



**HAL**  
open science

# RÉDACTION DE SPÉCIFICATIONS FORMELLES : ÉLABORATION À PARTIR DES SPÉCIFICATIONS ÉCRITES EN LANGAGE NATUREL

Alain-Jérôme Fougères, Philippe Trigano

► **To cite this version:**

Alain-Jérôme Fougères, Philippe Trigano. RÉDACTION DE SPÉCIFICATIONS FORMELLES : ÉLABORATION À PARTIR DES SPÉCIFICATIONS ÉCRITES EN LANGAGE NATUREL. In Cognito - Cahiers Romains de Sciences Cognitives, 1997, 1 (8), pp.29-36. hal-00569735

**HAL Id: hal-00569735**

**<https://hal.science/hal-00569735v1>**

Submitted on 25 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# RÉDACTION DE SPÉCIFICATIONS FORMELLES : ÉLABORATION À PARTIR DES SPÉCIFICATIONS ÉCRITES EN LANGAGE NATUREL

**Alain-Jérôme Fougères\* et Philippe Trigano\*\***

\* IUT de Lannion

rue Edouard Branly, 22302 Lannion

Tel : 96 48 57 63 E-mail : fougères@iut-lannion.fr

\*\* Université Technologique de Compiègne

HEUDIASYC-CNRS UMR 6599

B.P. 529, 60205 Compiègne

Tel : 03 44 23 45 02 E-mail : trigano@utc.fr

---

## Résumé

*L'activité de spécification devient considérable ; une multitude de pages sont écrites tous les jours et la plupart du temps en langage naturel. Pour le CNET (Centre National d'Etudes des Télécommunications), qui réalise des études de services et d'équipements de France Telecom, et qui possède la maîtrise des étapes de spécification et de validation, la nécessité de réduire les temps de développement des services est une priorité. Une condition pour atteindre cet objectif consiste à formaliser le maximum de spécifications produites. Dans ce contexte, nous essayerons de montrer, la possibilité d'une certaine automatisation du passage de l'informel au formel, grâce à des méthodes et outils fiables, susceptibles d'assister un expert humain en spécifications. Nous proposons pour cela un processus de formalisation qui s'appuie sur une représentatif intermédiaire des spécifications avec le formalisme des graphes conceptuels, avant de dériver une description formelle en Z de la spécification initiale.*

## 1. Présentation

Pour formuler des besoins liés à la création d'un nouveau service de télécommunication, on utilise généralement le langage naturel. Celui-ci semble bien adapté, puisqu'il est connu de tous les acteurs investis dans le développement du service. Et pourtant, de nombreux problèmes, ayant parfois de lourdes conséquences sur le développement du service et son avenir, apparaissent, révélant une transmission inadéquate des connaissances :

- . problèmes de communication entre le rédacteur du cahier des charges, le donneur d'ordre et le soumissionnaire, dus à une interprétation multiple des spécifications ;
- . manque de rigueur du document ;
- . difficultés d'exploitation pour des traitements automatiques, etc.

Afin de remédier à ces inconvénients, on recommande l'utilisation d'un langage formel pour la spécification [Wirsing 1993]. Et quand l'ampleur de la tâche ne permet pas d'envisager de spécifier formellement la totalité de l'application, on ne peut qu'insister sur les avantages indiscutables que procure la spécification formelle des parties les plus critiques de l'application.

Notre travail porte principalement sur la pertinence d'utiliser des outils de traitement automatique du

langage naturel et de représentation des connaissances dans le processus de formalisation. Le lecteur trouvera dans [Toussaint 1992], [SPECS 1993] et [Biebow et al 1995] [Vadera et Meziane 1994], des approches antérieures de cette problématique. Deux cas de figure sont à envisager :

- . soit on considère la démarche séquentiellement - écriture de la spécification informelle (langage naturel), puis production de la spécification formelle correspondante ;
- . soit on se situe dans un contexte d'élaboration de spécifications dans lequel on respecte une méthodologie interactive de formalisation.

Notre approche du problème s'est trouvée imposée par deux contraintes : le support est fixe, car les spécifications informelles sont déjà rédigées ; d'autre part, nous avons été confronté à l'impossibilité de disposer d'une spécification à la fois dans sa version initiale et dans sa version formalisée<sup>1</sup>. Cependant, même si nos conditions d'études s'en sont éloignées, il est primordial de garder en perspective le second cas de figure, dont le cadre nous paraît excessivement pertinent et prometteur - il est de toute évidence certain que cette élaboration interactive serait la conséquence logique de toute suite apportée à notre travail.

Nous consacrons cet article à l'exposé de nos objectifs d'aide à la formalisation. La section suivante

---

<sup>1</sup>En fait, très peu de spécifications sont formalisées, et, lorsqu'elles le sont, ce n'est souvent que partiellement.

fixe ces objectifs, par la formulation et la discussion de l'ensemble des hypothèses servant de base à notre travail. Dans une troisième partie, nous décrivons le rôle essentiel de chaque phase entrant dans le processus de formalisation tel que nous l'avons échafaudé à l'aide des différentes hypothèses. Cette description couvrira les étapes suivantes :

- . acquisition des connaissances, comme étape préliminaire ;
- . analyse linguistique : constitution de corpus de connaissances lexicales et sémantiques ;
- . représentation des connaissances, pour laquelle nous évoquerons les éléments qualitatifs attendus d'un bon modèle ;
- . formalisation, avec la nécessité de construire une description logique, à partir des éléments de représentation des connaissances, servant de pivot à la traduction vers des langages de descriptions formels.

