

Assurer la qualité sémiotique des données d'une communauté scientifique

Daniel. GALARRETA

CNES, 18, avenue Edouard Belin, 310401, Toulouse cedex 4, France.

Galarreta@cnes.fr

Résumé : Les données produites ou utilisées dans le cadre de projets scientifiques véhiculent une signification que la norme ISO 11179 définit comme une « représentation formelle de faits, concepts ou instructions convenant à la communication, l'interprétation ou le traitement par des êtres humains ou des automatismes ».

Si on se préoccupe d'assurer le maintien de la signification de ces données dans le temps de telle sorte qu'elles « soient reconnaissables demain comme aujourd'hui par le plus grand nombre », il est légitime de se pencher sur le problème de la signification, c'est-à-dire sur la manière dont le sens est produit ou saisi par un être humain.

L'analyse que nous faisons des modèles de signification existants nous conduit à recommander une description des données fondée sur une approche linguistique au lieu d'une approche conceptuelle. Les données elles-mêmes peuvent être définies dans le cadre d'un modèle de communication qui ne présuppose pas l'existence d'un niveau conceptuel autonome vis-à-vis du langage. Elles s'inscrivent alors dans un cadre de gestion des connaissances.

Le dispositif informationnel que le schéma de communication emprunte pour se réaliser peut-être qualifié du point de vue de la gestion des connaissances par une qualité sémiotique. Nous examinons cette qualité dans le cas de données scientifiques.

Abstract : Data which produced or used within the frame of scientific projects support a signification that the norm ISO 11179 defines as "A representation of facts, concepts, or instructions in a formalised manner suitable for communication, interpretation, or processing by humans or by automatic means".

As far as we are concerned by the maintaining of the signification of data, in such a way that "they can be recognised and used in the future, as today, by as many users as possible", it is justified to examine the problem of signification per se, viz. the way meaning is produced or grasped by people.

The analyse that we make of signification models lead us to recommend a data description based upon a linguistic approach instead of a conceptual one. Data themselves can be defined within the framework of a communication model, which does not presuppose the existence of an independent conceptual level with regard to natural languages. Therefore data can be set within a framework of knowledge management.

The information system that is used by the communication schema to be implemented, can be characterised with respect to knowledge management by a semiotic quality. We examine this quality in the case of scientific data.

1. Objectif de cet article

Les projets scientifiques qui utilisent des moyens spatiaux ont cette caractéristique de nécessiter une infrastructure complexe, de concerner une communauté scientifique importante, de s'étendre sur des périodes longues (de 10 à 20 ans) et de produire de très grandes quantités de données numériques.

De ce fait il est primordial afin d'assurer la viabilité économique de ces programmes, d'assurer que les données produites soient reconnaissables aujourd'hui comme demain par le plus grand nombre.

Cet article se propose d'examiner les dispositions pratiques qu'il faut envisager du point de vue organisationnel et communicationnel pour répondre à cet impératif. Ces dispositions vont prendre la forme d'une démarche d'assurance et de contrôle qualité d'un type nouveau, propre à garantir la restitution de la signification des données sans perte ni distorsion et d'une manière adaptée aux besoins des scientifiques.

2. Position du problème

L'approche actuelle pour s'assurer que des données produites, disons au cours d'une expérience scientifique, demeurent exploitables pour d'autres scientifiques que ceux qui sont à l'origine de l'expérience, est de décrire ces données d'une manière qui soit suffisamment précise pour rendre compte de l'expérience, adaptée au domaine d'étude concerné et standardisé pour permettre une même interprétation au plus grand nombre¹.

Les éléments qui assurent cette fonction descriptive sont appelés des métadonnées. La démarche qualité à mettre en œuvre pour garantir la pérennité « sémantique » des données consiste à assurer et

¹ Norme ISO 11179

contrôler la qualité des métadonnées correspondantes au travers de l'assurance et du contrôle du niveau conceptuel auquel les métadonnées se situent. C'est-à-dire que la qualité des données s'assure et se contrôle au travers de la « chaîne » : qualité des concepts => qualité des métadonnées => qualité des données ; chaque maillon contrôlant le suivant.

Cette chaîne est justifiée in fine, par la hiérarchie implicite concepts > métadonnées > données et garantie en dernière analyse par la relation liant le phénomène étudié par l'expérience et les concepts décrivant ou élucidant² le phénomène.

On résumera de manière très simplifiée cette approche par le schéma suivant (figure 1):

