



**HAL**  
open science

# Elements pour une typologie des unites textuelles de textes techniques

Stéphane Caro

► **To cite this version:**

Stéphane Caro. Elements pour une typologie des unites textuelles de textes techniques. [Rapport de recherche] RR-1930, 1993. inria-00074744v1

**HAL Id: inria-00074744**

**<https://inria.hal.science/inria-00074744v1>**

Submitted on 24 May 2006 (v1), last revised 25 Jan 2010 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Éléments pour une typologie  
des unités textuelles  
de textes techniques*

Stéphane CARO

N° 1930  
Juin 1993

PROGRAMME 3

Intelligence artificielle,  
systèmes cognitifs et  
interaction homme-machine

*R*apport  
*de recherche*

1993



## ÉLÉMENTS POUR UNE TYPOLOGIE DES UNITÉS TEXTUELLES DE TEXTES TECHNIQUES

Stéphane CARO

INRIA  
Décembre 1992

---

Cette étude a été réalisée au sein de l'action de recherche "TEXTE" (Grenoble) du projet de psychologie ergonomique de l'I.N.R.I.A, sous la direction d'André Bisseret.  
Elle a été en partie supportée par un contrat du Plan Construction & Architecture, dans le cadre d'une collaboration avec le C.S.T.B. et le C.N.R.S. Elle a également bénéficié de moyens du LSD2 (Laboratoire de Structures Discrètes et Didactiques - I.M.A.G.).

### *Les métaphrases*

1. L'auteur présente TG par Mpr	Mpr ∈ Epr
2. L'auteur illustre tp par Mil	Mil ∈ Eil
3. L'auteur récapitule tp à l'aide du procédé Mre	Mre ∈ Ere
4. L'auteur indexe TG par Mind	Mind ∈ Eind
5. L'auteur guide la lecture par Mguid	Mguid ∈ Eaa
6. L'auteur conduit l'action du lecteur par Mcdt	Mcdt ∈ Eaa
7. L'auteur identifie TG avec Mide	Mide ∈ Eide

### *Ensembles associés*

Eaa= {commentaire, annotation, glose, précision, explication, interprétation, justification, clarification, etc.} [Pascual, 91]

Epr = {composition du texte, articulation, statut de l'information, Sujet, Objet, But, Domaine d'application, Descripteurs (mots clé), Statut du document etc.}

Eil = {exemple, analogie etc.}

Eind = {pointeur, sommaire, n° de page, titre, sous-titre, citation, renvoi, mention etc.}

Ere = {tableau, figure, énumération, liste}

Eide = {Nom de l'émetteur, Coordonnées de l'émetteur, Membres du groupe de travail, Nom ou code du groupe de travail, Nom des membres du groupes, Date d'édition, Version, Lieu de l'édition, Imprimeur, Support, Titres/sous titre du texte, N° ISSN /ISBN, Type de document, Statut du document, N° dans la famille du document, Pays, Dates, Evolution du document, Copyright, Forme de l'ouvrage, Nom du destinataire ; Nom du groupe ; Nom d'un sous groupe etc.}

## CLASSIFICATION DES UNITES PAR CHAQUE JUGE

Liste des catégories avec leur numéros

N°	Catégorie
212	Identité Destinataire
211	Identité Texte
210	Identité Edition
209	Identité Emetteur
208	Adjonctions Interprétatives
207	Adjonctions Instructionnelles
206	Organisateur Inter-textuel
205	Organisateur Intra-textuel
204	Synthèse
203	Illustration
202	Programme Thématique
201	Composantes Structurelles
100	1er Plan

