

**DESCRIPTION ET MODELISATION DE LA
THÉMATIQUE (HYPER-)TEXTUELLE A L'AIDE DE SCRIPTS
THEMATIQUES**

Une esquisse

**Peter Stockinger,
MSH-INALCO**

1995

Table des Matières

INTRODUCTION.....	3
LE PLAN THÉMATIQUE DU DOCUMENT.....	7
LE SCRIPT THÉMATIQUE.....	12
1) LE SCRIPT COMME "VISION" OU "HYPOTHÈSE"	12
2) LA CONFIGURATION THÉMATIQUE	13
3) SCRIPT THÉMATIQUE CANONIQUE ET SCRIPT THÉMATIQUE DERIVEE.....	14
4) CONDENSATION THÉMATIQUE ET EXPANSION THÉMATIQUE.....	20
5) L'IDENTIFICATION D'UNE SITUATION À L'AIDE D'UN SCRIPT THÉMATIQUE.....	22

INTRODUCTION

Le développement de systèmes hypertextuels constitue, actuellement, un des axes de recherche le plus important dans la mesure où ce type de systèmes est censé non seulement de rendre accessible les informations (textuelles ou autres) demandées par l'utilisateur, mais aussi de rendre possible, à partir d'une information donnée, l'exploration des informations qui en constituent le contexte (potentiellement) pertinent.

D'un point de vue de l'utilisateur, un système hypertextuel se présente, avant tout, comme un "guide" à l'aide duquel il peut explorer un univers textuel ou multimedia selon ses besoins d'information. D'un point de vue du système, un hypertexte est constitué, outre une interface utilisateur d'interrogation et de navigation, par une base de données (de textes, d'images, de vidéos, ...) et un ensemble de liens à l'aide desquels sont définis les "chemins" selon lesquels l'utilisateur peut explorer, à partir d'une donnée d'information sélectionnée, un ensemble d'autres données. Une gamme de plus en plus importante de logiciels d'hypertexte est mis à la disposition de concepteurs-réaliseurs d'applications particulières.

Ceci étant, les actuels logiciels et systèmes hypertextuels souffrent de trois faiblesses principales qui les rendent difficilement opérationnels soit pour de grandes bases de données multimedias soit encore pour de bases de données dont la caractéristique principale est la mise à jour permanente.

La première faiblesse concerne la sémantique très rudimentaire dont sont pourvus les liens à l'aide desquels les différentes données d'une application sont connectées entre elles. En effet, le type principal de liens hypertextuels se réduit tout simplement à une relation formelle orientée entre deux, voir plusieurs entités (i.e. données). Bien sur, on reconnaît que le rapprochement de données contenant des informations

pertinentes à la requête d'un utilisateur, ne peut pas être toujours réduit à ce type élémentaire de relations.

La deuxième faiblesse concerne le fait que les chemins entre les différentes données à explorer, sont, habituellement, prédéfinis. Autrement dit, c'est le concepteur d'une application hypertextuelle (ou encore hypermedia si cette application englobe, outre des données textuelles, des données image, video, son, ..) qui introduit et fixe les chemins qu'un utilisateur peut emprunter pour "naviguer" entre un ensemble de données. A part du fait que la pré-définition de liens contraint l'utilisateur de se contenter des chemins qui lui sont imposés entre les données contenant des informations pertinentes, cette procédure devient, d'un point de vue pratique, de plus en plus difficilement à gérer pour des bases de données quantitativement importantes et, enfin, ne quasiment plus praticable pour des bases de données qui évoluent quasiment tous les jours comme c'est, par exemple, le cas pour les bases de données où sont stockés les articles provenant de quotidiens ou encore pour celles dédiées au stockage de documents rapportant les activités diverses d'une entreprise.

La troisième faiblesse à trait au fait qu'un index (et plus généralement, un thesaurus) utilisé pour accéder et identifier une information, est habituellement confondu soit avec l'information elle-même soit avec la définition-description de celle-ci. A strictement parler, un index de "mots" (mais aussi de "descripteurs") n'est rien d'autre qu'un ensemble de "chaînes de caractères" ou encore de "processing units" selon la terminologie utilisée dans le cadre du PDP. Ce n'est qu'en les référant à de définitions-descriptions (a de "visions" ou "points de vue") concernant un certain type de situations, qu'on peut conférer à de "mots" ou "descripteurs" un sens plus précis. Le fait qu'un index possède un sens plus ou moins précis pouvant être utilisé à la fois pour le dialogue entre l'utilisateur et le système ainsi que pour l'identification d'un certain type d'informations, est dû simplement à de conventions linguistiques collectivement plus moins bien partagées quant à l'emploi des expressions écrites ou orales dans un domaine d'expertise donné (domaine d'expertise qui est d'ailleurs en général de nature technique ou scientifique).

