendre la liaison qui est établie entre les creux de tension et le blindage. Si la phrase est complétée en : (4 bis) *si des creux de tension sont constatés, alors il s'agit probablement d'un court-circuit et il faut blinder les fils*, il devient clair que l'action du blindage trouve sa justification dans le contexte reconnu à partir du constat des creux de tension. Il n'existe pas de continuité entre un processus physique qui entraîne des creux de tension et la décision de blinder ; la nécessité de l'enchaînement est entièrement conditionnée par la capacité de l'expert à formuler une hypothèse explicative, et indirectement par la justesse de cette hypothèse. Cette fois, le contenu déclaratif de cet énoncé sera formalisé dans une règle.

A contrario, les énoncés 2 et 3 semblent pouvoir être représentés tous deux à partir du modèle des traitements ; tout dépend du lien qui peut être établi entre la lettre de réclamation et le réexamen du dossier.

Si les différents types d'incidents qui peuvent intervenir dans le traitement d'un dossier sont connus d'un expert et si l'ajustement des conduites est à apprécier dans le contexte de chacun de ces types, nous sommes ramenés à la situation du court-circuit et nous formaliserons l'énoncé par une règle :

si incident X
alors lettre de réclamation
et réexamen du dossier

Si la réception d'une lettre de réclamation constitue la forme ordinaire de rattrapage du dossier, sans préjuger de la situation qui pourra être interprétée lors du réexamen, alors on peut formaliser cet énoncé comme un traitement.

Les opérateurs logiques remplissent d'ailleurs d'autres fonctions que la connexion logique entre propositions, fonction argumentative ou même phatique (support et relais de la communication) :

ce qu'on peut faire là, c'est aller regarder sur la pompe si on les trouve ; donc sur la pompe on les a pas ; c'est net donc

Le premier “donc” de cette phrase sert à ponctuer une série de remarques et d'observations, tandis que le second accompagne logiquement la conclusion.

4. Extraction et normalisation

Le “texte”, qui constitue finalement la référence cumulée des entités extraites en jouant sur les marques utilisables et les gabarits pré-définis, se présente comme une série de propositions protocolaires. Ce point devient évident lorsqu'on édite après coup les références au texte associées aux entrées du dictionnaire, en prenant pour fil conducteur l'organisation interne de ce dictionnaire. L'information retenue se plastifie dans la formalisation proposée en éliminant les marques parasites de l'énonciation. L'exemple suivant (Vogel, 1988a) montre comment la charge péjorative des parasites “mauvaise”, “foiré”, “foireuses”, “éliminer”, “résidus”, etc., est réduite au profit d'une information neutralisée :

Là aussi c'est le cas typique routine d'un lancement. Dans un cas de routine, on connaît bien l'orbite, puisqu'on l'avait calculée la veille donc on l'a bien. Donc si une mesure est mauvaise, c'est que la station a foiré pratiquement. Par contre dans le cas lancement qui nous intéresse, puisque le système expert, c'est plutôt pour les lancements, si tu lances le calcul d'orbite et qu'il y a des mesures foireuses dedans, il peut très bien arriver le phénomène suivant : c'est que tu n'arrives pas à les éliminer, et que tu as des gros résidus partout. C'est-à-dire que tu essaies de passer dans toutes les mesures, tu y arrives, parce que certaines sont foireuses, mais comme tu

n'as pas identifié ni éliminé les mauvaises, tu élimines rien, mais ici tu te retrouves avec des résidus énormes de l'ordre de ... Au lieu de 33 mètres, tu te retrouves avec 200, 300 mètres. Mais à ce moment là, c'est là qu'au niveau des diagrammes, tu n'as plus ici que des étendues; tu as tout un passage relativement concentré en un point, et tu as des points comme ça un peu partout dans ton orbite. Ce sont des points relativement concentrés. Alors quand ça arrive, en général, on identifie vite un problème de mesures, que la restitution proprement dite n'a pas réussi à éliminer, parce que ça s'est mal passé, et elle a trouvé une orbite moyenne qui passait un peu dans tout mais qui n'est pas la bonne. Alors quand c'est ça, en général, on revient à SEMEL et on trie station par station. On va supposer que telle station est mauvaise, ou que c'est elle qui a fait de mauvaises mesures, on l'élimine et on recommence avec uniquement 2 stations par exemple, ou 3 stations si on avait 4 et on regarde ce qui se passe (document CNES).

