

ANNEXE V.1

Etude descriptive
de la structure
des textes d'Action
(1991-1996)

Application à l'édition électronique
et au traitement automatique

Table des matières de l'Annexe V.1

A. INTRODUCTION.....	557
B. PRÉSENTATION DU CORPUS DES TEXTES D'ACTION.....	558
1. L'ordonnancement DER.....	558
2. La forme d'une fiche d'Action, et sa version électronique	559
3. Les parties du texte d'Action : référentiel pour cette étude	561
4. Le corpus de travail.....	562
C. POINTS DE VUE SUR LA STRUCTURATION DU TEXTE DES ACTIONS : EXISTANT	563
1. Vue normative : plan-type de référence	563
2. Vue technique : modèle de document SGML.....	563
3. Vue algorithmique : automate de reconnaissance des titres de parties	564
4. Vue théorique (linguistique computationnelle) : l'analyse de l'organisation générale des textes d'Action	565
D. ELÉMENTS D'ANALYSE DE L'ORGANISATION DES TEXTES : FONCTIONS DES PARTIES.....	569
1. Contexte.....	569
2. But.....	570
3. État d'avancement.....	571
4. Programme (et étapes).....	571
5. Objectifs ultérieurs	573
6. Documents de référence	574
7. Partenariats extérieurs.....	574
8. Autres.....	575
E. VERS UNE RECONNAISSANCE AUTOMATIQUE DE L'ORGANISATION SÉMANTIQUE DES TEXTES D'ACTION	577
1. Un point d'appui : les intitulés	577
a) Fonctions des titres	577
b) Titre plein et titre creux	578

c) <i>La forme des intitulés</i>	578
d) <i>Le jeu des variations dans les titres creux : de la variante à la rectification</i>	579
e) <i>Eléments pour la description de la forme des titres</i>	580
f) <i>Règles sur le vocabulaire des titres</i>	581
2. Les rapports entre parties	583
a) <i>Succession</i>	583
b) <i>Unité (fusion, inclusion)</i>	583
c) <i>Répétition : factorisation</i>	584
d) <i>Présence</i>	584
e) <i>Longueur</i>	585
3. Les indices linguistiques dans le contenu, au fil du texte	585
4. Intégrer les indicateurs locaux et globaux	586
F. PERSPECTIVES	588
1. Utilisation de la structure des Actions	588
a) <i>Application à la consultation (Livre Electronique)</i>	588
b) <i>Application à la construction de corpus</i>	588
c) <i>Application à un traitement automatique d'analyse des textes</i>	589
2. Portée théorique et caractère généralisable de l'étude	589
a) <i>Les genres textuels</i>	589
b) <i>Les corpus de textes avec plan-type</i>	590
c) <i>Et pourquoi pas : l'Action comme un récit</i>	591
G. ANNEXES	593
1. L'algorithme de repérage des intertitres pour le Livre Electronique	593
a) <i>L'algorithme existant</i>	593
b) <i>Observations</i>	593
c) <i>Première proposition : enrichissement des indices</i>	594
d) <i>Deuxième proposition : considérer davantage les groupes de paragraphe</i>	594
2. Le vocabulaire des intitulés des parties : proposition de définition d'unités paradigmatiques	595
a) <i>Première colonne : mots recherchés</i>	595
b) <i>Deuxième colonne : nom d'une classe paradigmatique de rattachement</i>	595

A. INTRODUCTION

L'analyse d'un corpus exige d'avoir connaissance de la pratique dans laquelle ses textes interviennent. Ce cadre permet de préciser les fonctions générales que doit remplir le texte, et d'en esquisser le type¹. Ici, il s'agit, avec les textes d'Action, de la description générale en une à deux pages d'une démarche de recherche scientifique et technique, rédigée par le chercheur à l'intention de sa hiérarchie, et ayant un rôle décisif pour la détermination du contenu de son travail et des moyens qui lui sont accordés. On perçoit déjà que tout cela a une incidence sur le choix du vocabulaire (technique mais pas trop spécialisé), les types d'informations attendues (justifications, calendrier,...), la forme (longueur « convenable » (suffisante) mais « raisonnable » (limitée) du document, articulation systématique par points de certaines parties), le point de vue porté sur le sujet (positif, engageant) et le « ton » de l'exposé (rapport hiérarchique)².

Nous nous intéressons ici à l'organisation du texte d'Action, à son articulation en parties. L'existence et la fermeté de consignes (plan-type), les échanges avec les collègues proches, l'image que l'on a des informations que doit fournir le texte, se combinent et façonnent des parties. Celles-ci deviennent peu à peu autonomes (sans exclure des affinités particulières avec telle ou telle), et prennent consistance en réalisant une fonction d'information.

L'identification des parties et des fonctions d'information qu'elles réalisent permet ensuite de paramétrer un traitement automatique, sachant le type d'information que l'on recherche et connaissant, grâce à l'analyse précédente, les modes de parcours propres à chaque partie.

¹ Ce sont bien effectivement les pratiques (sociales, professionnelles, culturelles, etc.) qui déterminent les genres textuels : voir par exemple (Rastier, Cavazza, Abeillé 1994 - § VII.4.1.).

² En filigrane on retrouve les quatre composantes sémantiques du texte proposées par François Rastier (Rastier, Cavazza, Abeillé 1994 - § II.7. et VII.4.) : le vocabulaire renvoie à la *thématique*, la chronologie (explicite dans un calendrier d'étapes) à la *dialectique*, le point de vue et le ton à la *dialogique*, la forme à la *tactique*.

B. PRÉSENTATION DU CORPUS DES TEXTES D'ACTION

1. L'ordonnement DER

La DER, pour assurer un pilotage précis de son programme de travail, le découpe chaque année en Actions de recherche élémentaires (environ 2 000 Actions). Elles s'organisent en une petite dizaine de types, principalement en fonction des clients qui les financent (recherche plus ou moins immédiatement appliquée, interne ou externe, etc.). Les ARD (*Action de Recherche et Développement*) et AID (*Action Initiée à la Demande*) sont les types les plus courants, si bien que quand nous parlerons ici des ARD/AID, il faudra entendre en fait l'ensemble des Actions.

Un calendrier de l'ordonnement rythme annuellement la publication de documents permettant de suivre l'avancement des recherches : *fiches* descriptives puis *comptes-rendus* (Aureille, Baby 1990) (Allard 1995). Nous avons travaillé sur les fiches : elles ont un point de vue prospectif, anticipateur, qui est tout à fait intéressant pour DECID ; en outre, les comptes-rendus soulèvent davantage de problèmes de confidentialité et de déontologie. La chronologie de parution des fiches comporte deux moments. Une version provisoire des fiches d'Action concernant les recherches de l'année N est soumise vers le mois de juin de l'année N-1. La version définitive est établie à la fin de l'année (N-1), après examen et avis de la hiérarchie et des Directions opérationnelles. Dans une fiche d'Action, le chercheur (responsable d'Action) expose donc un projet d'activité qui, pour prendre effet, doit recevoir l'aval d'une consultation des instances supérieures et voisines : on veut s'assurer de cette manière que les projets répondent à de réels besoins et s'inscrivent dans la stratégie d'ensemble de l'entreprise. Ce circuit de relecture conditionne évidemment le style du document : il doit être précis et informatif, mais ni trop long (en général une à deux pages) ni trop technique. La fiche explique l'intérêt du sujet, les personnes impliquées, l'avancement des connaissances et des réalisations, les moyens à mettre en œuvre (humains, techniques et budgétaires), et les résultats programmés dans l'année. Tous ses éléments correspondent au besoin d'information rappelé dans le document de doctrine :

L'ordonnement est un acte de management essentiel qui répond à plusieurs objectifs fondamentaux :

- se donner un programme de travail précis et motivé [...]
- en assurer la transparence [...]
- en suivre et contrôler le déroulement, et donc en particulier identifier les éventuelles difficultés, permettre les ajustements et arbitrages nécessaires
- formaliser la délégation aux équipes techniques des moyens nécessaires pour atteindre les objectifs fixés, en assurer le contrôle et, de façon générale, de contractualiser la relation entre les responsables d'Action et les interlocuteurs principaux (clients).

(Allard 1995 - p.4)

Chaque Action génère donc un certain nombre de tâches :

- suivi (explicitation du projet, justification, relecture, rédaction d'un compte-rendu...)
- engagements, formalisés par les étapes (objectifs d'avancement, résultats scientifiques, émission de note, production d'un système...)
- décisions (viabilité, moyens alloués...)

Pour ne pas multiplier ces tâches, la tendance serait de regrouper dans le cadre d'une même ARD plusieurs thèmes voisins, plutôt que de les déclarer indépendamment dans des fiches distinctes.

Motifs pour définir plusieurs Actions	Motifs pour regrouper plusieurs activités
<ul style="list-style-type: none"> - Bien différencier les ressources, découpler les décisions - Mieux représenter le volume d'activité à la charge d'un responsable - Respecter la vocation du concept d'Action, qui est de délimiter une unité minimale, élémentaire, de recherche. 	<ul style="list-style-type: none"> - Alléger la somme des procédures liées à la gestion des Actions. - Ne pas isoler artificiellement des études qui dépendent étroitement les unes des autres (origine, réalisations).

Il est prévu un type d'Action, les « Actions à buts multiples », pour rendre compte des Actions conjuguant plusieurs thèmes d'étude. Néanmoins, ce cas s'écarte évidemment de la volonté générale de décomposer la recherche en briques élémentaires, aussi est-il instamment déconseillé. Le document de doctrine a le commentaire suivant :

ARD/AID, variante à buts multiples : cas normalement rare des Actions recouvrant différents buts, à éviter autant que faire se peut. (Allard 1995 - p.17)

2. La forme d'une fiche d'Action, et sa version électronique

Depuis 1986, une organisation a été mise en place pour disposer d'une version électronique des fiches et des comptes-rendus d'Action. Ces données sont gérées au sein de la base de données textuelles SPHERE, sous forme de tables Oracle. Des interrogations SQL permettent des extractions plus ou moins élaborées et des vues variées sur les différents champs. Ce fonds d'informations peut être utilisé non seulement ponctuellement et individuellement (recherche d'un renseignement archivé), mais aussi globalement pour analyser la répartition et l'évolution de l'activité à la DER : classements, cartographie... (Hébrail, Marsais 1993)

Depuis 1991, un support de consultation hypertexte³ est mis à disposition de la hiérarchie sous forme de *Livre Electronique*. Il est non pas basé directement sur SPHERE, mais exploite les données d'un fichier au format SGML automatiquement généré à partir des tables Oracle (Le Gall, Odone 1995). Un *Livre* regroupe toutes les fiches d'Action d'une année. Plusieurs modes de parcours sont prévus :

- lecture séquentielle,
- accès par la table des matières, reflétant l'organisation hiérarchique en Services, Départements et Groupes,
- recherche sur le texte intégral : équation booléenne sur chaînes de caractères (troncature possible).
La recherche peut ne porter que sur une zone particulière, comme le nom du responsable, ou le code (identificateur) de l'Action.

L'Action comprend un en-tête et un corps. Au moment de la saisie par exemple, l'en-tête est visualisé par des zones dans un cadre. Ces zones représentent des champs prédéfinis, accueillant des éléments administratifs et budgétaires. Le texte descriptif de l'Action débute à la suite du cadre. C'est un texte libre, à savoir que son vocabulaire n'est pas imposé, les formulations ne sont pas contraintes, et sa structure n'est pas rigide (du moins dans les faits, pour 1991-1996).

³ Interface *DynaText*, puis Web (Intranet) depuis 1997.

FICHE D'ACTION

Données administratives :

- rattachement dans la structure Service-Département-Groupe,
- code identificateur, titre,
- nom du responsable,
- nom et fonction des différents interlocuteurs

Données budgétaires :

- ressources humaines mobilisées,
- coût

Données textuelles :

- texte rédigé, comprenant souvent
- des parties avec un titre (« But de l'action », « Programme », etc.),
- des paragraphes numérotés, des listes,
- et quelquefois des tableaux (pour l'instant mal enregistrés électroniquement⁴).

Sur les années 1991-1996, on voit évoluer la forme de la partie rédigée (texte) des Actions.

Il ressort très nettement qu'au début de cette période, il y a peu de prescriptions sur la manière d'organiser les informations, voire même sur le type de contenu attendu. Le document de doctrine paru en 1990 (Aureille, Baby 1990) est centré sur le circuit de relecture et de validation de l'ordonnancement, l'unique indication sur le contenu du texte des fiches d'Action est qu'y est défini « le programme de travail ». Ce n'est qu'avec le document suivant (Allard 1995) que l'on s'intéresse au texte de la Fiche d'Action : un plan-type de référence est introduit et commenté.

En revanche, trois autres facteurs contribuent dès le début à homogénéiser ces écrits :

- le rédacteur s'inspire des Actions existantes, de son entourage (collègues, prédécesseur).
- le secrétariat de l'équipe, qui rassemble les textes et transmet le fichier aux instances pour relecture, a une vue d'ensemble des textes et peut intervenir sur leur mise en forme pour faire ressortir leur cohérence.
- quelquefois un modèle de structure, un plan-type, est utilisé (avec plus ou moins de variantes) par tous les membres d'une entité, sans doute en application d'une consigne à l'initiative du chef de Groupe ou de l'état-major du Département.

L'ensemble des Actions comporte alors des textes de longueur et de structure variées, avec des petits groupes homogènes de textes analogues correspondant de fait à l'organisation structurelle de la DER.

Dans une lettre du 7 avril 1994, M. Merlin, Directeur adjoint, donne des directives pour que le contenu des textes des ARD/AID soit harmonisé au niveau de la DER (Vuldy, Poncet 1995, p.5). L'idée se fait jour de donner un plan-type de référence, suivi dans toute la DER : un tel plan est prescrit dans un document de doctrine diffusé en septembre 1995⁵, pour entrer en vigueur dès 1996. Tel qu'il est conçu, ce plan opère une synthèse équilibrée des pratiques précédentes : il reprend les idées d'organisation les plus intéressantes qui figuraient déjà ici ou là. La structuration normée des textes d'Action est bénéfique au niveau du circuit de l'ordonnancement, avec, en jeu, un triple objectif :

Chaque Action fait l'objet d'une fiche décrivant le contenu technique et les moyens associés. Le texte est structuré suivant un plan type dont l'objectif est triple :

- faciliter le travail du lecteur en harmonisant les contenus
 - s'inscrire systématiquement dans une logique objectifs / résultats
 - aider le rédacteur en lui indiquant explicitement ce que le lecteur (client) souhaite y trouver.
- (Allard 1995 - p.5)

⁴ La DTD SGML ne les prévoit pas. On retrouve donc le contenu des cellules mais leurs rapports réciproques sont perdus.

⁵ C'est la Note (Allard 1995).

Cette décision de structuration générale a aussi des répercussions positives du point de vue des données qui entrent dans la base SPHERE. D'une part, un codage sémantique des parties du texte⁶ peut être envisagé, et enregistré dans le modèle du document (DTD) (Toulotte, Sanson 1993). La vocation du langage SGML est ainsi mieux respectée, au profit de sa puissance : les balises n'ont pas à s'en tenir à la présentation, la mise en page du texte, (tous les paragraphes sont dans ce cas mis à plat), mais ont à repérer l'architecture logique du document. D'autre part, la régularité de la structure, en favorisant l'homogénéité du corpus, possible un gain de qualité, de représentativité et de finesse des traitements automatiques globaux.

La difficulté pratique reste de pouvoir retrouver effectivement, et de retrouver sans erreur d'interprétation, la structure dans les textes. Sauf à mouler dès leur production les données dans le cadre prévu (par une structuration à la source comme celle proposée par Vuldy et Poncet), il s'agit de négocier le décalage entre la norme et les faits, entre la doctrine et ce qu'on observe.

Avant 1996, une vue rétrospective pourrait essayer de retrouver des contours sémantiques, des parties réalisant une fonction prévue (description du contexte, énumération des objectifs de l'année...). En effet, le plan-type s'inscrit dans la continuité des pratiques observées les années précédentes, en reprenant les cadres informatifs qui se sont forgés, ont été expérimentés, appréciés, repris.

En 1996, le plan de référence désormais explicité n'est pas toujours strictement suivi. Peut-être déjà n'a-t-il pas été communiqué partout avec le même appui ou la même exigence de rigueur (il a pu être perçu comme un plan indicatif ou bien comme un plan strictement imposé). On observe des variantes dans la formulation des titres des parties et dans leur nombre (certaines parties semblent accessoires ou contextuelles, et sont souvent omises).

A partir de 1999, une feuille de style sera utilisée lors de la saisie des fiches d'Actions. On disposera donc d'un marquage sûr du découpage en parties et des intertitres (au lieu de les retrouver par un calcul heuristique). L'information ainsi enregistrée sera structurelle et non pas sémantique : les natures différentes des parties ne seront pas distinguées. Cependant, cela suffira pour une visualisation propre, et cela facilitera considérablement les procédures de reconnaissance automatique.

3. Les parties du texte d'Action : référentiel pour cette étude

Pour décrire la structuration interne des textes d'Action, nous avons établi un inventaire des parties qui permet de rendre compte du plan-type institutionnel et des observations à partir de corpus.

Comme nous sommes amenés à faire constamment référence aux différentes parties, nous introduisons dès maintenant l'inventaire que nous en faisons et les notations choisies pour les désigner. La validité et la consistance de cette grille d'analyse seront établies rétrospectivement, par son adéquation descriptive et les résultats qu'elle permet d'acquérir.

Numéro	Nom	Notation
Partie 1	<i>Contexte</i>	Contexte
Partie 2	<i>But</i>	But
Partie 3	<i>Etat d'avancement</i>	Etat
Partie 4	<i>Programme</i>	Programme

⁶ Exemple de structuration sémantique : un groupe de mots qui était étiqueté « intertitre » (dans une structuration directement liée à la mise en page) peut être précisé en « titre de la section décrivant le but de l'Action », et associé à la zone de texte qu'il couvre. Et même, si cet intertitre ne consiste en rien d'autre qu'annoncer la nature de la partie par un intitulé conventionnel (comme *But de l'Action*), sa formulation peut purement et simplement disparaître au niveau du codage, et être généré automatiquement pour l'affichage en ouverture de la partie *But*.

