



FR9403256

CEA DES

R
A
P
P
O
R
T

RAPPORT DES/183

**AVIS :
UNE METHODE D'ANALYSE
DE LA COHERENCE DES DOCUMENTS**

**LIPPOLD B., POMIAN J.,
HENRY J.Y., ELSENSOHN O.**



**CEA INSTITUT DE PROTECTION ET DE SURETE NUCLEAIRE
DEPARTEMENT D'EVALUATION DE SURETE**

RAPPORT DES/183

**AVIS :
UNE METHODE D'ANALYSE
DE LA COHERENCE DES DOCUMENTS**

**LIPPOLD B.* , POMIAN J. * ,
HENRY J.Y.** , ELSENSOHN O.****

**9e colloque international
"Fiabilité et maintenabilité"
LA BAULE, 30 mai - 03 juin 1994**

* CISI/Rungis
** DES/SAMS

Juin 1994

**AVIS : UNE METHODE D'ANALYSE
DE LA COHERENCE DES DOCUMENTS**

**Beate LIPPOLD *, Joanna POMIAN *
Jean-Yves HENRY **, Olivier ELSENSOHN ****

** CISI, Rungis*

*** Institut de Protection et de Sûreté Nucléaire (IPSN) - CEA Fontenay-Aux-Roses*

RÉSUMÉ - Ce document se propose de donner un bref aperçu d'AVIS qui est une méthode de validation de la qualité des documents techniques. L'exposé comprend une présentation de la démarche adoptée qui se fonde sur la méthode K.O.D., la définition des critères de qualité d'un document technique, ainsi qu'une description des moyens d'appréciation de ces critères.

ABSTRACT - *The present document intends to give a short insight into AVIS, a method which permits to verify the quality of technical documents. The paper includes the presentation of the applied approach based on the K.O.D. method, the definition of quality criteria of a technical document, as well as a description of the means of valuating these criteria.*

SOMMAIRE

1. INTRODUCTION	1
2. ANALYSE ACTUELLE	2
3. LE DOCUMENT	3
3.1. LE DOCUMENT : UN FACTEUR DE RISQUE	3
3.1.1. Les connaissances	4
3.1.2. La mémoire cognitivo-sémantique.....	4
3.1.3. Le rédacteur.....	4
3.1.4. Le lecteur.....	4
3.1.5. Le langage.....	5
3.2. LE DOCUMENT : UN MOYEN DE CONTRÔLE	5
4. DEMARCHE	6
4.1. PRÉSENTATION GENERALE	6
4.2. MODELISATION	6
4.2.1. Les concepts	7
4.2.2. Les structures	8
4.2.2.1. La taxinomie.....	8
4.2.2.2. La décomposition	8
4.2.2.3. Les connexions	9
4.2.2.4. L'actinomie	9
4.3. LES CRITERES D'ANALYSE	9
4.3.1. DÉFINITIONS	9
4.3.1.1. La complexité.....	10
4.3.1.2. La clarté	10
4.3.1.3. Les normes.....	10
4.3.1.4. La précision.....	11
4.3.1.5. La complétude	11
4.3.1.6. La pertinence	11
4.3.1.7. L'exactitude.....	11
4.3.2 - Moyens d'évaluation	11
4.3.2.1. La complexité.....	12
4.3.2.2. La clarté	13
4.3.2.3. La précision.....	13
4.3.2.4. La complétude.....	14
4.3.2.5. La pertinence	14
4.3.2.6. L'exactitude.....	16
5. CONCLUSION	17
BIBLIOGRAPHIE	18

1. INTRODUCTION

En France, l'Institut de Protection et de Sûreté Nucléaire (IPSN) est chargé par la Direction de la Sûreté des Installations Nucléaires (DSIN) des analyses de sûreté relatives aux réacteurs à eau sous pression (REP).

L'IPSN évalue la sûreté d'une installation nucléaire à partir des dossiers requis pour la démonstration de sûreté et des notes complémentaires qui sont élaborées par l'exploitant de l'installation ou de ses sous-traitants. Dans le cas des systèmes programmés et programmables du contrôle-commande d'un REP qui sont classés de sûreté, le logiciel fait l'objet d'une analyse particulière qui examine, entre autres, les documents de spécification, de conception, de programmation et d'essais.

Le développement de ces logiciels est organisé par un Plan d'Assurance Qualité (PAQ), qui donne lieu à l'élaboration d'une documentation volumineuse, tout au long du cycle de développement. Cette dernière comprend des documents écrits en langage naturel (spécifications, conception, tests) et le programme lui-même. L'ensemble de la documentation est produit sur une période de temps importante, du fait de l'ampleur de ce type de logiciel.