(i) Là aussi c'est le cas typique routine d'un lancement

On peut caractériser “le cas typique routine” comme un objet, si on considère l'ensemble de ses traces matérielles (diagrammes, etc.), ou comme un schéma. Lorsque le cas étudié est une instanciation d'un cas qualifié de “routine”, le système expert devra effectuer la même reconnaissance d'un contexte connu et inscrire ensuite ses actions à l'intérieur de ce contexte identifié :

objet ou schéma : le cas typique routine d'un lancement

(ii) Dans un cas de routine, on connaît bien l'orbite

Ce schéma d'interprétation associe une situation connue (le “cas de routine”) à une propriété manifeste, la bonne connaissance de l'orbite :

schéma : un cas de routine

on connaît bien l'orbite

(iii) *puisque'on l'avait calculée la veille donc on l'a bien*

L'action “calcul” de l'orbite implique un état de connaissance de l'orbite. La relation d'antécédence causale est marquée par “puisque” et “donc” :

action : calcul

destinateur : ?

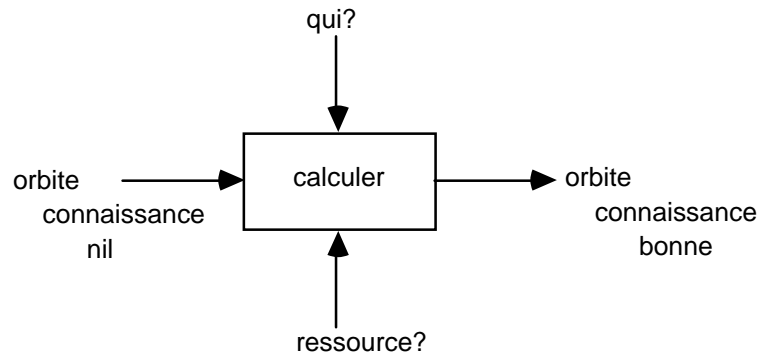
destinataire : orbite

état antérieur : <connaissance, nil>

état postérieur : <connaissance, bonne>

instrument : ?

temps : la veille



(iv) *Donc si une mesure est mauvaise, c'est que la station a foiré pratiquement*

Ce schéma associe le contexte “la station a foiré” à la propriété “une mesure est mauvaise”. Ce schéma est construit hors contexte sur la forme :

contexte implique manifestation prévisible,

et il est utilisé en contexte par l'expert sous la forme :

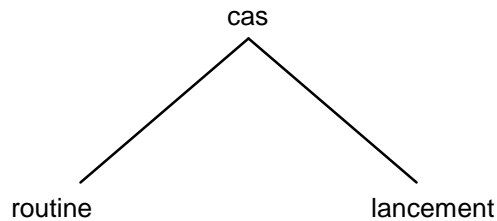
signe induit phénomène explicatif.

schéma : la station a foiré

une mesure est mauvaise

(v) *Par contre dans le cas lancement qui nous intéresse*

Le “cas lancement” est un autre type de cas, on peut identifier ce nouveau contexte et commencer à assembler une taxinomie des cas (schémas ou objets) connus de l'expert :



Ce début de classification devra être consolidé plus tard puisque la place du “cas lancement qui nous intéresse” par rapport au “cas de routine” n'est pas très claire ;

objet ou schéma : cas de lancement

(vi) *puisque le système expert, c'est plutôt pour les lancements*

Cet énoncé peut être caractérisé comme un schéma, mais il s'agit d'un schéma sans rapport étroit avec le thème central, qui joue plus ici le rôle d'un commentaire. Ce type de schéma est écarté.