Partie 4b	<i>Etapas</i>	Etapas
Partie 5	<i>Objectifs ultérieurs</i>	Ultérieur
Partie 6	<i>Documents de référence</i>	Référence
Partie 7	<i>Partenariats extérieurs</i>	Partenariat
Parties secondaires :		
<i>Partie n0</i>	<i>Préalable</i>	
<i>Partie n1</i>	<i>Actions liées</i>	
<i>Partie n4</i>	<i>Agents participant</i>	
<i>Partie n5</i>	<i>Période de déroulement</i>	
<i>Partie n</i>	<i>Annexe</i>	

4. Le corpus de travail

Notre corpus comprend les fiches d'Action complètes,

- au format SGML (les données administratives et budgétaires sont identifiées ; dans le texte, la division en paragraphes est enregistrée, et des algorithmes ont permis de repérer automatiquement, non exhaustivement mais avec peu d'erreurs, certains renvois à d'autres Actions, et certains titres de parties⁷)
- année par année, de 1991 à 1996, soit 6 années,
- pour l'ensemble de la DER, sauf les entités qui n'ont pas à proprement parler une activité de recherche : le Service AGT (son activité est centrée sur l'administration et sur la gestion du site) et l'état-major de la Direction (ses Actions se rapportent au management concret de la recherche et à son orientation stratégique).
- à l'exclusion des Actions de statut « Confidentiel ».

Ce corpus est produit par une extraction de la base de données SPHERE, à l'aide du programme SPHERE2LE décrit dans (Le Gall, Odone 1995). Le corpus est donc fourni au format utilisé pour le *Livre Electronique* (même DTD).

	1991	1992	1993	1994	1995	1996	1997
Volume du fichier SGML (MégaOctets)	4,4	6,1	6,9	7,4	10,6	13,2	14,9
Nombre d'Actions dans le corpus (hors AGT, hors Direction, hors confidentielles)	1 675	2 119	2 105	2 066	2 179	2 299	2 483
Nombre d'Actions du corpus avec un texte vide	87	95	105	262	167	5	4

⁷ Le repérage des renvois d'une Action à une autre se fait par la reconnaissance du code de l'Action mentionnée. Les codes d'Action ont une forme régulière (en principe), combinant des lettres, chiffres, et séparateurs. On en trouve plusieurs exemple dans la bibliographie de cette étude. Le repérage automatique des titres de parties est décrit dans la section suivante, concernant l'existant en matière de structuration des Actions.

C. POINTS DE VUE SUR LA STRUCTURATION DU TEXTE DES ACTIONS : EXISTANT

1. Vue normative : plan-type de référence

Le document de doctrine (Allard, 1995) fournit les plans pour les documents d'ordonnancement, assortis de brèves recommandations.

Le plan des ARD/AID prévoit 7 parties, dans un ordre déterminé, avec un titre fixé (qui « doit être libellé exactement comme le modèle en utilisant des caractères gras »).

Rien n'indique le caractère obligatoire ou facultatif de chaque partie. Implicitement donc, l'Action n'est convenablement décrite que si tout est présent. Néanmoins, une consigne de présentation considère le cas où une rubrique n'est pas renseignée (« Au cas où une rubrique ne serait pas renseignée, il n'est pas nécessaire de la faire figurer »). L'étude attentive des recommandations associées à chaque partie révèle que :

- quatre parties sur 7 reçoivent une description du contenu attendu qui explique le point de vue à adopter. Cela se présente comme le cadre de mise en forme d'une information *a priori* existante et demandée.
- deux parties sur 7 introduisent un critère pour décider s'il existe des éléments à indiquer : par exemple, dans la partie 6, on ne mentionnera des documents que s'ils apportent un supplément d'information sur le contenu, le programme de l'Action ; ou bien, en partie 7, on fera état des (éventuels) partenariats si leur enregistrement au niveau DER se justifie. Pour ces deux parties donc, l'Action n'a pas nécessairement les éléments particuliers visés.
- une partie n'attire aucune remarque particulière, celle concernant l'état d'avancement. Son usage est un peu moins fréquent, dans la mesure où cette rubrique est potentiellement inopérante lorsque l'Action est nouvelle.

Le noyau minimal d'information commun à toutes les Actions serait donc formé des parties 1, 2, 4 (la partie 5 étant peut-être moins assurée, jusque dans sa présentation : « objectifs pressentis pour les années ultérieures », sans autre précision). Les parties 3, 6, 7 seraient conditionnelles par nature.

2. Vue technique : modèle de document SGML

Une DTD sert actuellement pour le *Livre Electronique*. Dans cette DTD, le corps de l'Action (c'est-à-dire le texte qui nous intéresse ici) est beaucoup plus faiblement structuré que l'en-tête⁸. Il n'y a pas de structure sémantique applicable aux paragraphes et aux parties du texte de l'Action. Le modèle code principalement des informations utilisées pour l'affichage (il a une présentation adaptée pour les titres de paragraphes, les paragraphes, les listes, etc.) et pour la navigation hypertexte (les renvois à d'autres Actions se traduisent par des liens).

Une autre DTD existe mais n'est pas (encore) utilisée actuellement. C'est une DTD plus « sémantique », et c'est elle qui nous intéresse ici.

Elle a été écrite dans le cadre d'une expérimentation technique de saisie « à la source » de la structure de l'Action selon le plan-type. Le texte de l'Action est un document Word, avec la particularité de se doter de styles prédéfinis qui correspondent aux différentes parties du plan. Chaque paragraphe est ainsi affecté du style qui convient. Un logiciel du commerce (*SGML Author*), paramétré pour l'application, permet de générer automatiquement le document structuré sous forme SGML.

Ce contexte applicatif et expérimental a joué au moment de la conception de la DTD. Par exemple, elle est assez souple pour enregistrer des variantes de titres de parties (sinon, le titre pourrait être généré automatiquement au besoin, connaissant la partie dans laquelle on se trouve et la formulation donnée par le plan-type). Elle permet aussi de donner le statut « confidentiel » à

⁸ Dans tous les cas, la finesse de la structuration est conditionnée par l'architecture des données gérées dans les bases (données textuelles dans SPHERE, données administratives dans SAGACE).

n'importe quel élément (titre, paragraphe...) indépendamment de son contexte, tout en prévoyant la possibilité de parties entièrement confidentielles, hors du plan-type.

La DTD est calquée sur le plan-type quant à la nature des parties et à leur ordre. Ceci est encore vrai pour la variante à buts multiples, la seule partie factorisée étant alors le *Contexte*.

En revanche, à la différence de ce que nous avons observé dans le document de doctrine, la DTD est obligée de se prononcer sur le caractère obligatoire ou non d'une partie. Le modèle reste prudent, une seule partie est imposée : le *But*. C'est donc ici la partie 2 (*But*) qui constitue le noyau dur de l'Action. Un tel modèle permet de rendre compte avec justesse d'à peu près toutes les Actions des années 1991-1996, dont un nombre notable n'ont qu'une partie intitulée « But de l'action » (ou variantes).

Sur la base de nos observations pour 1991-1996, les limites pour l'application rétrospective de cette DTD seraient :

- l'ordre des parties, qui n'est pas constant sur ce corpus ;
- le cas d'Actions à buts multiples comprenant un but général : la DTD n'empêcherait pas d'étiqueter le paragraphe comme « but », mais situerait dans l'arborescence ce paragraphe au même niveau que les buts particuliers.
- le cas des Actions avec 1 ou 2 parties : il arrive que le paragraphe coiffé d'un titre comme « but de l'action » traite plutôt du programme de l'Action (partie 4), voire de son contexte (partie 1). Le but semble à première vue présent (annoncé par le titre de la partie), mais en réalité ne figure pas (à la lecture).
- les listes d'un seul niveau : en particulier, le détail du programme (partie 4) a souvent une structure très analytique et arborescente, avec plusieurs niveaux d'énumération imbriqués.

3. Vue algorithmique : automate de reconnaissance des titres de parties

Les textes des Actions sous forme électronique ne comportent à l'origine (dans SPHERE) aucun codage des parties, alors que généralement les titres qui introduisent les parties sont repérables par une lecture attentive. Pour le *Livre Electronique*, un algorithme a été mis au point pour détecter automatiquement ces titres de partie, afin de les mettre en valeur typographiquement. Ainsi les Actions sont beaucoup plus lisibles, leur présentation est plus claire et permet de s'orienter plus efficacement dans chaque texte et d'un texte à l'autre.

Le fonctionnement de l'algorithme est simple et relativement efficace. Il se base sur les hypothèses suivantes :

- un titre fait une longueur limitée (il ne dépasse pas 55 caractères).
- l'ensemble des mots qui peuvent commencer un titre est bien défini (une quinzaine de mots sont recensés pour les fiches et les comptes-rendus).
- quelquefois, selon le mot qui débute la ligne, il faut vérifier une troisième condition pour pouvoir reconnaître un titre : présence d'un autre mot donné, ou une contrainte plus serrée sur la longueur.

Ici, on désigne par mot une chaîne de caractères non nécessairement suivie d'un blanc, c'est-à-dire que le mot « objectif » sera reconnu si l'on a « objectifs ».

Cet algorithme donne de bons résultats pour la mise en page du *Livre Electronique*. S'il n'identifie pas tous les titres, il est rare qu'il se trompe pour ceux qu'il reconnaît. Assez souvent il sélectionne les sous sections *objectif n°1*, *objectif n°2* etc. décomposant la partie *Programme*. A l'affichage ces deux niveaux de titre sont donc confondus, ce qui empêche de se dégager clairement les parties du plan type, mais cette division marquée des objectifs du programme est en accord avec le rythme de l'exposé.

Si l'on entre dans le détail de l'algorithme, on s'aperçoit qu'il repère en outre différents genres de paragraphe, dits « groupes ». Des titres reconnus sur des critères différents sont ainsi associés, s'ils relèvent d'un même genre. Ces genres peuvent être mis en correspondance avec le plan-type comme suit :

Partie de l'Action	Genre de paragraphe
Partie 1 : <i>Contexte</i>	CONTEXTE
Partie 2 : <i>But</i>	BUT, OBJECTIF
Partie 3 : <i>Etat d'avancement</i>	ETAT
Partie 4 : <i>Programme (et Etapes)</i>	OBJECTIFETAP, OBJECTIF, ETAPE, PROGRAMME
Partie 5 : <i>Objectifs ultérieurs</i>	OBJECTIFULT, ORIENTATION
Partie 6 : <i>Documents de référence</i>	DOCUMENT, REFERENCE
Partie 7 : <i>Partenariats extérieurs</i>	PARTENARIAT
Partie n5 : <i>Période de déroulement</i>	PERIODE
Partie propre au Compte-Rendu d'Action	RESULTAT, RAPPEL

Deux types d'écarts se manifestent :

- d'une part, l'algorithme considère plusieurs genres de paragraphe pour une seule et même partie du plan-type (par exemple dans le cas des parties 4 ou 5).
- d'autre part, un même genre de paragraphe peut correspondre à plusieurs parties du plan-type (notamment le groupe OBJECTIF).

En conclusion, cet algorithme s'avère très satisfaisant pour un travail de mise en page (en l'occurrence pour le *Livre Electronique*) ; il pourrait être un préliminaire, mais serait à compléter, pour l'identification des différentes parties d'une Action. Autrement dit, il peut convenir pour un premier découpage des Actions, fournissant des zones visuelles et contextuelles, mais manquerait de précision pour un traitement ajusté au contenu de chaque partie.

4. Vue théorique (linguistique computationnelle) : l'analyse de l'organisation générale des textes d'Action

Florence Pugeault a travaillé sur l'extraction de connaissances structurées dans les textes d'Action, dans le cadre d'une thèse à l'Institut de Recherche en Informatique de Toulouse. Son but est « d'aller au-delà de l'indexation classique par mots-clés en représentant l'information retenue comme pertinente sous une forme structurée. » (Pugeault 1995 - résumé). Cette forme structurée est celle de la relation prédicat-argument : il s'agit de « représenter les actions par le biais de prédicats et les objets par les arguments des prédicats concernés. Les relations entre un prédicat et ses arguments sont exprimées par le biais de rôles thématiques qui sont assignés aux arguments du prédicat. » (Pugeault 1995 - p.5).

Son corpus de travail est beaucoup plus restreint que le nôtre. Voici ce que l'on peut en dire à partir de la thèse (Pugeault 1995) :

- ce sont des textes d'ARD (p.5)
- de l'année 1993 (exemple p.72)
- en version provisoire (certains exemples présentent quelques différences par rapport aux versions définitives dont nous disposons ; par ailleurs, la disponibilité du corpus en juin 1992 correspond mieux au calendrier de la thèse)
- du Service IMA, et accessoirement du Groupe N46 qui a lancé cette thèse (on le déduit de l'ordre de grandeur du nombre de textes (200), des exemples cités que l'on a pu identifier (I2309F, I2011R, I2101R, I2102R, N4603R, I2305R), et de l'explication de la régularité des textes par l'appartenance à une même entité)

- uniquement la première partie des textes, correspondant à l'exposé du contexte (« Les textes d'ARD dont nous disposons sont des textes relativement courts, moins d'une quarantaine de lignes (nous considérons la partie « Contextes et Objectifs » des ARD) »).
- on ne dispose pas du titre général de l'Action.
- chaque texte est donné sous la forme d'une liste de phrases, dans l'ordre dans lequel elles apparaissent dans le texte.
- il n'y a pas les limites des paragraphes (ce qui correspondrait aux retours-chariot dans le texte d'origine).
- la ponctuation est conservée l'intérieur de chaque phrase en ce qui concerne les virgules et les guillemets.

Le traitement automatique pour obtenir l'indexation structurée se déroule en trois phases successives. Ces phases sont appelées respectivement *niveau pragmatique*, *niveau linguistique* et *niveau conceptuel*. Seule la première phase (*niveau pragmatique*) nous intéresse ici, car les deux suivantes se déroulent dans les limites de chacune des quelques phrases retenues. L'analyse qui suit ne porte donc que sur une partie du travail de thèse.

Florence Pugeault affirme que les textes de son corpus contiennent des informations de quatre sortes : *thème*, *motivations*, *problèmes rencontrés*, *réalisations*. Elle les nomme *articulations* du texte.

Voici les définitions précises du contenu des différentes articulations telles que nous les avons élaborées à partir de l'analyse linguistique des textes.

Le <THEME> caractérise l'objet du discours. Il correspond le plus souvent à un sujet assez général et central au texte : c'est ce sur quoi se focalisent les auteurs du texte. Dans le cas d'Actions de Recherche, par exemple, le <THEME> correspond au sujet de l'ARD, à ce sur quoi les chercheurs travaillent, à leur intérêt principal.

Dans un texte, le <THEME> est considéré comme obligatoirement présent, car c'est le point central et la base de la définition de l'activité relatée par le texte. [...]

Les <MOTIVATIONS> sont une articulation qui exprime les objectifs généraux de l'ARD-AID : les besoins nécessités par l'état de l'art courant en la matière, les buts, les objectifs à atteindre ou les projets à venir qui motivent le développement de cette Action. Nous considérons que cette articulation définit le dessein majeur de l'auteur du texte. Dans le cas d'une Action de Recherche cette partie cherche à justifier les objectifs et les besoins de l'équipe de recherche, sans pour autant préciser un quelconque moyen de réaliser ces objectifs. [...]

Cette articulation est optionnelle dans les textes d'ARD. Cependant, cette articulation est présente en règle générale puisqu'elle justifie le travail des auteurs et répond aux attentes des rapports d'ARD, ce qui est un point capital de ce genre de texte. [...]

Les <PROBLEMES> correspondent à l'énoncé des problèmes qui sont liés à l'état de l'art courant, et à la situation actuelle du projet. Il s'agit en fait dans cette partie d'explicitier la situation dans laquelle on se trouve à présent, afin de justifier ce qu'il nous reste à faire pour atteindre l'objectif décrit dans l'articulation <MOTIVATIONS>. Dans les ARD, lorsqu'il existe des problèmes liés à l'état d'avancement du projet, c'est ici qu'ils sont mentionnés afin de proposer d'éventuelles solutions réalisables à moyen terme pour la bonne continuation du projet de recherche. L'articulation <PROBLEMES> est facultative. Dans le corpus d'ARD dont nous disposons cette articulation est peu renseignée : les problèmes sont en fait longuement décrits dans une partie qui leur est spécifique (partie « état de l'art courant » et que nous ne prenons pas en compte dans notre corpus (nous avons déjà dit que nous traitons la partie « contexte et objectifs » des ARD). Lorsque cette articulation existe, elle se situe le plus souvent entre les articulations <MOTIVATIONS> et <REALISATIONS>. [...]

L'articulation <REALISATIONS> introduit les différentes tâches, effectuées ou en prévision, de l'Action de Recherche que l'on traite. L'organisation du travail et des tâches à réaliser sont les points principaux qui sont exprimés dans cette articulation. Cette articulation correspond en fait au niveau concret de la partie <MOTIVATIONS> : elle explicite ce qu'il reste à faire pour atteindre l'objectif qui a été fixé. Ceci revient donc le plus souvent à faire une liste exhaustive d'actions à réaliser à court terme pour concrétiser l'objectif de l'ARD à long terme. Dans les textes d'ARD, cette articulation concerne en général les travaux à réaliser dans le courant de l'année suivante. Cette articulation peut être absente, mais elle est généralement très détaillée. Elle est parfois complétée dans la partie « Objectifs pour l'année à venir » des documents. Les phrases relatives à cette articulation se trouvent en général à la fin des textes que nous étudions.

Florence Pugeault donne en exemple le repérage qu'elle fait de ces parties dans quatre textes (comme les articulations s'enchaînent, nous avons seulement indiqué où chacune commence) :

Texte 1 :

<THEME> Les mesures destructives (ou assimilables) posent toujours des problèmes concernant le faible nombre de données disponibles ou encore leur coût qui s'associe généralement à la nécessité d'une bonne précision.