L'examen conduit par l'IPSN, permet de juger de la pertinence des informations contenues dans ces documents, par rapport à la connaissance technique des experts et à la réglementation en vigueur pour ce projet.

Chaque document est lu non seulement pour comprendre les fonctions réalisées, mais aussi pour vérifier s'il existe des informations superflues (sources de complexité), manquantes ou incohérentes en regard de l'ensemble de la documentation concernant ces logiciels.

La méthode AVIS et son support AVISO aident à réaliser cet examen par une démarche systématique et rigoureuse. De plus, l'analyse d'une documentation produite sur une période longue et faisant l'objet de modifications à chaque nouvelle version de logiciel sera facilitée par la capacité à mémoriser les informations par AVISO.

La méthode s'appuie sur l'analyse linguistique pour constituer des graphes qui représentent les informations décrites dans le document. Cette opération met en relation chaque élément du texte avec le graphe. Par rapport à un langage discursif ou mathématique, la modélisation ainsi opérée présente l'avantage de mettre en évidence les relations entre les informations portées par le texte. La vision globale, qui en découle, permet de focaliser l'attention aussi bien sur le sens que sur les détails associés à chaque information. Cette représentation facilite d'une part, l'examen de la cohérence et de la complétude de ces informations. D'autre part, les références établies entre le texte et le graphisme permettent de lier les anomalies de modélisation au texte originel.

2. ANALYSE ACTUELLE

L'évaluation menée par l'IPSN en ce qui concerne la documentation du SPIN (Système de Protection Intégré Numérique) porte sur la totalité de la documentation associée aux spécifications techniques.

Cette documentation est analysée par plusieurs experts qui confrontent leur point de vue et en discutent avec l'exploitant, pour vérifier que ce dernier a bien pris en compte toutes les règles de sûreté.

3. LE DOCUMENT

Dans le cadre de la Sûreté du Fonctionnement, l'intérêt porté au document se justifie par deux faits.

Premièrement, le document est un moyen de communication et doit, de ce fait, obéir à un certain nombre de règles pour que cette communication se passe bien. Vu sous cet angle, le document présente un facteur de risque qu'il faut surveiller.

Deuxièmement, il est lui-même un moyen de contrôle du bon déroulement d'un projet et ainsi garant de sa qualité.

3.1. LE DOCUMENT : UN FACTEUR DE RISQUE

Dans le cas des documents techniques, le but est de transmettre des informations élaborées par des concepteurs à des spécialistes d'un même domaine. Cette transmission met en lumière le rôle important que joue l'intelligibilité du document. Autrement dit, la compréhension du lecteur dépend en premier lieu de la qualité informationnelle du document.

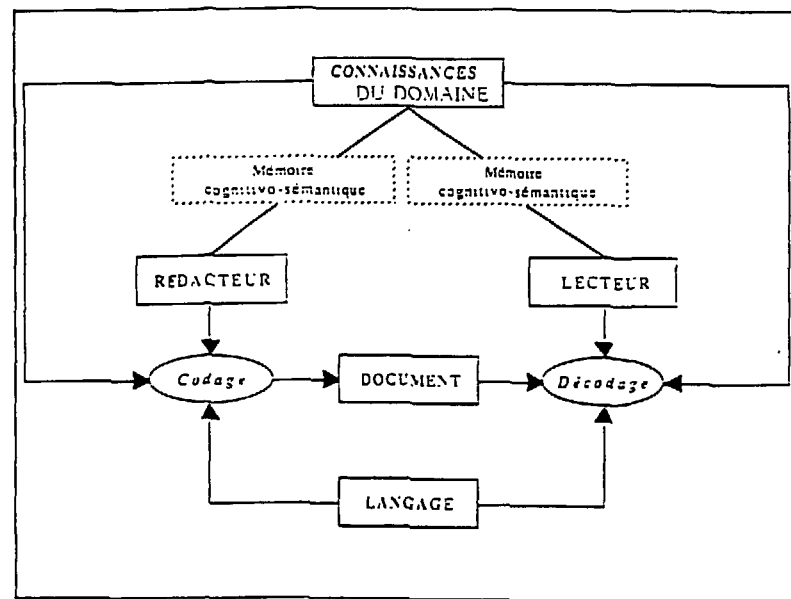


Fig.1 : Schéma de communication par document

Le processus de communication par document se compose en deux phases : le codage de l'information et le décodage du document.

Le codage est réalisé par le rédacteur qui traduit une partie des connaissances techniques représentées dans sa mémoire cognitivo-sémantique personnelle en un document.

Le décodage est effectué par le lecteur qui perçoit le sens des informations à travers sa mémoire cognitivo-sémantique et enrichit sa propre base de connaissances techniques.

3.1.1. Les connaissances

Dans le cas de notre étude, les connaissances sont de type technique et incluent les règles logiques. On manipule donc principalement des objets concrets.