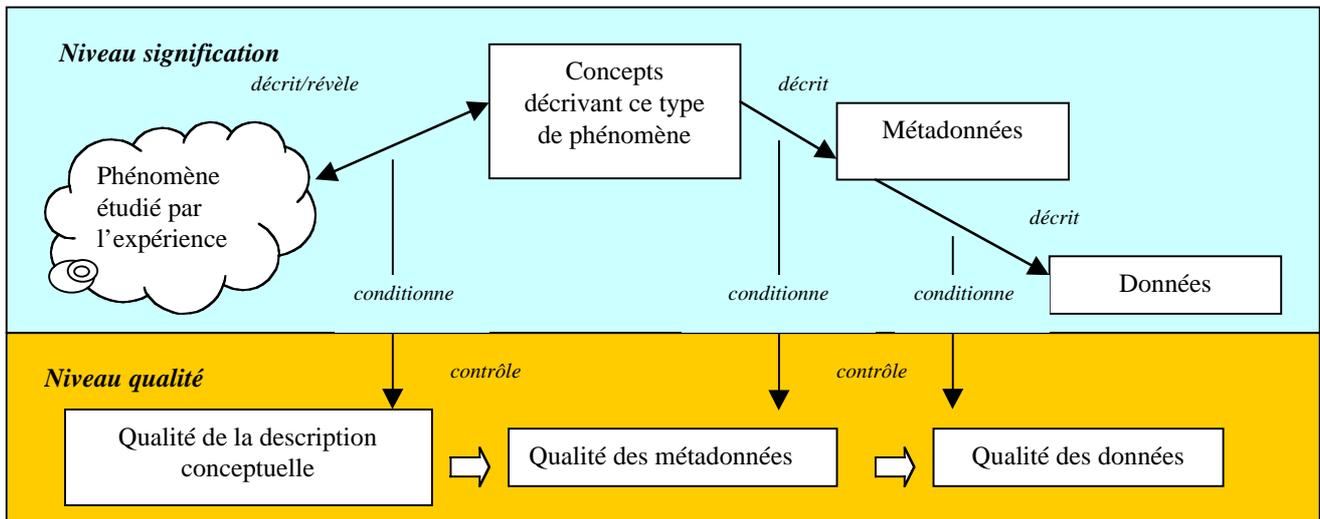


FIG. 1 : Dépendance de la qualité des données vis-à-vis de la description des données et des métadonnées

Ce schéma justifie qu'on étudie le fonctionnement du niveau conceptuel ; ce fonctionnement doit être examiné dans son rapport au problème général de la signification pour tenir compte de la relation qui se manifeste aux scientifiques entre le phénomène et les données. Plus spécifiquement le problème du maintien de la signification de ces données dans le temps « de telle sorte qu'elles soient reconnaissables demain comme aujourd'hui par le plus grand nombre », rend légitime de réfléchir sur le problème de la signification, c'est-à-dire sur la manière dont le sens est produit ou saisi par un être humain.

3. Discussion sur les modèles de signification

En disposant d'un modèle conceptuel explicitant les conditions de production et de saisie du sens, on peut ainsi espérer être plus efficace vis-à-vis du problème de la diffusion et de la conservation du sens des données scientifiques qui nous intéresse ici. Nous examinons maintenant cette question.

Dans les quatre premières sections qui suivent nous reprenons, en la simplifiant à l'extrême, la présentation que François Rastier consacre au chapitre 3 de son livre à la question ([1], p.73) : « faut-il distinguer le niveau sémantique des langues et le niveau conceptuel ?³ »

3.1 La triade

« Depuis Aristote (et particulièrement le début du *Péri hermèneias*), la philosophie du langage se fonde sur un modèle triadique ; cf. I, 16 a, 3-8 : « la parole est un ensemble d'éléments symbolisant les états d'âme, et l'écriture un ensemble d'éléments symbolisant la parole. Et, de même que les hommes n'ont pas tous le même système d'écriture, ils ne parlent pas tous de la même façon. Toutefois, ce que la parole signifie immédiatement, ce sont des états de l'âme qui, eux, sont identiques pour tous les hommes ; et ce que ces états de l'âme représentent, ce sont des choses, non moins identiques pour tout le monde. »

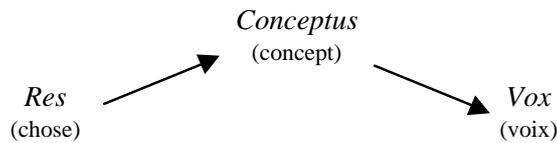
[...]

Après Boèce, Thomas d'Aquin reformule ainsi cette triade : « les paroles sont les signes des pensées et

² Cette distinction dépend de la position épistémologique qu'on adopte.

³ Cette question est équivalente à la question suivante : le signifié linguistique se distingue-t-il du concept, logique ou psychologique ?

les pensées des similitudes (similitudines) des choses. d'où il suit que les paroles se réfèrent aux choses désignés moyennant les concepts » (*Somme théologique, I-ap, 2-13, al, resp.*). Soit :



[...] » ([1], p.75)

3.2 L'indice

« La triade aristotélicienne n'est pas la seule conception antique de la signification [...] Reformulant la théorie rhétorique de l'indice, Aristote définit ainsi le séméion : « le signe (*το σημιον*), entend être une prémisses démonstrative, nécessaire ou probable [figure 2]. La chose, dont l'existence ou la génération entraîne l'existence ou la génération d'une autre chose qui lui est antérieur ou postérieur, c'est ce qui constitue le signe de la génération ou de l'existence. » (*Premiers analytiques, II, 27 ; 70 a, 7*) [...] il s'est poursuivi dans la théorie des signes naturels chez saint Augustin (*signa naturalia*, cf. notamment *De doctrina christiana, II,1, 2*) jusqu'aux grammaires générales des lumières (de Condillac à Tracy), voire dans la phanérosopie de Pierce. » ([1], p.80)

