

Exposé de soutenance de la thèse : *L'apport de la modélisation des connaissances à la codification et à la simplification des textes normatifs.*

Christian Tremblay – 28 mai 2002

Cette thèse consiste en une analyse préalable au cahier des charges d'un programme de développement relatif au traitement assisté par ordinateur des textes normatifs dans un but de codification et plus généralement d'accessibilité du public aux textes normatifs.

Cette analyse préalable tente de répondre à des questions essentielles :

- peut-on modéliser sans appauvrir le sens ?
- peut-on modéliser directement à partir de l'analyse textuelle ?

Nous répondons positivement à ces deux questions sous deux conditions :

- de l'application d'une approche systémique respectant la complexité fondamentale
- du recours aux ressources de la linguistique

L'analyse textuelle apparaît comme le niveau incontournable , qu'elle soit manuelle ou assistée par ordinateur, qui mène du texte à sa modélisation. Elle sera le fil directeur de toute la démarche.

L'exposé aura une orientation résolument opératoire plus que théorique.

Il est sensé être compréhensible sans qu'il soit nécessaire d'avoir lu la thèse.

Cette orientation a comme contrepartie de devoir commencer par rappeler quelques notions de base, outils fondamentaux sur lesquels s'appuie la démonstration.

Nous traiterons ensuite de cas pratiques d'interprétation, de cas pratiques de compréhension et d'interaction, et de cas pratiques d'opérations normatives.

Enfin nous concluons par quelques conséquences au plan des développements informatiques.

Plan

- Notions de base
- Trois cas pratiques d'interprétation
- Un cas pratique de compréhension et d'interaction en langage naturel
- Cas pratiques d'opérations normatives
- Quelques implications au plan informatique

Notions de base et définitions

Toutes les notions de base énoncées ci-après sont issues à titre principal de l'œuvre de Bernard Pottier, complétés, s'agissant en particulier de la micro-sémantique, par les travaux de François Rastier.

- Notion de schème

- Théorie des voix
- Micro-sémantique
- Sémiogramme

Notion de schèmes

Schèmes conceptuels

- Toute phrase, tout énoncé comporte un support conceptuel. Toute phrase, tout énoncé est la manifestation formelle d'une ou plusieurs relations sous-jacentes, se situant à un niveau appelé conceptuel.

- **Niveau conceptuel**

Le niveau conceptuel se définit comme indépendant des langues : c'est le niveau où s'opèrent tous les phénomènes de compréhension profonde, tous les phénomènes de mémorisation, et tous les transferts de langue à langue.*

- **Le propos, la base et le prédicat**

Chacune des relations fondamentales qui est sous-jacente à un énoncé ou un texte est appelée un **propos** (ce que l'on se propose de communiquer).

Le propos comprend deux éléments : la **base** (ce sur quoi l'on se propose de communiquer) et le **prédicat** (ce que l'on communique à propos de la base)

schèmes linguistiques ou schèmes d'entendement

Partant du niveau conceptuel, l'énonciateur procède à un certain nombre de choix qui vont modeler l'énoncé.

Opérations obligatoires

- Schème d'entendement (SE)
choix des lexèmes \longrightarrow modules actantiels
- Schème prédiqué (SP)
choix prédicatif = détermination de la base de vision

Opérations facultatives

Nous n'en donnons ici qu'une énumération

- Topicalisation
- Focalisation
- Impersonnalisation
- Hiérarchisation phrastique (juxtaposition, coordination, subordination)
- Réduction d'actance
- Détermination-Modalisation
- Aspectualisation

Le résultat de ces transformations successives mais instantanées dans l'acte d'énonciation est le Schème résultatif (SR).

Ces opérations existent dans toutes les langues, et relèvent de la linguistique générale. Elles sont codifiables.

Théorie des voix

Dès que l'on quitte le domaine conceptuel, c'est-à-dire dès le choix des lexèmes, on entre dans des contraintes sémantico-syntactiques propres à chaque langue, car chaque lexème entraîne un certain type de module actantiel.

La théorie des voix donne une classification sémantique des modules actantiels.

La théorie des voix permet de distinguer entre :

- | | | |
|-----------------------------|----|--------------------------|
| – Pierre traverse la rue | ou | Pierre arrive de Paris |
| et | | et |
| – La chat regarde la souris | | Pierre est âgé de 30 ans |

La théorie des voix comporte une nomenclature de base de cinq voix : existentiel, équatif, situatif, descriptif, subjectif, lesquelles possèdent trois statuts: statif, évolutif, causatif, lesquels statuts sont eux-mêmes décomposables et traversent des champs sémantiques variés. Il est impossible d'en dire plus à ce niveau.