3.1.2. La mémoire cognitivo-sémantique

La mémoire cognitivo-sémantique correspond à la représentation mentale des connaissances d'un individu. Elle constitue le modèle représentant le domaine technique et toute expression faite sur celui-ci. Autrement dit, avant toute formulation, l'individu sollicite sa mémoire cognitivo-sémantique. On peut la représenter comme un réseau de relations complexes entre des concepts. Les règles de mise en relation des concepts sont étudiées par les psycho-linguistes qui eux fournissent des mécanismes universels de conceptualisation.

A part ces règles de mise en relation, la mémoire cognitivo-sémantique est propre à chaque individu. Or, aucune communication n'est possible sans qu'il existe des intersections entre la mémoire de l'émetteur (rédacteur) et celle du récepteur (lecteur).

3.1.3. Le rédacteur

Le rédacteur est donc l'émetteur d'informations, c'est à dire qu'il décrit une partie du domaine qu'il partage avec le lecteur. Pour traduire ses idées, le rédacteur produit un texte en agençant des mots du lexique à l'intérieur des phrases selon les règles syntaxiques et sémantiques propres à une langue. Afin d'obtenir un texte, il doit en plus organiser les phrases à l'intérieur d'unités textuelles plus complexes (paragraphe, sous-chapitres, chapitres, etc.).

M. Fayol [Fay 86] appelle ce niveau de traduction l'arrangement superficiel des propositions et estime qu'il détermine largement la possibilité pour le lecteur de construire une organisation cohérente.

3.1.4. Le lecteur

Le lecteur doit avoir un minimum de connaissances sur le sujet traité par le rédacteur pour pouvoir accrocher les nouvelles informations à l'intérieur de sa mémoire

cognitivo-sémantique, ou alors connaître le domaine en entier pour pouvoir vérifier les informations.

3.1.5. Le langage

On considère appartenant au niveau du langage tout élément propre à une langue et à la constitution d'un document. Pour qu'une communication soit possible le rédacteur et le lecteur doivent partager le même lexique et les mêmes règles morpho-syntaxiques et de structuration de texte.

3.2. LE DOCUMENT : UN MOYEN DE CONTRÔLE

Le document est un des moyens qui permettent d'apprécier la réalisation d'un projet depuis la phase de spécifications jusqu'aux tests. L'intérêt porte uniquement sur le contenu. Les documents produits durant chaque phase du projet ont chacun une forme spécifique, mais le contenu doit être relatif au même domaine (système, équipement,...), même s'il se trouve de plus en plus détaillé au cours de l'avancement du projet et que le point de vue change d'une phase à une autre.

La méthode AVIS introduit un formalisme de modélisation d'une base de connaissance, que l'on appelle référentiel et, qui permet d'évaluer au fur et à mesure de l'avancement du projet les critères portant sur le contenu des documents produits.

4. DEMARCHE

4.1. PRÉSENTATION GENERALE

Le document est dans un premier temps modélisé sous forme de graphes et d'annotations, afin de représenter les éléments contenus dans celui-ci, tels que les objets, les actions, les liens entre les objets, les propriétés associées à ces objets, l'enchaînement des actions, les événements initiateurs de ces actions, ainsi que les contraintes portant sur ces actions.

La vision globale et synthétique offerte par la représentation des informations sous forme de graphique facilite pour l'analyste, la vérification de la cohérence des informations, entre elles et par rapport à sa connaissance du système. Il est à noter que des incohérences ou des problèmes de compréhension peuvent être révélés déjà lors de la modélisation, soit par la présence d'informations contradictoires par rapport à celles déjà modélisées, soit parce qu'il n'est pas possible de modéliser ces informations.

Dans un second temps, une analyse plus détaillée (analyse interne du document, analyse par rapport à un référentiel) est effectuée à l'aide de critères linguistiques, décrits dans le paragraphe suivant. Ces derniers vont permettre de mettre en évidence des incohérences, des incomplétudes, ou des informations non pertinentes, soit directement par une analyse du contenu, soit indirectement par une analyse de la forme du document. Cette analyse va s'appuyer pour certains critères sur la modélisation qui a été effectuée.

Cette approche a été appliquée sur un document avant projet pour donner lieu à une modélisation qui une fois validée sert de référentiel pour l'analyse des autres documents. Une recherche automatique des termes existants à la fois dans les documents à analyser et dans le référentiel est faite, et permet ainsi de comparer plus facilement les informations.

4.2. MODELISATION

La notion de cohérence est fortement liée à celle du contenu, élément variant d'un document à l'autre. Ainsi, une méthode d'analyse de la cohérence doit disposer d'outils qui permettent d'accéder au contenu, de le stocker au fur et à mesure de l'avancement de la lecture et, de l'évaluer à partir d'une certaine grille de critères. Ces critères dépendent du type de document. La méthode AVIS se restreint à l'analyse des documents techniques.