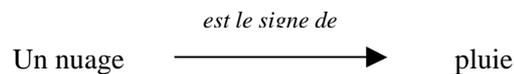


FIG. 2 : exemple d'inférence

3.3 La prééminence du niveau conceptuel

Ces modèles de signification conduisent à deux types de sémantiques :

Une sémantique de la référence : L'étude des relations des concepts aux référents constitue la théorie extensionnelle de la signification. F. Rastier remarque que cette théorie « convient aux langages formels mais qu'il reste à montrer comment elle peut s'appliquer aux langues. Elle n'est d'ailleurs même pas spécifique aux systèmes de signes, puisque à strictement parler, ce sont les concepts qui sont dotés d'une extension (qu'ils soient ou non exprimés par des signes). » ([1], p.82)

Une sémantique de l'inférence : Les schémas (frames), scénarios (scripts) et plans qui sont utilisés en intelligence artificielle se présentent comme des supports d'inférences. Ils matérialisent un contexte propre à permettre l'inférence. « Les modèles mentaux (voir notamment Johnson-Laird, 1983)

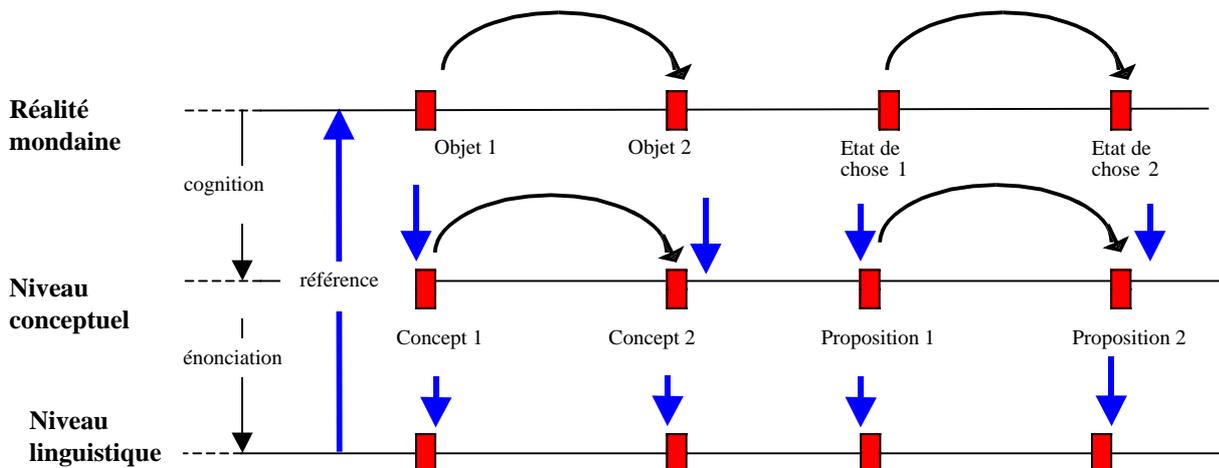


FIG 3 : La référence comme l'inférence repose sur l'existence d'un niveau conceptuel (après Rastier p.)

transposent cette problématique dans le domaine de la psychologie. Ces modèles sont en fait des schémas construits pour rendre compte de la signification des énoncés. » ([1], p.86)

On peut résumer ce qu'on vient de présenter brièvement sous la forme du schéma précédent (figure 3) (situé p. 89 de [1]) qui montre la complémentarité entre les deux types de sémantiques et la prééminence du niveau conceptuel :

« Dans ce dispositif, le niveau conceptuel dispose d'une fonction prééminente : il médiatise la référence des expressions ; il détermine les inférences que l'on peut effectuer sur elles. [...] Enfin ce dispositif ne laisse pas de place à la sémantique linguistique, car le sens n'y est pas autre chose qu'une représentation. Il n'existe pas d'autres signifiés que les concepts ou les propositions, et bien entendu ils sont indépendants des langues. » ([1], p.89)

3.4 Une sémantique de la différence

1. Une autre voie est possible pour définir une *sémantique linguistique* autonome par rapport aux modèles précédents et à leurs présupposés logiques ou psychologiques, et reposant sur le paradigme de la *différence*.

« Le problème de la différenciation est à l'évidence un problème philosophique fondamental qui touche la distinction des objets et celle des concepts. Quant aux objets, Xénophane notait le caractère différentiel de leur perception : « si Dieu n'avait pas fait le miel doré, les figues nous paraîtraient plus douces » (fr. 38) » ([1], p.98)

C'est à partir de la réflexion sur la synonymie, que le problème de la différenciation a été trouvé dans la langue une productivité toute particulière ; c'est de cette réflexion « qu'allait se former la problématique propre à la sémantique linguistique. Systématisant pour les dépasser des observations éparses chez Varron, Donat, Servius et, parmi les modernes, chez le père Vavasseur, Scioppius, Henri Etienne, l'abbé Girard ose écrire dans son *Traité de la justesse de la langue française*. Qu'il « n'y a point de mots synonymes en aucune langue » (1718, p. 28) et ouvre ainsi à nos yeux le paradigme différentiel en sémantique.