Micro-sémantique 1 : Sémèmes et ensembles de définition

•Sémème

Contenu d'un morphème (signe minimal ou signifiant). Le sémème se compose de sèmes ou traits sémantiques.

•Taxème

Classe sémantique au sein de laquelle les sémèmes sont inter-définis.

•Domaine

Classe sémantique composée de taxèmes, tel que dans un domaine donné il n'existe pas de polysémie. Le domaine est lié à des normes sociales et à l'expérience du groupe (ex.: alimentation, musique, politique, droit, etc.)

•Dimension

Classe sémantique de généralité supérieure, indépendante des domaines. Ex.: matériel/immatériel, animé/inanimé

•Dans une représentation objet un sémème appartient à un taxème, lequel appartient à un domaine et à une dimension. Dans un même domaine, on rencontrera des taxèmes de dimensions différentes.

Micro-sémantique 2 Sèmes

•Générique

- Le sème générique marque l'appartenance du sémème à une classe sémantique (taxème, domaine, dimension)
- L'ensemble des sèmes génériques constituent le classème
- Un même sème générique rapproche deux sémèmes par une relation d'équivalence

•Spécifique

- Sème permettant à l'intérieur d'un même taxème de distinguer les sémèmes entre eux.
- L'ensemble des sèmes spécifiques constituent le sémantème
- Les sèmes spécifiques (différents par définition) permettent d'opposer des sémèmes voisins par une relation d'incompatibilité.

•Inhérent

- Un trait inhérent est un trait codifié en langue, c'est-à-dire indépendant du contexte.

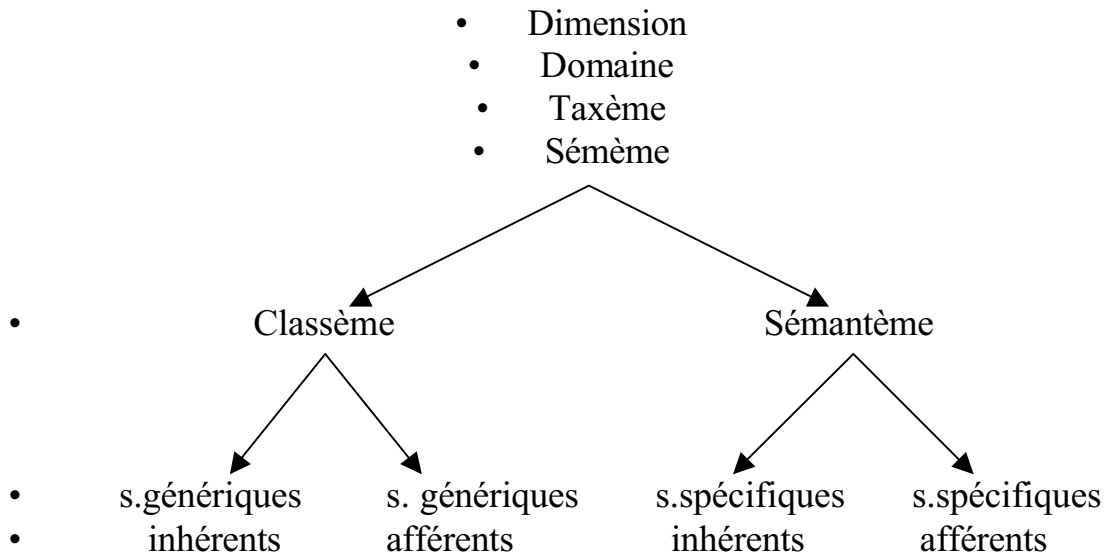
–Un trait inhérent peut être générique ou spécifique

•Afférent

–Un trait afférent est déterminé par le contexte textuel (cotexte) ou situationnel (pragmatique).

–Un trait afférent peut être générique ou spécifique

Schéma récapitulatif



Micro-sémantique 3 : polysémies

- Polysémie

Pour un même signifiant, on peut avoir plusieurs sémèmes. Ex. bureau

- Polysémie d'emploi

Il y a polysémie d'emploi lorsque deux sémèmes diffèrent par au moins un sème localement afférent. Ex.: convoi militaire et convoi funéraire

- Polysémie d'acception

Il y a polysémie d'acception, lorsque deux sémèmes diffèrent par au moins un sème afférent socialement normé. Ex. : minute et minute

- Polysémie de sens

Il y a polysémie de sens, lorsque deux sémèmes diffèrent par au moins un sème inhérent.