(vii) *Si tu lances le calcul d'orbite et qu'il y a des mesures foireuses dedans, il peut très bien arriver le phénomène suivant*

Il s'agit d'un schéma articulé avec la sortie de l'action “calcul d'orbite” déjà définie ci-dessus; le “phénomène suivant” va lister les signes prévisibles :

schéma : il y a des mesures foireuses dedans (le calcul d'orbite)

(viii) *c'est que tu n'arrives pas à les éliminer, et que tu as des gros résidus partout*

La suite de l'entretien complète le schéma; l'action négative “tu n'arrives pas à les éliminer” n'est pas prise en compte comme une action :

*schéma : il y a des mesures foireuses dedans (le calcul d'orbite)
tu n'arrives pas à les éliminer
tu as des gros résidus partout*

(ix) *C'est-à-dire que tu essaies de passer dans toutes les mesures*

L'action "passer dans toutes les mesures" est partiellement décrite; la formule "c'est-à-dire" permet de penser que cette action concerne élimination et résidus, donc un trait sémantique "qualité" des mesures; la qualité de sortie demande à être précisée :

*action : passer dans toutes les mesures
destinateur : ?
destinataire : mesures
état antérieur : <qualité, nil>
état postérieur : <qualité, ?>
instrument : ?
temps : la veille*

(x) *tu y arrives, parce que certaines sont foireuses*

"sont foireuses", introduit par "parce que", est interprété comme une contrainte sur l'action "passer dans toutes les mesures" :

*action : passer dans toutes les mesures
contrainte : certaines sont foireuses*

(xi) *mais comme tu n'as pas identifié ni éliminé les mauvaises, tu élimines rien, mais ici tu te retrouves avec des résidus énormes*

Cet énoncé permet de reprendre la formulation du schéma précédent en associant "il y a des mesures foireuses dedans (le calcul d'orbite)" et "tu n'as pas identifié ni éliminé les mauvaises", et en isolant "tu te retrouves avec des résidus énormes" comme une propriété manifeste introduite par "comme" et "mais" :

schéma : il y a des mesures foireuses dedans (le calcul d'orbite) et tu n'as pas identifié ni éliminé les mauvaises tu te retrouves avec des résidus énormes

(xii) de l'ordre de ... Au lieu de 33 mètres, tu te retrouves avec 200, 300 mètres

Le résidu qui est observable est un objet, la quantité du résidu, normale ou “énorme”, est évaluée ; la référence de 33 demande à être validée;

objet : résidu

quantité normale 33 mètres

quantité énorme 200, 300 mètres

(xiii) Mais à ce moment là, c'est là qu'au niveau des diagrammes, tu n'as plus ici que des étendues

Ce que l'on observe “au niveau des diagrammes” est une autre manifestation des mauvaises mesures qui n'ont pas été éliminées; cette manifestation est introduite par le marqueur “à ce moment-là” qui note plus la consécution des observations que la conséquence “des résidus énormes” ; de plus, ces résidus, étant observables, ne peuvent avoir un statut d'explication :