Dans cette esquisse, nous exposons en grandes lignes une approche thématique pour la conception de systèmes hypertextuels en particulier et de systèmes d'information en général. Cette approche se caractérise par le fait qu'il repose sur une base de *scripts (ou configurations) thématiques* décrivant à un niveau relativement général les informations provenant d'un domaine d'expertise.

Un script thématique (ou encore une conjonction de plusieurs scripts thématiques) gère à la fois la requête d'une information, l'exploration contextuelle d'une information donnée, l'indexation de documents, la traduction (rapprochée) sous forme de résumés d'une information donnée ainsi qu'une partie de la production/création d'un nouveau document. Nous pensons qu'un système d'information intégrant une base de scripts thématiques permet de dépasser certaines restrictions qui pèsent sur les systèmes existants; il permet notamment:

- contrairement à la plupart de systèmes de documentation existants, de tenir compte du contenu même développé dans un document donné;
- contrairement aux systèmes d'information actuels de proposer une interprétation du flux d'information ou encore de clusters de "mots" au sens de "chaînes de caractères";
- contrairement aux systèmes hypertextuels de considérer chaque document comme une unité indépendante qui sera mis en relation avec un ou plusieurs autres documents en tenant compte à la fois de la requête de l'utilisateur et de l'information sélectionnée.

L'approche que nous esquissons ici fait l'objet d'une coopération entre plusieurs centres de recherche visant le développement d'un système d'information et de communication pour la presse économique internationale. Il y a trois axes de recherche qui sont définis dans cette coopération:

- conception et développement d'une base de scripts thématiques;

- développement d'un parseur lexical valencié pour l'instanciation de scripts thématiques à la surface linguistique de documents;
- analyse statistique du flux d'information et interprétation du flux par de scripts thématiques.

Les objets pris en consideration dans la presse économique sont constitués, plus particulièrement, par:

- les transactions entre sociétés (achat/ventes de sociétés, participations au capital, fusions de sociétés, ...),
- les activités économiques des sociétés (exploitation, production, distribution, investissement, recherche et développement, ...)
- les acteurs économiques (types d'acteurs, structure juridique, états financiers, localisations évolutions économiques, évolutions financières, ...)
- les secteurs d'activités (types de secteurs, structures actorielles, états et évolutions économiques, ...)
- les marchés (pays et régions géo-politiques, institutions à caractère économique et politique, ...)
- les produits (types et classes de produits, caractéristiques de produits, ...)
- les périodes temporelles (années/mois/jours; périodes nommées; périodes saisonnières, ...)
- la presse (journal: nom, structure, localisation, diffusion, ...).

LE PLAN THÉMATIQUE DU DOCUMENT

Pour pouvoir mieux situer la problématique de la description du plan thématique d'un document, nous présentons, dans ce chapitre, d'une manière fort condensée les différents plans de l'analyse d'un document en général, et puis, plus particulièrement, certains aspects centraux dont l'analyse thématique doit tenir compte.

En nous tenant au cadre de la gestion électronique de documents (multimédias), nous distinguons quatre plans principaux qui caractérisent un document ou encore selon lesquels un document quelconque peut être, en principe, analysé:

- le *plan thématique* qui recouvre les informations véhiculées par un document (ou un ensemble de documents);
- le *plan des codes d'expression* qui recouvre les signes sémiotiques à travers lesquels les informations sont exprimées, transférées et reçues par un destinataire (notons que la notion "langage" renvoie, au moins dans la tradition structuraliste, à une corrélation entre un certain *type* d'informations - appelé "contenu" - et un certain *type* de codes d'expression - appelé "expression" sans restreindre pour autant les types d'expression aux seuls types d'expression graphique ou orale);
- le plan de l'*organisation du document* qui décrit les instances de différents types d'objets dont est constitué un document ainsi que la répartition (généralement) planaire et séquentielle entre ces instances;
- le plan *du support physique* dans lequel un document se réalise.

Le plan thématique peut être approché par le biais de trois composants principaux:

- le composant topique,
- le composant énonciatif et
- le composant discursif.