- Homonymie

Deux sémèmes diffèrent par au moins tous les sèmes spécifiques

Sémiogramme

- Structure intermédiaire entre le sémème et la connaissance encyclopédique.
- Le sémème ne contient que les sèmes qui sont strictement nécessaires à son identification.
- Le sémiogramme comprend toutes les « relations sémantiques essentielles gravitant autour d'un sémème : association, inclusion, opposition, participation ».
- Il constitue l'environnement sémantique du sémème ou encore la « visualisation sémantique du sémème » dans un domaine donné.

Nota : notion distincte de celle de domaine notionnel selon Culioli, qui inclut tous les sens, acceptions et emplois (cf. p. 256 Thèse), alors que le sémiogramme se situe dans un domaine sémantique défini comme excluant la polysémie.

Trois cas pratiques d'interprétation

- Sur la notion de « bureau »
- « Il est interdit de marcher sur la pelouse »
- Loi PML et conseils d'école

1^{er} cas pratique : Sur la notion de « bureau »

Parmi les nombreux sens du mot bureau (sans distinguer pour la simplicité de l'exposé les différences de sens, d'acception ou d'emploi), nous en retenons trois : le bureau, synonyme de table de travail, le bureau, pièce où l'on travaille, le bureau désignant le lieu de travail, travail tertiaire par opposition à l'usine ou à l'atelier.

Nous sommes dans un cas de polysémie de sens caractérisée dans la mesure si le mot travail est un sème inhérent commun, donc générique, aux trois taxèmes "meuble", "pièce-local", "lieu de travail", les sèmes "meubles", "pièce-local", "lieu de travail tertiaire", sont bien des sèmes distinctifs, donc spécifiques. Il est intéressant d'observer qu'en espagnol, à ces trois sens du mot bureau correspondent des mots différents : escritorio, despacho, oficina. Il en est de même en anglais : desk, study, office.

D'autres correspondances pourraient être identifiées au niveau des polysémies d'acception ou d'emploi (ex. : bureau de poste : post office, bureau de vote : poll station).

Il est donc essentiel qu'au niveau du texte source on puisse par l'analyse textuelle identifier le sème jouant le rôle de sélecteur de sens permettant de déterminer la bonne interprétation dans la langue source et le mot correspondant dans une langue cible.

Un nombre plus ou moins grand de sèmes peut, selon le contexte, jouer ce rôle. Les exemples qui suivent s'en tiennent au cotexte le plus immédiat. Ainsi :

Recherche d'un sélecteur :

1. Il travaille à son bureau (sème spécifique : extériorité)
il est assis à son bureau (mais « il est assis dans son fauteuil »)
il est à son bureau
le dossier est posé sur son bureau
il est endormi sur son bureau
il lit à son bureau
il écrit à son bureau

2. Il travaille dans son bureau (sème : intériorité)
il est assis dans son bureau
il est dans son bureau
le dossier est dans son bureau
il est endormi dans son bureau
il reçoit dans son bureau
il lit dans son bureau

il écrit dans son bureau

3. Il travaille au bureau (sème : lieu)

il est parti au bureau

il reçoit au bureau

il est en réunion au bureau

sa vie se passe au bureau

Choix du sélecteur

- « extériorité » pour bureau(1)
- « intériorité » pour bureau(2)
- « lieu (flou) » pour bureau(3)

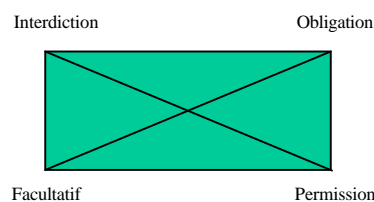
On peut discuter du choix du sème faisant office de sélecteur. Il est évident que si le sème inhérent spécifique est déjà présent, la sélection est immédiate. Le cas traité ici est un peu plus subtil.

En tout état de cause, nous dégageons de cette rapide analyse la conclusion que la problème de la polysémie n'est pas soluble, aussi bien au niveau de l'interprétation dans une seule langue, qu'à celui de la traduction automatisée par ordinateur, sans recours à l'analyse sémique, et notamment à la mise en évidence des isosémies ou des isotopies, faits scientifiques aussi certains et opératoires que le sont les courbes de niveau en géographie physique.