La notion de systématisme [Cha 86] est la base de la démarche adoptée. Cette notion ne se définit pas seulement comme un comportement du rédacteur adopté lors de la

production d'un document, mais aussi comme un comportement du lecteur lors du "décodage" du document.

Le principal apport de la méthode AVIS réside dans la proposition au lecteur d'une approche systématique du texte. L'acquisition des informations ne sera plus liée à la structure formelle du texte qui est linéaire, mais revient à une recherche d'informations dirigée.

Cette recherche est guidée par une base de connaissances, appelé référentiel contenant la représentation conceptuelle du domaine traité dans le document. Le contenu de la base doit être accepté par les lecteurs potentiels du texte, c'est à dire validé par les experts du domaine.

Se basant sur l'hypothèse qu'un sujet élabore une représentation cognitivo-sémantique du domaine au fur et à mesure de l'avancement de la lecture d'un texte, on propose par la méthode K.O.D. (Knowledge Oriented Design) une sorte de macro-structure universelle sous-tendant toute représentation individuelle de différents lecteurs. En outre, on présume qu'un texte est considéré comme bien formé ou *cohérent* quand sa structure reflète la manière dont nous avons l'habitude d'organiser les connaissances du domaine traité. Il sera donc d'autant plus lisible et mémorisable qu'il cadrera avec un schéma cognitif existant.

Tout écart de la structure prédéfinie est identifié comme une incohérence, aussi bien par rapport à l'organisation de l'information qu'à son expression.

De plus, on considère comme nuisible à la compréhension les incomplétudes. Ce sont des informations attendues, mais non verbalisées, qui se manifestent par des champs vides à l'intérieur du réseau sémantique K.O.D.

Le modèle de représentation K.O.D. comprend un ensemble de concepts de base qui sont mis en relation par des liens conceptuels établis *a priori*. Les concepts présentent les points d'entrée dans une structure donnée, les liens permettent de vérifier si tout est dit sur un concept donné.

4.2.1. Les concepts

On distingue deux concepts élémentaires, les objets (éléments statiques) et les actions (éléments dynamiques).

L'objet est une entité physique ou abstraite. Il est défini par un ensemble de propriétés inhérentes représentant le prototype de l'objet et par des propriétés occasionnelles. L'attribution de propriétés occasionnelles permet de décrire des situations identifiées à travers différents états d'un ou de plusieurs objets.

L'action n'est pas un concept simple, mais représente déjà une relation conceptuelle complexe. Une action n'est rien d'autre qu'une mise en relation de plusieurs concepts-objets dont la co-existence poursuit un but (événement désiré).

4.2.2. Les structures

Les structures sont des réseaux conceptuels complexes qui permettent de faire des hypothèses sur des états ou des événements. C'est leur présence qui nous permet de gérer l'implicite.

Le modèle K.O.D. inclut deux types de structures : les structures statiques (classificatoires) et la structure dynamique.

Parmi les structures statiques nous distinguons:

- o la taxinomie
- o la décomposition
- o les connexions

La structure dynamique est appelée actinomie.

4.2.2.1. La taxinomie

La taxinomie est une organisation des objets en classes et espèces. Les classes d'un même niveau s'excluent mutuellement. C'est à dire, un objet ne peut jamais appartenir à deux classes en même temps, ce qui ne va pas sans problème. A l'intérieur d'un même domaine d'activités deux experts peuvent regrouper les objets dans différentes classes, selon leur point de vue. Ceci nous amène à dire, que l'opération de classer n'est pas une opération objective. Or, pour qu'il y ait communication sans ambiguïté, il paraît indispensable de se mettre d'accord sur une représentation commune. Une organisation taxinomique, qui se base sur le principe de l'héritage des propriétés, permet de faire des inférences sur la description des objets.

4.2.2.2. La décomposition

La décomposition est une forme d'organisation d'objets, à l'intérieur de laquelle les relations entre un objet (représentant le "tout") et les sous-objets le composant ("partie du tout") sont explicitées. Ce type de classification semble plus objectif, car il gère des entités concrètes directement observables. Le comportement des propriétés dans un modèle de décomposition n'est pas le même que dans une relation de type classe-instance. En effet,

- o toutes les propriétés du tout ne se transmettent pas à la partie,
- o toutes les propriétés de la partie ne sont pas incluses dans celles du tout.

L'état actuel de nos recherches nous ne permet pas de transmettre des propriétés du tout à chaque partie (pas d'héritage automatique).