## 2. Notre approche

### 2.1. L'objectif

L'objet de notre recherche est l'étude de méthodes, techniques et outils permettant d'offrir une aide aux rédacteurs de spécifications de services de France Télécom dans leur activité de formalisation. Un résumé de notre problématique peut se formuler au moyen de l'interrogation suivante :

*comment participer au processus de formalisation (i.e. passage de l'informel au formel), à partir de spécifications écrites en langage naturel ?*

### 2.2. Les hypothèses

Avant de proposer une solution à la question précédente, nous dressons ci-dessous la liste des principales hypothèses qui ont guidé notre approche et nous ont ainsi permis d'élaborer l'architecture d'un système de formalisation de spécifications [Fougères et Trigano 1996].

**1<sup>ère</sup> hypothèse** : *il n'est pas envisageable de traduire directement une spécification informelle en une spécification formelle.*

Cette hypothèse est directement corrélée à la problématique de traduction d'un langage naturel en un langage formel. De plus, l'observation de l'activité de spécification corrobore cette hypothèse, puisqu'il est trivial d'énoncer qu'une spécification formelle s'élabore progressivement (i.e. par *raffinement*).

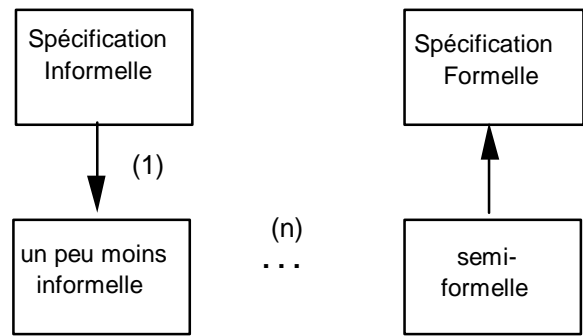


Figure 1 : Méthode de raffinement successif d'une spécification.

**2<sup>ème</sup> hypothèse** : *il est pertinent d'utiliser une représentation intermédiaire (de type graphes conceptuels [Sowa 1984]) pour passer de l'informel au formel.*

Cette hypothèse est induite par l'hypothèse précédente. Nous venons de présenter le paradigme de raffinement pour la construction de spécifications formelles, et aurions pu nous en inspirer, mais celui-ci n'est pas suffisamment défini, pour établir de façon univoque le nombre de passes de raffinement nécessaires pour une spécification donnée. Aussi, nous sommes convaincus que le passage par une représentation intermédiaire, conçue comme un pivot pour la formalisation, répond à notre problématique (fig. 2).

Notre choix pour la représentation des connaissances s'est orienté vers le formalisme des *graphes conceptuels*, considérant qu'il respecte la définition des critères préalablement fixés, notamment : le pouvoir de représentation de la sémantique profonde de phrases écrites en langage naturel, un formalisme proche de la logique des prédicats et une manipulation efficace des connaissances permettant de réaliser des inférences.

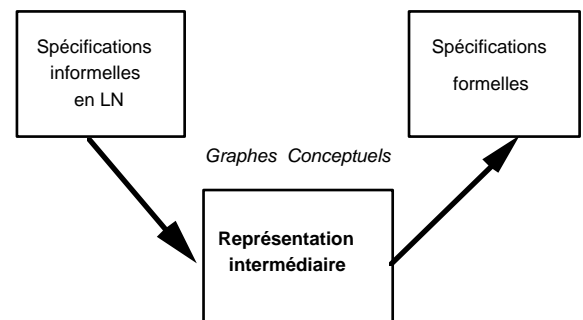


Figure 2 : Illustration de la deuxième hypothèse de notre approche.

Le passage par une représentation intermédiaire ne lève pas pour autant toutes les difficultés énoncées ci-dessus, il permet cependant de mieux les caractériser :

- . en amont de la construction de la représentation intermédiaire : la traduction du langage naturel dans le formalisme de représentation retenu, où

l'on retrouve la majorité des problèmes liés au traitement du langage naturel ;

- en aval de la représentation sémantique intermédiaire : l'interprétation logique du contenu représenté et l'adéquation aux langages de spécifications formelles cibles.

Cette hypothèse ne résoud donc pas notre problème ; la complexité de la traduction d'une sémantique naturelle en une sémantique formelle demeure, mais elle se dédouble, ce qui rend le processus plus facile à maîtriser.

**3ème hypothèse** : construction d'une description logique à partir de la représentation intermédiaire, qui puisse servir de pivot pour la traduction vers les différents langages de spécification formels.

Cette hypothèse est le résultat d'une constatation unanime chez les experts en spécification de protocoles de communications : il est rare, surtout lorsqu'une spécification est volumineuse, qu'un langage de spécification formel se prête correctement à la modélisation de la totalité de la spécification. Si les langages formels s'avèrent inadéquats pour formaliser l'ensemble d'une spécification, cela provient de la spécificité de chacun d'eux qui les rend aptes à modéliser un aspect plus qu'un autre. Il n'est pas interdit de penser qu'un langage futur aura les propriétés de généralité, mais tel n'est pas encore le cas. Une des conséquences de ceci est l'apparition du paradigme de *multi-spécification*, qui consiste à utiliser, dans une même spécification, le langage le plus propice à fournir une description d'un module donné.