Le composant topique identifie les informations ou encore les situations dont un document fait état. Dans la presse économique, une grande partie de ces informations concerne notamment :

- les acteurs économiques (les entreprises, sociétés, companies, ...),
- les activités de ces acteurs comme celles de la production, de l'investissement, du développement, ... ainsi que les rapports conflictuels et/ou consensuels qui se nouent entre deux ou plusieurs acteurs,
- les marchés et secteurs d'activités,
- l'évolution économique et financière d'un acteur, d'un secteur d'activité ou encore d'un marché,
- et ainsi de suite.

L'identification de ces informations présuppose de schémas de description que nous appelons "scripts thématiques" (ou encore, plus précisément, "scripts topiques") et qui constituent, en fin de compte, de sorte d'hypothèses d'interprétation dont la validité est, en principe, toujours négociable, voir révisable.

Le *composant énonciatif* tient compte de la source d'information, c'est-à-dire de l'actant qui prend en charge la veracité d'une information donnée ainsi que d'un ensemble d'autres paramètres dits contextuels spécifiant davantage la validité relative d'une information assertée.

L'introduction du composant énonciatif complique considérablement la description du plan thématique pour au moins deux raisons principales.

La première concerne le fait que dans un article de presse il y a, outre la source d'information qu'est le journaliste lui-même assertant un ensemble d'informations, d'autres sources d'information (nommées ou pas, explicites ou seulement implicites) qui sont introduites par le journaliste pour prendre en charge des informations complémentaires, similaires ou contraires à celles prise en charge par le journaliste.

Tout en mettant à part les problèmes de l'identification des traces linguistiques pouvant se poser pour reconstituer l'ensemble de sources d'information intervenant dans un article de presse, une description voulant tenir compte d'une telle "orchestration" des sources d'information doit obligatoirement recourir à de schémas de description pourvus d'au moins trois niveaux différents de pertinence descriptive:

- un premier niveau tenant compte du journaliste comme auteur d'un article de presse;
- un deuxième niveau tenant compte du journaliste comme source d'information d'un ensemble d'informations dont il prend en charge la véracité au même titre que les autres sources d'information qu'il cite en tant qu'auteur de l'article;
- un troisième niveau, enfin, tenant compte de l'identification des informations à l'aide de schémas ou scripts topiques.

La distinction entre le premier et le deuxième niveaux est importante dans la mesure où elle permet non seulement de comprendre les stratégies de sélection de sources d'information effectuée par le journaliste comme auteur mais aussi de comparer et d'évaluer différents articles de presse non seulement par rapport à la sélection des sources d'information mais aussi par rapport à l'attribution, par le journaliste comme auteur, des "responsabilités" d'information aux différentes sources d'information (y incluant lui-même).

La deuxième complication de l'analyse thématique, due à l'introduction du composant énonciatif concerne le fait suivant: deux sources d'information peuvent être en contradiction tout simplement par le fait qu'elles fournissent des informations incompatibles tout en se référant au même schéma (au même script) thématique. Il est souvent le cas que deux articles donnent des informations incompatibles à propos, par exemple, du montant d'une transaction entre deux sociétés ou encore à propos de l'identité d'une marque de produits sur laquelle porte l'investissement d'une société. Cependant ce genre d'incompatibilités ne mettent pas (ou ne mettent très vraisemblablement pas) en cause le schéma d'interprétation (de description) lui-même à l'aide duquel deux articles identifient leurs informations.

Par contre, comme nous l'avons déjà dit plus haut, un script thématique (ou plus spécifiquement topique) est une sorte d'hypothèse (i.e. une "vision", un "point de vue") à l'aide de laquelle un ensemble d'informations est décrit ou interprété. D'où donc une double source d'incertitudes:

- i) on ne peut jamais exclure la possibilité que deux sources d'information détectent et décrivent des informations *en se référant en effet à de schémas différents*;
- ii) un schéma de description utilisée par l'analyste lui-même ne constitue qu'un "*point de vue*" (qu'une "vision") à l'aide duquel les informations présentées dans un article de presse et prises en charge par une ou plusieurs sources d'information, sont interprétées.

Ces deux problèmes ne font qu'un seul, à savoir que toute description (linguistique ou non) est intrinsèquement limitée d'un point de vue épistémique et ne peut acquérir qu'une certaine valeur probabiliste.

Le troisième composant - le *composant* dit *discursif* - concerne plus particulièrement ce qu'on a l'habitude d'appeler les "actes de discours" à l'aide desquels les informations prises en charge par une source d'information sont développées dans un article de presse.