<MOTIVATIONS> Il est donc nécessaire d'optimiser les campagnes de mesure pour mieux connaître, voire améliorer, les incertitudes de mesure, et, lorsque cela est possible, réduire les coûts induits.

<PROBLEMES> Ces problèmes sont d'autant plus difficiles à traiter que les paramètres en jeu ont des comportements non-linéaires.

<REALISATIONS> Il est donc nécessaire, au préalable, d'étudier les méthodes permettant de prendre en compte cette non-linéarité.

Texte 2 :

<THEME> L'automatique est une discipline en constante évolution, ce qui lui permet d'aborder des problèmes de plus en plus complexes.

<MOTIVATIONS> En retard, la complexité croissante des modèles traités oblige à des avancées dans les méthodes mais aussi dans les outils.

De ce fait, on constate que les méthodes nouvelles, telle que la mu-synthèse par exemple, peuvent être mises en œuvre aisément grâce à des outils informatiques d'ores et déjà commercialisés.

Les approches neuronales, basées sur les méthodes floues, peuvent également être mises en œuvre aisément.

<PROBLEMES> A cette facilité de mise en œuvre, il faut opposer la difficulté à obtenir un modèle réaliste et adapté à la méthode retenue, et surtout le choix de la méthode elle-même.

En effet, il n'est pas encore possible de passer dynamiquement d'une méthode de résolution à l'autre, comme on peut le faire dans le domaine de la mécanique, avec un solveur comme ASTER par exemple.

<REALISATIONS> De ce fait, un soutien méthodologique aux utilisateurs reste indispensable.

Texte 3 :

<THEME> L'objectif est d'assurer au sein de l'établissement une formation et une diffusion minimale centrées sur l'outil Genesis2.

<REALISATIONS> D'autre part, le groupe participera à la diffusion des connaissances par l'organisation de séminaires, la participation à des groupes de travail et par la poursuite de l'action de veille technologique.

Texte 4 :

<THEME> Il s'agit de la continuation d'une action pluriannuelle visant à construire un système, baptisé Descartes, et une base de connaissances dédiée à la programmation sous ses divers aspects.

<MOTIVATIONS> Cette action est liée aux actions d'application de la synthèse de programme à différents problèmes de l'entreprise : quittancement élaboré, Darts (temps réel), solveurs, refonte CAO.

La procédure de reconnaissance des articulations est automatisée. A chaque articulation sont associés un ensemble de *déclencheurs*. La présence d'un déclencheur permet de classer la phrase qui le contient dans l'articulation correspondante.

Les déclencheurs sont donnés en annexe de la thèse. Ce sont des mots du lexique (noms, verbe ou construction verbale, adjectif, adverbe, préposition), des connecteurs (préposition, locution, conjonction, expression), la position de la phrase dans le texte (première, dernière), les temps des verbes, la négation. Des paradigmes se dégagent, et permettent de décrire une articulation avec quelques familles de déclencheurs : expressions temporelles, expressions de but, de possibilité, d'action, etc.

L'heuristique de reconnaissance des articulations est la suivante. On considère l'une après l'autre les quatre articulations : d'abord le *thème*, puis les *motivations*, puis les *problèmes*, enfin les *réalisations*. On retient pour chacune au plus une phrase : la première phrase (dans l'ordre du texte) qui comporte un déclencheur de l'articulation cherchée et qui n'a pas déjà été retenue pour une autre articulation. Le *thème* est la seule articulation toujours présente : si aucun déclencheur n'est trouvé, c'est la première phrase du texte qui est sélectionnée comme *thème*. Quand la recherche des quatre articulations a été faite, et donc a abouti au repérage d'au plus quatre phrases, toutes les phrases

restantes prennent le statut de *développement* de l'articulation qui les précède immédiatement dans le texte.

Ce travail s'écarte du nôtre sur beaucoup de plans, mais peut néanmoins apporter quelques éléments à notre étude.

- La limitation au tout début du texte d'Action exclut d'emblée l'étude de l'organisation des parties des textes d'Action. En revanche, pour l'analyse du contenu des premières parties (*Contexte, But*), on peut étudier la validité descriptive des quatre articulations proposées : *thème, motivations, problèmes, réalisations*.
- Une seconde limitation définit ici le corpus, celle à un seul Service (le service IMA). Or nous avons vu que l'appartenance à une même entité était un facteur d'homogénéisation des textes. Le corpus de Florence Pugeault est considérablement plus homogène que celui de l'ensemble des Actions de 1991 à 1996. Les faiblesses que nous pressentons quant à la stabilité et la généralisabilité d'une analyse quasiment strictement locale à la phrase et par règles ponctuelles (présence d'un élément déclencheur dans une phrase) d'une part confirme la difficulté de la structuration du corpus, et d'autre part encourage l'adoption d'une approche différente, davantage globale et contextuelle.
- Le traitement appliqué par Florence Pugeault obéit à une logique de réduction quantitative du nombre de phrases. En effet, elle utilise des théories linguistiques computationnelles qui ne fonctionnent qu'au niveau de la proposition (centrée sur le verbe). Le volume de la représentation sémantique obtenue est proportionnel au nombre de phrases considérées. Par la force des choses, son objectif est donc de s'en tenir à quelques phrases sélectionnées : si bien que les premières opérations consistent à passer du texte d'Action complet à un extrait court (par construction du corpus), puis de l'extrait (correspondant à peu près à la partie contexte) à au plus quatre phrases, élues comme représentant les quatre articulations. Une telle approche, dominée par la syntaxe de la proposition, commence par démembrer le texte pour s'attacher à quelques bribes isolés. Elle se prive d'emblée d'une vision globale et synthétique. Une étude de la structuration d'ensembles de textes est donc appelée à sortir des sentiers battus de la linguistique computationnelle axée sur les phrases.
- Une prise en compte (trop) stricte de la linéarité du texte est discutable, comme par exemple le fait de rapporter les phrases non sélectionnées à l'articulation précédente, avec le statut de développement. Il y a par contre une piste plus intéressante dans le fait de hiérarchiser qualitativement les articulations (ordre de priorité pour le balayage du texte). Florence Pugeault observe d'ailleurs qu'un algorithme traitant les textes articulation par articulation est supérieur à un algorithme qui procéderait d'abord phrase par phrase : il y a moins d'erreurs dans les résultats produits (Pugeault 1995 - p.113).
- C'est sur le plan méthodologique que nous nous trouvons des affinités avec le travail de Florence Pugeault : (i) le point de départ est l'observation des textes, elle est fondée sur l'étude du corpus ; (ii) l'analyse linguistique ne se cantonne pas à un type d'éléments (les noms par exemples), mais relève des données de natures diversifiées et de différents paliers (mot, phrase, texte) ; (iii) des paradigmes de dégagent et s'interprètent par rapport au contenu attendu de chaque partie.

D. ELÉMENTS D'ANALYSE DE L'ORGANISATION DES TEXTES : FONCTIONS DES PARTIES

Ce chapitre récapitule, partie par partie, les observations faites sur le corpus, en intégrant l'orientation donnée par les commentaires associés au plan-type du document de doctrine. L'objectif est de donner une analyse réaliste et motivée du contenu propre à chaque partie.

1. Contexte

Cette rubrique élargit l'horizon par rapport à ce que décrit le reste de la fiche : enjeux pour l'Entreprise, décisions prises, évolution des techniques... Elle peut également comporter des indications sur le caractère pluriannuel de la recherche (période de déroulement, coût total estimé, organisation...). (Allard 1995)

L'intitulé *Contexte* résonne en fait sur quatre modes, susceptibles d'apparaître conjointement dans les textes.

- *Contexte de référence* : cadre le sujet dans une discipline, cite la théorie, la méthode, l'outil, ou toute autre entité connue et reconnue autour de laquelle s'articule l'Action. Tout se joue ici sur ce « connu et reconnu » : *connu*, qui rapporte à un cadre familier, éclaire, donne un premier degré de compréhension ; *reconnu*, donc digne de confiance, car la solidité de la discipline, l'expérience acquise ou la notoriété publique se font les garants du sérieux du projet.
- *Contexte justificatif* : décrit l'existant, met en exergue une difficulté à résoudre (éventuellement soumise par une Direction opérationnelle), précise le gain à attendre de l'étude⁹. Le ressort de ce contexte justificatif est de convaincre de l'intérêt du thème, de par son caractère novateur, son utilité directe et immédiate pour l'entreprise, ou son apport à moyen terme. Le contexte justificatif révèle un besoin autour du thème de recherche proposé.
- *Contexte structurel* : explicite la portée de l'Action. Il inscrit l'Action dans un projet construit, montre son articulation avec d'autres Actions, esquisse un réseau de collaborations internes : rattachement à un même PPRD, Actions liées, contribution à un axe stratégique fédérateur... La force du contexte structurel est d'insérer l'Action au sein d'une architecture plus large : elle prend sa part propre dans la réalisation du tout, et affiche un esprit de coopération au bénéfice d'une œuvre d'envergure.
- *Contexte chronologique* : trace la présence de l'Action au fil des ans. Le point de vue peut être rétrospectif, en prenant du recul pour faire ressortir quelques étapes majeures : il donne une assise au projet de l'année, inscrit la recherche dans une dynamique prometteuse. Un autre point de vue est celui qui embrasse toute la durée de l'Action : il mesure le cheminement et annonce l'accomplissement du projet. C'est une gestion contrôlée, une maîtrise du temps, qui donne assurance.

Sur ces quatre modes vont se réaliser les trois fonctions que nous repérons pour cette partie.

- La première fonction est une *fonction unificatrice*. A savoir, le contexte assoit, instaure, la cohérence de l'Action. Il constitue l'unité de l'Action « de l'extérieur », en en dessinant les contours, en en dégageant la base et les fondements. Une manifestation forte de cette fonction unificatrice est le fait que le *Contexte* soit l'unique rubrique commune, centrale, à partir de laquelle se déploie l'éventail des buts, dans le plan-type prévu pour l'Action à buts multiples : c'est donc bien ici le contexte qui institue l'unité de l'Action.

⁹ Si elles sont présentes dans la partie contexte, les articulations *motivations*, *problèmes* et *réalisations* proposées par Florence Pugeault se retrouveraient dans ce contexte justificatif. Quant à l'articulation *thème*, elle serait plutôt à rapporter au contexte de référence. Néanmoins, *thème* et *motivations* sont préférentiellement développés dans la partie suivante (*But*) ; et les *réalisations* seraient plutôt l'affaire de la partie 4 (*Programme*). L'analyse de Florence Pugeault s'était basée sur un corpus très restreint et relativement homogène d'Actions d'un même Service, et pouvait en outre avoir porté sur des textes peu structurés, dont la première partie était en fait une fusion de plusieurs parties du plan-type (1, 2 et 4 par exemple).

- La seconde fonction de la partie *Contexte* est la *fonction introductive*. C'est la passerelle initiale qui fait le lien entre le lecteur et le projet spécifique et technique exposé. Le lecteur destinataire est plus ou moins familier du sujet mais veut être en mesure de situer l'Action et son impact par rapport à la stratégie de l'entreprise. Pour conduire son lecteur sur cette passerelle, le rédacteur est amené à adopter un style ménageant la transition : phrases construites, s'enchaînant dans des paragraphes synthétiques (plutôt par exemple que des énumérations analytiques), en évitant la multiplication des termes trop pointus.
- Enfin, il y a une *fonction argumentative* sous-jacente. Tous les éléments convergent pour « bien présenter » le projet de recherche. L'Action doit mériter d'être proposée. C'est encore le cadre, mais au sens de l'encadrement soigné d'une peinture destiné à la mettre en valeur.

2. But

Description qualitative et dans un cadre en général pluriannuel de ce qui est visé au titre de l'Action. Vision plus globale que celle de la rubrique « Objectifs et principales étapes ». (Allard 1995)

Dans la pratique, le but doit souvent être défini par rapport aux objectifs, du fait de la proximité de sens entre les mots « but(s) » et « objectif(s) ». Ces mots sont d'ailleurs couramment utilisés l'un comme l'autre pour l'une ou l'autre des parties 2 ou 4.

La partie *But* expose ce qui est visé par l'Action de façon générale et à long terme, alors que la partie 4 (*Programme*, souvent intitulée « Objectifs de l'action ») s'attache à détailler les objectifs que l'on s'assigne pour l'année, et qui ne sont souvent que des étapes intermédiaires, concourant à la réalisation générale annoncée par 2 (le *But*).

Cette partie *But* a deux fonctions majeures.

- La *fonction définitoire* explicite et cerne de façon nette et précise la teneur de l'Action. La formulation est bien analogue à celle d'une définition : (i) Utilisation fréquente d'une tournure conventionnelle présentative, objectivante, comme « Cette action a pour but de », « La présente action a pour objet de », « Cette action vise à », « L'action consiste à », « L'objectif de l'action est », « Il s'agit de ». (ii) Mention d'une action (ou plusieurs, sous forme d'une liste¹⁰), exprimée par un verbe soit à l'infinitif soit substantivé. Il est remarquable que le but n'est jamais un état (visé à terme) mais une action, vue comme un accomplissement. (iii) Instanciation : on fixe la couverture, le domaine d'activité de l'Action, dans sa mise en œuvre concrète ; quels travaux sont par nature ou par convention inclus ou exclus. Un fil conducteur peut être donné en signalant quelques tâches permanentes, de fond. La partie hérite finalement de la propriété de stabilité de la définition : le but est en quelque sorte trop constitutif de l'Action pour connaître des révisions majeures sans compromettre l'identité de l'Action.
- La seconde fonction est la *fonction de synthèse*, au sens où cette partie prend valeur de résumé de l'ensemble. Elle va tendre à respecter les propriétés classiques de brièveté, lisibilité, représentativité. De ce fait, ce qui est retenu dans cette partie forme le cœur, le noyau de l'ensemble de l'Action.

On retrouve donc en corollaire une fonction d'unification, mais cette fois-ci unifiant « par l'intérieur », en donnant l'essence de l'Action (et non par l'extérieur, comme la fonction unificatrice du contexte qui donnait les contours). Cette *fonction unificatrice* peut être contrecarrée par l'éclatement du but dans la variante à buts multiples.

On perçoit bien l'importance de cette partie pour le lecteur qui doit prendre connaissance rapidement d'un grand nombre d'Actions. Ce paragraphe en devient central, essentiel, incontournable.

Et pourtant il peut s'amenuiser jusqu'à disparaître. Le stade minimal de présence est l'intertitre plein : le but est explicité par un verbe à l'infinitif ou substantivé, suivi de ses compléments. Cette forme semble même celle prévue dans le plan-type de la variante à buts multiples. En effet, les indentations du plan tel qu'il est mis en page dans le document de référence (Allard 1995) placent chaque but non pas au même niveau que les autres parties, mais comme chapeau des

¹⁰ L'analogie avec les définitions d'un dictionnaire ne doit pas être poussée trop loin. A la différence des articles de dictionnaire enregistrant les acceptions d'un mot, et qui sont conçus comme des alternatives (« ou »), l'énumération des actions détaille les différentes facettes concourant conjointement à l'Action (« et »).

parties qui en dépendent. En outre, la formulation *But n°1*, *But n°2*, etc., fait un parallèle avec les sous-parties de la partie 4 (*Objectif 1*, *Objectif 2*, etc.) dont il est dit : « ces paragraphes peuvent être libellés directement avec un titre d'objectif ou de résultat ».

La disparition du *But* peut se comprendre en raison de sa superposition avec le titre général de l'Action. En effet, le titre est à même de présenter une réalisation des deux fonctions du *But* : la fonction synthétique (de toute évidence), et éventuellement la fonction définitoire. Le *But* serait intrinsèquement redondant avec le titre s'il n'en constituait un développement éclairant. Le *But* s'exprime aussi davantage en termes d'actions à accomplir que de thème général.

Le caractère central de la partie *But* lui confère une importance ambivalente : elle est « obligatoire », au risque de perdre son caractère. Ainsi, un certain nombre de textes d'Action de 1991-1996 ne présentent qu'une seule partie, pas très longue mais un peu fourre-tout (contexte, but, programme) : elle est alors le plus souvent intitulée « But de l'action » (ou équivalent).

3. Etat d'avancement

Aucun commentaire n'est apporté sur le contenu de cette partie dans le document de doctrine.

L'*Etat d'avancement* consiste en un bilan rétrospectif du travail de l'année précédente effectué dans le cadre de l'Action, travail que poursuit le projet proposé.

Cette partie n'a donc pas toujours lieu d'être complétée quand l'Action est nouvelle. Dans ce cas, on a quelquefois simplement une mention comme « nouvelle action », ou « l'action démarre », voire un renvoi au Compte-Rendu. Mais il peut aussi être opportun d'indiquer des apports d'autres Actions ou d'études précédentes, qui vont être réutilisés ou sur lesquels est basée l'Action nouvelle.

Le calendrier de rédaction des fiches est tel que certains des résultats qui doivent être consignés ici sont encore hypothétiques, car la recherche est en cours. La première version des fiches est faite à la mi-année, la seconde en novembre-décembre. D'où les précautions et les *réserves* de l'exposé, l'emploi fréquent du futur antérieur au lieu du passé composé (dans les versions provisoires, de mi-année). Le registre reste celui du mode indicatif : le calendrier de l'Action suit son cours, on enregistre ou on prévoit des faits.

Cette partie peut être conçue sur un mode analytique ou synthétique.

- La *présentation analytique* reprend une liste de pôles de recherche actuels de l'Action et indique les résultats acquis, les questions à creuser pour chacun d'eux, voire des objectifs qui n'ont pu être abordés et figureront à nouveau au *Programme*¹¹. Elle peut suivre le plan du programme de l'année précédente (il s'agit alors d'une sorte de compte-rendu des principales pistes suivies), ou suivre le programme envisagé présenté juste après dans l'Action (c'est alors repérer le point de départ des objectifs que l'on se donne ensuite).
- La *présentation synthétique* ne détaille pas point par point l'avancée des recherches, mais résume les principaux acquis en quelques phrases. Cela ne l'empêche pas d'être précise, en marquant les dates d'étapes importantes, les noms propres et les sigles, permettant de se référer à des résultats consignés de façon développée ailleurs (par exemple dans le compte-rendu de l'Action).