Dans notre cas, la construction d'une description logique respecte le paradigme de multi-spécification en offrant la possibilité de traduire son contenu dans le langage formel cible - moyennant la disponibilité du traducteur adéquat. Dans [Fougères 1997], nous donnons les éléments algorithmiques de construction de ce type de traducteur pour le langage Z.

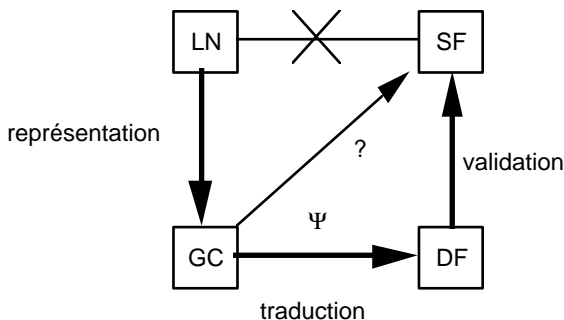


Figure 3 : Synthèse schématique des 3 hypothèses<sup>2</sup>.

<sup>2</sup>LN : Langage Naturel ; GC : Graphe Conceptuel  
 DF : Description Formelle ; SF : Spécification Formelle  
 LP : Logique des Prédicats  
 $\Psi = \phi \circ \phi'$  ;  $\phi : GC \rightarrow LP$  ;  $\phi' : LP \rightarrow DF$

Le schéma précédent visualise la construction de notre conception du processus de formalisation de spécifications en fonction des hypothèses évoquées plus haut. Nous définissons une traduction ( $\Psi$ ), à partir de graphes conceptuels, qui permet d'obtenir une description formelle de ceux-ci ; si elle est jugée satisfaisante elle peut alors être validée. Une question reste cependant posée : est-il possible d'obtenir une spécification formelle directement à partir des graphes conceptuels ? Une des réponses à cette interrogation fait l'objet des travaux de G. Ellis [Ellis 1995]. Pour notre part, nous estimons que les bases formelles du modèle ne s'appliquent pas encore à un ensemble suffisamment large d'éléments de représentation des connaissances, pour prétendre dans l'immédiat atteindre cet objectif.

**4ème hypothèse** : les éléments de la représentation intermédiaire sont trop près du LN (sémantique littérale), il est souvent nécessaire de les transformer au moyen de règles pour les rendre pertinents au niveau de la spécification formelle.

Cette quatrième hypothèse est apparue après une première série d'expérimentations, dans lesquelles nous construisions une spécification formelle issue de la représentation intermédiaire sous forme de graphes conceptuels qui contenaient la sémantique littérale des textes de spécifications. Cette formalisation n'était pas fautive en soi, mais était assez éloignée de celle produite par les experts en spécification. La raison principale de cette distance tenait dans la trop forte exploitation de relations casuelles (de type *Agent*, *Objet*, etc.) propres à représenter le sens des phrases mais nullement évocatrices pour des spécifications techniques.

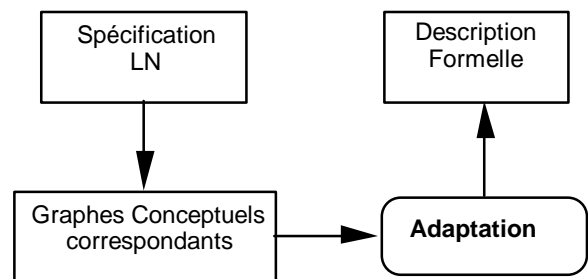


Figure 4 : Transformation des graphes conceptuels pour les adapter à la spécification.

Une des principales transformations mises en évidence est la nominalisation des verbes d'action. Par cette transformation, il est possible de supprimer des relations de type casuel (*Agent*, *Objet*) obtenues lors de l'analyse linguistique d'une phrase et de les remplacer par une nouvelle relation (nominalisation du verbe) pour laquelle les relations supprimées sont implicites. Par exemple, les deux prédicats de relation *AGNT(transmettre, émetteur)* et *OBJ(transmettre, message)* peuvent se réécrire avec le nouveau prédicat fonctionnel *Transmission(émetteur, message)*.

### 3. Un processus de formalisation

Nous inscrivant dans cette démarche de formalisation, nous proposons un enchaînement de traitements sur les spécifications informelles initiales, susceptibles de fournir une description formelle (fig. 5).