2^e cas pratique : « Il est interdit de marcher sur la pelouse »

L'exemple suivant, emprunté à Paul Amselek, vise à illustrer la difficulté d'une interprétation logiciste de la norme telle qu'on pourrait la tenir à partir de l'article 5 de la DDHC selon laquelle tout ce qui n'est pas défendu par la loi ne peut être empêché, et selon les préceptes du maître de Vienne, Hans Kelsen.

L'énoncé « Il est interdit de marcher sur la pelouse » peut être interprété au moyen du carré déontique, selon le schéma ci-après.



Nous laissons de côté l'obligation et le facultatif qui donnent des énoncés logiquement valide, mais linguistique non valide. Limitons-nous à la permission.

En fait nous avons deux éléments dans la proposition qui doivent combinés : « marcher » et « pelouse ».

Donc, l'interdiction de marcher sur la pelouse peut avoir trois conséquences au regard de la permission :

Pour « pelouse » :

- marcher en dehors de la pelouse : OK

Pour « marcher » :

Il est permis de :

- ne pas marcher sur le pelouse : OK (il suffit de s'abstenir de marcher)

- faire autre chose que marcher sur la pelouse: NON OK. En effet, il est clair que s'il est interdit de marcher sur la pelouse, il est également interdit d'y faire du vélo ou du roller, c'est-à-dire toute activité susceptible de détériorer la pelouse, et à laquelle celle-ci, naturellement fragile, ne pourrait résister longtemps.

Autrement dit, la glose sur les termes « interdiction, obligation, permission, facultatif », à laquelle conduit l'application du carré déontique est d'un intérêt médiocre, car le problème d'interprétation ne concerne pas ces termes, mais le sens qu'il faut donner à « marcher sur la pelouse ».

Si « pelouse » est physiquement parfaitement circonscrit, il n'en va pas de même de « marcher ».

Ici, « marcher » ne désigne par seulement l'activité de « marcher », mais toute activité à laquelle se livreraient les utilisateurs du jardin public et qui aurait le même effet destructeur. Nous concluons donc qu'il est nécessaire de se livrer à un processus interprétatif pour déterminer la portée de l'interdiction de « marcher sur la pelouse » = « ...et tout activité similaire » et que pour ce faire, il y a lieu de faire appel à la notion de prototype bien connue en sémantique.

Accessoirement, nous soulignerons que si le « principe de complétude » de Kelsen dont l'application la plus typique est le principe selon lequel « tout ce qui n'est pas interdit » est permis » n'est pas contestable, il n'est néanmoins opératoire qu'au prix d'un détour interprétatif complexe.

3^e cas pratique : Loi PML et conseil d'école

Le problème posé vient de la récente loi du 27 février 2002 sur la démocratie de proximité qui a soulevé une question de compatibilité avec le décret du 6 décembre 1990, pris pour l'application de la loi du 10 juillet 1989 sur l'éducation, et qui régit en son article 17 les conseils d'école.

Ainsi le décret dispose que le conseil d'école comprend le Maire ou son représentant et un membre du conseil municipal.

La loi modifiant la loi PML (art. L.2511-19 du CGCT) introduit la disposition que le conseil d'arrondissement procède, en son sein, à la désignation des représentants de la commune dans les organismes dont le champ d'action est limité à l'arrondissement, et précise que cette disposition est applicable aux conseils d'école.

Se pose en premier lieu un

• Problème de hiérarchie des normes juridiques

Une loi d'application territorialement limitée l'emporte a priori sur un décret même de portée nationale. Il faut néanmoins pour cela qu'il y ait une contradiction non susceptible d'être levée.

La question porte alors sur l'interprétation de la notion de « représentant de la commune ».

Le représentant de la commune est-il :

- L' élu du conseil municipal

ou

- L' élu du Conseil municipal et le Maire ou son représentant

Il n'y a en fait qu'une seule réponse : l' élu du conseil municipal. Le Maire qui est légalement à la fois chef de l'exécutif municipal et représentant de l'Etat ne peut être remplacée par un représentant de la commune désigné en son sein par le conseil d'arrondissement. La loi n'a pas pu davantage vouloir substituer par la disposition en question le maire d'arrondissement au Maire de la commune.

Solution : un problème de hiérarchie des normes; un processus interprétatif complexe concernant la notion de représentant du maire. Tout système permettant de traiter ce genre de

problème devrait être doté d'une capacité de raisonnement sur la hiérarchie des normes relativement puissant.