Le ton est tantôt descriptif, tantôt évaluatif. Le *ton descriptif* est celui d'un état des lieux rendant compte des avancées de l'année passée et établissant la base de départ de l'Action. Le rédacteur adopte un *ton évaluatif* s'il juge les acquis, en tire des enseignements, et discute sur cette base des suites à donner : poursuite de telle piste, abandon de telle autre, réorientations, nouvelles études.

4. Programme (et étapes)

Les différents objectifs décrivent les principaux résultats mesurables attendus au cours de l'exercice. Il est fait explicitement référence à toutes les étapes d'ordonnement (numéro entre parenthèse dans le corps du texte). Des étapes non « officielles » (points de rendez-vous intermédiaires) peuvent également être mentionnées. (Allard 1995)

¹¹ Ce cas ne fait pas l'objet d'une rubrique spécifique : ce serait avaliser, régulariser une situation qui doit plutôt rester exceptionnelle. Mais aussi ce serait épingler sans nuance les Actions dans ce cas.

La partie expose le programme de l'année. Nous avons vu qu'elle se différencie du *But* déjà quant à son contenu. Le *But* donne la visée générale et à terme de l'Action, alors que le *Programme* fixe des objectifs précis et détaillés pour l'année. Une difficulté subsiste pour bien distinguer ces parties, si l'Action se décompose en activités permanentes vs activités particulières (activités du moment, pour l'année). Il n'est pas évident de savoir où consigner les activités permanentes : elles peuvent être vues comme constitutives de l'Action (et être rapportées dans le *But*), ou être simplement des activités que l'on poursuit d'année en année (et appartenir alors au *Programme*).

Dans le plan du document de doctrine cette partie se décompose en sous-parties numérotées : les objectifs. Ceci annonce déjà le caractère analytique du texte de cette partie, et nous verrons que cela participe à une fonction, la fonction schématique. Certains rédacteurs reprennent littéralement le plan-type en intitulant les paragraphes *objectif n°1*, *objectif n°2*, etc., d'autres choisissent des titres qui indiquent la teneur de l'objectif. La structure liste sous-jacente n'est pas obligatoirement marquée : il arrive qu'il n'y ait pas de numérotation, de points ou de tirets.

La première chose à noter est que le programme se déploie selon trois dimensions. L'une ou l'autre peut être adoptée préférentiellement comme fil conducteur. Ce sont : la dimension thématique, la dimension chronologique, et la dimension évaluative.

- La *dimension thématique* répartit les activités de sorte à former des thèmes de recherche. L'Action s'organise ainsi autour de quelques pôles d'intérêt, qui sont traités plus ou moins en parallèle. C'est sans doute la logique de présentation qui correspond le mieux à la division du programme en objectifs. Les intertitres qui désignent ces thèmes sont très synthétiques et précis : ils demandent, pour être compris, d'être situés dans le contexte général de l'Action, par exemple résumé par le titre de l'Action.
- La *dimension chronologique* est celle qui s'occupe du déroulement au fil des mois de l'année. Elle correspond plutôt à la division en étapes, puisqu'aux étapes sont associées des échéances.
- La *dimension évaluative* présente les travaux envisagés par ordre d'importance ou de priorité. Elle peut en fait doubler la dimension thématique, par exemple en classant les pôles des pôles dominants aux pôles subsidiaires.

En revanche, la dimension thématique et la dimension chronologique ne sont *a priori* pas superposables. Aussi arrive-t-il que la partie *Programme* se scinde en deux : d'abord les objectifs (thématiques), puis un récapitulatif des étapes (chronologique). Cette scission est plus ou moins marquée : depuis le simple enchaînement au sein de la partie *Programme* (les étapes succédant aux objectifs), jusqu'à l'éclatement en deux parties autonomes, dont la seconde fonctionne comme un rappel des étapes (normalement fixées dans un autre document), et peut alors se situer tout à la fin du texte d'Action (telle une annexe).

La partie *Programme* remplit trois fonctions.

- La première fonction est la *fonction schématique*. Elle comporte trois aspects : simplification, organisation, annotation.

Il y a une *simplification* au sens où le rédacteur dégage les grandes lignes de son projet, il sélectionne un certain nombre de points clés, notamment les transitions (début, bilan intermédiaire donnant lieu à des choix d'orientation, bilan conclusif).

Cette partie tient aussi beaucoup du schéma pour l'*organisation* qu'elle propose en regroupant et structurant les activités mentionnées : ceci aide à la lecture et à l'intelligence du programme. La recherche à mener est analysée et décomposée, pour être exposée systématiquement, point par point, en suivant un plan éventuellement arborescent (à plusieurs niveaux : structure de listes et sous-listes). Le document de doctrine fixe bien dans cette partie un découpage en sous-parties qui sont autant d'objectifs numérotés. Nous avons vu que cette organisation peut être conduite selon trois axes : pôles thématiques, succession chronologique, ordre d'importance.

Enfin, et c'est là le troisième aspect schématique, on est ici plutôt dans le registre de l'*annotation* que dans celui de la rédaction. Chaque objectif peut se résumer à la seule mention d'une action exprimée par un verbe, à l'infinitif ou substantivé, suivi de ses compléments. Il peut aussi être indiqué par un libellé, commenté dans un paragraphe ou détaillé par une énumération de points. Tout ceci concourt à une forme de présentation analytique : décomposition, numérotation, hiérarchisation.

- La seconde fonction est la *fonction de concrétisation*. Autant le *But* se devait d'avoir une vue générale et, si nécessaire, de prendre du recul par rapport aux réalisations immédiates, autant les objectifs du *Programme* donnent corps à la recherche. On veut du concret, du mesurable. On parle en termes de délai, d'acquis, de tâche à accomplir, de produits, d'outils. Les noms et les sigles abondent et contribuent fortement à l'effet de réel : projets, matériel, équipes, partenaires, lieux. La partie complète utilement de ce point de vue les *Etapas*, car l'espace d'une partie permet d'exposer le déroulement des opérations, leur contenu (qui fait quoi, comment, où, pourquoi). Les *Etapas*, elles, se présentent sous forme de tableau, et s'attachent à un descriptif formel : intitulé, responsable, nature, échéance. L'intitulé est laconique et la nature s'en tient au support ou à la forme du résultat (par exemple : note, cours, code informatique, logiciel, documentation d'un outil).
- La troisième *fonction* est celle d'*engagement*. Cette partie détaille en quelque sorte le contenu du contrat qui règle l'activité du chercheur, et sur lequel les différents partenaires s'engagent. C'est un repère, vis-à-vis duquel le travail annuel du chercheur est ensuite évalué. D'où la focalisation sur des « résultats mesurables » (pour reprendre les mots du document de doctrine) : il est clair de voir si la date a été respectée, le produit livré ou la note écrite. Il n'est plus question d'une « description qualitative » comme celle du *But*.

5. Objectifs ultérieurs

Il s'agit d'objectifs pressentis pour les années ultérieures. (Allard 1995)

Chacun sait qu'il est aventureux de présager du déroulement intégral et des orientations d'un projet, ce qui donne ici un statut hypothétique aux actions envisagées, au contraire de la partie *Programme* où les actions prennent une valeur contractuelle (fonction d'engagement).

Il y a en fait deux manières de remplir cette partie : soit directement en évoquant le contenu de la suite de l'Action, soit en reportant ou annulant cette information : c'est trop tôt, cela dépend d'autres instances (« à définir en fonction des besoins »), l'Action se termine (« arrêt de l'action »).

Deux fonctions pourraient être accordées à cette partie.

- La première fonction est une *fonction de chute*. « chute » est un mot aux multiples consonances, nous en retenons deux.
La chute est d'abord la fin, la terminaison brève et conclusive d'un développement important. Le développement ici est le *Programme*, les *Objectifs ultérieurs* en sont la chute. C'est à cette partie qu'il revient éventuellement de rappeler le terme de l'Action à la fin de l'année. Si des perspectives sont évoquées, cela reste concis. Si on reporte l'information, une formule brève convient. La lecture en est rapide car on retrouve des tournures stéréotypées (« poursuite de », « à définir en fonction des besoins », « arrêt de l'action »).
On appelle aussi « chute » un reste inutilisé (par exemple, une chute de tissu). C'est ici ce qui déborde du cadre de l'année, qui n'entre pas dans l'exposé du programme. A ce titre, une étape plus lointaine mais déjà préparée peut être mentionnée.
- La seconde fonction est une *fonction diagnostic*. Le ton de cette partie donne une idée de la vigueur du projet (plus ou moins fortement soutenu), de son caractère (pérenne ou ponctuel, tout tracé ou très ouvert), de sa maturité (entre la phase initiale et le stade avancé), de son ambition (évolution et envergure). La formulation peut être assurée ou réservée, précise ou évasive, riche de pistes ou ténue, prudente ou bouillonnante, soignée ou quelconque. Les différences se creusent d'autant plus que cette partie apparaît comme mineure à côté du programme, et qu'elle reçoit *a priori* moins d'attention.

Ces deux fonctions sont orientées vers des effets opposés. A l'inverse de la fonction de chute, qui est un facteur d'homogénéisation (formules conventionnelles, brièveté), la fonction diagnostic joue des écarts dans la manière de traiter la partie. Quant à leur fonctionnement narratif, la chute se pose comme une clôture alors que le diagnostic engendre un éclairage rétrospectif.

6. Documents de référence

Documents permettant de préciser le contenu de l'action : plan d'action, programme de travail, texte de PPRD... (Allard 1995)

Le document de doctrine distingue les *documents de référence*, qui figurent dans la fiche d'Action, et les *documents produits*, qui appartiennent au compte-rendu. Ces derniers sont présentés comme des « documents (notes H, publications,...) fournissant le détail des résultats obtenus. »

Ces rubriques sont donc appelées à se spécialiser chacune pour un type de document : les *documents de référence* pour les documents descriptifs du projet (la délimitation de son contenu, sa portée, son organisation), et les *documents produits* pour les documents scientifiques et techniques exposant les réalisations.

Cette spécialisation des rubriques n'est pas toujours perçue dans les textes de 1991-1996. Les documents de référence sont entendus comme ceux auxquels le texte d'Action peut faire référence (au moyen quelquefois de renvois numérotés), y compris souvent les principales notes techniques récentes qui développent les études réalisées et sur lesquelles se fonde la suite des activités.

La rubrique constitue en quelque sorte une bibliographie choisie des documents de l'équipe (documents internes) rendant compte de l'état actuel du contenu de l'Action. On a donc une fonction bibliographique, mais bien particulière : quelquefois limitée à des documents de type programme (plan de travail, PPRD, manuel d'assurance qualité), et conventionnellement restreinte aux documents internes EDF émis par l'équipe ou par un regroupement (PPRD) dont elle dépend. Cette fonction bibliographique se réalise classiquement comme une liste de références. C'est plus rare que la partie comporte des phrases rédigées. Elle peut également être déclarée vide par une expression brève comme « Non », « néant », « sans objet ». Les documents de type programme peuvent n'être recensés que par leur nature et leur code d'identification. En revanche, les notes techniques sont toujours présentées à la manière de publications, avec, outre le code d'identification, les noms d'auteur, et le titre complet. Le titre est un titre plein qui informe sur la question abordée dans la note ; c'est un bon indicateur de l'objet de l'Action et des préoccupations courantes (ou récentes) du chercheur. La ponctuation est aussi relativement régulière et conventionnelle : par exemple, les titres d'articles ou de Notes internes sont généralement entre guillemets.

La partie se situe à la croisée de cette *fonction bibliographique*, et d'une *fonction d'officialisation interne*. Les documents mentionnés marquent le rattachement à un texte institutionnel (par exemple un contrat de PPRD), ou montrent l'adhésion à une démarche maison (plan qualité, manuel de sécurité). Ce sont aussi les documents enregistrés dans le fonds documentaire interne, la *Mémoire Technique Electronique* de l'entreprise, reflétant donc officiellement l'activité interne de recherche et développement.

La fonction bibliographique est dominante, et fait conventionnellement figurer la partie vers la fin du texte d'Action. Mais quelquefois l'influence de la fonction d'officialisation interne est telle que les documents se trouvent référencés à la suite de l'exposé du contexte, ou comme éléments tangibles et reconnus pour marquer l'état d'avancement.

7. Partenariats extérieurs

Organismes extérieurs (laboratoires, sociétés,...) intervenant dans l'action à un titre ou un autre (CERD,...) et qu'il est intéressant de repérer au titre de nos partenariats. (Allard 1995)

Les chercheurs sont incités à externaliser une partie de l'activité de recherche sous forme de sous-traitance (expérimentations, études, développements), de partenariats, ou de participation à des projets nationaux ou internationaux. Des collaborations scientifiques peuvent se nouer avec des Universités, des Ecoles, des Laboratoires, des partenaires industriels.

Cette partie n'a pas toujours lieu de figurer. Elle peut être vide, manquante, ou avoir une mention comme « sans objet », « RAS ».

- La fonction la plus évidente de cette partie est celle d'index. C'est un récapitulatif complet, recensant les partenaires mentionnés dans le programme de l'Action, ainsi que ceux restés jusque là implicites. Comme tout index, il est à la fois systématique et sélectif, car il fait acte d'enregistrement.

D'où l'importance de donner un statut précis à cet enregistrement : le plan-type s'intéresse aux partenariats envisagés, mais le rédacteur précise parfois que les partenariats sont engagés et participent déjà au programme de l'Action. Ce sont les partenariats en cours. Il y a alors superposition avec la rubrique équivalente prévue dans le compte-rendu de l'Action.

Toute entité est représentée par son nom : appellation courante, intitulé développé exact, ou sigle, qui identifie sans équivoque le partenaire concerné. Cette partie va ainsi concentrer les noms propres et les sigles, l'emploi des majuscules souligne chaque référence. Cette régularité typographique aide la lecture.

Enfin, de par cette *fonction d'index*, la rédaction peut être minimale voire inexistante, pour ne s'en tenir qu'à l'énumération des partenaires.

- La seconde fonction est une *fonction d'appui*.

Autant les documents de référence donnaient de l'Action une image interne à la DER, autant cette rubrique est explicitement consacrée au déploiement externe du réseau de relations. Elle traduit le rayonnement de l'Action. Il arrive que soient consignées des collaborations dans l'entreprise mais pour des secteurs d'activité éloignés.

Le fait d'avoir noué des relations de travail avec d'autres organismes témoigne en faveur de la consistance et du sérieux du projet. Ces relations s'interprètent comme un signe tangible de sa mise en œuvre.

De plus, la réputation, voire le prestige de certains partenaires, rejaillit sur l'Action. On apprécie que l'Action contribue à l'image de marque de l'entreprise.

- La troisième fonction est une *fonction de repérage, de sectorialisation*. Les relations externes sont analysées quant à leur nature et à leur contenu. Par exemple, à défaut d'indiquer le nom d'un partenaire, le réacteur peut définir le rôle de la collaboration envisagée dans le programme, la teneur de sa participation, le type de partenaire recherché pour l'effectuer : équipe de recherche spécialisée dans tel domaine ou ayant telle expérience, constructeur, etc. Ou encore, la liste des partenaires peut être organisée logiquement, avec regroupement des entités d'un même secteur ou ayant un même type de relation.

La fonction d'index entraîne conventionnellement cette partie à la fin du texte d'Action. Mais les deux autres fonctions peuvent motiver d'autres positionnements, proches du *Contexte*, de l'*Etat d'avancement* ou du *Programme*.

8. Autres

Il s'agit de parties non représentées dans le plan-type, mais qui ont été localement employées au cours des années précédentes. Nous allons voir leur contenu, et comment elles s'intègrent ou non dans ce plan-type.

- *Période de déroulement* : quand elle figure, cette rubrique se borne habituellement à la donnée des dates de début et de fin de l'Action. Elle est donc à la limite d'un contenu textuel, elle fonctionne plutôt comme un champ enregistrant des données factuelles (en l'occurrence des dates). La *Période de déroulement* ne figure pas explicitement comme une partie dans le plan-type, et de fait n'a jamais été employée de façon générale dans la DER. Par contre, elle a pu être consignée très régulièrement à l'échelle de certains Services, et donc être assez représentée. Le plan-type la reconnaît et lui donne sa place comme élément du *Contexte*. Semi-autonome, elle se détache généralement à la fin de la partie *Contexte*.
- *Actions liées* : cette partie donne les codes des Actions avec lesquelles l'Action est en rapport, à un titre ou à un autre (sujets connexes, rattachement à un même PPRD). Cela peut être repris par le *Contexte* sur le mode du contexte structurel.
- *Agents participant à l'Action* : l'ordonnancement se contente de recenser les contacts autour de l'Action : le responsable, le client, les autres interlocuteurs. Ce paragraphe est ajouté par quelques rédacteurs pour citer la composition de l'équipe.
- *Planification budgétaire* : le rédacteur évoque dans un développement la question du financement, récapitule les moyens à prévoir, estime les coûts, répartit les ressources. Les parties chiffrées de l'évaluation trouvent plutôt leur place dans les données budgétaires de l'en-tête de l'Action.

- *Valorisation* : propriété industrielle et dépôt de brevets. C'est sans doute trop particulier (et parfois trop confidentiel ?) pour figurer systématiquement en tant que partie du plan-type. Certains brevets pourraient être enregistrés au titre d'acquis (partie *Etat d'avancement*) ou d'objectif (partie *Programme*). Mais l'enregistrement de ces données revient plutôt à une autre base.
- *Glossaire* : une annexe répertorie le lexique du domaine, donnant les définitions du vocabulaire technique. La non institutionnalisation de cette partie (au demeurant rarement attestée auparavant : Département E2 en 1994) oblige le rédacteur à contrôler l'usage qu'il fait de la terminologie technique et à donner accès à son sens au sein même du texte.