Parmi ces actes, les plus récurrents sont les actes dits *expositifs*, c'est-à-dire les actes de la description et de la narration au sens simple (i.e. le développement des informations sous forme d'événements et suivant un schéma plus ou moins chronologique) ou plus complexe (i.e. le développement des événements dans un cadre polémique-consensuel au sens qu'en a donné Greimas à ce terme). à ces actes expositifs s'ajoutent plusieurs types d'actes de nature argumentative comme l'explication, l'évaluation (et le pronostique) ainsi que le débat (ou la discussion) pour trancher entre les propos divergents tenus par différentes sources d'information.

Il est entendu que la prise en compte de ces actes présuppose, de nouveau, de schémas de description (i.e. de scripts que nous appelons

discursifs) dont il faut, naturellement, distinguer les formes manifestées dans un article de presse. Une simple comparaison entre plusieurs articles de presse peut nous montrer que la manifestation d'un type d'actes de discours peut se faire:

- d'une manière condensée ou "expansée" (i.e. un acte comme la description peut se réduire à une seule phrase ou, au contraire, s'étendre sur un ou plusieurs paragraphes);
- d'une manière directe ou indirecte (i.e. un acte comme l'explication peut se manifester comme tel, mais il ne peut être également qu'inférable);
- d'une manière continue ou discontinue (i.e. un acte comme la description peut être achevée "en bloc" ou se distribuer sur l'ensemble de l'article en étant coupé par d'autres actes de discours).

Il va de soi que ces trois modes de manifestation valent également pour les composants topique et énonciatif. La comparaison de quelques articles de presse développant, par exemple, des informations relatives au bilan annuel d'une société, montre bien que tel article tient compte in extenso des informations relatives à l'actionnariat tandis que tel autre article n'en fait que très peu état tout en mettant l'accent sur les investissements effectués par cette même société; elle peut montrer également que telle source d'information est nommée et caractérisée par tel article tandis que tel autre article y réfère sous le manteau de l'anonymat en tant que source non-identifiée mais, bien sûr, bien informée; et ainsi de suite.

LE SCRIPT THÉMATIQUE

1) Le script comme “vision” ou “hypothèse”

D'une manière générale, les scripts thématiques peuvent être considérés comme des hypothèses d'interprétation de caractéristiques saillantes d'un domaine donné (ou encore, pour parler avec Barwise & Perry (1983), comme de schémas ou *types de situations* à l'aide desquels un agent cognitif catégorise les informations provenant de situations concrètes, c'est-à-dire des occurrences d'un type de situations).

Il est entendu qu'un script thématique entendu comme hypothèse d'interprétation d'une situation ou d'un ensemble de situations concrètes, présuppose toujours un actant qui l'asserte (qui le "prend en charge"), un contexte - des circonstances selon Barwise (1987) - dans lequel il est tenu valide par l'actant ainsi que le fait objectif ou réel qu'une situation donnée possède ses propres caractéristiques pouvant évoluer ou être modifiées et dont un script thématique réussit plus ou moins bien à en tenir compte. Dans ce sens, nous parlons de standard thématique correspondant à ce que Lewis (1969) a appelé une convention.

Dans une terminologie plus usuelle, un standard thématique décrit un point de vue ou encore une "vision" qui est plus ou moins partagée par une collectivité d'actants et qui possède une stabilité plus ou moins forte mais qui, en principe, est toujours négociable, voir révisable.

En faisant abstraction de terminologies parfois fort différentes, il nous semble que cette conception est assez communément acceptée en recherches sémantiques (formelles et/ou descriptives) actuelles qui, tout en gardant la notion sémantique centrale de la vérité, la paramétrise à la fois par rapport à de *contextes épistémiques* (cf. à ce propos, par exemple, Stalnaker (1986)) ainsi qu'à la *probabilité de la stabilité* (de la récurrence) propre aux caractéristiques de situations concrètes (cf. à ce propos le travail déjà cité de Barwise & Perry (1983)).

Ajoutons encore que cette conception a été très clairement exposée par Eikmeyer (1982) qui examine, en se référant à la notion du stéréotype chez Putnam (1976), différentes définitions-descriptions d'un objet comme la saccharine pour proposer, ensuite, une théorie de la révision de standards (i.e. de définitions-descriptions).

Il n'est pas notre intention de développer, dans cet article, plus en détail cette conception ni ses conséquences qui nous semblent être importantes dans ce vaste domaine de recherche qu'on appelle maintenant la représentation des connaissances. Nous ne pouvons pas non plus prendre position par rapport aux controverses que celle-ci suscite notamment en sémantique componentielle et structurale. En présentant en quelques lignes cette conception sémantique, nous n'avons voulu qu'identifier le cadre théorique qui sous-tend, pour nous, l'analyse thématique de documents et l'élaboration de scripts thématiques.