4.2.2.3. *Les connexions*

Contrairement aux structures décrites ci-dessus, qui se présentent sous forme d'arborescences, les "connexions" correspondent à une structure de réseau. Elle implique des relations fortes entre les parties. Les objets sont mis en relation par des liaisons physiques directes. Toutes les parties du réseau ont donc une réalité matérielle. Cette structure statique trouve sa justification au niveau dynamique. Les connexions sont choisies afin de garantir le fonctionnement d'un système. Chaque partie du réseau de connexions peut être considérée comme un micro-système qui a besoin d'une liaison physique avec un autre micro-système pour pouvoir fonctionner.

Tout ce que l'on sait sur le comportement des propriétés dans un modèle réseau est qu'il est différent de celui que l'on trouve dans des relations de type classe-instance ou partie-tout. Une étude approfondie ultérieure sera nécessaire pour décrire l'interaction des propriétés des micro-systèmes et leur comportement par rapport aux caractéristiques du macro-système.

4.2.2.4. *L'actinomie*

L'actinomie est une structure qui permet de mettre en relation des actions. Les actions s'articulent sur deux axes : vertical et horizontal. L'axe vertical suit les relations de décomposition entre une action générique et ses actions élémentaires. L'axe horizontal permet de représenter des liens chronologiques entre des actions d'un même niveau de décomposition (enchaînement des actions). Il existe en plus un axe paradigmatique qui met en relation les actions qui peuvent se substituer.

Ce type de structure peut nous servir à vérifier le bon développement thématique à l'intérieur d'un document par analyse de l'exhaustivité thématique et de l'agencement des informations.

4.3. LES CRITERES D'ANALYSE

4.3.1. DÉFINITIONS

Suite aux réflexions exposées ci-dessus, on distingue dès le début deux niveaux d'analyse: la forme et le contenu du document. Or, on se trouve face à une dichotomie complexe.

Bien que les deux concepts soient analysés indépendamment l'un de l'autre, nous sommes obligés de passer par une analyse de la forme pour pouvoir identifier le contenu.

On retrouve ces deux aspects définis plus précisément chez H.Jelitte [Jel 86] qui introduit la différence entre la structure formelle externe d'un texte et sa structure cognitivo-sémantique interne. L'analyse de la structure formelle correspond à l'évaluation de la cohésion du texte. L'analyse de la structure cognitivo-sémantique repose sur la comparaison du contenu informationnel du texte au contenu référentiel.

Compte tenu du but de communication qui découle du type de document, nous avons déterminé un certain nombre de critères qui peuvent parfois qualifier et la forme et le contenu du document. Ces critères sont soit normatifs, soit justifiés par le profil du lecteur final qui doit appliquer le document.

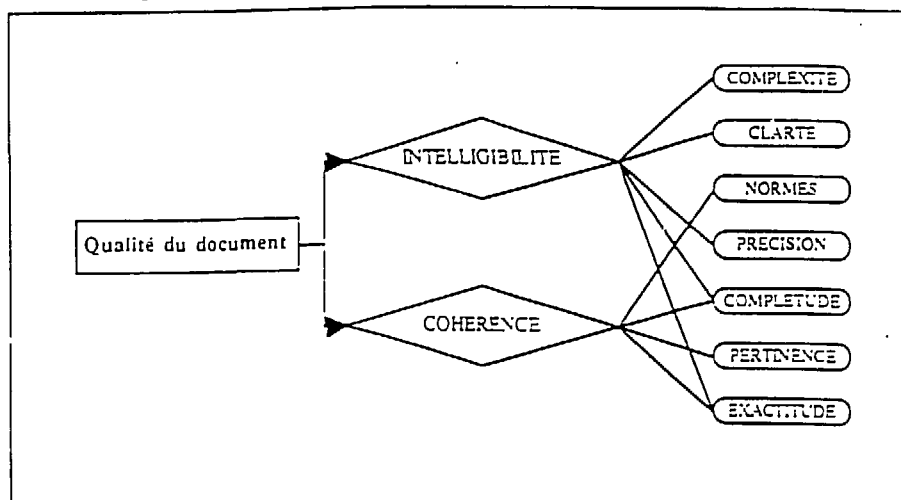


Fig.2 : Classification des critères d'analyse du document

4.3.i.1. La complexité

La complexité est analysée pour juger de l'intelligibilité d'un document. Elle est appréciable au niveau de la forme et du contenu. Dans la méthode AVIS, on se restreint aux aspects formels. Le principe retenu dans le cas de ce critère considère que moins les moyens linguistiques utilisés sont complexes, plus le degré de la compréhension augmente. On recherche donc la simplicité dans la formulation.