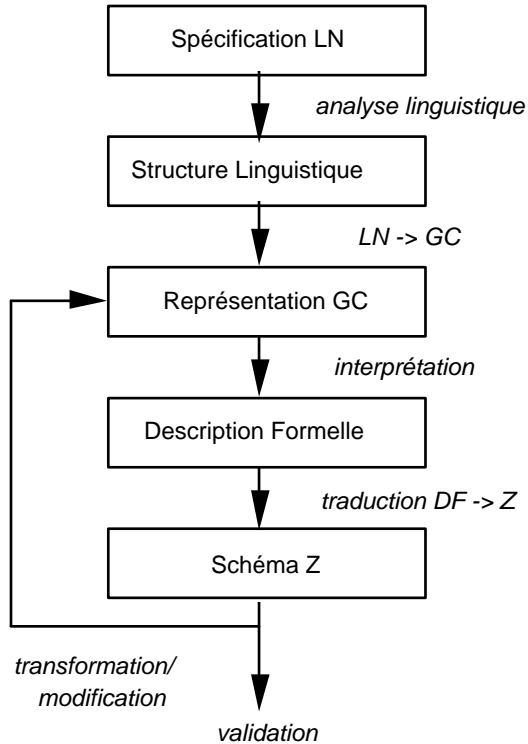


Figure 5 : Chaîne de traitements des spécifications.

Le séquençement des différentes étapes qui constituent ce processus, faisant suite à une étape préalable d'acquisition de connaissances, représente la mise en oeuvre de notre objectif, pondéré par l'ensemble des hypothèses énoncées ci-dessus. Chacune de ces phases de traitement ayant sa propre problématique, nous leur consacrerons les quatre sections suivantes : les aspects liés à l'acquisition des connaissances, à l'analyse linguistique, à la représentation des connaissances et à la formalisation (obtention d'un modèle en langage Z). Puis nous terminons par la présentation d'un court exemple illustrant le processus de formalisation proposé.

#### 3.1. Acquisition de connaissances

Les procédures de traitement du langage naturel reposent sur un corpus de connaissances. Il est donc indispensable de mettre en oeuvre un certain nombre de traitements avant même de commencer la phase d'analyse des phrases du texte à considérer. Cette étape préalable consiste en l'extraction d'informations lexicales contenues dans le texte, afin de déterminer les liens privilégiés que les mots ont entre eux. Une simple étude de co-occurrence de mots, basée sur une analyse de proximité lexicale, permet ainsi de révéler la

présence de mots composés, de locutions, de relations prédicatives et même de schémas de phrases propres au domaine en question. L'utilisation conjointe à cette analyse fréquentielle de techniques de filtrage statistique, telle que l'*information mutuelle*<sup>3</sup>, permet d'affiner et d'augmenter la pertinence des résultats obtenus.

La deuxième phase d'acquisition des connaissances, spécifique à notre modélisation, consiste à extraire, du dictionnaire, des définitions de termes retenus comme concepts, afin de les décrire dans un dictionnaire sémantique sous la forme de graphes conceptuels. Le but de ce traitement est d'accroître les connaissances de la base en y intégrant des *informations sémantiques lexicales* issues d'un dictionnaire de langue et ainsi d'élargir l'ontologie de notre modèle grâce à l'apport de définitions pour chaque concept. Pour automatiser cette tâche, nous avons adapté les algorithmes de [Hernert 1993] qui permettent de détecter les *relations hyperonymiques* contenues dans les définitions et de les moduler en fonction de la présence éventuelle de modificateurs définitionnels. Une fois analysé le contenu d'une définition, il est alors possible de construire le graphe conceptuel correspondant et de l'inclure dans une *base canonique*.

#### 3.2. Une analyse linguistique

Il est habituel en linguistique de considérer cinq niveaux de la langue écrite [Coulon et Kayser 86] : morphologie, lexicale, syntaxe, sémantique, pragmatique. Les procédures de traitement des langues naturelles reposent sur un corpus de connaissances dans chacun de ces cinq niveaux linguistiques. Le fonctionnement d'un système de traitement automatique du langage naturel consiste alors à établir une correspondance entre les phrases à analyser et les différentes catégories de connaissances dont il dispose. On peut certes discuter de l'interdépendance de ces différents niveaux, mais nous considérons que cette stratification a le mérite d'être claire à exposer et qu'elle permet de présenter sans complexité tous les détails de nature linguistique.

##### 3.2.1. L'analyse morpho-lexicale

Il s'agit lors de cette phase de séquençer les phrases analysées afin d'obtenir une suite de *lexèmes*, après identification des mots simples, des mots composés et des expressions figées. Cette analyse est rendue possible par l'utilisation d'un lexique apparenté

<sup>3</sup>Lorsque le nombre de couples d'unités lexicales à observer devient élevé, on estime les probabilités d'associations pertinentes par une méthode de vraisemblance :  $\hat{l}(x,y) = \log(n_{x,y}/n_x n_y)$ , avec  $n_x$  et  $n_y$  le nombre d'occurrences de  $x$  et  $y$ , et  $n_{x,y}$  le nombre d'occurrences du couple  $(x,y)$  [Church et Hanks 1990].

au DELAS<sup>4</sup> et d'un module d'analyse flexionnelle pour la description des formes fléchies (conjugaison, genre, nombre). En plus de la reconnaissance des mots de la langue, cette étape permet la distribution de ceux-ci en classes (catégories syntaxiques), telles que *nom*, *verbe*, *adjectif*.

### 3.2.2. L'analyse syntaxique

Les stratégies d'analyse syntaxique sont très nombreuses : analyses en constituants, grammaires formelles, étiquetage grammatical probabiliste, etc. Notre propos, n'est pas ici de faire un panorama de ces méthodes, mais plutôt d'exposer succinctement le formalisme LFG [Kaplan et Bresnan 1982] que nous avons choisi.