Le contenu, cependant, n'est rien d'autre qu'une certaine "vision", qu'un schéma à l'aide duquel un actant tient (ou essaie de tenir) compte d'une réalité. On ne voit pas bien ce qui permettrait de dissocier un contenu, disons, linguistique d'un contenu, disons, non-linguistique. Il y a des "visions" (ou de schémas) relevant plutôt du sens commun que d'autres qui sont plus spécialisées, il y a certaines qui sont historiquement parlant plus stables que d'autres, il y a de "visions" qui sont plus fondamentales que d'autres dans ce sens qu'en leur absence certaines visions ne pourraient pas exister, et ainsi de suite.

Toutes ces "visions", pour qu'elles peuvent être perçues, communiquées, interprétées, discutées ou encore révisées, doivent être exprimées par de signes perceptibles.

2) La configuration thématique

La définition d'un type particulier de situations nous conduit à la mise en place de ce que nous appelons un *script thématique* (ou, plus précisément, un script topique) et l'identification d'une occurrence

spécifique d'un type particulier de situations à celle d'un *script thématique (partiellement) instanciée*. Script thématique et script thématique (partiellement) instanciée sont de *schémas de description*. Dans une terminologie actuellement en usage, le script thématique correspond à ce qu'on appelle une *connaissance* et le script thématique (partiellement) instanciée à ce qu'on appelle une *information*.

Un script thématique (instanciée) se compose essentiellement d'un ou de plusieurs *thèmes* (que nous appelons parfois également *dimensions thématiques*), de *relations* permettant de positionner les thèmes les un par rapport aux autres, de *valeurs* ou *instances* auxquelles un script thématique réfère, ainsi que d'un ou de plusieurs contextes organisant "en profondeur" (ou encore, comme on dit, "hypotaxiquement") un script thématique.

Comme nous l'avons développé ailleurs, un script thématique est isomorphe à un *graphe conceptuel* - le premier constituant un élément d'une théorie descriptive d'un objet, le second constituant un élément du langage de représentation graphique d'un formalisme permettant de traiter ou encore de manipuler les connaissances et/ou informations que nous avons d'un objet à l'aide du recours à une théorie descriptive de celui-ci (Stockinger 1992).

Prenons comme exemple la situation de la démission. Sa définition-description présuppose que nous tenons bien compte des deux oppositions suivantes:

- *script thématique canonique vs script thématique dérivé* (ou encore défini d'une manière inductive à partir d'une base de scripts canoniques);
- *condensation thématique vs expansion thématique*.

3) Script Thématique Canonique et Script Thématique Derivée

Un script thématique canonique de la situation de la démission tient compte d'une définition minimale (cf. Martin 1987) de cette situation, c'est-à-dire de critères dont nous pensons ou jugeons qu'ils doivent être

remplis si tel ou tel événement concret peut être interpréter comme appartenant au type particulier d'événements appelées "démission".

Ceci dit, dans notre article, la reconnaissance de ce type de situations n'est pas particulièrement problématisée comme cela pourrait être le cas, par exemple, dans un ouvrage de politologie. Le journaliste de l'article présuppose chez son lecteur que celui-ci sait de quoi il s'agit quand on parle d'une démission, que le lecteur sait faire la différence entre une démission et, par exemple, une destitution de quelqu'un de ses fonctions.

Une définition minimale ou canonique dépend donc d'un *contexte épistémique* (cf. à ce propos Eikmeyer 1983) et peut être *révisé* ou encore "*négocié*" par des actants épistémiques (des "experts") qui assurent sa validité.

Étant donné donc le peu d'informations que nous sont données dans cet article à propos du type lui-même de la situation de démission, nous nous contentons d'une définition minimale qui possède toutes les allures d'une définition lexico-sémantique stipulant quelque chose comme suit: une démission est une situation dans laquelle un acteur quitte ses fonctions institutionnelles qu'il a occupé préalablement. Bien sur, c'est une définition assez frustrée mais elle correspond à un contexte épistémique putatif qu'est celui de l'article en question (et, comme nous présumons, de l'ensemble des articles politiques dont la tâche de communication n'est pas tellement de problématiser *l'interprétation même* de la situation de démission mais plutôt de relater tel ou tel événement particulier qui est identifié comme une occurrence spécifique de ce type de situation qu'est la démission).