4.3.1.2. La clarté

La clarté est également un critère qui permet d'évaluer l'intelligibilité d'un document. Le rédacteur doit utiliser les expressions les plus claires possibles pour permettre au lecteur d'identifier aisément et sans ambiguïté les concepts évoqués dans le document qu'il doit appliquer. L'analyse de la clarté repose également sur une appréciation de la forme.

4.3.1.3. Les normes

L'évaluation du respect des normes correspond au contrôle des règles qui ont été définies par le métier et auxquelles un rédacteur est soumis *a priori*. Il s'agit de conventions d'écriture et de l'énonciation obligatoire de certains thèmes. Le respect de l'application des normes peut donc être évalué au niveau de la forme et du contenu du document. Elles sont strictement dépendantes des habitudes de la communauté au sein de laquelle sont rédigés les documents. De ce fait, elles peuvent ne pas exister.

4.3.1.4. *La précision*

Le but du document technique étant de transmettre des informations nécessaires à l'exécution d'une tâche, chaque imprécision constitue une source d'erreur potentielle dans le projet. Soit elle laisse au lecteur le libre choix entre plusieurs modes d'exécution possibles, ce qui est rarement le but recherché. Soit elle nuit à la compréhension de l'information transmise. Le critère de précision est évalué au niveau de la forme et du contenu du document.

4.3.1.5. *La complétude*

La complétude se décompose en réalité en deux sous-critères. Le premier est évalué par une analyse du document hors contexte, c'est à dire sans connaissance du domaine dont traite le document.

En revanche, le second critère dépend en grande partie du but fixé à cette communication.

Le premier critère est celui de la complétude définitoire. Il permet de contrôler l'intelligibilité du document.

Le deuxième critère est celui de l'exhaustivité thématique. Il permet de contrôler si toutes les informations attendues sont contenues dans le document.

4.3.1.6. *La pertinence*

Le critère de pertinence fait partie de l'analyse du contenu du document. Il s'associe au critère de complétude. Au lieu de détecter les lacunes informationnelles dans un document, il s'agit d'identifier les informations superflues. Le superflu paraît moins nuisible que l'incomplétude, mais représente tout de même une source d'incompréhension, car il augmente inutilement la charge informationnelle du document. Comme le critère de complétude, la pertinence peut être évaluée à deux niveaux. Soit on analyse la pertinence par rapport aux informations attendues, soit on l'évalue sans aucune référence.

4.3.1.7. *L'exactitude*

L'exactitude est un critère qui concerne à la fois la forme et le contenu du document. Au niveau du contenu on traite l'incompatibilité d'une information par rapport à une autre, c'est à dire le cas où il y a contradiction. L'exactitude au niveau de la forme correspond à la conformité de la rédaction aux règles de la langue française.

4.3.2 - Moyens d'évaluation

Après avoir défini les critères d'évaluation de la qualité informationnelle d'un document, on présente ci-après, à partir de quelles caractéristiques on peut les mesurer.

Le tableau suivant semble tracer une frontière nette entre les problèmes de forme et de contenu. Or, cette frontière reste artificielle. Parfois, il est difficile de décider définitivement si on se trouve dans un cas de non-intelligibilité ou de véritable incohérence. Les deux phénomènes se manifestent souvent d'abord par un problème de compréhension. Autrement dit, le lecteur détecte qu'une information dans le texte ne correspond pas à celle représentée au sein de sa propre structure cognitivo-sémantique ou bien il a des difficultés d'identification des concepts ou de mise en relation de concepts.

Le contrôle de l'intelligibilité du document correspond à une analyse des phénomènes de surface, c'est à dire des marqueurs linguistiques. A ce niveau, on ne se préoccupe plus de l'exactitude des propositions, mais plutôt de l'évaluation de la formulation.

Pour l'évaluation de la cohérence le contenu doit être clairement identifié, afin de le modéliser. Pour ce faire, la méthode AVIS utilise le formalisme K.O.D..

	INTELLIGIBILITÉ	COHÉRENCE
Complexité	<ul style="list-style-type: none"> o Absence de synonymie o Syntaxe simple o Bonne structuration du texte 	
Clarté	<ul style="list-style-type: none"> o Bonne structuration du texte o Absence d'ambiguïté lexicale o Absence d'ambiguïté syntaxique 	
Précision	o Absence de marqueurs de flou	o Définitions complètes
Complétude	o Syntaxe complète	o Exhaustivité thématique
Pertinence		o Absence de superflu
Exactitude		o Absence de contradiction logique

4.3.2.1. La complexité

Comme la complexité n'est évaluée qu'au niveau de la forme, l'analyse porte sur des formulations dites "de surface". Pour garantir la simplicité du document, il faut:

- o l'absence de synonymie ou quasi-synonymie,
- o une syntaxe simple,
- o une bonne structuration du texte.