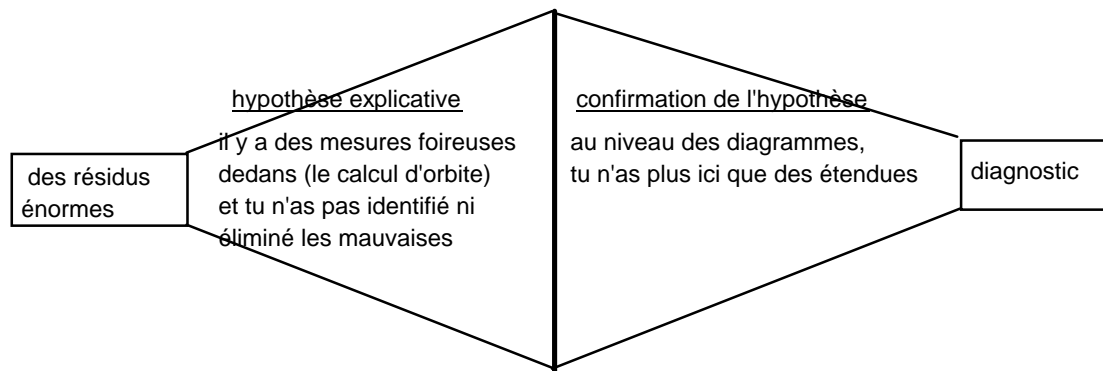
schéma : il y a des mesures foireuses dedans (le calcul d'orbite)

et tu n'as pas identifié ni éliminé les mauvaises

tu te retrouves avec des résidus énormes

au niveau des diagrammes, tu n'as plus ici que des étendues

La consécution des observations correspond, dans la stratégie du diagnostic mené en contexte, à une amplification des signes observés à partir d'une interprétation admise pour expliquer les premiers signes rencontrés :



Génération et confirmation de l'hypothèse

(xiv) *tu as tout un passage relativement concentré en un point, et tu as des points comme ça un peu partout dans ton orbite. Ce sont des points relativement concentrés*

Les “étendues”, pour être qualifiées dans le diagramme, introduisent les objets “passages” et “points” caractérisés par une propriété de concentration :

objet : passage

concentré relativement en un point

objet : point

position un peu partout dans ton orbite

concentrés relativement

(xv) *Alors quand ça arrive, en général, on identifie vite un problème de mesures*

En contexte, la caractérisation des “étendues” sous leurs différentes formes permet d'identifier “un problème de mesures”.

(xvi) *que la restitution proprement dite n'a pas réussi à éliminer, parce que ça s'est mal passé*

Une action positive, “la restitution”, est énoncée :

action : restituer

destinateur : ?

destinataire : mesures

état antérieur : <qualité, ?>

état postérieur : <qualité, problème éliminé>

instrument : ?

contrainte : à préciser à partir de “ça s'est mal passé”

Lorsque cette action échoue, dans des conditions à préciser, l'état de sortie des mesures se manifeste par les “résidus” et les “étendues”.

(xvii) et elle a trouvé une orbite moyenne qui passait un peu dans tout mais qui n'est pas la bonne

La restitution cause également un changement d'état sur l'orbite qui devient “trouvée”, cette orbite étant caractérisée par deux attributs qui la rendent acceptable, et une restriction introduite par “mais” :

action : restituer

destinateur : ?

destinataire : orbite

état antérieur : <existence, nil>

état postérieur : <existence, trouvée>

instrument : ?

objet : orbite

qualité moyenne

passé un peu dans tout

validité n'est pas la bonne

(xviii) Alors quand c'est ça, en général, on revient à SEMEL {SEMEL signifie SElections des MEsures de Loc} et on trie station par station

“quand c'est ça” fait référence à l'interprétation complète de la situation par le schéma : il y a des mesures foireuses dedans (le calcul d'orbite) et tu n'as pas identifié ni éliminé les mauvaises, qui déclenche la succession de deux actions : revenir à SEMEL et trier station par station :

action : revenir à SEMEL

destinateur : ?
destinataire : données considérées
état antérieur : <source, ?>
état postérieur : <source, SEMEL>
instrument : ?
action : trier station par station
destinateur : ?
destinataire : <données, source, SEMEL>
état antérieur : ?
état postérieur : ?
instrument : ?