Les grammaires lexicales fonctionnelles (LFG) se décomposent en deux niveaux :

- . la *c-structure* (analyse en constituant) décrite au moyen de règles de productions d'une grammaire hors-contexte, représente les structures syntaxiques sous forme d'arbres ;
- . la *f-structure* (description fonctionnelle), composée de paires fonction-valeur, met en évidence les fonctions grammaticales telles que *Sujet*, *Objet*, etc. Une fonction particulière, nommée *Pred*, met en correspondance les fonctions syntaxiques et les rôles sémantiques d'un prédicat, facilitant ainsi l'interprétation sémantique ultérieure.

Les grammaires LFG ajoutent donc à la construction de la structure syntaxique la formation d'une structure fonctionnelle de la phrase (les liens entre les groupes : analyse logique), spécifiée au moyen de schémas associés aux règles de grammaire. La validité de ces structures est régie par des principes généraux de bonne formation, qui permettent de rejeter certaines analyses syntaxiques incorrectes. La deuxième caractéristique du formalisme LFG est son aspect lexical. Il donne en effet une place importante au lexique, en effectuant directement sur les interprétations des mots toutes les transformations pertinentes (passivation, forme pronominale de certains verbes, etc.). Ces transformations sont décrites au moyen de règles lexicales et la grammaire elle-même ne connaît pas la notion de transformation.

### 3.2.3. L'analyse sémantique

La sémantique du langage naturel concerne les relations qui existent entre les phrases d'une part, les objets, les relations, les actions et les événements du monde d'autre part. L'analyse sémantique a donc pour rôle d'identifier les différents acteurs, objets, instruments intervenant dans une phrase, ainsi que les compléments circonstanciels accompagnant le verbe.

Mais produire une bonne représentation de la phrase suppose aussi être capable de traiter un certain nombre de phénomènes présents dans toutes les langues : les ambiguïtés, les paraphrases, les références, les métaphores, les figures du discours (tropes). Ce sont autant de contraintes dont le formalisme de représentation doit rendre compte.

Le formalisme de représentation des connaissances sémantiques retenu étant le modèle des graphes conceptuels, cette analyse consiste donc en la traduction sémantique de la structure syntaxique sous forme de graphes conceptuels. Pour cela, nous nous inspirons des grammaires de cas [Fillmore 1968] qui déterminent les différents rôles thématiques remplis par les constituants d'une phrase à l'aide des informations acquises sur l'ordre des mots, les prépositions, les verbes et le contexte. En d'autres termes, l'analyseur détermine la façon dont les syntagmes nominaux d'une phrase sont reliés aux verbes - le rôle sémantique spécifiant comment un objet participe à la description d'une action.

Les sources d'informations activées pour les besoins de cette analyse sont : le dictionnaire sémantique des termes, regroupant la sémantique lexicale du domaine, et la description fonctionnelle produite lors de l'analyse LFG. Des méthodes de traduction du langage naturel sous forme de graphes conceptuels sont explicitées dans [Sabah 1990] et [Zweigenbaum et al 1995].

## 3.3. Une représentation des connaissances

La représentation des connaissances est un sujet d'études dans diverses disciplines (la philosophie, la psychologie cognitive, la logique, l'informatique et l'intelligence artificielle) mais elle constitue un objet différent dans chacun de ces domaines. Le problème fondamental de la représentation informatique des connaissances est l'élaboration d'une notation suffisamment précise et formelle (cadre représentationnel) permettant cette représentation [Haton et al 1991]. L'utilisation de ce cadre consiste alors à spécifier une *base de connaissances* contenant un ensemble de *faits* au format établi.

Le monde décrit par un *cadre représentationnel* est une collection d'individus et de relations entre ces individus qui spécifient un état susceptible de subir des transformations. Trois familles de cadres représentationnels peuvent être évoquées suivant les critères de leur construction :

- . à partir d'assertions vraies sur les états, il s'agit de cadres logiques ;
- . à partir de la représentation des individus et de leurs relations, le cadre est du type *réseau sémantique* ;

---

<sup>4</sup>Dictionnaire du LADL contenant environ 80000 mots simples avec leurs références catégorielles [Gross et Silberstein 1995].

- . à partir de la description de changements d'états, les cadres sont dits procéduraux.

Dans le modèle des graphes conceptuels [Sowa 1984], les objets élémentaires sont les *concepts* et les *relations* (individus et relations), ce qui apparente le formalisme à la famille des réseaux sémantiques<sup>5</sup>. Chaque proposition (fait) est représentée par un graphe conceptuel construit au moyen d'arcs dirigés reliant concepts et relations. Des règles offrant la possibilité de joindre ou de dissocier des graphes conceptuels (joints et projections) sont données. Une correspondance formelle (isomorphisme) avec la logique des prédicats du premier ordre est établie pour un noyau de base.

Cette description sommaire des graphes conceptuels met en évidence la mixité des qualités de ce formalisme de représentation des connaissances : un modèle assez "psychologique" dans sa forme, ce qui lui confère une grande lisibilité, et un système de preuves mathématiques ayant des fondements axiomatiques solides, ce qui le rend formel.