Comment, en effet, apprécier la différence entre les mots synonymes ? On dira qu'elle réside dans des idées accessoires. Mais si chaque mot a un sens différent, le modèle triadique de la signification ne fonctionne plus. En effet, deux mots synonymes au sens large – disons *automobile* et *voiture* – (figure 4) renvoient au même objet. Peut-on alors soutenir qu'ils renvoient à deux concepts différents ? Non, mais à un seul concept ou idée principale. Si toutefois on prend la synonymie au sens strict, on doit considérer qu'ils renvoient à deux concepts différents par leur idées accessoires. Mais deux concepts différents devraient renvoyer à leur tour à deux objets différents. Donc les voitures ne seraient pas des automobiles. » ([1], p.99)

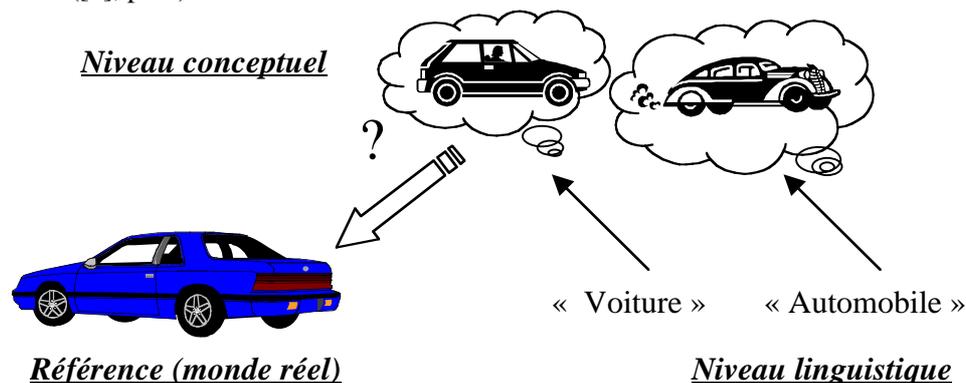


FIG.4 : le problème de la synonymie dans le langage

« Une autre voie consisterait à admettre l'irréductibilité des langues les unes aux autres, et la spécificité de leur sémantique, dont témoigne en premier lieu leur lexique. » ([1], p.100)

« Le sens linguistique n'est pas (ou pas seulement) constitué par la référence à des choses, ou par l'inférence entre concepts, mais aussi et d'abord par la différence entre des unités linguistiques. » ([1], p.101)

3.5 Conséquence pour l'approche conceptuelle

La conséquence la plus importante pour notre propos vis-à-vis de la notion de concept est que la signification ne nécessite plus l'existence d'un niveau conceptuel autonome vis-à-vis de la langue ;

autrement dit, elle n'est plus dépendante d'une ontologie qui serait seule capable de relier les mots au monde par l'intermédiaire des concepts.

3.6 Une sémiotique du discours

Le paradigme différentiel que nous avons introduit dans le cadre d'une sémantique linguistique doit être généralisé quand :

- a. au lieu de se situer au sein d'une langue donnée on veut pouvoir considérer toutes sortes de langages possibles ou ensembles signifiants (par exemple une image satellite ou un tableau de peinture);
- b. au lieu de se cantonner au signifié lexical, on veut pouvoir appréhender le sens d'unités plus importantes, et même au-delà de la phrase, d'un texte entier, ou d'une image entière ;

Cette généralisation conduit à la formulation d'une théorie sémiotique générale.

Une telle théorie à la suite de Saussure puis de Hjelmslev concevra tout ensemble signifiant comme la juxtaposition de deux plans : un plan d'expression et un plan de contenu. Cette caractéristique « biplane » des langages, veut dire que ce par quoi ils se manifestent ne se confond pas avec le manifesté : la langue parlée est faite de sons, mais son propos n'est pas de parler de sons ; les sifflements du dauphin signifient autre chose que les bruits qu'il émet, etc. (p. 203) article langage.

Le paradigme différentiel évoqué pour mettre à jour le signifié lexical est repris de manière plus générale à propos de la *signification* comme saisie des écarts –qui définit d'après Saussure, la nature même du langage. « Entendue comme la mise en place de relations –ou comme leur saisie – la signification s'inscrit comme « sens articulé », dans la dichotomie sens /signification. » p. 353 (article signification)

Historiquement (à la suite des travaux de Jakobson et de Hjelmslev) cette saisie d'écarts a d'abord pris la forme d'une caractérisation de type logique mais a été étendue à tout le discours par Algirdas Julien Greimas : en faisant d'abord une synthèse des travaux de Vladimir Propp sur la *Morphologie du conte*, Greimas a montré qu'un ensemble d'opérations *oppositives* pouvait rendre compte du schématisme narratif mis en lumière par Propp.