En tenant compte donc à la fois de notre "moule" méthodologique introduite auparavant et de la définition minimale de la situation de démission, le script thématique canonique décrivant celle-ci possède, approximativement, l'allure suivante:

```
scriptDemission (x)
  est une [[Occupation
    à comme source Acteur:+
    à comme objet Poste: +
    dans Institution
    jusqu'a Date]
  est suivi de [Rénonciation: *x
    à comme source Acteur:+
    à comme objet Poste:+
    à comme moyen Moyen
    au moment de Date]
  est suivi de _[Occupation
    à comme source Acteur:+
    à comme objet Poste:+
    dans Institution
    à partir de Date]].
```

Ce script stipule alors que la démission est une situation particulière qui réfère à un état initial où un acteur occupe un poste dans une institution; cet état est suivi d'un processus de rénonciation à ce poste par le même acteur ayant comme résultat un nouveau état dans lequel l'acteur en question n'occupe plus sa fonction dans l'institution.

D'un point de vue méthodologique, les étiquettes commençant avec une lettre majuscule (cf. Acteur, Institution, Date, ...) correspondent à des thèmes ou dimensions thématiques, les formulations comme "est caractérisé par", "est suivi de", "dans", etc. correspondent aux relations thématiques dont la tâche est de positionner les thèmes les uns par rapport aux autres.

Le symbole ":" correspond au séparateur entre le champ du thème et le champ de la valeur (ou de l'instance) à laquelle un thème réfère.

L'expression "scriptDemission: (x)" représente le type du thème Démission - appelé aussi genre ou genus (la procédure définitionnelle utilisée ici est celle appelée, traditionnellement, "per genus proximum et differentiam specificam").

Le symbole "+" tient compte de la co-référence, c'est-à-dire du fait que les différentes instances d'un thème donné sont identiques (si on avait, par exemple, une situation du type "destitution de quelqu'un de ses fonctions", l'acteur du procès serait différent de l'acteur occupant un poste dans une institution).

Les symboles "[" et "]" correspondent aux contextes qui structurent hypotaxiquement un script thématique.

Le symbole "_", enfin, désigne l'opérateur de la négation.

D'un point de vue sémiotique, le script thématique décrivant la situation de démission n'est rien d'autre qu'une *configuration thématique*. On obtient une telle configuration par:

- le *déploiement syntagmatique* à partir de la base d'une structure considérée comme élémentaire qu'est ledit programme narratif (Greimas 1970; Greimas & Courtes 1979; Stockinger 1985);
- la *projection* de la structure élémentaire et/ou déployée dans un treillis de types de thèmes (ou dimensions thématiques) et de relations thématiques; et
- - possiblement - la *conjonction* avec une ou plusieurs autres configurations thématiques.

Nous ne voulons pas développer ici les aspects techniques et formels de la construction d'une configuration thématique (cf. à ce propos Stockinger 1993).

Notons néanmoins que dans le script "démission" sont contenus de critères temporels (i.e. "jusqu'à Date", "au moment de Date" et "à partir de Date" ainsi que "est suivi de Thème") qui font appel à la configuration thématique - fort générale - de la localisation temporelle.

Cette configuration définit les critères dont nous avons besoin pour pouvoir d'une part localiser temporellement les situations ou encore les "faits" les uns par rapport aux autres et d'autre part "raisonner" sur et à partir de leurs localisations temporelles. Dans le script décrivant la

situation de la démission, certaines "branches" de la configuration de la localisation temporelle sont introduites justement à cet effet. Ce sont elles qui assurent le lien entre les deux configurations ainsi que la possibilité d'avoir recours à un "raisonnement" temporel dont nous pourrions avoir besoin pour des inférences spécifiques concernant certains "faits" qui concernent une démission spécifique.

La définition donnée ci-dessus de la démission n'est, cependant, pas tout à fait appropriée au type de démission qui est thématisée dans notre article de presse. En effet, celui-ci ne thématise pas n'importe quel type de démission mais plutôt une démission au sens politique du terme mettant en jeu une institution politique qu'est le gouvernement (russe) et une fonction plus spécifique qu'est celle du premier vice-premier ministre.

Par rapport à ces "données", notre définition est quelque peu trop générale et il convient de la restreindre thématiquement. Cette restriction - ou "spécification" - thématique exige justement une description plus explicite de treillis thématiques Institution et Poste. Bien sur, celle-ci ne peut pas se contenter par la simple mise en place de hiérarchies (partielles) de thèmes telles que:

- | |
|---|
| <ul style="list-style-type: none">- Institution > Gouvernement- Poste > Premier Vice-Premier Ministre. |
|---|

Un tel procédé n'est pas satisfaisant par la simple raison que les formes et fonctions à l'intérieur d'un gouvernement changent d'un pays à un autre, d'une période à une autre. Cela signifie qu'il convient, de nouveau, de procéder à la définition d'un script thématique canonique dont l'objet est Gouvernement ou, plus spécifiquement, Gouvernement en Russie.