### 3.4. La formalisation

La complexité croissante des spécifications d'un domaine quelconque rend de plus en plus évident le besoin en techniques et outils permettant d'une part d'aider à l'élaboration de ces spécifications et d'autre part d'en vérifier la correction et la cohérence. Il s'agit bien entendu d'un sujet d'étude extrêmement vaste : le terme "spécification" englobe des documents au contenu de natures très diverses se prêtant plus ou moins bien à une formalisation. De plus, au-delà de cette étape de formalisation, se pose le problème de la cohérence d'un ensemble de spécifications et de son évolution dans le temps : à ce jour, il faut bien reconnaître qu'on se trouve totalement démuné face à ces problèmes.

La raison majeure qui pousse à formaliser une spécification est de disposer d'une description assimilable par la "machine" en vue d'objectifs bien précis, notamment :

- . aider à la conception d'un protocole en rendant les spécifications "exécutables" ;
- . valider la spécification, c'est-à-dire vérifier que le protocole décrit rend bien le service qu'on attend de lui ;
- . dans le cas de spécifications d'interfaces, produire automatiquement les fiches de tests destinés à

contrôler la conformité d'un équipement vis-à-vis de cette interface [Groz et Phalippou 1991] ;

- . déduire de la description une implantation de référence, c'est-à-dire engendrer directement un programme dans un langage approprié à la machine cible sur laquelle on veut implanter un protocole - on parle d'implantation de référence dans le sens où, si le programme engendré n'est pas forcément performant, il est censé respecter fidèlement la spécification.

Outre ces objectifs, on peut aussi concevoir une description formelle comme un document de référence non ambigu, non sujet à interprétation et exhaustif, complémentaire du document informel lisible (sinon compréhensible) par tout un chacun. Pour cela, nous proposons de construire une description logique à partir de graphes conceptuels - véhicules du sens et du contenu des spécifications informelles - puis de traduire cette description logique dans un langage de description formel, le langage Z [Spivey 1988] en l'occurrence.

Une spécification en Z est constituée par une séquence de *paragraphes* comportant des *schémas*, des *variables* et des *types de base*. À toute expression apparaissant dans une spécification en Z est associé un type unique (ensemble, produit cartésien ou schéma). Un schéma est constitué d'une *signature* (collection de variables typées) et d'une propriété sur cette signature, appelée *partie prédictive*.

Pour établir la description formelle correspondante à une spécification informelle, nous avons procédé par analogie avec le mécanisme de construction déployé par un expert humain. On commence par extraire les éléments de la description formelle, on identifie et on insère les éléments indispensables qui ne sont pas compris dans le module à spécifier, mais auxquels on fait référence, puis on établit les formules logiques (pré-conditions, post-conditions, propriétés) correspondant aux éléments collectés et à leurs liens, tels qu'ils sont définis en langage naturel.

La phase finale de la formalisation est constituée par la modélisation en Z des graphes conceptuels construits à partir de la spécification informelle. On retiendra essentiellement, pour cette traduction, les deux traitements successifs opérés sur les concepts et les relations. Le référent d'un concept devient un élément de l'ensemble représenté par le *label de type* ; quant à la relation elle fait l'objet d'une définition fonctionnelle.

### 3.5. Un exemple simple

Pour illustrer la démarche décrite (fig. 5), nous présentons dans la figure suivante un exemple appliqué à la spécification de la phrase :

<sup>5</sup>Un réseau sémantique est un ensemble de noeuds (les concepts) représentant chacune des entités du monde, entre lesquels sont tissés des liens (les relations sémantiques) représentant les rapports qu'entretiennent les concepts entre eux.

L'émetteur transmet un message au récepteur par l'intermédiaire d'un canal.

qui se représente par le graphe conceptuel :

$u$  : [émetteur:#1]-  
 →(INST)→[canal:\*],  
 →(Transmission)→[message:\*]→(DEST)→[récepteur:#2].

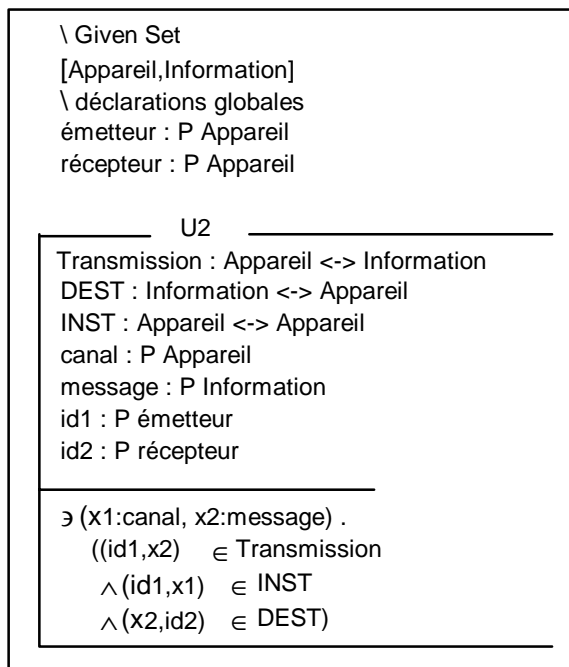


Figure 6 : Traduction en Z du graphe conceptuel  $u$ .