La sémiotique se généralise à tout type de discours entendu comme un processus. « Le discours ne se contente pas d'utiliser les unités d'un système ou d'un code préétabli ; cette des choses n'est applicable que dans un petit nombre de situations marginales et finalement peu intéressantes (le code de la route, par exemple). Tout au contraire, le discours invente sans cesse de nouvelles figures, contribue à infléchir ou à déformer le système, que d'autres discours avait nourri auparavant. [...] D'où l'intérêt, pour la sémiotique d'aujourd'hui de ne jamais perdre de vue la production des formes signifiantes, la manière dont le discours schématise nos expériences et nos représentations en vue de les rendre signifiantes et de les faire partager par autrui. » (pp. 81-82 in [2])

Ce qu'il faut retenir à ce point c'est que « dans un discours le sens n'est saisissable qu'à travers ses transformations » (p. 83, in [2]).

3.7 Conséquence sur les schémas de communication

L'approche sémiotique que nous avons nous amène à une conception particulière de la communication.

« Dans la mesure où la communication est située entre des sujets et que les valeurs investies dans les objets mis en circulation (valeurs pragmatiques ou cognitives, descriptives ou modales) sont considérées comme l'être du sujet (celui-ci se trouvant constamment en augmentation ou en diminution de son être) il est évident que le destinataire et le destinataire ne peuvent plus être traitées comme des positions vide d'émetteur et de récepteur, qu'ils sont au contraire des sujets compétents, saisis à un moment de leur devenir, inscrit chacun dans son propre discours » (voir [3], p. 47 article *communication*)

Un dialogue sera dans ce contexte considéré comme un discours à deux (ou plusieurs) voix.

Plutôt de considérer le schéma suivant : (figure 5)

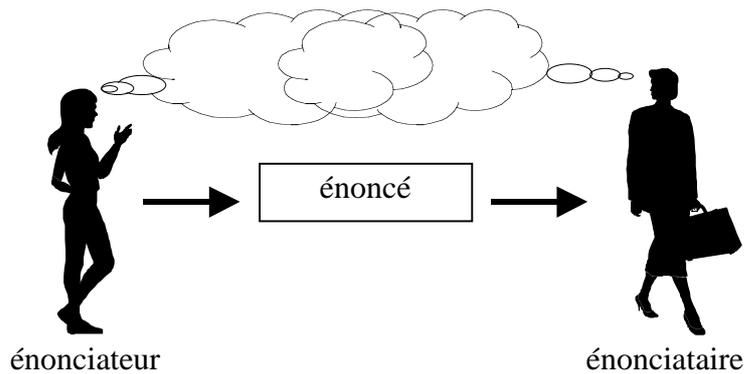


FIG. 5 : Schéma classique de communication. L'existence du niveau conceptuel commun garantit la communication

On préfère le schéma suivant (figure 6) où énonciateur et énonciataire produisent des énoncés dans le cadre d'une activité sémiotique plus « symétrique », où la lecture des significations d'un texte est aussi bien une construction de signification que l'écriture elle-même.

Mutuellement présents l'un à l'autre dans le discours (même si ce n'est qu'en tant que présupposés), l'énonciataire et l'énonciateur sont susceptibles de produire ou d'interpréter des énoncés qui peuvent être incompatibles ou compatibles selon la dynamique du discours.

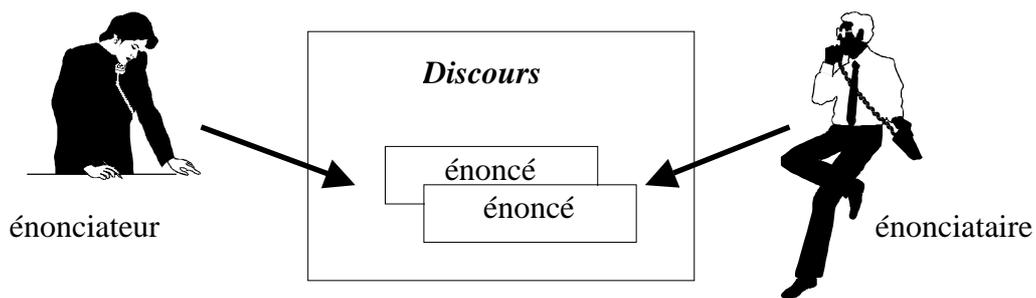


FIG. 6 : schéma sémiotique de communication ne nécessitant pas l'existence d'un niveau conceptuel. Chaque acteur est doté d'une compétence linguistique propre.

Dans le contexte de projets spatiaux cette dynamique prend une forme particulièrement complexe compte tenu qu'il ne s'agit plus simplement d'un dialogue « polyphonique ».

En effet la fréquentation des projets de réalisation de systèmes complexes comme les systèmes spatiaux atteste à l'évidence de la coexistence de langues de spécialités partiellement étanches les unes aux autres et plus généralement de langages spécifiques suivant le découpage des "métiers" en présence.

Ces "métiers" se confrontent autour de la définition —spécification, conception ou réalisation— d'objets techniques dont il faut à chaque fois reconnaître qu'ils ne sont pas donnés a priori mais au contraire progressivement construits au travers de négociations du sens qu'ils doivent avoir.

Autrement dit ces objets sont —au moins avant leur réalisation, mais également après— des objets sémiotiques, appartenant à différents ensembles signifiants [il s'agit par exemple de descriptions combinant modélisations thermiques, électrique, mécaniques ...] susceptibles d'être appréhendés, informés et articulés par une théorie sémiotique.