E. VERS UNE RECONNAISSANCE AUTOMATIQUE DE L'ORGANISATION SÉMANTIQUE DES TEXTES D'ACTION

1. Un point d'appui : les intitulés

a) Fonctions des titres

Les titres (titre général, titre de partie, de paragraphe, etc.) ont un rôle évident et essentiel dans la structuration interne des documents textuels. Ils remplissent au moins cinq fonctions :

- Une *fonction de découpage* : par leur positionnement dans la linéarité du texte, ils marquent une frontière, une transition. Ils coupent le fil du texte, indiquent un seuil dans son déroulement. C'est déjà un premier stade pour appréhender ce qui sinon pourrait être un flot continu, peu « maniable ».
- Une *fonction de création d'unités* : le titre a une portée, il couvre une certaine portion du texte. En la délimitant, il instaure une unité textuelle.

Ceci n'est pas anodin : cette décomposition, conforme à un cadre logique agencé par le rédacteur, s'impose pour rendre l'ensemble d'un texte appréhendable, et cognitivement gérable (charge mémorielle). Ce serait à mettre en lien avec l'universalité de ce procédé de segmentation des textes, faisant que chaque passage du texte trouve sa place dans une architecture d'ensemble. La relative brièveté des titres s'accorde aussi avec ces propriétés mnémoniques.

D'autre part, constituer une unité, c'est indiquer sa singularité et sa consistance. La délimitation distingue un intérieur d'un extérieur. L'interprétation du texte est poussée vers la recherche de la cohérence de l'unité sur ces deux plans : facteurs communs qui homogénéisent son contenu (interne), et caractéristiques qui le différencient du reste (externe).

Alors que le découpage est directement matérialisé par la position du titre, l'indication de la portée du titre est plus complexe. Du fait de la linéarité syntagmatique de la langue écrite, repérer la portée c'est trouver les limites de début et de fin. Le titre marque conventionnellement le début, il est en position d'annonce. Pour saisir la fin, le lecteur interprète des éléments typographiques, le système de numérotation éventuel et l'insertion dans un sommaire, peut-être le type de la partie (place, longueur habituelle, etc.), la nature spécifique de son contenu, et sans doute potentiellement bien d'autres repères encore.

- Une *fonction d'annonce* : les conventions de présentation veulent que le titre soit antéposé, c'est-à-dire précède juste ce sur quoi il porte. Ceci est tout à fait en correspondance avec la réalisation de la fonction d'annonce : le titre oriente ainsi la lecture, crée des attentes sur ce qui devrait être trouvé, conditionne le mode de parcours du texte (par exemple lecture intégrale, si l'on s'attend à devoir suivre une progression logique, ou partielle, si la partie se construit sur une famille d'alternatives dont l'une périmé les autres).

Le titre reste en général indicatif, sans être réellement informatif : il donne le sujet traité sans préciser ce qu'on en tire¹². Les développements et les propositions restent l'objet du texte lui-même.

- Une *fonction d'accroche* : les titres sont des repères significatifs dans un balayage rapide d'un ensemble de textes ou d'une suite de pages. Ils médiatisent l'accès au texte : le titre a servi d'intermédiaire au texte pour capter l'attention du lecteur.

Dans certains cas, la fonction d'accroche occulte la fonction d'annonce : le titre s'écarte du contenu du texte, et n'y paraît plus fidèle. Cela peut engendrer la déception du lecteur, mais aussi créer un effet rhétorique de surprise qui au contraire réactive l'attention.

- Une *fonction de repérage* : les titres balisent le texte, lui donnent un cadre stable. La régularité de leur forme syntaxique (souvent un nom ou un groupe nominal, avec un usage restreint de la

¹² Le genre de l'article de journal pourrait faire exception sur ce point : l'usage journalistique veut que le cœur de l'information à transmettre soit délivré par le titre, le chapeau et la première phrase. Le titre joue aussi beaucoup comme facteur d'accroche, pour capter l'attention du lecteur.

ponctuation), qui tranche sur l'enchaînement des phrases, les met en évidence et leur permet de remplir leur rôle de repères. Bien sûr, la typographie amplifie encore notablement cet effet. De plus, le fait que les titres soient généralement des formes nominales ou substantivées (et non des propositions, l'exception notable étant encore une fois le titre d'article de presse), pourrait concourir à la réalisation de cet effet de stabilité.

Ces fonctions modèlent la forme et le comportement des titres, surtout leur caractère métatextuel (ils sont du « texte à propos du texte »), et leur brièveté. Mais pour poursuivre notre analyse, il faut distinguer deux natures de titres.

b) Titre plein et titre creux

On appelle *titre plein* un titre qui indique le contenu de la partie, son thème, par opposition au *titre creux*, titre qui désigne le type, la fonction de la partie, son insertion dans la structure conventionnelle du document. Par exemple, dans les textes d'Action, les titres des parties sont généralement des titres creux (comme ceux prévus par le plan-type), ils pourraient être invariants d'une Action à l'autre quelque soit le domaine considéré, tandis que le titre général de l'Action est un titre plein.

Le titre plein constitue comme un résumé minimal du texte, sous forme d'un syntagme nominal. De façon synthétique, il donne le pôle central. Ouvrant la lecture, il influence le rapport du lecteur au texte. Il oriente l'attention et les attentes du lecteur sur un thème particulier. Il peut aussi motiver le lecteur à entrer dans les développements qui suivent en rejoignant ses centres d'intérêt et en suscitant sa curiosité. Point d'accès au texte, il en devient un identifiant privilégié : il doit donc être à la fois assez spécifique et singulier pour distinguer le texte introduit, mais assez général pour refléter le contenu du texte et aider à le situer par rapport aux connaissances du lecteur et à d'autres références connues.

Le titre creux souligne les articulations du texte, sa construction selon un plan (ce pourrait être : *résumé, introduction, ... , résultats, ...*). Son efficacité repose sur le fait qu'il adopte une forme conventionnelle qui permet immédiatement de le repérer et de l'interpréter. Il guide aussi la lecture dans la mesure où il annonce le statut et les fonctions des paragraphes qu'il introduit. Par exemple : « ceci a le statut de résultat » (décret qui est lui-même un acte scientifique), « ceci est un résumé » et donc reprend les idées essentielles du document. Le titre creux est en fait reconnu comme un titre standard d'un certain type de partie, ce qui lui permet d'activer un ensemble d'attentes quant aux fonctions que remplit le type de partie en question (en tant que type de partie, il a des spécificités prévues).

c) La forme des intitulés

Quels sont les éléments dans la forme même du titre, qui permettent de le reconnaître ?

A partir de notre observation des intertitres d'Actions, nous en recensons quatre classes :

- *éléments lexicaux*, dans le cas de titres creux : on reconnaît par exemple les formules des rubriques du plan-type des Actions. De même, *Introduction, Résumé, etc.*, employés isolément, sont conventionnellement des titres de parties.
- *éléments syntaxiques* : il s'agit généralement d'un groupe nominal, fréquemment sans déterminant. La ponctuation est rare, et limitée : virgules, deux-points, parenthèses, tirets, peuvent le rythmer, mais pas les principales ponctuations fortes (point, point-virgule, points de suspension). En revanche, certains signes de ponctuation servent couramment à articuler le système de numérotation (point, tiret, parenthèse fermante). Cette construction nominale contraste avec un paragraphe rédigé, dans lequel s'enchaînent des propositions. On trouve cependant aussi des constructions nominales indépendantes dans les énumérations, et sous forme de listes. La structure nominale et la ponctuation parcimonieuse vont dans le sens de la brièveté du titre.
- *éléments typographiques* : les titres sont mis en valeur de façon uniforme selon leur niveau dans la structure dans le texte. Pour les Actions 1991-1996, un seul niveau est à peu près toujours présent : celui des parties que nous étudions. Il y a quelquefois un niveau au-dessus (le découpage des

différents buts de la variante à buts multiples), et des niveaux en dessous (l'exposé par objectifs et sous-objectifs dans la partie *Programme*).

- *éléments de numérotation* : le système de numérotation traduit à la fois le niveau d'emboîtement (partie, sous-partie, etc.), la succession (ordre alphanumérique), et le caractère initial ou non de la partie (le point de départ de la numérotation est I., 1., a, A, α, etc.). La numérotation guide le repérage des titres dans la mesure où elle permet de s'apercevoir d'oublis (si on trouve un § 2 et un § 4, alors il doit y avoir un § 3) de préciser la zone dans laquelle se trouve un titre que l'on cherche (on veut trouver le § 3 : il doit être entre les §§ 2 et 4).

L'algorithme de repérage des titres conçu pour la mise en page du *Livre Electronique* utilise uniquement la forme des titres, et s'appuie seulement sur deux classes d'éléments :

- les éléments lexicaux : la présence de certains mots sert d'indicateur
- les éléments syntaxiques (indirectement), dans la mesure où la limitation de la longueur est corrélée aux constructions nominales avec pas ou peu de ponctuation, et où le premier mot est ici souvent le nom central (tête grammaticale) de l'intitulé (absence de déterminant).

d) Le jeu des variations dans les titres creux : de la variante à la rectification

Sauf dans le cas où les titres de partie seraient générés automatiquement, ou strictement imposés —ce qui n'est pas le cas dans notre corpus 1991-1996—, un degré de liberté apparaît, et les familles de variantes se déploient pour chaque intitulé de partie.

L'observation montre que ce phénomène n'est pas du tout marginal. De fait, non seulement la forme mais aussi l'existence de certains titres ne sont pas stabilisées. D'autre part, les titres sont saisis pour chaque texte d'Action, leur insertion n'est pas mécanique : il peut y avoir des erreurs, des confusions. Enfin, le rédacteur introduit un intitulé de la même façon qu'il compose son texte, il l'intègre dans un ensemble qu'il assume en tant qu'auteur : nous verrons que cela peut avoir quelque incidence au niveau des intitulés.

Nous pouvons ainsi dresser une typologie des variantes, en les classant des écarts involontaires ou inconscients aux écarts sciemment introduits.

- *fautes de frappe* : elles se traduisent couramment par un caractère manquant, un caractère en trop¹³, l'intervention de deux caractères successifs¹⁴, ou le remplacement d'un caractère par un autre¹⁵. Une seule fois nous avons relevé un cas plus complexe, de confusion de syllabes. Quant aux fautes d'orthographe, les erreurs d'accord ne sont pas rares, sans doute par distraction, et modification d'un terme en oubliant d'ajuster l'autre ; les erreurs d'usage sont à peu près exclues, sur le vocabulaire restreint, connu, et sans difficulté particulière des titres.

partenariats extérieurs envisgés
objectifs ultérieurs
 obje**c**ifs et étapes principales de l'année 1995
bjectif de l'action
 documents de réfère**ncee**
 document**d** de référence
 objectifs et principales ét**ap**tes de l'années 1995
 objet**t**ifs ultérieurs
 partenariats extr**é**rieurs prévus
 état d'avanc**me**ent à la fin de l'année 1994
 période **preu**ve pour le développement

¹³ Notamment la répétition due à une mauvaise saisie au clavier.

¹⁴ Certaines erreurs s'analysent à la fois comme l'ajout d'un caractère et comme l'intervention de deux caractères successifs : le caractère ajouté est en fait une occurrence anticipée du caractère deux positions plus loin (voir exemples).

¹⁵ L'utilisation d'un caractère pour un autre n'est quasiment jamais la superposition des opérations d'omission et de d'ajout. Le cas typique est l'échange entre *a* et *q*, qui correspond au passage d'un clavier azerty à un clavier querty ; la proximité des touches peut également jouer, le doigt frappant sur une touche immédiatement voisine de celle qui convient.

partarianats extérieurs envisagés
partenariats extérieurs envisagés

- *variantes mineures* : les formes polaires que sont les noms ne bougent pas, il y a seulement des variantes grammaticales modestes au niveau des prépositions et des articles, sans réorientation du sens général du titre (on obtient plutôt des formes plus ou moins abrégées ou développées). De la même façon, on peut enregistrer des variantes de flexion. Par exemple, le rédacteur emploie indifféremment le singulier et le pluriel pour *objectif* / *objectifs*, sans rapport avec le contenu de la partie ou sa définition (le *But*, décrivant la visée générale de l'Action, demanderait plutôt le singulier, alors que le *Programme*, énumérant les étapes intermédiaires de l'année, appellerait le pluriel) : il n'est pas exceptionnel de relever un intitulé de *But* « Objectifs de l'Action », puis un intitulé de *Programme* « Objectif pour 199... », quand bien même plusieurs étapes sont explicitées.
- *échanges lexicaux et glissements* : un mot est employé pour un autre. Plusieurs facteurs favorisent le glissement : en général une proximité entre les deux mots (proximité sémantique ou plus souvent appartenance commune au vocabulaire des intitulés d'Action), et une analogie phonétique, ou le parallélisme entre les plans des Fiches d'Action et celui des Comptes-Rendus d'Action.
Exemples : Etapes d'avancement (par analogie avec Etat d'avancement)
- *accord avec le contenu de la partie* : dans un souci de cohérence de son texte, le rédacteur harmonise parfois son titre avec le contenu de la partie. Ainsi, il met un pluriel à l'intitulé du *But* si dans sa présentation il place l'Action à la convergence de plusieurs visées à long terme. Ou encore, le très courant « Objectifs et principales étapes » est corrigés, certains termes passés au singulier, s'il s'avère que l'Action n'annonce qu'un objectif ou ne déclare qu'une étape.
- *qualification et statut (modalisation)* : le titre ne correspond pas à ce que le rédacteur assume dans la partie. Les modifications vont dans le sens de réserves (« sous réserve »), ou au contraire précisent ce qui est déclaré.

La frontière entre écarts involontaires et écarts contrôlés sépare les trois premiers points des deux derniers. Appelons les trois premiers points des *variantes*, les deux derniers des *rectifications*.

L'opposition entre variantes et rectifications joue évidemment pour une reconnaissance et un classement fin des intertitres.

Ainsi, pour la reconnaissance automatique des parties, les différences entre variantes devraient être neutralisées (lemmatisation, comparaison floue entre chaînes de caractères, classes sémantiques de mots associés à une partie,...), alors que les rectifications pourraient être enregistrées (choix de qualificatifs d'une certaine valeur modale, annotations).

De plus, variantes et rectifications vont modifier la nature du titre dans des directions opposées.

La variante renforce de caractère creux du titre, qui ne fait que renvoyer à un titre-type auquel sont associées les propriétés de la partie qu'il annonce. En effet, les mots perdent de leur motivation sémantique, leur spécificité s'efface, puisqu'il y a confusion. A la limite, la déformation, même légère, est telle, que le libellé considéré isolément n'aurait plus vraiment de sens.

A l'inverse, la rectification remotive le choix des termes, le détail de leur arrangement. Elle fonctionne en prêtant attention à l'interprétation de la formulation du titre lui-même, si bien que c'est se rapprocher du mode de lecture des titres pleins.

e) *Éléments pour la description de la forme des titres*

La description de la forme des titres passe par les quatre points recensés plus haut : éléments lexicaux, syntaxiques, typographiques, et éléments de numérotation.

La ligne du titre de partie commence souvent par une numérotation, de rang 1 ou 2, pas plus. La syntaxe de ces numéros se décrit comme une suite de paires formées d'un indicateur d'ordre (un chiffre, une lettre¹⁶) et d'un ou plusieurs séparateurs de type blanc, point ou tiret¹⁷.

¹⁶ Les chiffres romains devraient faire l'objet d'une description complémentaire, qui de toutes façons reste limitée car on ne s'attend guère à trouver plus de dix parties.

¹⁷ Expressions régulières : $[0-9] + [\ . -] +$ et $[A-Z] [\ . -]$ pour le rang 1.

La longueur (courte) du titre est un autre indicateur utile, comme cela a été démontré par l'algorithme mis en place pour le *Livre Electronique*. Cette longueur pourrait être détaillée en différenciant :

- la *longueur totale*, en nombre de mots (plutôt qu'en nombre de caractères), jusqu'au retour à la ligne suivant (c'est à dire la prochaine balise *paragraphe* pour le codage SGML) ;
- la *longueur utile* (et son complémentaire, le *reste*), en arrêtant le décompte des mots à la première ponctuation si celle-ci est deux-points, des guillemets ou une parenthèse ouvrante (sinon, la longueur utile est prise égale à la longueur totale). Ceci doit permettre de considérer le cas où l'intitulé du plan-type est suivi d'un titre plein ou d'un commentaire.
- la *longueur des inconnus*, c'est-à-dire le décompte des mots non prévus dans le lexique des intitulés, pour la zone correspondant à la longueur utile (première partie de la ligne). Un seuil pourrait être placé à trois : un mot non reconnu ; deux mots non reconnus ; trois mots non reconnus ; plus de trois mots non reconnus.

Le lexique utilisé dans l'intitulé est le premier indicateur de la nature de la partie. L'intitulé se présente comme une combinaison de termes, pris dans certains paradigmes. Par exemple, « But de l'action » recourt à un terme caractéristique de la partie 2 indiquant la visée (« but », mais on aurait pu trouver aussi *buts*, *objectif*, *objectifs*), à une *articulation* (« de l' »), et au vocabulaire général de désignation de l'Action (« action », mais cela aurait pu être *activité*, *étude*, *travail*). Un inventaire de ces indicateurs est donné en annexe.

Une règle quasiment jamais enfreinte est le commencement du titre par un nom (ou un qualificatif), et non par un déterminant. Le paradigme de l'élément en tête de ligne, après la numérotation éventuelle, n'est pas une *articulation*, une *parenthèse*, une *ponctuation*, un *temps*¹⁸. Plus finement, on peut indiquer pour chaque indicateur s'il peut éventuellement se trouver en tête d'un intitulé.

Les indicateurs sont à chercher dans la *signature* de la partie, c'est-à-dire le début de l'intitulé en s'arrêtant le cas échéant à une ponctuation marquant la fin de la longueur utile, ou au troisième terme non reconnu.