Les éléments algorithmiques de la traduction d'un graphe conceptuel en Z étant présentés dans [Fougères 1997], nous ne décrivons ci-dessous que les grandes lignes de cette traduction. Les concepts {canal, émetteur, message, récepteur} sont traduits dans le modèle Z suivant leur référent respectif (individuel, générique ou variable) – leur super-type dans le treillis des types de concepts<sup>6</sup>, permet d'établir une relation d'appartenance à un ensemble, conformément à la sémantique du langage Z. La distinction apportée à la traduction des référents n'est pas uniquement syntaxique, mais essentiellement d'ordre sémantique. En effet, la déclaration d'un concept dont le référent est individuel signifie certes son existence, mais, de plus, qu'il a pu faire l'objet d'une précédente déclaration, soit dans le schéma en cours, soit dans un autre schéma, auquel cas il est nécessaire de l'inclure.

Les relations {Transmission, DEST, INST} sont traduites successivement dans les parties déclarative et prédicative, en termes de définition typée et de

<sup>6</sup> Pour notre exemple le treillis des types de concepts est constitué des éléments suivants (où "<" symbolise la relation *est-un*) :

ENTITE, ACTION < T; Appareil, Information < ENTITE;  
 émetteur, récepteur, canal < Appareil; message < Information.

relation d'appartenance. Pour la description de l'état d'un système en Z, on s'intéresse davantage aux relations entre variables qu'aux types auxquels elles appartiennent ; ceci nécessite une vérification de conformité des types au niveau de la spécification informelle. Il est donc indispensable de procéder à cette vérification avant de traduire les relations contenues dans les graphes conceptuels. Ainsi, la signature de la relation DEST étant, dans notre modèle des graphes conceptuels, [Entité]→(DEST)→[Entité], il faut s'assurer que la généralisation de Information et Appareil au type de concept Entité, soit valide dans le treillis des types de concepts.

À l'aide de cet exemple simple, nous pouvons énoncer les deux remarques suivantes qui confirment les hypothèses que nous avons formulées :

- . il est possible de décomposer et de simuler le déroulement de la description formelle d'une spécification rédigée en langage naturel ;
- . le langage Z est bien adapté pour une traduction à partir de graphes conceptuels.

Pourtant la traduction des graphes conceptuels en Z n'est pas aussi simpliste qu'il n'y paraît au premier abord. Certains éléments du formalisme des graphes conceptuels ne trouvent pas de correspondance dans une modélisation en Z (notamment les modalités). D'autre part, la prise en compte du temps, pour la synchronisation de protocoles par exemple, n'est pas permise en Z. Nous trouvons là confirmation de l'hypothèse émise sur la nécessité d'une représentation intermédiaire (ou pivot) pour prétendre traduire des spécifications informelles dans des langages de descriptions formels adaptés, ayant chacun de fortes spécificités.

## 4. Conclusion

Cette communication a privilégié une présentation globale de la formalisation de spécifications informelles rédigées en langage naturel. Les perspectives de mise en oeuvre de cette démarche de formalisation nous ont conduit à établir quatre hypothèses préalables. On retiendra plus particulièrement la deuxième hypothèse, celle qui postule l'utilisation d'une représentation intermédiaire pour passer de la spécification informelle à sa description formelle. Le choix du modèle des graphes conceptuels pour cette représentation sémantique intermédiaire a été motivé notamment par le fait qu'ils sont bien adaptés pour représenter la sémantique du langage naturel, par les possibilités qu'offre la dérivation directe d'un graphe conceptuel simple en une formule logique du premier ordre et par l'ensemble des travaux réalisés depuis 1984 sur l'interprétation logique de ces graphes, y compris les extensions vers des modèles logiques du second ordre et modaux [Mugnier et Chein 1996].



Nous avons ainsi mis en évidence quatre niveaux de représentation d'une spécification : le texte source de la spécification écrite en langage naturel, la représentation contenant les structures linguistiques des phrases du texte, la représentation sémantique sous forme de graphes conceptuels et la représentation finale dans un modèle de description formelle tel que le langage Z. À partir de la première représentation en langage naturel, nous pouvons définir un processus dont les phases successives de traitement construisent les trois autres représentations de la spécification - ce que nous nommons processus de formalisation.

Un système, conçu selon les critères définis précédemment, peut se présenter comme une composante intelligente d'un atelier de développement de logiciels. Il en constituerait le module d'élaboration des spécifications formelles, en assistant autant que possible de façon interactive un rédacteur de spécifications. Idéalement, l'entrée du système serait constituée par une spécification informelle, document rédigé en langue naturelle, et le système fournirait en sortie une spécification formelle, après coopération avec le rédacteur. Avant de satisfaire à cette perspective, il reste néanmoins de nombreuses difficultés à résoudre, notamment les trois types de problèmes suivants :

- . gérer les irrégularités sémantiques intrinsèques du langage naturel ;
- . retrouver une grande quantité d'informations expertes, dont il est fait abstraction dans les spécifications sources [Toussaint 1992] – c'est toute la complexité de la caractérisation des niveaux de l'explicite et de l'implicite.
- . compenser la part irréductible du passage de l'informel au formel – Newell souligne clairement l'existence d'un niveau spécifique de la connaissance situé hors de tout système formel, et d'un problème de transfert de cette connaissance informelle en une connaissance formelle (le "symbol level") [Newell 1982].