Cependant toute sémiotique qu'on voudrait mettre en oeuvre doit pouvoir rendre compte du caractère objectif de la plupart des objets conçus et/fabriqués dans ce contexte. Objectif doit s'entendre ici au deux sens suivants : (a) comme résultant d'un accord collectif et (b) relevant d'un caractère extra-linguistique.

Cette obligation oriente vers une problématique que nous allons maintenant examiner.

4. Concevoir une sémiotique des données

4.1 Sémiotique multi-points de vue

A partir de ce nouveau schéma de communication qu'on vient de présenter, on introduit la notion de point de vue qui va correspondre à une compétence collective ou individuelle à communiquer dans le cadre de ce schéma. La description de la dynamique de communication selon ce schéma est désignée par sémiotique multi-points de vue⁴.

Nous illustrerons cette description sémiotique à partir d'un exemple.

Considérons une image SPOT (figure 7) de la région de Krasnodar (Russie) en envisageant **différentes situations de communication.**

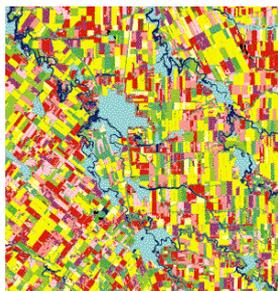


FIG. 7 : image satellite de la région de Krasnodar (Russie). (Spot image)

a) **un seul point de vue considère l'image** : par exemple la personne en charge des algorithmes de traitements d'image. Pour elle cette image est une matrice de pixels. Une question qu'elle peut se poser portera par exemple sur la « dynamique » des niveaux de gris à l'intérieur de l'image, ou le calcul de fonctions de corrélation sur des zones de l'image

b) **deux points de vue au moins considèrent en même temps l'image ; un des points de vue communique sa vue aux autres points de vue en présence** : par exemple il s'agit d'un dialogue entre le thématicien de Spot-Image et le commanditaire de l'image à propos des cultures pratiquées dans la zone agricole considérée. Une question sur cette image à laquelle il s'agit de répondre pourra concerner sur les types de cultures qu'on peut observer sur la région de Krasnodar. Il n'est pas dit que le commanditaire s'attende à cette réponse, ni que le thématicien connaisse les attentes de son client.

c) **deux points de vue au moins considèrent en même temps l'image ; les points de vue en présence s'accordent sur les vues respectives de chacun** : par exemple dans le cadre d'une négociation entre les représentants de la région et ceux de l'état portant sur la définition d'une politique agricole ; il s'agit d'établir éléments consensuels. Cette image permettra par exemple de répondre à une question du type : pourquoi la culture du colza est-elle en régression dans la région de Krasnodar ?

Un énoncé produit par un point de vue sera appelé une **vue**. Cette vue sera selon les circonstances de communication :

- Une donnée dans le cas a)
- Une information dans le cas b)
- Une connaissance dans le cas c)

Ainsi dans l'exemple choisi, l'image SPOT sera une donnée dans le cas a) : matrice de pixels ; une information dans le cas b) : informations sur les cultures de Krasnodar ; une connaissance dans le cas c) : un fait se rapportant à la culture du colza et expliquant sa régression.

⁴ Une présentation plus détaillée de la sémiotique multi-points de vue peut être trouvée dans « Supporting the semiotic quality of data of a scientific community »

Une donnée n'est pas moins bien qu'une information, qui serait moins bien qu'une connaissance, simplement elles ne sont pas produites ou utilisées dans le même contexte de communication et de ce fait, ne permettent pas de répondre aux mêmes types de questions⁵.

4.2 Sémiotique des données

En nous plaçant, selon le schéma de communication adopté, une sémiotique des données va étendre le paradigme différentiel de la sémantique linguistique (cf. section 3.4) repris par la sémiotique du discours (cf. section 3.6) au cas où :

- a. on prend en compte un point de vue particulier et non pas le « point de vue » indifférencié de tout le monde
- b. on admet des données mêlant des signifiants hétérogènes, par exemple des textes et des formules, des tableaux etc.

Précisons. Il s'agit de :

Prendre en compte un point de vue particulier et rendre compte de la solidarité des données entre elles :

L'intelligibilité de données produites par un même point de vue repose sur l'existence –au moins– d'un système sémiotique. De la même manière que l'intelligibilité d'un énoncé en français résulte de certains énoncés qui ont déjà été produits, de même l'intelligibilité d'une donnée particulière dépend de certaines données qui ont déjà été produites. Nous y insistons : Le sens apparemment « autonome » vis-à-vis d'une langue particulière, c'est-à-dire conceptuel, d'un énoncé ou d'un terme, dépend selon nous, d'un système de différences.

Appréhender l'hétérogénéité de l'expression des données :

Dans le cadre du nouveau schéma de communication adopté, un point de vue n'est pas limité à s'exprimer en adoptant de manière exclusive soit l'anglais ou le français, soit des formules mathématiques, soit des images. Le recours à des « textes » mêlant expressions naturelles et expressions mathématiques est fréquent dans des situations de communication où se sont des données qui sont exprimées. Par exemple dans le cas très particulier des textes mathématiques, la signification n'est pas véhiculée uniquement au travers des expressions mathématiques : le texte en expression naturelle participe également à la signification d'ensemble⁶.