Absence de synonymie

Pour contrôler la synonymie, on dispose de dictionnaires de synonymes et quasi-synonymes.

La désignation d'un concept doit se faire par une seule et même expression. Si pour des raisons stylistiques ou anaphoriques, le rédacteur utilise tout de même des synonymes, il est obligé d'explicitement cet emploi, pour que le lecteur n'ait aucune difficulté à identifier le concept.

Syntaxe simple

Les phrases construites doivent être courtes, c'est à dire ne pas comporter beaucoup de membres principaux ou subordonnés. Cependant, les limites de la longueur des phrases restent floues.

Bonne structuration du texte

Une bonne structuration du texte pour garantir la simplicité du document correspond à un découpage du texte en un nombre de niveaux de chapitres limité. Une hiérarchie de chapitres incluant plus de cinq niveaux, rend un document complexe à aborder. Pour contrôler cette règle, on procède à une analyse du sommaire. Cependant, suivant la nature du document, le seuil peut être différent.

4.3.2.2. La clarté

La clarté dépend des phénomènes suivants :

- o une bonne structuration du texte,
- o l'absence d'ambiguïté lexicale,
- o l'absence d'ambiguïté syntaxique.

Bonne structuration du texte

Selon M. Fayol [Fay 86], on ignore tout des critères utilisés pour estimer si une suite de phrases est ou non cohérente. M. Charolles [Cha 86] remarque qu'il n'existe pas de règles de combinaison d'unités textuelles au delà de la phrase, mais que la seule consigne implicite est de combiner les éléments du contenu de façon à ce que l'interlocuteur comprenne. Il estime que la notion de cohérence peut tout au plus servir à désigner les phénomènes de systématisme dans la mise en oeuvre par un locuteur de certains processus énonciatifs.

Ainsi, la bonne structuration d'un document est garant de la lisibilité. Plus la structuration reflète les schémas de représentation mentale de la connaissance, plus le décodage par le lecteur est facilité.

Les schémas de représentation de la connaissance sont définis et formalisés par K.O.D. La méthode AVIS propose donc un contrôle de l'organisation des idées contenues dans le document par une comparaison avec ces schémas.

Absence d'ambiguïté lexicale

L'ambiguïté lexicale est définie comme l'incapacité d'identifier clairement le concept désigné par un mot.

Absence d'ambiguïté syntaxique

L'ambiguïté syntaxique est définie comme l'incapacité d'identifier clairement la mise en relation de concepts verbalisée dans une phrase.

4.3.2.3. La précision

Pour apprécier cette caractéristique dans le document, la méthode s'intéresse à :

- o l'absence de marqueurs de flou,
- o la complétude des définitions.

Absence de marqueurs de flou

Cette évaluation de forme est effectuée par une recherche des marqueurs linguistiques indiquant le flou. Ces marqueurs appartiennent soit au lexique (exemples : les adjectifs certains, plusieurs) soit au système de ponctuation (exemples : les trois points qui indique une suite). L'ensemble de ces marqueurs est stocké dans un dictionnaire exploitable informatiquement.

L'analyse de la précision correspond donc à la détection automatique de ces marqueurs dans le texte.

Définitions complètes

La cohérence d'un document est largement déterminée par une définition suffisante des concepts traités. Tout nouveau concept introduit dans le document analysé doit être défini de façon précise. La méthode AVIS se base sur deux méta-concepts, l'objet et l'action, dont les définitions minimales sont constituées de certains traits caractéristiques. Il est vérifié l'existence de liens sémantiques du concept avec un autre concept déjà défini (dans le référentiel constitué lors de l'analyse de documents amonts ou dans le document même). Le document doit expliciter au moins deux des trois traits suivants pour la définition d'un objet :

- | | | |
|--------------------|---|--|
| 1) père / fils | : | l'objet appartient à une <i>classe d'objets</i> |
| 2) tout / partie | : | l'objet est partie d'un <i>objet</i> |
| 3) acteur / action | : | l'objet joue le rôle de destinataire, sujet ou ressource d'une <i>action</i> |

La définition minimale d'une action se fait à travers les trois traits suivants :

- 1) action / sujet : l'action est effectuée par un *objet*
- 2) action / destinataire : l'action "transforme" un *objet*
- 3) action / actinomie : l'action est une étape d'une *action générique*

4.3.2.4. La complétude

L'analyse de la complétude s'effectue au niveau de la forme et du contenu et porte respectivement sur:

- o la syntaxe,
- o l'exhaustivité thématique.

Syntaxe complète

Pour garantir l'intelligibilité du document, les phrases doivent avoir une structure minimale (sujet, verbe, complément) qui laisse moins de possibilités d'interprétations. L'incomplétude est identifiable avec un analyseur syntaxique qui stocke les anomalies rencontrées au cours de son traitement. Ainsi, un style télégraphique apparaîtrait comme une source d'incomplétude.