Nous ne développerons pas davantage ce point mais retenons néanmoins l'enseignement suivant dont nous en tirerons les conséquences plus tard: la description du plan topique (et, plus généralement, thématique) d'un document tel qu'un article de presse, dépasse forcément ce qu'est dit ou exprimé dans ce document, dans cet

article de presse. En d'autres termes: pour comprendre et interpréter de ce qu'est dit dans un texte, il faut forcément recourir à un "modèle" thématique plus riche de nature foncièrement culturelle.

Nous avons affirmé plus haut que le script thématique décrivant la situation de la démission est canonique dans ce sens que nous stipulons qu'il détermine les critères à l'aide desquels nous reconnaissons tel ou tel événement particulier comme une démission. En outre, nous avons également vu que ce script contient de thèmes qui figurent non seulement dans celui-ci mais aussi dans d'autres tels que le script de la localisation temporelle, le script définissant une institution politique comme le gouvernement (russe) ou encore le script définissant un acteur (politique).

Tous ces scripts sont à la fois indépendants (ils constituent de sorte de modules) et en interactions les un par rapport aux autres. En tant que modules indépendants, ils constituent de voies d'accès spécifiques à un champ thématique qui, tout en étant intrinsèquement limité par l'existence d'une base de scripts canoniques, se structurent dynamiquement suivant les interactions entre un script choisi et les autres. Le résultat d'une telle interaction est justement ce qu'on appelle un *script thématique dérivé*.

En choisissant le script thématique "démission", la thématique d'un article de presse (pour rester dans cet univers) peut en effet, ensuite, explorer le champ topique lié au script "institution politique" (en thematisant, par exemple, à partir d'une démission donnée les avatars d'une institution politique); de même, en choisissant le script thématique "institution politique", un autre article de presse peut ensuite explorer le champ thématique lié au script "démission", et ainsi de suite.

Nous aurons encore plusieurs fois l'occasion de revenir sur le problème de la structuration dynamique d'une thématique qui s'articule, dans ce cas précis, par ce qu'on appelle, en sémiotique, un changement de l'isotopie (thématique). Retenons, pour le moment, qu'elle explicite assez bien, à notre avis, l'image de l'*entrelacement thématique (partiel)* entre différents textes constituant une des multiples facettes de l'*intertextualité*. Il est, cependant, vrai aussi que son explicitation

présuppose une activité descriptive dont le coût à la fois théorique et pratique semble être très élevé.

4) Condensation Thématique et Expansion Thématique

En stipulant qu'une démission est une situation qui réfère à la renonciation d'une fonction institutionnelle par le même acteur qui l'a occupé auparavant, nous avons, d'une certaine façon, introduite une "histoire" canonique qui la sous-tend typiquement. Exprimé d'une manière informelle, tant qu'on s'intéresse davantage à une démission, on la voit se déployer en quelques étapes principales qui nous montre - un peu comme dans un film - son déroulement: au début on voit un acteur occuper sa fonction, puis on le voit s'en défaire, et, enfin, on le voit encore comme acteur qui en est dépourvu. Bien sur, tel quel, c'est un "film" bien peu sophistiqué mais, disons, c'est au metteur en scène (a l'analyste, dans notre cas) d'elaboree davantage ce petit scénario.

Le point intéressant ici, c'est le fait que le script canonique expose ce qu'est comprimé ou condensé dans le thème Démission. Nous avons donc ici la deuxième opposition introduite auparavant et dont nous devons tenir compte dans la description d'un champ topique; il s'agit de l'opposition entre la *condensation thématique* et l'*expansion thématique*. C'est bien à Greimas à qui nous devons d'avoir reconnu le rôle tout à fait central de cette opposition dans le fonctionnement du discours:

"Exprimé de façon naïve, ce principe (i.e. le principe d'équivalence d'unités linguistiques inégales, P.S.) veut simplement dire qu'une chose peut être présentée tout aussi bien de façon simple que de façon compliquée, qu'un mot simple peut être expliqué par une séquence plus large, et qu'inversement un seul mot peut souvent être trouvé pour désigner ce que l'on a d'abord conçu sous forme d'un développement. L'expansion n'est donc pas cette propriété syntaxique du discours qui permet l'adjonction de déterminations successives grâce aux termes manifestés les uns après les autres: c'est le propre du fonctionnement du discours. Elle ne prend toute sa signification que si une séquence en expansion est reconnue comme équivalente d'une unité de

communication syntaxiquement plus simple qu'elle. C'est cette équivalence, théoriquement toujours possible - bien qu'elle ne soit pas toujours manifestée lexicalement -, qui constitue l'écart structurel définissant le fonctionnement métalinguistique du discours." (Greimas 1986: 73).