Cas pratique de compréhension et d'interaction en langage naturel

1 – problème de paraphrase

Le nombre de paraphrases pour un énoncé varie de zéro à l'infini, l'élément incontrôlable étant le niveau de mobilisation des connaissances encyclopédiques qui dépend de l'énonciateur. Toutefois, pour un niveau déterminé de mobilisation de ces connaissances, on peut contrôler la création paraphrastique.

Prenons l'exemple du premier alinéa de l'article 6 de la Constitution française pour lequel nous proposons 4 formulations s'écartant très peu du texte original :

• Art. 6C – Le Président de la République est élu pour cinq ans au suffrage universel direct.

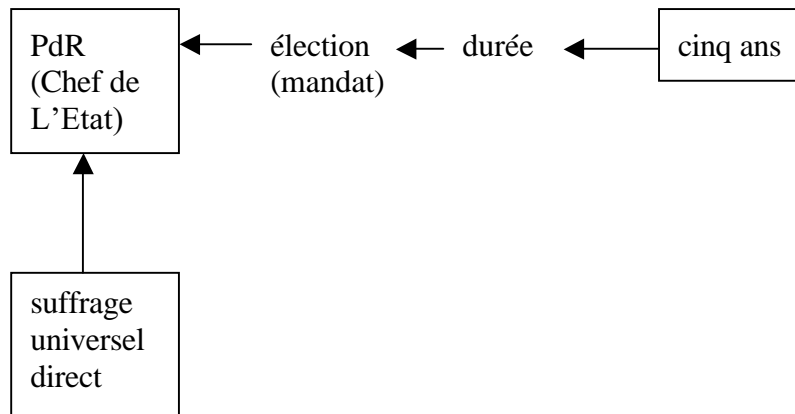
- Formulation 1 : Le PdR est élu au suffrage universel direct pour cinq ans
- Formulation 2 : Le PdR est élu au suffrage universel direct pour une durée de cinq ans
- Formulation 3 : Le PdR est élu au suffrage universel direct. La durée de son mandat est de cinq ans.
- Formulation 4 : L'élection du PdR a lieu au suffrage universel direct. Le mandat du chef de l'Etat est de cinq ans.

On peut mesurer les écarts de signification :

- F0 ⇒ F1 : simple déplacement de « pour cinq ans » non significatif.
- F1 ⇒ F2 : remplacement de « pour cinq ans » par « pour une durée de cinq ». Le remplacement n'apporte aucun élément significatif supplémentaire, le sème /durée/ étant déjà présent dans « élection » et dans « pour cinq ans »
- F2 ⇒ F3 : césure, anaphore (« son »), introduction de la notion de mandat. Or, la notion de mandat, doit être associée à « élection » par une relation d'association dans le sémiogramme.
- F3 ⇒ F4 : topicalisation sur le verbe qui substantivé devient la base. Coréférence « chef de l'Etat » ⇒ « PdR » (le sémiogramme doit spécifier que le Président de la République est par définition Chef de l'Etat dans une République). Effacement de « durée », le sème /durée/ étant de toute façon présent dans le sémème de « mandat ».
- F4 ⇒ F1 : pour revenir à F1 il faut opérer 3 réductions :
 - « Mandat » à « élection » (sémiogramme)
 - « son » à « pdR »
 - « durée » à « 5 ans »

Un système de compréhension doit être capable de calculer l'équivalence de ces différentes formulations. On observera que les deux termes nouveaux, mais dont la présence s'impose dans le sémiogramme, sont « mandat », attaché à élection, et « Chef de l'Etat », indissociable de « Président de la République ».

Le schème correspondant à ces 5 formulations est le suivant.



2 – Forme interrogative

Nous proposons 6 questions tournant autour de l'énoncé initial :

1. Comment est désigné le PdR?
2. Comment est élu le PdR?
3. Selon quel mode de scrutin le PdR est-il élu? Depuis quand le PdR est-il élu pour 5 ans?
5. Pour combien de temps le PdR est-il élu?
6. Quelle est la durée du mandat du PdR?

Nous pouvons observer 2 caractéristiques des questions :

- La question porte sur une partie seulement du schème (Q1, Q2, Q3 ⇒ type de désignation, Q4, Q5, Q6 ⇒ durée)
- La question peut être plus large que le schème. En effet, Q1 évoque le mode de désignation qui peut ne pas être électif, ce qui renvoie à l'arborescence du taxème auquel appartient « élection » ; Q3 évoque le mode de scrutin qui ne se résume pas au type d'élection, ce qui renvoie à des d'autres dispositions de la Constitution ; Q4 évoque la durée mais fait appel à une connaissance purement encyclopédique qui est la date à partir de laquelle la durée du mandat est passée de 7 à 5 ans.