(xix) On va supposer que telle station est mauvaise, ou que c'est elle qui a fait de mauvaises mesures, on l'élimine et on recommence avec uniquement 2 stations par exemple, ou 3 stations si on avait 4 et on regarde ce qui se passe

L'action "trier par station" qui était très générique, est maintenant expansée et précisée par les actions "éliminer", "recommencer", "regarder ce qui se passe" :

action : éliminer
destinateur : ?
destinataire : station
état antérieur : considérée
état postérieur : éliminée
instrument : ?
contrainte : supposer que telle station est mauvaise

"Recommencer" et "regarder ce qui se passe" relancent explicitement une interprétation par le schéma : il y a des mesures foireuses dedans (le calcul d'orbite) et tu n'as pas identifié ni éliminé les mauvaises; on peut supposer un déclenchement récursif de "trier par station" si le diagnostic réussit. Pour qualifier complètement cet

algorithme, il faudrait encore définir la conduite à tenir lorsque le schéma se déclenche avec succès sur la dernière station.

V. Problèmes de référenciation

1. La référence au modèle

Lorsque l'expert interprète les phénomènes qui le concernent, il se situe tour à tour dans deux systèmes de référence. Le premier est directement porteur de l'interprétation, c'est celui des "schémas d'interprétation" (Garfinkel, *ibid.*), totalités extratemporelles rassemblant des pans complets d'expérience du sujet, et simplifiant le réel pour n'en retenir que les liaisons logiques essentielles. Le second sert de justification au premier, c'est celui des formes causales de l'action, englobant ses dimensions spatio-temporelles :

L'aspect causal de l'action propre englobe ses dimensions spatio-temporelles, ses vitesses et son dynamisme, tandis que les liaisons logico-mathématiques font abstraction de ces conditions physiques pour ne retenir que la forme de ces coordinations [...] Ces différenciations et ces coordinations, soit opératoires, soit causales, aboutissent à une dissociation progressive des plans : celui du réel, donc des contenus et des objets, et celui des formes opératoires du sujet, construites de manière à n'être jamais plus contredites par les faits, ce qui revient en définitive à les situer au palier des liaisons hypothético-déductives, donc des liaisons directes et extratemporelles entre le possible et le nécessaire (Piaget, 1971 : 123).

La causalité est caractérisée de façon canonique (ou "prototypique", Lakoff et Johnson, 1985 : 80) par un complexe de propriétés : pouvoir de décision exercé par un sujet utilisant l'énergie d'une ressource pour transformer un destinataire, lien spatio-temporel entre le sujet, la ressource et le destinataire.

Pour produire l'interprétation, l'expert se dégage de la succession causale pour considérer des formes stables de représentation, mais ces

formes restent déterminées et légitimées par la dynamique des processus explicatifs qui leur servent de référence :

schéma d'interprétation (format K.O.D.)

SI

couplage *perturbation d'un fil de contrôle commande par*
coactif ou inductif

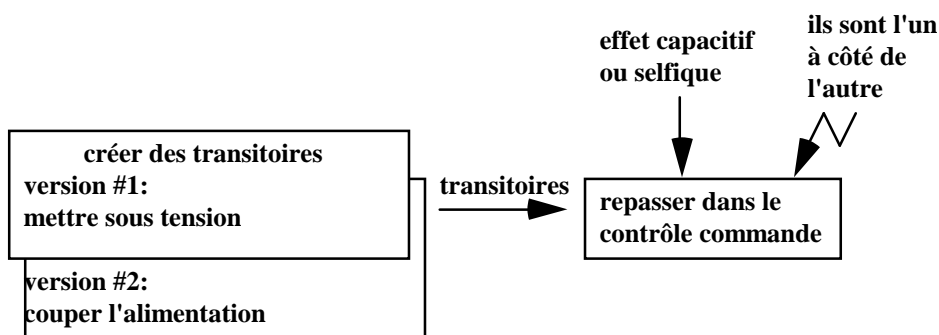
ALORS

fil d'alimentation relié à alimentation

alimentation mise sous tension

fil de contrôle commande perturbé par des volts

Autre exemple : quand vous mettez en tension et quand vous coupez l'alimentation vous créez des transitoires très violents qui vont repasser dans le contrôle commande par effet capacitif ou selfique s'ils sont l'un à côté de l'autre (EDF, 1990). Actinomie (format K.O.D.) :