5. Une démarche qualité adaptée

En adoptant la perspective multi-points de vue que nous venons de présenter, autrement dit en nous plaçant selon le nouveau schéma de communication, nous sommes amenés à opérer des choix : sélectionner les bonnes méthodes et les bons outils adaptés aux contextes de communication. Par exemple il vaut mieux éviter d'utiliser une solution adaptée pour traiter efficacement une « information » quand le contexte correspond à un besoin de « connaissance » et inversement ...

Nous sommes là devant une attente dont la prise en compte rationnelle relève d'une démarche qualité. ***Mais de quelle qualité s'agit-il exactement ?***

On trouve déjà depuis quelques années dans la littérature spécialisée dans le management de l'information des listes de critères d'évaluation de la qualité d'un système d'information⁷.

⁵ Pour un développement de ce point on consultera [4]

⁶ On constate qu'à la place du triade signe/concept/référent (cf. section 3.1) on peut se référer à une double expression naturelle et notationnelle en relation avec un plan de contenu (cf. section 3.6). Pour une analyse sémiotique de textes mathématiques voir [5]

⁷ Voici à titre d'exemple ce que propose la société EoExchange (www.eoexchange.com) :

- Disponibilité : « Indique à quel degré l'information est dispersée dans l'entreprise. En supposant l'existence d'informations spécifiques dans l'entreprise, quelle est son accessibilité en terme de droits d'accès, d'accès à distance et de politique de publication ? ».

- Sélectabilité : « Indique à quel point une information peut être facilement sélectionnée et trouvée au moyen d'attributs spécifiques ».

- Contexte : « Indique à quel degré des documents en relation se réfèrent mutuellement et peuvent être associés au sein de multi-documents ».

- Intégrité : La crédibilité d'une information est-elle clairement discernable ? Des lecteurs peuvent-ils facilement différencier une information faisant autorité d'une information moins sûre ? Si des informations sont distribuées en plusieurs endroits sont mises à jour en même temps ? »

- Convénience : « Possibilité d'accéder facilement à l'information. Existe-t-il des outils ou des technologies pour faciliter le partage et la récupération d'information documentée. »

Un défaut fréquent de ces listes c'est leur hétérogénéité et par conséquent leur arbitraire. Par exemple dans l'exemple donné en note, le critère de *disponibilité* qui fait référence à la localisation de l'information et à son accessibilité c'est à dire au « comment » ne se situe pas sur le même plan que le critère *d'intégrité* qui concerne la crédibilité de l'information, qui va délimiter la portée de l'information. On peut légitimement s'interroger sur l'indépendance des critères de *sélectabilité* et de *convénience*. Cette liste est-elle exhaustive ? Est-elle minimale ?

5.1 La qualité sémiotique : Un concept de qualité adapté à la gestion de la connaissance

La gestion de la connaissance peut être définie comme « fournir la bonne information à la bonne personne au bon moment ». Elle apparaît alors une ingénierie reposant d'une part sur des moyens de traitement de l'information et d'autre part sur une démarche qualité spécifique. Il est donc nécessaire de préciser cette qualité.

Toutes les solutions actuelles qui sont mises en œuvre sous l'étendard de la gestion des connaissances, (portails d'informations, moteurs de recherche, référentiels de connaissances, etc.) visent à améliorer la qualité « sémantique » de l'information produite ou acquise. En réalité il ne s'agit pas d'une qualité sémantique mais de la *pertinence* avec laquelle telle ou telle représentation ou information se manifestent en tant que *signe pour quelqu'un, de quelque chose*. L'importance que représente alors le contexte pour l'évaluation de cette pertinence, indique que ce n'est pas une qualité sémantique qu'il s'agit d'évaluer ou de prescrire mais bien une qualité sémiotique.

On désigne par **qualité sémiotique**, l'adéquation d'un dispositif informationnel à un contexte de production ou d'utilisation de représentations en tant que signes pertinents d'un univers technique donné

Ce dispositif pourra être général. Cette adéquation se décline selon trois aspects (figure 8) :

- **une qualité mémoire** du dispositif : capacité du dispositif à préserver les normes de lectures et d'écriture régissant cet univers technique ;
- **La qualité heuristique** du dispositif : capacité du dispositif à retrouver, consolider, découvrir l'identité d'un objet technique dans cet univers technique ;
- **La qualité épistémique** du dispositif : capacité du dispositif à assurer un niveau de crédibilité et d'adéquation suffisant vis-à-vis de la criticité d'une ensemble de problèmes de cet univers technique ;