Dans une analyse stimulante du lexème français "colère", Greimas soutient l'hypothèse qu'il est

"notoire que les lexèmes se présentent souvent comme des condensations recouvrant, pour peu qu'on les explicite, des structures narratives et discursives complexes. ... (...) les descriptions lexématiques peuvent constituer, de façon économique, des modèles de prévisibilité pour des analyses discursives ultérieures. " (Greimas 1983: 225)

Ayant défini la démission comme une situation qui est typiquement caractérisé par une "histoire" de renonciation à une fonction institutionnelle par l'acteur lui-même qui l'a occupé auparavant, nous pouvons donc établir entre le thème de la démission et sa définition une sorte d'équivalence en affirmant que le premier condense ce que le second articule en expansion. Les scripts suivants:

1) scriptDémission
a comme source Acteur

2) scriptDémission
a comme source Acteur
a comme Objet Poste
au moment de Date

sont deux scripts plus ou moins condensés par rapport au script canonique que nous avons introduit auparavant. C'est dans Sowa (1984) ou on trouve des algorithmes à l'aide desquels on peut manipuler ces deux opérations qui sont en effet centraux pour une sémantique du discours.

Nous avons affirmé ci-dessus que la compréhension et l'interprétation de la thématique d'un document comme celle d'un article de presse présuppose nécessairement un "modèle" thématique plus riche. En tenant compte à la fois de ce que nous avons dit à propos de scripts thématiques canoniques et dérivés ainsi que des deux opérations de la condensation et de l'expansion, nous sommes maintenant en mesure de préciser cette affirmation.

Dans notre article de presse, sont thematisees plutôt de scripts tels que les scripts 1 et 2 que nous venons de mentionner. Ces scripts constituent ce que nous appelons un *standard thématique actualisé* par le document et qui "puise" dans un *standard thématique de référence* qu'est constituée par le script canonique ou encore par une de ses formes dérivés à partir d'une base de scripts thématiques (pour une discussion plus approfondie de cette notion, cf. le chapitre 2 consacrée à la notion du standard).

Le "modèle" plus riche auquel nous avons fait allusion auparavant, n'est donc rien d'autre que le script canonique ou encore une de versions dérivés de celui-ci constituant simultanément l'"expansion thématique maximale" de scripts plus ou moins condensés qui forment le plan thématique actualisé par un document tel qu'un article de presse.

5) L'identification d'une situation à l'aide d'un script thématique

Revenons à notre exemple. Jusqu'à maintenant, nous avons discuté un ensemble de questions relatives à la définition-description d'un script thématique tenant compte d'un type particulier de situations qu'est, dans notre cas, celui de la démission. Nous avons dit que celles-ci concerne un aspect de la description topique (et, plus généralement, thématique) d'un document ou d'un ensemble de documents.

L'autre aspect concerne l'identification d'une ou de plusieurs occurrences d'un type particulier de situations. Dans notre article, nous pouvons constater plusieurs occurrences de la situation de démission:

- une première démission de Gaidar en septembre 1992;
- la démission de Gaidar dont est principalement question dans cet article;
- la démission d'Ella Pamfilova du poste du ministre de la protection sociale;
- la démission (probable) de Boris Fiodorov du poste du vice-premier ministre chargé des finances.

L'*identification* (et la distinction) de ces différentes occurrences procède principalement par la spécification du champ référentiel d'un ou de plusieurs thèmes dans un script thématique actualisé par l'article, cf.:

1) scriptDémission
a comme source Acteur: "Gaidar"
dans Gouvernement_Russe: #1
au moment de Date: "septembre 1992"

2) scriptDémission
a comme source Acteur: "Gaidar"
a comme objet Premier_Vice_Premier_Ministre
dans Gouvernement_Russe: #2
au moment de Date: "janvier 1994"

3) etc.

Notons que les signes "#1" et "#2" ont pour fonction d'indiquer qu'il s'agit ici de deux gouvernements différents (c'est-à-dire que Gaidar à occupé un poste dans deux gouvernements différents dont il à renoncé ensuite).