On pourrait très bien imaginer que le schéma ci-dessus soit complété par une action “éloigner l'alimentation du contrôle commande” ou “isoler les composants”. Ces actions seraient alors à leur tour déclinées dans des actinomies précisant phases successives du processus d'isolation, ressources afférentes, etc. Lorsque le système expert pourra diagnostiquer cette situation, il conseillera d'exécuter ces consignes et pourra indiquer leur mode d'exécution. Il s'agit bien d'actions *exécutable*s. La formalisation du processus de création des transitoires et de perturbation par effet selfique utilise la même grammaire : des formes causatives organisées autour d'un complexe de

destinateur, destinataire et instrument, mais *le résultat de cette formalisation n'est jamais exécutable*. Il s'agit cette fois d'une explication qui constitue dynamiquement, dans un monde théorique de référence, la justification causale du schéma utilisable au premier degré, dans le monde réel.

Les recherches qui s'intéressent aux raisonnements qualitatifs portant sur les systèmes physiques se placent dans une perspective exactement convergente :

It was extremely surprising that LOCAL, augmented with only 39 rules, would be able to localize the same set of faulty components as would a human expert. Subsequently, we analyzed each of the rules to understand what model-based reasoning was missing. Consider one such rule: 'If transistor Q5 is off, then the external symptom cannot be caused by the voltage reference being low'. The rationale behind this rule is as follows (notice the qualitative form of the argument, not the content):

The power supply outputs a voltage and current limited by the minimum of its two control settings. This minimum is computed through a feedback path in which transistor Q5 is a central part. If the voltage reference were low enough to cause a symptom, then Q5 would be on, so there is no way that a low voltage reference could be contributing to the power supply's faulty behavior if Q5 were off. If Q5 were off, then the output might be high, but the output could not be too low because the reference was too low.

Notice that this argument is both qualitative (e.g., too high and too low) and reliant on an underlying causal model of the feedback operation of the power supply (de Kleer, 1990 : 5).

Modélisation par la physique qualitative ou par les règles expertes sont souvent présentées comme des stratégies alternatives dans le développement des systèmes experts. Il s'agit en fait de deux attitudes mentales complémentaires, l'une directement liée à la pratique du diagnostic, l'autre à la référence documentaire du spécialiste.

2. La référence à l'exemple

Il arrive fréquemment que l'expert s'explique en faisant référence à un exemple. Les recherches conduites sur l'apprentissage automatique (après Michalski, 1983) ont popularisé les techniques de généralisation inductive de catégories et de classification automatique. Ces méthodes s'appliquent généralement sur des séries homogènes de propriétés. Les exemples rencontrés "naturellement" dans les entretiens constituent un mode d'énonciation en soi, et la généralisation qu'ils appellent peut porter aussi bien sur une série de propriétés, sur une procédure, sur une contrainte ou sur un schéma d'interprétation : tous les types d'entités constitutifs des modèles K.O.D. peuvent être énoncés sous forme d'exemples.

(SNCF, 1990) : *Alors qu'est ce qu'on peut avoir comme régimes de circulation ?*

Et bien on peut avoir ce qui est le plus simple, c'est-à-dire que le train est quotidien et il roule tous les jours. Lundi, mardi, mercredi, jeudi, vendredi, samedi, dimanche. Alors là c'est super, si on avait que des trains comme cela, je dirais que c'est... Là il y aurait un suivi ou un montage qui serait facile à faire.

Où cela commence à se corser, c'est quand on veut toucher à ce régime de circulation, et je vais vous en donner un, brut de fonderie, comme cela, on l'analysera après, je prends le 604. Le 604, alors c'est un Lyon Perrache, Lyon Part-Dieu, à un certain moment ou c'est un Lyon Perrache, Lyon Part-Dieu, Paris à d'autres...