Exhaustivité thématique

La détection des incomplétudes à ce niveau demande une consultation préalable du référentiel pour savoir ce qui est attendu. Dans ce but, on analyse les structures d'organisation de la connaissance définies dans la méthode.

On peut également effectuer une évaluation de ce critère hors contexte en analysant l'ensemble d'une structure comme par exemple la décomposition d'un objet. Il s'agit d'apprécier l'homogénéité des descriptions de type père/fils ou tout/partie (arbres structurants). Dans le cas où une branche se trouve moins développée que toutes les autres, on peut soupçonner une incomplétude informationnelle.

4.3.2.5. La pertinence

La pertinence est un critère d'évaluation du contenu. Pour garantir la cohérence du document, il faut éviter le superflu. On distingue, comme c'est le cas pour l'exhaustivité thématique, l'analyse de la pertinence dans ou hors contexte.

Dans le premier cas, il s'agit d'une pertinence *relative*, l'évaluation se fait par rapport au domaine formalisé à l'intérieur d'un modèle (référentiel).

Dans le deuxième cas, on évalue une pertinence *absolue* qui est indépendante de toute connaissance explicitée dans un modèle prédéfini.

Absence de pertinence relative

L'évaluation se fait par consultation d'un référentiel. Si une description contient plus d'éléments que d'autres à un même niveau de détail, certains éléments devraient être jugés non-pertinents, ou alors il convient de modifier le référentiel. Dans ce dernier cas, cela signifie que le contenu du projet a évolué et une mise à jour des documents en amont devient nécessaire.

Absence de pertinence absolue

La pertinence absolue est évaluée par l'analyse de l'ensemble d'une structure d'organisation. Il s'agit d'observer l'homogénéité des arbres structurants. Dès qu'une branche est plus développée que les autres, elle contient vraisemblablement des éléments superflus.

4.3.2.6. L'exactitude

L'analyse de l'exactitude s'effectue au niveau du contenu. Cette analyse s'intéresse respectivement à :

- o l'absence de contradiction logique.

Absence de contradiction logique

Le seul cas de contradiction logique qui est actuellement traité par la méthode est celui de l'incompatibilité d'une information par rapport à une autre, c'est à dire le cas de la contradiction. Les informations comparées sont les suivantes :

- o information document vs information référentiel,
- o information document vs information document.

5. CONCLUSION

La méthode AVIS, par le systématisme qu'elle introduit et par sa capacité de modélisation du contenu informationnel des textes rédigés en langage naturel, est une aide à l'examen critique de documents.

AVIS s'applique dans tout contexte de vérification de contenu de documents, spécialement lorsque le vérificateur est extérieur à la structure qui produit le document, par exemple les services d'Assurance Qualité de grands organismes. Dans le cas où le vérificateur appartient à la même structure, les critères d'AVIS peuvent être utilisés en amont en tant qu'aides à la rédaction.

La validité de la méthode a été confirmée dans certains contextes d'analyse. La facilité de son application peut être améliorée par des investissements en matière de support informatique. En outre, l'avancement dans la recherche des formalismes de modélisation de bases de connaissances de plus en plus proche du langage naturel, promet d'arriver à une maniabilité encore plus aisée de ce genre de méthode.

BIBLIOGRAPHIE

- [Bie 93] B. Biébow, S. Szulman: *Disert: a tool for the detection of incoherences in software engineering requirements texts*, 1993.
- [Czu 88] A.J. Czuchry & D.R. Harris: *KBRA: a new paradigm for requirements engineering*; IEEE Expert, Winter 88.
- [Ros 77] D.T. Ross & K.E. Schomann: *Structured analysis for requirements definition*; IEEE Trans. on Softw. Eng., special issue on requirements analysis, vol. SE-3, January 1977.
- [Sac 89] M. Sacki, H. Horai & H. Enomoto: *Software development process from natural language specification*; Proc. of the 11th ICSE, Pittsburg, USA, May 1989.
- [Cha 86] M. Charolles: *Le problème de la cohérence dans les études françaises sur le discours durant 1965-1975*; Research in Text Connexity and Text Coherence, Helmut Buske Verlag, Hamburg, 1986.
- [Dra 86] E. Dragos: *Cohesion and coherence in Romanian linguistics*; idem.
- [Fay 86] M. Fayol: *Cohérence et cohésion: une revue des travaux français de psychologie expérimentale*; idem.
- [Jel 86] H. Jelite: *Die Textkohärenz als Forschungsgegenstand der Sowjetischen Textlinguistik*; idem.
- [Vog 88] C. Vogel: *Génie cognitif*, Masson, Paris, 1988.