Le calendrier est un constituant fréquent des intitulés : la structure interne des date, période, etc. pourrait être reconnue afin d'identifier ces sous-parties des intitulés comme une unité. En revanche, les années ne sont pas un indicateur complètement fiable de la nature de la partie, il y a des erreurs de date : le *Programme* peut être donné pour l'année N-1 (il faut comprendre N), ou inversement l'*Etat d'avancement* peut être situé à l'année N (au lieu de l'année N-1, comme c'est la pratique normale). Il faut donc utiliser les dates avec prudence, en s'appuyant aussi sur d'autres indicateurs.

f) Règles sur le vocabulaire des titres

En utilisant les paradigmes fournis en annexe à ce document, des règles peuvent être construites pour identifier les formulations s'apparentant à un intertitre de partie. Par exemple, les règles associées à la partie 2 se lisent comme suit :

« une formulation qui se compose essentiellement
d'un élément de *But_Focalisation* (à savoir : *caractéristique*, *caractéristiques*, ou *résumé*), ou
d'un élément de *But_Visée* (*but(s)*, *objectif(s)*, *objet*), ou encore
du mot *But* suivi d'une forme de numérotation (un chiffre, éventuellement précédé de *n°*, par exemple),
et qui comporte éventuellement aussi
un élément de *Action* (défini ci-dessous), ou/et
un élément de type *Parenthèse_Plein* (une infinitive, un groupe nominal), ou/et
un élément de *Portée* (un adjectif éventuellement fléchi parmi *général*, *global*, *particulier*, *permanent*, *spécifique*)
s'apparente à un intitulé ouvrant une partie *But*. »

Voici donc l'ensemble des règles proposées pour décrire les intitulés des parties des Actions :

¹⁸ Sont ici reprises les étiquettes données aux paradigmes détaillés en annexe.

- *Partie 1 : Contexte*
Contexte_Argument, Contexte_Introduction, Contexte_Situation
(+ Action) (+ Portée)
 - *Partie 2 : But*
But_Focalisation, But_Visée
« But » + Numérotation
(+ Action) (+ Parenthèse_Plein) (+ Portée)
 - *Partie 3 : Etat d'avancement*
Etat_Bilan, Etat_Résultats, Etat_Rétrospectif
Etat_Moment_Qualification + Temps
Contexte_Situation + Temps
(+ Action + Etat_Action_Qualification)
(+ Action) (+ Portée) (+ Temps)
 - *Partie 4 : Programme*
Programme_Etape, Programme_Pôles, Programme_Polysémie
But_Visée + Temps
(+ Programme_Action) (+ Programme_Chronologie) (+ Programme_Etape_Qualification)
(+ Action) (+ Parenthèse_Plein) (+ Portée) (+ Temps)
- Section 4 :
« Objectif » + Numérotation
- *Partie 5 : Objectifs ultérieurs*
Ultérieur_Evolution, Ultérieur_Futur
Action_Mention + Ultérieur_Futur
Programme + Ultérieur_Futur
(+ Temps)
 - *Partie 6 : Documents de référence*
Référence_Document
(+ Référence_Qualification)
(+ Rapport)
 - *Partie 7 : Partenariats*
Partenariat_Externe, Partenariat_Lien
 - *Partie n : Divers*
Divers_Appendice, Divers_Brevet, Divers_Financement, Divers_Rubrique, Divers_Terminologie
Action_Mention + Rapport
 - *Elément Temps*
Temps_Date
(+ Temps_Articulation) (+ Temps_Qualification)
[date] = (((j ou jj) (s) +) (m ou mm) (s) +) (aa ou aaaa)
avec s = point (.) ou barre oblique (/) ou tiret (-)
 - *Elément Action*
Action_Mention
Action_Sigle
(+ Action_Renvoi) (+ Action_Qualification)
 - Pour toutes les parties :
(+ Articulation_Coordination) (+ Articulation_Liant) (+ Parenthèse_Contour)

Les contrôles sur le mot en première place (éventuellement après une numérotation) sont limités : l'algorithme peut vérifier que le premier mot ne fait pas partie d'*Articulation_Liant*. En effet, les déterminants et prépositions sont les plus susceptibles de commencer un paragraphe qui n'est pas un intertitre. Pour les autres paradigmes, leurs éléments sont soit de bons candidats à débiter un

intitulé, soit n'ont guère de raisons (grammaticales ou autres) de débiter quelque paragraphe que ce soit. Pour donner un exemple, « antérieur » a peu de chance de commencer un titre ou une phrase. Ils ne sont donc pas source de confusions.

2. Les rapports entre parties

La sélection des intertitres uniquement sur leur forme réserve des surprises : ainsi, on sélectionne un « Contexte charbonnier » au cœur du *Programme*, ou le « But » d'un partenariat. Or, ne serait-ce que l'ordre conventionnel des parties aurait pointé l'invraisemblance de ces propositions.

Nous avons trouvé cinq dimensions selon lesquelles les parties sont en rapport les unes avec les autres. Une sixième pourrait s'ajouter, dans un contexte de reconnaissance automatique des parties : il s'agirait de leur *repérabilité* respective (telle partie étant associée à des indices plus efficaces que telle autre).

a) *Succession*

Les parties respectent les règles d'enchaînement suivantes :

- 1 et 2 sont avant 4, 4b et 5.
1 et 2 ont une tonalité introductive ; 4, 4b et 5 entrent dans le détail de l'Action et s'appuient sur le contexte dressé par 1 et 2.
2 peut précéder 1, en jouant le rôle de chapeau (1995E4104R, 1995E4150B, 1995E4207R, 1995E5411S, 1993T5350R et suivantes).
- 3 est après 1 et avant 5.
3 est plus concret et spécifique que 1, puisque c'est le contexte de l'année passée et propre à l'Action. Quant à 5, l'organisation chronologique du descriptif des activités, et le caractère prospectif des *Objectifs ultérieurs*, fait que 5 se place très naturellement après 3.
3 peut précéder 2 (1995E2623O). 3 peut aussi passer après 4 (1996N3301R)
- 4b est après 4.
4b fait figure de récapitulatif.
4b peut se trouver en toute fin de l'Action (1995I2303R et suivantes).

Les parties 6 et 7 seraient conventionnellement placées à la fin. Dans les faits, elles se comportent indépendamment l'une de l'autre, et peuvent s'insérer n'importe où : il arrive de trouver la 6 juste après la 1, ou de trouver la 7 entre la 4 et la 5. La 7 peut très bien figurer avant la 6 (1995E5206R), mais l'habitude (instituée par les consignes) devrait rendre ce cas de plus en plus rare.

Les parties *n*, à savoir les parties diverses non reprises dans le plan type (la plus fréquente étant la *Période de déroulement*), apparaissent avant 2 ou 3, ou bien après 4b ou 5. Autrement dit, si toutes les parties du plan type sont présentes (cf. la dimension *présence* ci-après), les parties *n* évitent le voisinage immédiat avec la partie 4, positionnée au cœur de l'Action. Elles fonctionnent comme des pièces liminaires ou des annexes et rejoignent la périphérie du texte.

b) *Unité (fusion, inclusion)*

Certaines parties ont des affinités, si bien qu'elles peuvent fusionner. Si l'une domine l'autre, cela se traduit par une inclusion de la seconde dans la première, la seconde disparaissant donc au niveau du plan général de l'Action.

Concrètement, on a les affinités suivantes : 1 et 2¹⁹ ; 2 et 4²⁰ ; 3 et 4²¹ ; 4 et 5²². Quant aux parties 6 et 7, elles sont en affinité avec les parties 1, 3 et 4, ce qui se traduit par leur positionnement

¹⁹ Exemples de fusion des parties 1 et 2 : 1995E4305B, 1995E4505B, 1995T2647B, 1995TA466R.

²⁰ Les tâches permanentes se situent à la frontière entre 2 et 4, et figurent indifféremment dans l'une ou l'autre partie.

²¹ L'*Etat d'avancement* peut même être distribué au fil des objectifs, cf. 1995E5103R.

²² En général 4 domine 5, mais 4b peut être inclus dans 5 comme dans 1995P1401A.

variable et leurs inclusions faciles. A titre d'illustration, le tableau ci-après donne des exemples des associations variées attestées dans notre corpus (enchaînements ou inclusions) :

	1	3	4
6	1993I7418D, 1994I7411D	1995I7401R, 1992I7101R, 1993I7401R	1995E4202R
7	1993TB113R	1995N4604R	1996E1110R

Cette même dimension pourrait connaître une variante, qui serait au contraire l'*incompatibilité* entre certaines parties, l'une excluant l'autre (ou les autres, formant un ensemble d'alternatives). Nous ne faisons que mentionner ce cas, qui n'est pas utile ici pour la description des Actions.

c) Répétition : factorisation

Chaque partie apparaît une fois au plus. Il n'y a qu'un cas dans lequel une partie peut avoir plusieurs occurrences, c'est celui d'une Action à buts multiples : tout se passe alors comme si plusieurs Actions avaient été mises bout à bout, mais que l'ensemble des premières parties avaient été fusionnées et placées en tête de l'ensemble.

Le plan de l'Action à buts multiples prévoit, pour un contexte commun, plusieurs buts, chacun suivi de leur développement. Soit donc l'enchaînement : 1, (2, 3, 4, 5, 6, 7), (2, 3, 4, 5, 6, 7), (2, 3, etc.

La coupure entre ce qui est fusionné et ce qui est ensuite répété n'est pas uniforme dans les faits. Cependant, ce sont les différents sous buts qui découpent l'Action ; un intertitre les rappelle, qui peut être suivi d'un paragraphe *But*, complément éventuel d'un *But* général. Soit donc les variantes suivantes :

- pas de factorisation : (1, 2, 3, 4, 5, 6, 7), (1, 2, 3, 4, 5, 6, 7), (1, 2, ...)
- factorisation de 1 : 1, (2, 3, 4, 5, 6, 7), (2, 3, 4, 5, 6, 7), (2, 3, ...)
- factorisation de 1 et 2 : 1, 2, ((2,) 3, 4, 5, 6, 7), ((2,) 3, 4, 5, 6, 7), ((2,) 3, 4,
- factorisation de 1, 2 et 3 : 1, 2, 3, ((2,) 4, 5, 6, 7), ((2,) 4, 5, 6, 7), ((2,) 4, ...)

d) Présence

Examinons les parties, des plus essentielles aux moins régulièrement présentes.

Les parties 2 et 4 sont le noyau de l'Action ; comme elles ont des affinités, elles peuvent se fusionner en une. Mais il serait anormal de trouver une Action ne comportant ni une partie explicitant son but, ni une partie présentant son programme pour l'année²³.

La partie 1 est régulièrement présente ; mais si l'Action est courte et peu structurée, elle s'efface devant 2 ou/et 4.

Les parties 3 et 5 sont aussi bien connues et employées, seulement par la nature même de leur contenu elles peuvent n'avoir pas lieu de figurer : si l'Action est nouvelle, son *Etat d'avancement* peut au mieux mentionner d'autres Actions dont elle est le prolongement, ou quelques études exploratoires qui ont motivé le lancement de l'Action. Et si c'est la dernière année prévue pour l'Action, sa conclusion est l'achèvement de son programme et non la prévision d'autres objectifs à atteindre pour satisfaire le *But*.

Les parties 6 et 7 ne sont jamais imposées, elles ne sont même pas toujours souhaitables. La partie 6 recense des documents, mais le document de doctrine précise qu'elle ne devrait mentionner que les documents donnant une vue plus développée du programme de l'Action, quand ils existent. Cela ne devrait donc concerner que les Actions qui s'inscrivent dans un projet plus vaste (ayant lui-même son programme). Quant aux partenariats, le document de doctrine dit de s'en tenir à ne mentionner que ceux qu'il est « intéressant de repérer » au titre des collaborations.

²³ Exemples d'Actions n'ayant pas de partie 2 : 1996P2809R, 1997P1264F.

e) Longueur

Cette dimension est à la limite des dimensions de cohérence globale des parties : car si telle partie est habituellement plus longue que telle autre, la longueur de chaque partie est directement liée à la nature de son contenu propre. On peut cependant faire ressortir des rapports généraux :

- les parties 6 et 7, en fin d'Action ou au détour d'une partie, énumèrent quelques documents ou partenaires sans guère de commentaires : elles sont donc souvent brèves.
- la partie 2, par sa fonction de synthèse (elle peut être comprise comme un « résumé » général de l'Action), est plutôt courte, mais les rédacteurs prennent bien la longueur qu'ils jugent utile (il n'y a pas de limite fixée). La partie 4b, souvent sous forme de tableau récapitulatif, ne garde qu'une ligne par étape : c'est donc encore une partie courte ou moyenne. La partie 5 ébauche quelques pistes : elle est encore du même ordre de taille.
- la partie 1, présentant le contexte, est généralement de longueur moyenne (quelques paragraphes pas très longs). A la fois il convient de soigneusement situer l'Action (il en va de la viabilité de l'Action, cf. la *fonction argumentative* associée à cette partie), mais aussi l'équilibre d'ensemble veut que le contenu de l'Action soit premier, le Contexte n'étant qu'introductif.
- la partie 3 précise les réalisations précédentes : elle peut être du même ordre de grandeur que la longueur moyenne de la partie 1, voire un peu plus longue. Elle est importante car il s'agit du contenu concret de l'Action ; pour autant, ce n'est qu'un rappel (des objectifs des années précédentes et du compte-rendu).
- la partie 4 est *a priori* la plus longue : elle doit détailler les objectifs, et on attend qu'elle soit précise, de part sa fonction d'engagement. D'autre part, le rédacteur essaye de rendre compte du volume d'activité engendré par l'Action, et un Programme largement déployé permet de mieux percevoir la somme des travaux à réaliser.

3. Les indices linguistiques dans le contenu, au fil du texte

Chaque partie est un micro-genre ; ses fonctions (vues précédemment) se traduisent par certains usages linguistiques appropriés. Mentionnons, en récapitulant un certain nombre d'observations faites dans l'analyse des parties :

- le caractère plus ou moins rédigé : phrases construites, avec un verbe, des adverbes, des conjonctions de subordination et des connecteurs contribuant à la bonne liaison des phrases, par opposition à une juxtaposition, hâchée et morcelée, et avec peu de marques de temps, de groupes nominaux, de sigles, sous forme de listes ou d'énumérations.
Dans les parties rédigées, on observe, dans les flexions des verbes, les personnes d'une part, les modes et temps d'autre part :
- l'utilisation des personnes : le rédacteur d'une Action s'efface, les premières personnes sont rares (ce sont alors des premières personnes du pluriel : nous). La troisième personne est employée de différentes manières : désignation d'un acteur structurel (ex. : *le Département*), pronoms indéfinis (*on, c'(est)*), et constructions impersonnelles (*il faut, il s'agit*), et surtout abondance de constructions passives.
- un usage contrasté des temps : le *présent* de l'indicatif est préférentiellement employé dans le *Contexte* (davantage il semble que le passé composé), le *But* marque une certaine atemporalité (présent de l'indicatif et mais surtout infinitifs), l'*Etat d'avancement* connaît un usage privilégié des temps composés (*passé composé*, auquel se substitue volontiers le *futur antérieur* dans les versions provisoires des textes d'Action) ; le *Programme* et les *Objectifs ultérieurs* sont difficiles à distinguer entre eux, hormis les infinitifs il s'y trouve régulièrement quelques *futurs*.
- la distribution des modaux : elle reste à préciser, mais des formes comme *doit, devrait, peut* seraient à attendre dans les parties 1, 3, 4, 5, dans une moindre mesure 2 ou 7, exceptionnellement dans 6.
- l'utilisation abondante des tirets et numérotations de toutes sortes, organisant les structures de listes.

- la possibilité, voire la forte probabilité, de trouver des tableaux (typiquement ici, dans la partie 4b, établissant formellement des *étapes*).
- des formules conventionnelles ; un calcul de segments répétés identifierait des formulations qui appartiennent au style des Actions (*Cette action vise à* pour 2, etc.) et servent à introduire et à lier les éléments liés à la thématique propre de chaque texte.
- cas particulier du point précédent, des expressions « toutes faites » qui peuvent constituer à elles seules le contenu d'une partie « facultative » (ex. : *action nouvelle, nouvelle action, activité nouvelle* pour 3 ; à *définir* pour 4 ; *sans objet, néant* pour 6 ou 7).
- les désignations d'actions, qui sont pour une grande part dans l'emploi massif des phrases nominales ou infinitives.
- l'usage privilégié de certaines ponctuations : par exemple, la partie 6 *Documents de référence* se prête à une utilisation réglée des guillemets.
- l'apparition de certaines familles de notations²⁴ : dates, référence de note interne.
- les sigles et acronymes (se détachant par leur écriture en lettres capitales, et incluant parfois des chiffres) : la partie 7 (*Partenariats*) concentre les désignations d'entreprises et de laboratoires ; d'autres parties, voire l'Action tout entière, peuvent utiliser abondamment le nom d'un logiciel, d'un procédé, ou le code conventionnel d'un objet d'étude (un composant d'une centrale par exemple).

Toutes ces remarques valent comme éléments pour la construction d'indicateurs discriminants (pour le repérage et l'identification des parties), mais préalablement comme spécifications pour les outils de Traitement Automatique des Langues que l'on voudrait utiliser pour l'analyse.

4. Intégrer les indicateurs locaux et globaux

Les propositions précédentes doivent se combiner pour être efficaces. En effet, par exemple, la liste du vocabulaire des intertitres a été notablement enrichie par rapport à celle utilisée dans l'algorithme conçu pour le *Livre Electronique* : les effets de cet élargissement doivent être contrôlés. Quelques termes pourraient être simplement ajoutés aux listes utilisées par l'algorithme (voir nos propositions en Annexe), mais pour la plupart, les ajouter sans précautions dégraderait les performances de l'algorithme en lui faisant repérer des intertitres là où il n'y a pas lieu de le faire.