Alors en ce qui concerne le parcours de Lyon Perrache à Lyon Part-Dieu, en ce qui concerne le service d'été 90, il sera, sauf samedi jusqu'au 6 juillet et il ne circulera pas le 3 juin, il sera sauf samedi, dimanche et jours de fêtes du 9 juillet au 27 juillet, il sera sauf samedi à partir du 3 septembre et il circulera les 28, 29 et 30 août 90.

Et sur le parcours Lyon Part-Dieu Paris, il sera les dimanches jusqu'au 1er juillet, il circulera le 4 juin 90, il ne circulera pas le 3 juin 90, il sera sauf samedi, dimanche et fêtes du 9 au 27 juillet, il sera les dimanches à partir du 9 septembre

90, et il circulera le 28, 29 et 30 août 90. On a gagné le cocotier avec un train comme cela, on est sûr de s'amuser un sacré moment.

Mais si il n'y avait que cela...

C'est que les trains, en règle générale, devraient rouler tous seuls.

Alors un train TGV est composé soit d'un élément, deux motrices, huit remorques, 400 mètres plus quelques centimètres, soit lorsque la nécessité l'impose, que ce soit en service régulier, c'est-à-dire prévu à l'avance ou en service facultatif c'est-à-dire quelque chose qui vient se rajouter par suite d'une demande, on peut lui adjoindre un deuxième élément.

Donc, nous avons deux motrices, huit voitures, et accrochées derrière cela, deux autres motrices, huit voitures, le tout jumelé ensemble, tracté par le même mécanicien.

Jusque-là tout va bien, mais c'est que la plupart de ces trains, ils ont un élément, ils ont un deuxième élément mais ce deuxième élément a une gare qui sera située entre A et B, il va quitter le premier et partira dans une autre direction.

Donc, on a un train comme le 604, avec un régime de circulation qui est assez touffu. Et le 604, et bien on a de la chance, il a un autre train qui est derrière et celui qui est derrière, il s'appelle le 730, donc le 730 c'est soit Dijon, soit Grenoble, soit St Etienne. Bon là moi je sais que c'est un Grenoble.

Il faudra donc pour avoir la physionomie exacte de circulation du 604 et du 730, faire un mélange des deux ensembles de périodicité des deux trains sur les deux parcours, Grenoble Lyon Part-Dieu, Lyon Part-Dieu, Lyon Perrache, pour avoir, je dirais, le train qui roule complètement à chaque fois qu'il était prévu initialement de le faire rouler.

Alors il y a des trains qui sont..., moi je vous dis, j'avoue qu'à un certain moment, je fatigue énormément et dans cette période, il ne s'agit pas de m'agacer par un coup de téléphone, autre chose, parce que je perds simultanément tout ce que j'avais essayé de comprendre et je suis obligé de repartir.

Alors comment, comment arriver à faire ce mélange, tout au moins en sortir le produit définitif qui sera exploitable, et bien, par le biais d'un calendrier.

Par le biais d'un calendrier, je fais pour le service d'été par exemple 1990, une copie du calendrier, c'est-à-dire entre mai et septembre. Dessus je positionne mes journées, quels jours c'est. Je positionne mes jours de fêtes parce que ce serait trop beau s'il fallait encore pas tenir compte des jours de fêtes qui vont tomber en milieu de semaine.

Je prends un exemple, le 15 août, il va tomber un mercredi cette année, donc le 15 août on ne sera plus un mercredi, on sera un dimanche et puis le 16 août, on ne sera plus un jeudi, on sera un lundi.

Alors il faut inclure cela dans les périodicités et on tombe sur un document de base qui est comme cela. Cela c'est mon document de travail par excellence au départ.