C'est pourquoi, il n'est pas question pour nous d'envisager le "spécifieur automatique" [Balzer 1985], dans lequel on introduirait le titre d'une spécification et qui produirait les schémas de représentations formelles dépourvus d'ambiguïtés. L'outil doit donc rester au service de l'intelligence et de la créativité.

### Références bibliographiques.

- [Balzer 1985] : R. Balzer, (1985), A 15 year Perspective on Automatic Programming, *IEEE Transactions on Software Engineering*, SE-11(11), pp. 1257-1268.
- [Biebow et al 1995] : B. Biebow, T. Charnois & S. Szulman, (1995), Aide à l'élaboration de spécifications fonctionnelles, Rapport Technique du LIPN.
- [Church et Hanks 1990] : K.W. Church et P. Hanks, (1990), Word associations norms, mutual information, and lexicography, *Computational Linguistics*, 16(1).

[Coulon et Kayser 86] : D. Coulon et D. Kayser, (1986), Informatique et langage, présentation générale des méthodes d'interprétation des textes écrits, *T.S.I.*, 5(2).

[Ellis 1995] : G. Ellis, (1995), Object Oriented Conceptual Graphs, *Proceeding of the 3rd International Conference on Conceptual Structures (ICCS'95)*, Springer-Verlag Publisher, pp. 144-157.

[Fillmore 1968] : C. J. Fillmore, (1968), The case for case, *Universals in Linguistic Theory*, E. Bach & R. Harms Ed., New York.

[Fougères et Trigano 1996] : A-J. Fougères & P. Trigano, (1996), The formalisation of Specifications from Specification written in Natural Language, *Technology Transfer Series of IITT-International*, Expersys'96, Paris.

[Fougères 1997] : A-J. Fougères, (1997), Aide à la rédaction de spécifications formelles à partir des spécifications rédigées en langage naturel. Application aux spécifications de services de France Télécom, Thèse de Doctorat de l'Université Technologique de Compiègne.

[Gross et Silberstein 1995] : M. Gross & M. Silberstein, (1995), Outils de traitement linguistique, applications à l'analyse documentaire, *Ecole d'été Traitement des Langues Naturelles*, Trégastel.

[Groz et Phalippou 1991] : R. Groz & M. Phalippou, (1991), L'ingénierie des protocoles au CNET, *Actes du Colloque Francophone sur l'Ingénierie des Protocoles (CFIP'91)*, Omar Rafiq Editeur, Hermès, Pau.

[Haton et al 1991] : J-P. Haton, N. Bouzid, F. Charpillet, M-C. Haton, B. Lâasri, H. Lâasri, P. Marquis, T. Moudot et A. Napoli, (1991), Le raisonnement en intelligence artificielle. Modèles, techniques et architectures pour les systèmes à base de connaissances, InterEdition, Paris.

[Hernert 1993] : P. Hernert, (1993), Un système d'acquisition de définitions basé sur le modèle des graphes conceptuels, Thèse de l'Université Paris XIII - Institut Gallilée.

[Kaplan et Bresnan 1982] : R. Kaplan, J. Bresnan, (1982), Lexical-Functional Grammar: a Formal System for Grammar Representation, *The mental Representation of Grammatical Relations*, J. Bresnan Ed., MIT press, pp. 173-281.

[Mugnier et Chein 1996] : M-L. Mugnier & M. Chein, (1996), Représenter des connaissances et raisonner avec des graphes, *Revue d'intelligence artificielle*, vol. 10, n° 1, pp. 7-56.

[Newell 1982] : Allen Newell, (1982), The Knowledge Level, *Artificial Intelligence*, 18, pp. 87-127.

[Sabah 1990] : G. Sabah, (1990), Caramel : un système multi-expert pour le traitement automatique des langues, *Modèles Linguistiques*, 12(1), pp. 95-118.

[Sowa 1984] : J. F. Sowa, (1984), Conceptual Structures : Information Processing in Mind and Machine, *Addison-Wesley Publishing Company*, Reading, MA.

[SPECS 1993] : The SPECS Consortium, (1993), Spécification and Programming Environnement for Communication Software, Edited by R. Reed, W. Bouma, J. Evans, M. Dauphin & M. Michel, Amsterdam, North-Holland.

[Spivey 1988] : J. M. Spivey, (1988), , Understanding Z. A specification language and its formal semantics, Cambridge University Press.

[Toussaint 1992] : Y. Toussaint, (1989), Méthodes informatiques et linguistiques pour l'aide à la spécification de logiciel, Thèse de Doctorat de l'Université Toulouse III.

[Vadera et Meziane 1994] : S. Vadera et F. Meziane, (1994), From English to Formal Specifications, *The Computer Journal*, 37(9) pp. 753-763.

[Wirsing 1993] : M. Wirsing, (1993), Développement de logiciel et spécification formelle, *T.S.I.*, 12(4), pp-413-431.

[Zweigenbaum et al 1995] : P. Zweigenbaum, B. Bachimont, J. Bouaud, J. Charlet et J-F. Boisvieux, (1995), A multi-lingual architecture for building a normalized conceptual representation from medical language, *17th Annu. Symp. Computer Applications in Medical Care*, Reed M. Garner editor, New Orleans.