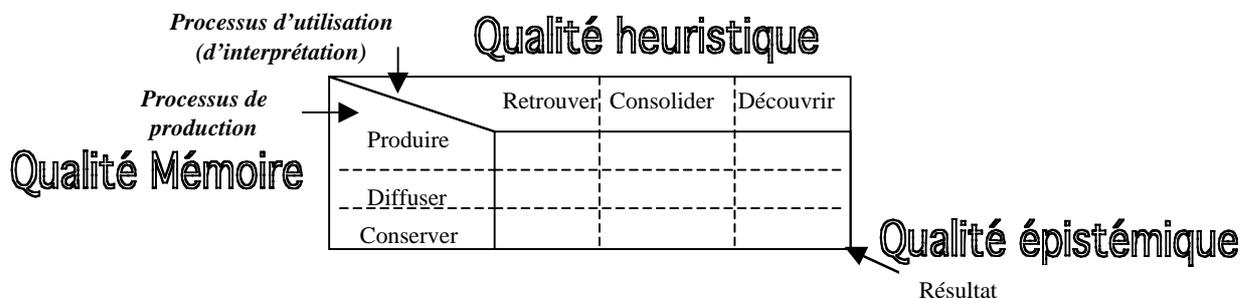


FIG. 8 : Déclinaison de la qualité sémiotique d'un dispositif informationnel en ses trois constituants

Une *qualité mémoire* est relative à un contexte de production de représentation (données, informations ou connaissances) pour : la production stricto sensu (ici et maintenant), la diffusion (reproduction ailleurs), la conservation (reproduction ultérieure).

Une *qualité heuristique* est relative à un contexte d'utilisation (d'interprétation) de représentations (données, informations, connaissances) pour les retrouver ou trouver à partir d'éléments définissant partiellement leur identité

Une *qualité épistémique* est relative à « l'effet de crédibilité » que ces représentations sont susceptibles de produire sur celui ou celle qui les appréhende. Il s'agit de la crédibilité du résultat de la production ou de l'interprétation d'une représentation.

5.2 Examen de la qualité sémiotique dans le cas de données

Ce qui importe pour un producteur de données, c'est que ses données soient manipulables d'abord par lui-même et ensuite par d'autres, avec une facilité suffisante ; autrement dit que son point de vue et éventuellement d'autres points de vue puissent s'y exercer. Si les « énoncés » définissant ces données

sont rapidement inexploitable, la « qualité mémoire » de ces données, plus exactement du dispositif utilisé pour la production des données sera médiocre.

Ainsi un responsable d'instrument embarqué sur un satellite scientifique qui est responsable de la production de données géophysiques (dites de niveau 2) doit disposer d'une interface adaptée à son activité, c'est-à-dire à son point de vue : accès aux mesures effectuées par l'instrument (dites de niveau 1), traitement de ces mesures pour obtenir les données de niveaux 2 correspondantes.

Un utilisateur devant produire des synthèses ou produire un modèle à partir de ces observations produira des données dites de niveau 3. Il sera placé dans une problématique analogue au producteur de données de niveau 2, sauf que son point de vue sera différent. Les grandeurs géophysiques précédentes seront pour lui des mesures d'entrées pour ses traitements. Il devra disposer à son tour d'une interface adaptée à son activité.

Sur cet exemple, on comprend que les mêmes termes scientifiques que manipulent deux catégories de scientifiques ne vont pas nécessairement jouer le même rôle dans leur activité, i.e. avoir la même participation à la signification des discours scientifiques.

Nous avons ici mis l'accent sur la « qualité mémoire » des dispositifs utilisés pour la production de données en adoptant la position des producteurs. Il faudrait maintenant considérer le problème selon la position des utilisateurs et examiner la question de la qualité heuristique des dispositifs existants ou envisagés⁸. C'est essentiellement la fonction « **Retrouver** » (voir figure 8) qui est concernée par des données au sens strict où nous les envisageons (cf. section 4.1). Les deux autres fonctions, « **Consolider** » et « **Découvrir** », correspondent essentiellement à des situations de communication mettant en scène des informations et des connaissances⁹.

6. Conclusion

La qualité sémiotique du dispositif informationnel de production et d'utilisation de données scientifiques dont nous avons ébauché la description dans cet article, repose sur une alternative à une approche conceptuelle. Cette alternative linguistique, permet d'introduire un schéma de communication au sein duquel la production et l'utilisation des données sont définies. La mise en œuvre de cette approche devrait, à terme, montrer sa faisabilité et son efficacité.

Références

- [1] Francois Rastier. Sémantique et recherches cognitives. Coll. Formes sémiotiques. Presses Universitaires de France, 1991. Paris.
- [2] Jacques Fontanille. Sémiotique du discours. Coll. Nouveaux actes sémiotiques. Presses Universitaires de Limoge. 1998. Limoges.
- [3] Algirdas J. Greimas, Joseph Courtès., Sémiotique. Dictionnaire raisonné de la théorie du langage. Hachette Université. 1979
- [4] Daniel Galarreta, Bernard Rothenburger, "Memory Quality : a proposal to manage the risks of memory loss" in Rosetta
- [5] Alain Herreman, « La topologie et ses signes. Éléments pour une histoire sémiotique des mathématiques ». L'Harmattan. 2000

⁸ La qualité épistémique d'un dispositif informationnel au sens large, recouvre dans le cas de données (relatives à un point de vue), des exigences classiques d'intégrité et « d'objectivité » pour le point de vue considéré.

⁹ Il faudrait examiner pour une vue quelconque, les conditions qui favorisent son passage d'un statut de donnée à celui d'information ou de connaissance (cf. section 4.1).