Pour répondre donc aux questions 1, 3 et 4, il convient que le système détermine si la réponse se trouve localement dans le schème ou s'il doit la chercher ailleurs dans la base de connaissances.

Cas pratiques d'opérations normatives

Il s'agit de ce que je considère comme étant l'apport principal de mon travail. L'opération normative est une structure sémantique typique des textes normatifs. J'ai isolé une vingtaine de ces structures depuis le principe général (PG), jusqu'à la norme prescriptive de la norme applicable (NPNA), en passant par la norme prescriptive d'une interdiction (NPI). Nous en traiterons ci-après à partir d'exemples.

- Ex. de norme descriptive de structure (NDS) :

Art. 6C – Le Président de la République est élu pour cinq ans au suffrage universel direct.

- Ex. de norme descriptive de procédure (NDP) :

Art. 7C – Le PdR est élu à la majorité absolue des suffrages exprimés. Si celle-ci n'est pas obtenue au premier tour de scrutin, il est procédé, le deuxième dimanche suivant, à un second tour.

- Ex. de Norme attributive de compétence (NAC) :

Art. 8C – Le PdR nomme le PM. Il met fin à ses fonctions sur présentation par celui-ci de la démission du gouvernement. **Pertinence externe/interne**

- P1 : Le PdR est élu pour cinq ans au suffrage universel direct (EQU EVO¹)
- P2 : Le PdR nomme le Premier Ministre (EQU CAU¹). Il met fin à ses fonctions sur présentation par celui-ci de la démission du Gouvernement (EXI CAU³)
- P3 : l'homme astiquait patiemment son fusil (DES CAU²). Soudain, la porte s'ouvrit (DES EVO¹)
- P4 : Depuis quelques années, la notion de politique linguistique est au cœur de nombreuses réflexions et de nombreuses pratiques. (SIT NOT) Intuitivement seuls P1

et P2 correspondent à des opérations normatives. Nous voudrions isoler les éléments permettant de les distinguer entre eux et d'établir que P3 et P4 ne peuvent relever du domaine des textes normatifs.

Interne

–P1/P2 : P1 = NDS = équatif évolutif; P2 = NAC = équatif causatif. ON peu déduire de P2 une NDS (« le Premier Ministre est nommé par le Président de la République, qui met fin à ses fonctions sur présentation par lui-même de la démission du gouvernement »)

Externe

- P1bis : le PdR est élu meilleur footballeur de l'année. Une telle transformation de P1 sort P1 du domaine des textes pouvant relever de celui des textes normatifs. En effet, dans P1bis, c'est la personne en tant que personne et non en tant que PdR qui est élue, on note une incongruité entre la fonction et le titre de meilleur footballeur, « de l'année » introduit un élément purement événementiel, non permanent qui le fait échapper à la normativité. Il en irait autrement si l'on disait : « chaque année il est procédé à l'élection du meilleur footballeur de l'année ».
- P3 par opposition à P1 présente un caractère personnel et événementiel, anecdotique qui fait échapper P3 au domaine de la normativité.
- P4 : des expressions telles que « depuis quelques années », « la notion de », « réflexion », sont étrangères à toute normativité.
- P4bis : « **La politique linguistique est une priorité nationale** ». Ainsi reformulé, P4 serait susceptible d'être considéré comme un texte normatif.

Nous devons tirer de ces quelques exemples une conclusion importante : aucun trait inhérent n'est suffisant pour établir le caractère normatif d'un texte. Aucun texte n'est par lui-même normatif. Il faut qu'il le soit de surcroît formellement, directement ou indirectement. Il faut un signe extérieur particulier : aujourd'hui, la signature et la publication au journal officiel, l'affichage sur les panneaux officiels, etc. Sous l'ancien régime, l'annonce par le héraut du roi sur la place du village, etc. Mais la citation conserve au texte cité sa valeur normative...

Externe : traits normatifs + afférence normative

Interne : traits spécifiques

Conséquences opératoires en informatique

Deux orientations sont particulièrement nécessaires :

- Extensibilité permanente de la base de connaissances
 - En intension (ajout ou différenciation de sèmes, insertion d'hyperonymes ou d'hyponymes)

- En extension : enrichissement du sémiogramme, insertion de nouveaux schèmes
- Normalisation des structures de description
 - Normalisation du dictionnaire sémémique
 - Normalisation des sémiogrammes
 - Normalisation des opérations normatives