Un traitement approfondi, exploitant l'ensemble des caractéristiques que nous avons mises au jour, pourrait se dérouler de la façon suivante :

1. Recherche sur critères locaux (forme des intertitres) : repérage de candidats intertitres grâce à (i) leur longueur totale ou leur longueur utile (inférieure à 80 caractères par exemple), (ii) leur numérotation de début (facultatif), (iii) l'absence de déterminant au début, (iv) la présence d'au moins un terme reconnu dans leur signature. Le ou les terme(s) reconnus permettent dans la plupart des cas de tirer des propositions quant à la nature de la partie introduite par l'intertitre.
2. Examen contextuel : recherche des caractéristiques linguistiques et des formulations figées caractéristiques des parties. Elimination en conséquence de certaines interprétations des candidats titres, voire de certains candidats titres.
3. Ajustement global : l'enchaînement des parties est confronté aux contraintes générales de présence et d'ordre des parties, pour valider la cohérence globale des candidats extraits. Elimination en conséquence de certaines interprétations des candidats titres, et repérage de trous : intertitres manquants car vraisemblablement non identifiés.
4. En cas de trous, recherche sur critères locaux, assouplie, mais avec des attentes précises : telle nature de partie, dans telle zone du texte. Puis retour en 2. puis 3. pour les nouveaux candidats sélectionnés. En cas de recherche infructueuse, avertissement à l'utilisateur (l'examen des textes difficiles à traiter automatiquement soit peut être fait à la main, pour autant que l'algorithme ait été efficace et ne soit pas indécis sur un grand nombre de cas, soit conduit à enrichir et ajuster les critères utilisés et à relancer le traitement).
5. Quand la cohérence d'ensemble est satisfaisante, la structure de parties calculée est adoptée.

²⁴ Elles pourraient, du fait de leur manifestation caractéristique dans les textes d'Action, mériter la mise au point de « grammaires locales » permettant de les identifier, voir par exemple (Maurel 1996).

Si la position des intertitres est déjà repérée (par exemple par un style « titre de section » dans le traitement de texte utilisé pour la saisie), il reste à déterminer la nature des parties associées. Le traitement est beaucoup plus simple :

1. Emission d'hypothèses en fonction des termes reconnus dans les intertitres ;
2. Ajout et élimination d'hypothèses en fonction du contenu des parties (caractéristiques discriminantes), et de la cohérence globale. Les cas litigieux sont reportés à l'utilisateur.

F. PERSPECTIVES

1. Utilisation de la structure des Actions

a) *Application à la consultation (Livre Electronique)*

La reconnaissance des intertitres sert actuellement à la mise en page des textes d'Action, pour leur affichage. Cette identification est absolument nécessaire pour une lisibilité normale (et non dégradée) de ces textes, elle fournit les repères pour construire la lecture. Si l'algorithme de repérage des intertitres s'appuyait sur des informations sur les parties, il pourrait sans doute encore gagner en qualité. Le respect par l'affichage du découpage en parties est un point d'appui essentiel pour le lecteur, pour parcourir le document (Dillon 1991).

A partir du moment où l'on reconnaît non plus seulement les intertitres, mais que l'on identifie les parties, sont ouvertes des possibilités démultipliées de vues et d'extractions. Le Livre Electronique pourrait, au gré du lecteur, présenter des pages allégées (par exemple uniquement les paragraphes *But*), ou agencées différemment, ou mettant en valeur certaines parties et réduisant typographiquement les autres, etc.

L'information apportée par les parties peut enfin être exploitée pour qualifier les recherches sur le texte intégral. Les parties deviennent autant de champs, au sens des fiches bibliographiques. Il serait par exemple possible de spécifier que l'on souhaite les Actions dans lesquelles tel mot figure dans telle partie (par exemple rechercher un nom de personne dans les *Documents de référence*, dans l'idée de le trouver en tant que rédacteur de note).

D'une manière générale, l'identification des parties est un gain pour la structuration du document, et en présente les nombreux intérêts (Futtersack 1995, §3.4).

b) *Application à la construction de corpus*

En fonction de l'application visée ou de la nature des résultats recherchés, les données textuelles exploitables doivent être étudiées pour construire un corpus adéquat. L'étude de la structure des Actions montre que leurs parties jouent des rôles différents, présentent des caractéristiques différentes ; certaines sont essentielles, d'autres ne figurent pas systématiquement.

Un corpus défini à partir des Actions d'une année peut être focalisé sur une ou quelques parties. C'est d'ailleurs ainsi qu'est défini le corpus sur lequel a travaillé Florence Pugeault (Pugeault 1995) : les textes d'Action qu'elle considère sont en fait les paragraphes *Contexte* (et sans doute *But* en cas d'absence de *Contexte*). Le choix de ces parties a semblé pertinent pour une application d'indexation automatique, sans doute parce qu'il présente le sujet de façon élargie (mentions d'autres éléments en relation), dans un vocabulaire pas trop spécialisé ni lié à des entités précises (logiciel, société) ignorées du Thesaurus. *Contexte* (ou *But*) sont présents dans toutes les Actions : toutes sont ainsi décrites et caractérisées. Le fait aussi que *Contexte* soit de l'ordre du paragraphe, au lieu d'une longueur de quelques pages, a pu être considéré avantageux, sachant que l'approche d'indexation adoptée est plus basée sur un principe d'extraction (extraction de phrases, puis de groupes nominaux) que sur un principe de synthèse (mise en relation et regroupement de termes, élection d'unités représentatives après une analyse globale).

En diachronie, l'examen des années 1991 à 1996 révèle une évolution dans la manière de rédiger les Actions : cela peut avoir une incidence sur l'homogénéité et la représentativité du corpus. Au fil des années, les textes sont plus nombreux, plus longs, et tendent à converger vers une même structure dictée par un plan type.

La structuration en parties est enfin un moyen d'entrer dans les textes, et de les parcourir de manière ciblée. Le corpus n'est plus considéré comme *du* texte (du matériau linguistique attesté), ni même simplement *des documents* (l'unité que forme chaque texte d'Action est reconnue), mais bien comme *des* textes, avec un déploiement et une organisation interne.

c) *Application à un traitement automatique d'analyse des textes*

Cette étude des textes d'Action a commencé à pointer leurs caractéristiques linguistiques, lexicales mais aussi morphosyntaxiques. Il est évident que cela se traduit concrètement dans le choix et le paramétrage d'outils de Traitement Automatique des Langues que l'on voudrait utiliser sur ces textes. A EDF, et plus largement chez tous les partenaires du consortium GRAAL²⁵, la nécessité d'adapter les traitements automatiques aux propriétés des corpus est explicitement reconnue, au point qu'une méthodologie est détaillée (Herviou, Coch, Leblond 1995) (Herviou 1994b) (Coch 1994) pour définir les spécifications techniques d'une application faisant appel au Traitement Automatique des Langues.

Clairement, il est de première importance que ces outils ne soient pas axés sur une analyse des propositions centrées sur un verbe conjugué, mais que l'analyse des phrases nominales et des infinitives, dans des structures d'énumérations ou de listes, fasse pleinement partie de leur domaine de fonctionnement. On s'aperçoit aussi que les indications sur le temps des verbes et la ponctuation sont dignes d'intérêt, et peuvent contribuer fortement à la reconnaissance automatique des parties. Pour ce faire, une heuristique s'appuyant sur les formes des verbes les plus courants, et en particulier de l'auxiliaire *être*, est une première tactique, compte-tenu de l'abondance des formes passives²⁶. L'aptitude à savoir identifier et gérer certaines notations (pour les dates, les références de notes) serait un plus, et peut justifier la construction de (plus ou moins) petits automates à états finis (Maurel 1996). La reconnaissance et l'emploi des sigles serait indispensable pour obtenir des représentations convenables de la partie 7 (*Partenariats extérieurs*), et serait utile pour la partie 4 (*Programme*), dans laquelle le recours à tel outil ou la référence à telle collaboration ou telle action sur site sont des précisions importantes. Enfin, un algorithme de reconnaissance de segments répétés serait certainement en mesure de repérer des « briques » de construction du style des textes d'Actions.

La mention d'actions joue un grand rôle. La majorité des phrases nominales, si fréquentes, sont en fait basées sur des verbes substantivés ; les noms correspondants ont gardé une signification d'acte qui transforme, crée, empêche, etc. Or le travail de thèse de Daniela Garcia a porté sur l'expression des actions et plus spécialement de la causalité (Garcia 1998) (Garcia 1996a), (Garcia 1996b), (Garcia, Aussenac-Gilles, Courcelle 1996), (Garcia, Jackiewicz, Gros 1995). Ce travail a été appliqué à des documents scientifiques et techniques, et a fait l'objet d'une implémentation (COATIS). Il vaudrait tout à fait d'être considéré dans le cadre des textes d'Action.

Pour le système DECID²⁷, d'autres enseignements encore peuvent être tirés de cette étude, en particulier sur les rôles et interactions des parties. De même que le titre générale de l'Action, la partie *But* est dans un rapport « méta » par rapport aux autres parties, elle les résume et sert de référence globale ; elle est en constant arrière-plan de la lecture, dans la pratique d'utilisation de ces textes. Autre remarque exploitable, un effort particulier pourrait être fait sur l'analyse des titres des publications citées en référence (partie 6) ; ils se présentent en effet comme des points d'entrée synthétiques sur une partie significative de l'activité décrite.

2. Portée théorique et caractère généralisable de l'étude

a) *Les genres textuels*

Cette étude sur le corpus des textes d'Actions constitue un cas d'analyse de corpus sensible à la question linguistique des genres. Elle s'appuie sur une connaissance des pratiques de rédaction et

²⁵ Le projet EUREKA GRAAL (*Grammaires Réutilisables pour l'Analyse Automatique des Langues*) est un projet européen (France, Italie, Suisse, Finlande, Portugal, Grèce) récent (1992-1996), qui réunit des laboratoires de recherche, des industriels spécialistes du traitement automatique des langues, et de grandes entreprises utilisatrices d'applications faisant appel à ces technologies.

²⁶ *est, sont* sont des indices d'un passage au présent ; (*a / ont / aura / auront*) *été*, d'un passage au passé composé ou au futur antérieur ; *sera, seront*, d'un passage au futur. Des confusions sur *été*, entre le nom (saison) et le participe passé (verbe ou auxiliaire *être*), sont ici très dommageables.

²⁷ *Diffusion Electronique Ciblée d'Informations et de Documents*, application sur l'Intranet EDF ; les profils des destinataires (sur la DER) sont calculés à partir des textes d'Action.

de lectures liées à ces textes. Ce contexte a une incidence évidente sur l'élaboration d'une organisation standard (matérialisée par un plan-type), et sur les fonctions attribuables à chacune des parties.

L'occasion a été donnée de remarquer des spécificités morphosyntaxiques, soit des textes d'Action en général, soit de telle ou telle de leurs parties. La liste des éléments observés (phrases verbales ou nominales, temps, etc.) donne des pistes pour la description d'autres genres²⁸, rejoignant en cela, mais avec une portée incontestablement plus modeste, les travaux beaucoup plus étendus de (Bronckart & al. 1985) et (Biber 1988).

b) Les corpus de textes avec plan-type

Le rédacteur d'un texte d'Action s'approprie assez librement le plan-type, d'ailleurs plus ou moins explicite et connu. Ce genre de textes à plan-type n'est pas réductible à un formulaire, avec des titres préétablis et invariants, et une organisation stricte et formelle. En cela, la problématique de l'étude de la structuration des textes d'Action s'apparente à un nombre important d'autres genres très représentés dans diverses pratiques professionnelles, par exemple le Curriculum Vitæ, le compte-rendu d'hospitalisation, les articles scientifiques (relevant d'une même spécialité, ou édités dans une même revue). Ces textes échappent à une description trop rigide de leur structuration, alors même que cette structuration conditionne fortement leur analyse linguistique et leur interprétation.

Un outil d'analyse qui serait capable de s'appuyer sur une description des rapports entre les parties (vus au § E.2), ou une tactique de reconnaissance des intitulés par delà leurs variantes (comme en § E.1), seraient réutilisables avec profit dans bon nombre de contextes. Les personnes qui ont à concevoir un traitement automatique sur ce genre de corpus reconnaissent en effet les approximations faites (et les erreurs qui en découlent) avec la mise en place d'heuristiques peu développées destinées à repérer la structure des textes.

Cet élargissement à d'autres corpus serait l'occasion de faire place à d'autres régularités et d'autres variables descriptives. Par exemple, le caractère « ouvert » ou « fermé » d'une partie, selon que son contenu se laisse décrire ou non comme une alternative de valeurs sémantiques prévisibles. Dans les textes d'Actions, les parties ont toutes été considérées comme ouvertes ; les *Partenariats extérieurs* cependant seraient semi fermés, dans la mesure où l'on n'y attend que la mention d'organismes (entreprises, laboratoires) potentiellement énumérables. En revanche, pour des compte-rendus hospitaliers liés à des cas de cancer de la thyroïde, la partie conclusive, prescrivant le traitement, voit son contenu limité aux soins et médicaments connus, disponibles et jugés efficaces.

L'exemple suivant présente la définition et l'utilisation d'une structuration interne, pour des textes de jugement :

[Diane Poirier a étudié] les particularités des jugements pour en dégager une structure d'information à reprendre dans les résumés. Elle a pu reconnaître trois catégories d'information toujours présentes dans un jugement :

- le *litige* qui regroupe « l'énoncé des procédures entreprises et des issues recherchées et obtenues. Pour utiliser le langage du métier, on parlera d'indicateur d'action et de dispositif, c'est-à-dire quelle est la nature de l'action intentée et quelle est la décision rendue sur cette action. C'est le noyau 'performatif' du jugement, ce en quoi il agit, il ordonne. »

- le *contexte* : « C'est le substrat concret sur lequel porte la qualification juridique. Le contexte comprend : la présentation des faits ou événements, des parties en cause et des relations entre celles-ci, des admissions ou documents nécessaires au litige. Il peut comprendre encore le cheminement du dossier jusqu'à l'instance qui est la sienne. »

²⁸ (Amitay 1997) étudie par exemple les caractéristiques des pages personnelles sur le Web. Les choix linguistiques convergent pour créer un ton affirmatif, positif, factuel, et informel : première et deuxième personnes ; déictiques (*click here*) ; formes contractées (négation en *-n't*) ; structure parfois plus de liste que de phrases construites et liées, avec alors un sous emploi des verbes, des déterminants, des conjonctions ; usage de l'impératif, du gérondif, du présent (présent simple, *present perfect*) et déficit de temps narratifs (*preterit*) ; concentration des noms propres ou des expressions avec initiale en majuscule, pour le texte des ancrés.

On le voit : les types d'observations faites par Einat AMITAY rejoignent les nôtres ; une grille de ces variables linguistiques serait une base pour définir une méthodologie de description de genres textuels.

- la *problématique* qui contient : « l'énoncé des prétentions respectives des parties, l'argumentation, la discussion, la recherche d'une solution et la solution elle-même, la règle de droit applicable. [...] La problématique peut être de diverses natures : elle peut porter sur la qualification juridique des faits, l'appréciation de la preuve, la détermination de la procédure appropriée, la détermination d'une responsabilité, l'évaluation d'un dommage, la recherche de la règle de droit applicable et son application, l'interprétation d'une règle de droit, etc. C'est dans la problématique que se trouve la matière juridique d'un jugement, la règle de droit qui pourra s'appliquer à d'autres cas analogues ou similaires. Alors que contexte et litige sont intimement liés à une affaire, c'est la problématique qui peut donner lieu à une généralisation » (Poirier 1985, pp. 120-122)²⁹

[...]

Les stratégies de lecture et de résumé varient [...] selon les individus et selon le domaine, car les jugements sont structurés différemment, contiennent des renseignements différents. On ne lit pas un jugement classé dans le domaine Famille comme on lit un jugement relevant du domaine Procédure civile. Dans certains domaines, ce qui est important ce sont les questions de droit discutées par le juge situées dans la problématique, dans d'autres domaines ce sont les argumentations des avocats localisées dans la partie contexte.

Les valeurs de propriétés surimposées au texte de départ jouent donc le rôle d'un paratexte orientant la lecture humaine et déterminant l'affichage sur écran ou l'impression papier.

[...]

Nous attribuons une valeur de propriété aux mots qui contribuent le plus à la distance du contexte (du litige, de la décision) [selon une mesure entre ensembles de parties] ; la recherche dans les textes intégraux de fortes concentrations de mots ou de propriétés discriminants peut servir à dépister les passages traitant du litige, du contexte ou de la problématique. Par exemple, les verbes qui caractérisent le contexte se rapportent aux parties par *alléguer, prétendre, demander, réclamer, soutenir, invoquer*, etc. et contrastent avec ceux de la problématique qui s'appliquent au juge : *devoir, falloir, démontrer, ordonner, appliquer, agir*, etc. Les lexèmes mes plus discriminants du contexte se rapportent au temps (rappel des événements à l'origine du litige) : *janvier, février, mars, avril*, etc., *mois, an*.

(Bertrand-Gastaldy 1993, pp. 8, 18-19)

c) Et pourquoi pas : l'Action comme un récit

A travers la description des différentes parties structurant un texte d'Action, enrichie par la lecture effective de textes de ce corpus, un connaisseur des analyses structurales de récit, dans la lignée des travaux de Greimas (Greimas 1966), ne manque pas d'y déceler des articulations fondamentales d'un récit. Le parallèle pourrait être, il est vrai, assez frappant, comme l'esquisse le tableau suivant :

²⁹ POIRIER Diane (1985) - *Pour des résumés adéquats de jurisprudence québécoise et canadienne : une étude du document jurisprudentiel, de sa structure, de ses citations, de son rôle et de sa spécificité*, Mémoire de maîtrise, Montréal : Université de Montréal, Ecole de bibliothéconomie et des sciences de l'information.

Partie	Rappel des fonctions trouvées	Modèle du récit
<i>Contexte</i>	fonction argumentative, (fonction introductive, fonction unificatrice)	qualification
<i>But</i>	fonction définitoire, (fonction de synthèse)	mandement, quête
<i>Etat d'avancement</i>		épreuves
<i>Programme</i>	fonction schématique, fonction de concrétisation	épreuves
	fonction d'engagement	contrat
<i>Objectifs ultérieurs</i>	fonction de chute, (fonction diagnostic)	épilogue
<i>Documents de référence</i>	fonction d'officialisation interne, fonction bibliographique	
<i>Partenariats extérieurs</i>	fonction d'appui, (fonction d'index)	(adjuvant)

G. ANNEXES

1. L'algorithme de repérage des intertitres pour le Livre Electronique

a) *L'algorithme existant*

Il a été conçu et mis au point par Anne Le Gall, puis entretenu par Véronique Jolly.