Les deux exemples donnés par l'expert sont utilisés dans des directions opposées. Le premier, celui des deux trains 604 et 730, sert à mettre en place une règle de conduite qui n'est jamais exprimée explicitement. L'analyste doit donc, pour la faire émerger, généraliser immédiatement l'exemple donné pour en étendre la portée à tous les trains qui se retrouveront dans la même situation. Le second, celui du 15 août, est le résultat de l'application d'une procédure qui vient d'être énoncée à une situation pratique. L'unification de la règle et des dates de l'exemple produit l'exemple proprement dit. Le deuxième exemple est redondant et même parasite par rapport à l'énoncé de la règle, puisqu'il va générer inutilement des références pratiques, le premier exemple est à retenir de plein droit puisqu'il permet de faire émerger la procédure de gestion du train combinant 604 et 730.

VI. Génie linguistique et génie logiciel

Les méthodologies de recueil d'expertise partagent un ensemble de contraintes qui placent leurs développements ultérieurs dans les champs combinés du génie linguistique et du génie logiciel : elles doivent s'inscrire dans un cycle de développement qui n'est pas

toujours bien préparé à les accueillir, elles amorcent le cycle à partir d'une collecte d'informations auprès des experts, elles découpent ensuite les entités utiles dans le texte que ces entretiens produisent, et finalement elles proposent une multitude de facilités d'accès à la structure élaborée qui rend particulièrement complexe le croisement des plans de référence et des points de vue.

Tous les problèmes qui sont évoqués ci-dessus sont critiques mais trouveront certainement, au moins sous la forme de règles de conduite et de gabarits externes, des réponses rapides. Il faudrait encore ajouter à ces questions urgentes mais immédiates les interrogations à plus long terme sur l'automatisation des premiers maillons de la chaîne d'acquisition : indexation automatique d'entités non descriptives, réutilisation de dictionnaires.

Claude VOGEL

CISI Ingénierie

3, rue Le Corbusier, SILIC 232

94528 Rungis Cedex (France)

Références

- COULON, A. (1987). *L'Ethnométhodologie*, PUF.
- de KLEER, J. (1990). "Qualitative Physics, a Personal View", in D.S. Weld et J. de Kleer, édés, *Readings in Qualitative Reasoning about Physical Systems*, Morgan Kaufmann Publishers, Inc.
- DUCROT, O. (1980). "Analyse de textes et linguistique de l'énonciation", in O.Ducrot et al., *Les mots du discours*, Paris, Editions de Minuit.
- GARFINKEL, H. (1967). *Studies in Ethnomethodology*, Englewood Cliffs, Prentice Hall.
- GARFINKEL, H., SACKS, H. (1970). "On Formal Structures of Practical Action", in J.C. McKinney et E.A. Tiryakan (édés), *Theoretical Sociology: Perspectives and Developments*, New-York, Appleton-Century-Crofts, pp. 337-366.
- HARRIS, Z.S. (1952). "Discourse Analysis", *Language*, Vol. 28, pp. 1-30.

- LAKOFF, G., JOHNSON, M. (1985). *Les métaphores dans la vie quotidienne*, Paris, Editions de Minuit.
- MICHALSKI, R.S. (1983). "A theory and methodology of inductive learning", R.S. Michalski, J.G. Carbonell, T.M. Mitchell édés, *Machine Learning: an artificial intelligence approach*, Los Altos, Tioga, pp. 83-129.
- PIAGET, J., GARCIA, R. (1971). *Les explications causales*, Paris, PUF.
- PINSKY, L., THEUREAU, J. (1985). *Signification et action dans la conduite de systèmes automatisés de production séquentielle*. Collection d'ergonomie et de neurophysiologie du travail, CNAM, n° 83.
- PIKE, K.L. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior*, La Haye, Mouton.
- VOGEL, C. (1988a). "Les systèmes experts dans le domaine spatial : l'innovation experte", *Actes du Colloque "Intelligence artificielle et Espace"*, Technospace, Bordeaux.
- VOGEL, C. (1988b). *Génie cognitif*, Paris, Masson.
- VOGEL, C. (1990). *KOD : la mise en œuvre*, Paris, Masson.