Sont sélectionnées comme intertitres les lignes (i) dont la longueur ne dépasse pas 55 caractères, (ii) et qui, modulo un passage en typographie pauvre (non prise en compte de la casse et des accents), respectent au moins une des contraintes prévues. Ces contraintes portent sur les premiers caractères de la ligne, et soit la présence d'une seconde chaîne de caractères dans la ligne, soit une contrainte plus serrée de longueur. Dans la pratique, cette contrainte de longueur sert à capter les titres brefs avec une indication d'année (ex. : *Etapes 1993*).

En fonction des chaînes de caractères trouvées, l'intertitre est reconnu comme faisant partie d'un groupe de paragraphe donné.

Groupe de paragraphe	Début de ligne	Suite de ligne contient	Longueur de la ligne
CONTEXTE	CONTEXTE		
RAPPEL	RAPPEL		
BUT	BUT		
ETAT	ETAT	AVANCEMENT	
OBJECTIF	OBJECTIF		
OBJECTIFETAP	OBJECTIF	ETAPE	
PROGRAMME	PROGRAMME	DE L'ACTION	
PROGRAMME	PROGRAMME	ETAPE	
PROGRAMME	PROGRAMME	PERSPECTIVE	
PROGRAMME	PROGRAMME	RESTANT	
PROGRAMME ³⁰	PERSPECTIVE		
ETAPE	ETAPE	ORDONNANCEMENT	
ETAPE	ETAPE	PREVU	
ETAPE	ETAPE	PRINCIPAL	
ETAPE	ETAPE	ULTERIEUR	
ETAPE	ETAPE		< 12
OBJECTIFULT	OBJECTIF	ULTERIEUR	
ORIENTATION	ORIENTATION		
DOCUMENT	DOCUMENT	PRODUIT	
DOCUMENT	DOCUMENT	REFERENCE	
REFERENCE	REFERENCE		
PARTENARIAT	PARTENAIRE	EXTERIEUR	
PARTENARIAT	PARTENARIAT		
PERIODE	PERIODE		
RESULTAT ³¹	RESULTAT	OBTENU	

b) *Observations*

Les règles où n'intervient que le début de la ligne CONTEXTE, BUT, OBJECTIF, PARTENARIAT, ou PERIODE, sont très productives et efficaces ; elles ne génèrent pratiquement pas d'erreurs (on peut éventuellement regretter que tous les sous-objectifs numérotés soient sélectionnés au même niveau que le titre du paragraphe les contenant : la présentation n'est alors pas tout à fait conforme à la structure logique sous-jacente, mais la lecture n'est pas gênée).

³⁰ ou plutôt OBJECTIFULT ?

³¹ se trouve dans les compte-rendus plutôt que dans les fiches d'Action.

L'algorithme est mis en difficulté par les majuscules accentuées (« État d'avancement », « Étapes principales »). Ce serait dû à des différences de codage entre les systèmes (Mac vs PC).

c) Première proposition : enrichissement des indices

En gardant le même algorithme, l'étude du corpus permet de compléter les règles existantes.

Groupe de paragraphe	Début de ligne	Suite de ligne contient	Longueur de la ligne
CONTEXTE	CADRE	ACTION	
ETAT	AVANCEMENT	FIN	
ETAT	POINT	FIN	
ETAT	POINT	AVANCEMENT	
PROGRAMME	PROGRAMME		< 15
PROGRAMME	PROGRAMME	ANNEE	
PROGRAMME	PROGRAMME	OBJECTIF	
PROGRAMME	PRINCIPAUX	OBJECTIF	
ETAPE	PRINCIPALE	ETAPE	
OBJECTIFULT	ACTION	ULTERIEUR	
OBJECTIFULT	DEVELOPPEMENT	ULTERIEUR	
OBJECTIFULT	ETAPE	ULTERIEUR	
OBJECTIFULT	PROGRAMME	ULTERIEUR	
PARTENARIAT	PARTENAIRE	ENVISAGE	
PARTENARIAT	PARTENAIRE	EXTERNE	

d) Deuxième proposition : considérer davantage les groupes de paragraphe

Les groupes de paragraphes semblent plus qu'un simple étiquetage de règles : ils s'approchent d'une identification des différentes parties du plan-type.

La partie 5 est peu prise en compte dans les groupes de paragraphe : lui correspondent les groupes OBJECTIFULT, ORIENTATION, et ce que sélectionne la règle ETAPE + ULTERIEUR.

Les règles énoncées au § E.1.f pourraient servir à enrichir systématiquement les indices lexicaux recherchés dans les titres de parties, avant même d'envisager une stratégie plus globale, esquissée par l'ensemble du § E.

2. Le vocabulaire des intitulés des parties : proposition de définition d'unités paradigmatiques

Le tableau ci-après recense l'ensemble complet des indicateurs lexicaux utilisables pour la reconnaissance et l'identification des intertitres des parties. Il a été établi à partir de l'examen minutieux du corpus des intertitres des Actions de 1991 à 1996.

a) Première colonne : mots recherchés

La première colonne donne l'unité lexicale à rechercher. Sans autre indication, il s'agit de la forme littérale à retrouver telle quelle, modulo des variations de casse (majuscules / minuscules). Les unités qui ne sont pas des mots (caractères alphabétiques) mais utilisent d'autres caractères (symboles, ponctuations, notations), et qui doivent aussi être recherchés littéralement, sont mentionnés entre guillemets. L'utilisation des crochets sert à indiquer la nature de l'unité (les cas rencontrés ici sont un syntagme construit autour d'un nom ou d'un infinitif, et une expression qui a la forme d'une date).

Le signe + qui suit une unité signale que l'on sous-entend, de façon équivalente, le paradigme « évident » de l'unité. Il s'agit la plupart du temps des variations flexionnelles (en nombre, et en genre le cas échéant), mais on résume aussi toutes sortes de modes de variation :

sous A.Q.+ se réalise par : *sous A.Q., sous AQ*
à partir de+ se réalise par : *à partir de, à partir d'*
jusqu'à+ se réalise par : *jusqu'à, jusqu'au, jusqu'aux*
janvier+ se réalise par : *janvier, février, mars,..., décembre.*
1er+ semestre se réalise par : *1^{er} semestre, 2^{ème} semestre, 2^{ème} semestre, 2nd semestre*
premier+ trimestre se réalise par : *premier trimestre, deuxième trimestre, second trimestre, troisième trimestre, quatrième trimestre*

C'est un raccourci commode pour abrégé les tables données, mais qu'il conviendrait de remplacer par l'explicitation des formes effectivement concernées pour passage à un traitement automatique. Ces formes sont rapidement énumérables par l'analyste, mais sont difficilement (ou à grands frais) calculables par la machine.

De la même façon, le chiffre *1*, ou la notation *n°1*, qui entre dans le paradigme des numérotations, vaut pour toute la série d'entiers raisonnablement envisageable dans le corpus des Actions (par exemple, de *1* à *20*).

b) Deuxième colonne : nom d'une classe paradigmatique de rattachement

Les noms des classes de rattachement s'analyse en champs, séparés par des barres obliques (caractère « / »).

Le premier champ précise la nature de la classe, en opposant les classes caractérisant les intitulés d'une partie déterminée (*partie, section*), et celles utilisée dans la description des intitulés de plusieurs parties (*voc*).

Le second champ indique la ou les parties concernées, à savoir pour lesquelles les indicateurs fournis par la classe sont utilisables. La notation des classes est celle donnée au § B.3, à la différence près que *n0, n1, n4* et *n* sont regroupés sous la notation *n*, et *n5* fait l'objet d'une description particulière sous l'appellation *m*.

Les troisième et éventuellement quatrième champs forment le nom de la classe. Dans le cas d'une classe propre à une partie, le champ trois est le nom de notation conventionnel de cette partie. Le champ quatre est utilisé pour séparer différentes classes de description qui se rapportent au repérage d'une même entité (indiquée en champ trois). Les champs trois et quatre sont en somme dans une relation générique / spécifique.

mot	classe
bénéfice+	partie / 1 / Contexte / argument
enjeu+	partie / 1 / Contexte / argument
justification+	partie / 1 / Contexte / argument
motivation+	partie / 1 / Contexte / argument
préambule+	partie / 1 / Contexte / introduction
cadre	partie / 1 / Contexte / situation
contexte+	partie / 1 / Contexte / situation
domaine	partie / 1 / Contexte / situation
origine	partie / 1 / Contexte / situation
point de départ	partie / 1 / Contexte / situation
présentation	partie / 1 / Contexte / situation
rappel+	partie / 1 / Contexte / situation
caractéristique+	partie / 2 / But / focalisation
résumé	partie / 2 / But / focalisation
but+	partie / 2 / But / visée
objectif+	partie / 2 / But / visée
objet	partie / 2 / But / visée
nouveau+	partie / 3 / Etat / action / qualification
bilan	partie / 3 / Etat / bilan
état+	partie / 3 / Etat / bilan
point	partie / 3 / Etat / bilan
situation	partie / 3 / Etat / bilan
début	partie / 3 / Etat / moment / qualification
fin	partie / 3 / Etat / moment / qualification
mi(-)	partie / 3 / Etat / moment / qualification
milieu	partie / 3 / Etat / moment / qualification
acquis+	partie / 3 / Etat / résultats
atteint+	partie / 3 / Etat / résultats
effectué+	Partie / 3 / Etat / résultats
obtenu+	partie / 3 / Etat / résultats
résultat+	partie / 3 / Etat / résultats
antérieur+	partie / 3 / Etat / rétrospectif
avancement	partie / 3 / Etat / rétrospectif
préliminaire+	partie / 3 / Etat / rétrospectif
à mener	partie / 4 / Programme / action
à réaliser	partie / 4 / Programme / action
restant	partie / 4 / Programme / action
démarche	partie / 4 / Programme / chronologie
déroulement	partie / 4 / Programme / chronologie
échéance+	partie / 4 / Programme / étape
étape+	partie / 4 / Programme / étape
ordonnancé+	partie / 4 / Programme / étape
ordonnancement	partie / 4 / Programme / étape
principal+	partie / 4 / Programme / étape
réalisation+	partie / 4 / Programme / étape
produits	partie / 4 / Programme / étape / qualification
délai+	partie / 4 / Programme / étape / qualification
contenu	partie / 4 / Programme / pôles
objectif+	partie / 4 / Programme / pôles

piste+	partie / 4 / Programme / pôles
programmation	partie / 4 / Programme / pôles
programme	partie / 4 / Programme / pôles
développement+	partie / 5 / Ultérieur / évolution
orientation+	partie / 5 / Ultérieur / évolution
perspective+	partie / 5 / Ultérieur / évolution
prévision+	partie / 5 / Ultérieur / évolution
tendance+	partie / 5 / Ultérieur / évolution
après	partie / 5 / Ultérieur / futur
au(-)delà	partie / 5 / Ultérieur / futur
final+	partie / 5 / Ultérieur / futur
futur+	partie / 5 / Ultérieur / futur
long terme	partie / 5 / Ultérieur / futur
moyen terme	partie / 5 / Ultérieur / futur
plus long terme	partie / 5 / Ultérieur / futur
prospective+	partie / 5 / Ultérieur / futur
suite+	partie / 5 / Ultérieur / futur
suivant+	partie / 5 / Ultérieur / futur
ultérieur+	partie / 5 / Ultérieur / futur
bibliographie	partie / 6 / Référence / document
bibliographique+	partie / 6 / Référence / document
document+	partie / 6 / Référence / document
documentation	partie / 6 / Référence / document
note+	partie / 6 / Référence / document
publié+	partie / 6 / Référence / document
référence+	partie / 6 / Référence / document
sous A.Q.+	partie / 6 / Référence / qualification
EDF	partie / 7 / Partenariat / externe
extérieur+	partie / 7 / Partenariat / externe
externe+	partie / 7 / Partenariat / externe
collaboration+	partie / 7 / Partenariat / lien
contact+	partie / 7 / Partenariat / lien
contrat+	partie / 7 / Partenariat / lien
coopération+	partie / 7 / Partenariat / lien
intervenant+	partie / 7 / Partenariat / lien
partenaire+	partie / 7 / Partenariat / lien
partenariat+	partie / 7 / Partenariat / lien
relation+	partie / 7 / Partenariat / lien
déroulement	partie / m / Période / période
durée	partie / m / Période / période
initié+	partie / m / Période / période
période+	partie / m / Période / période
annexe+	partie / n / Divers / appendice
observation+	partie / n / Divers / appendice
problème+	partie / n / Divers / appendice
remarque+	partie / n / Divers / appendice
propriété industrielle	partie / n / Divers / brevet
valorisation	partie / n / Divers / brevet
budget	partie / n / Divers / financement
coût+	partie / n / Divers / financement
délais	partie / n / Divers / financement

échancier	partie / n / Divers / financement
estimation	partie / n / Divers / financement
évaluation	partie / n / Divers / financement
facturation	partie / n / Divers / financement
modalités	partie / n / Divers / financement
moyens	partie / n / Divers / financement
paiement	partie / n / Divers / financement
répartition	partie / n / Divers / financement
priorité	partie / n / Divers / Rubrique
définition+	partie / n / Divers / terminologie
lexique	partie / n / Divers / terminologie
objectif	section / 4 / Objectif / objectif
européen+	voc / 1234 / action / qualification
pluriannuel+	voc / 1234 / action / qualification
sous A.Q.+	voc / 1234 / action / qualification
cf.	voc / 1234 / action / renvoi
dans le cadre	voc / 1234 / action / renvoi
voir	voc / 1234 / action / renvoi
voir aussi	voc / 1234 / action / renvoi
ADN	voc / 1234 / action / sigle
département+	voc / 1234 / action / sigle
groupe+	voc / 1234 / action / sigle
PPRD	voc / 1234 / action / sigle
PSE	voc / 1234 / action / sigle
service+	voc / 1234 / action / sigle
général+	voc / 1234 / portée
global+	voc / 1234 / portée
particulier+	voc / 1234 / portée
permanent+	voc / 1234 / portée
spécifique+	voc / 1234 / portée
"-"	voc / 1234567 / articulation / coordination
"&"	voc / 1234567 / articulation / coordination
","	voc / 1234567 / articulation / coordination
"/"	voc / 1234567 / articulation / coordination
"+"	voc / 1234567 / articulation / coordination
et	voc / 1234567 / articulation / coordination
à	voc / 1234567 / articulation / liant
au	voc / 1234567 / articulation / liant
aux	voc / 1234567 / articulation / liant
d'	voc / 1234567 / articulation / liant
de	voc / 1234567 / articulation / liant
des	voc / 1234567 / articulation / liant
du	voc / 1234567 / articulation / liant
l'	voc / 1234567 / articulation / liant
la	voc / 1234567 / articulation / liant
le	voc / 1234567 / articulation / liant
les	voc / 1234567 / articulation / liant
"()"	voc / 1234567 / parenthèse / contour
" : """	voc / 1234567 / parenthèse / contour
" : ()"	voc / 1234567 / parenthèse / contour
" : ."	voc / 1234567 / parenthèse / contour

style titre1	voc / 1234567 / parenthèse / erreur
"(*)"	voc / 1234567 / parenthèse / note
action+	voc / 12345m / action / mention
activité+	voc / 12345m / action / mention
contrat	voc / 12345m / action / mention
contribution	voc / 12345m / action / mention
étude+	voc / 12345m / action / mention
prestation	voc / 12345m / action / mention
projet	voc / 12345m / action / mention
sujet+	voc / 12345m / action / mention
travail+	voc / 12345m / action / mention
attendu+	voc / 1347m / statut / réserves
escompté+	voc / 1347m / statut / réserves
prévisible+	voc / 1347m / statut / réserves
prévision+	voc / 1347m / statut / réserves
prévu+	voc / 1347m / statut / réserves
probable+	voc / 1347m / statut / réserves
programmé+	voc / 1347m / statut / réserves
visé+	voc / 1347m / statut / réserves
"1"	voc / 24 / numérotation
"n°1"	voc / 24 / numérotation
"[infinitive]"	voc / 24 / parenthèse / plein
"[SN]"	voc / 24 / parenthèse / plein
à partir de+	voc / 345 / temps / articulation
au moment de+	voc / 345 / temps / articulation
en	voc / 345 / temps / articulation
jusqu'à+	voc / 345 / temps / articulation
jusqu'en	voc / 345 / temps / articulation
pour	voc / 345 / temps / articulation
sur	voc / 345 / temps / articulation
"[date]"	voc / 345 / temps / date
1er+ semestre	voc / 345 / temps / date
1er+ trimestre	voc / 345 / temps / date
année N	voc / 345 / temps / date
année N-1	voc / 345 / temps / date
année+	voc / 345 / temps / date
aujourd'hui	voc / 345 / temps / date
janvier+	voc / 345 / temps / date
mois de, mois d'	voc / 345 / temps / date
premier+ semestre	voc / 345 / temps / date
premier+ trimestre	voc / 345 / temps / date
néant	voc / 367 / parenthèse / annulation
sans objet	voc / 367 / parenthèse / annulation
établi+	voc / 367 / statut / existant
existant+	voc / 367 / statut / existant
produit+	voc / 367 / statut / existant
réalisé+	voc / 367 / statut / existant
à définir	voc / 457 / statut / prospectif
à établir	voc / 457 / statut / prospectif
envisagé+	voc / 457 / statut / prospectif
éventuel+	voc / 457 / statut / prospectif

indicatif+	voc / 457 / statut / prospectif
possible+	voc / 457 / statut / prospectif
potentiel+	voc / 457 / statut / prospectif
prévisionnel+	voc / 457 / statut / prospectif
proposition+	voc / 457 / statut / prospectif
sous réserve+	voc / 457 / statut / prospectif
abordé+	voc / 47 / statut / commencé
en cours	voc / 47 / statut / commencé
engagé+	voc / 47 / statut / commencé
mis en oeuvre	voc / 47 / statut / commencé
associé+	voc / 67m / rapport
autres	voc / 67m / rapport
avec	voc / 67m / rapport
connexe+	voc / 67m / rapport
lié+	voc / 67m / rapport
lien+	voc / 67m / rapport
participant+	voc / 67m / rapport
prestataire+	voc / 67m / rapport
rattaché+	voc / 67m / rapport