Expérience : Norme NF D 35-303

N° UT	Texte de l'UT	Juge 1	Juge 2	Juge 3	Juge 4	Juge 5
UT1	Norme française	211	211	211	211	211
UT2	homologuée par décision du Directeur Général de l'afnor le 20 octobre 1989 pour prendre effet le 20 novembre 1989.	211	209	211	211	209
UT3	Conduits de fumées composites métalliques rigides démontables ou non Classification des caractéristiques-Méthodes d'essais	211	204	100	211	211
UT4	E : Smoke pipe—Detachable or not detachable, metallic, rigid, compound smoke pipes— Characteristics classification—Test methods D : Verbindungsstücke für Hausschornsteine—Aus Steifen Metallrohren zusammensetzbar, abnehmbar oder nicht abnehmbar—Anforderungen Klassifikation—Prüfverfahren	211	200	211	211	211
UT5	La présente norme définit les caractéristiques fonctionnelles des conduits de fumées composites métalliques démontables ou non ainsi que les essais destinés à les vérifier, et le mode de désignation et de marquage.	202	100	202	202	100
UT6	Certaines caractéristiques sont exigibles pour tous les conduits de fumées quel que soit le type de foyer ou de structure d'immeuble. Ces caractéristiques sont définies dans des classes ;	202	100	208	202	100
UT7	pour le choix de la classe en fonction du type d'application,  voir NF D 35-304.	206	206	206	206	206
UT8	La présente norme sert de base aux normes d'application de tous les conduits de fumées composites métalliques rigides démontables ou non quels que soient le métal, le type de foyer industriel ou domestique, le type de combustible ou la nature de la construction.	202	100	202	202	100

UT9	DESIGNATION	205	204	205	205	205
UT10	Un conduit de fumées est désigné par les indices de classes	100	100	100	100	100
UT11	définis au paragraphe 5.1	205	205	205	206	100
UT12	dans l'ordre suivant: — classe de températures des gaz: T°1, T°2, T°3, T°4 ou T°5, — classe de démontabilité: D ou ND, — classe d'étanchéité au gaz, sans étanchéité interne et externe à l'eau: A1, A2 ou A3, — classe d'étanchéité au gaz, avec étanchéité interne à l'eau: A1 Ei, A2Ei ou A3Ei, — classe d'étanchéité au gaz, avec étanchéité externe à l'eau: A1 Ex, A2Ex ou A3Ex, — classe de résistance à la corrosion: C0, C1, C2, C3 ou C4, — classe de résistance thermique :11,12,13,14 ou 15, — la référence à la présente norme.	100	100	100	100	100
UT13	Exemple :	205	203	205	203	205
UT14	un conduit démontable dont la température des gaz en fonctionnement normal est d'au plus 150 °C, le débit de fuite de gaz est inférieure à 2000 I/h par mètre carré de surface intérieure de conduit, sans étanchéité intérieure ou extérieure à l'eau et sans prescriptions particulières pour la corrosion est désigné:  T° 1 D A2 C0 NF D 35-303	203	203	203	203	203
UT15	ESSAI DE COMPRESSION	205	204	100	100	205
UT16	On recherche la force maximale admissible sans déformation que peuvent supporter deux éléments de conduit. Cette force est exercée dans l'axe des éléments de conduit et parallèlement à celui-ci.  De cette force, est déduit le nombre d'éléments qui peuvent être empilés.	100	100	100	100	100
UT17	TECHNIQUE DES ESSAIS ET MESURES	205	204	205	100	205
UT18	Le tableau ci-dessous résume les nombres d'éléments de conduits devant subir les essais, la nature des essais applicables aux éléments de conduits et la référence aux paragraphes dans lesquels la méthode des essais est définie.	204	205	205	205	205

UT19	N° d'éléments à tester	Essais	Technique des essais					
	2	Démontabilité	7.2					
	3	Température de contact extérieur	7.4.2					
	3	Résistance aux chocs extérieurs	7.5.1.					
	3	Essai de sécurité à la température t3	7.5.2					

## Expérience : Manuel Omnipage

N° UT	Texte de l'UT	Juge 1	Juge 2	Juge 3	Juge 4	Juge 5
UT20	<b>2 - Mise en route Configuration nécessaire</b>	205	204	205	205	205
UT21	Ce chapitre décrit l'installation d'OmniPage Pro sur votre Macintosh.	201	202	202	202	202
UT22	Il comprend quatre sections :	201	201	201	201	202
UT23	<ul style="list-style-type: none"> <li>• Configuration nécessaire</li> <li>• Contenu du coffret</li> <li>• Installation du logiciel</li> <li>• Démarrage d'OmniPage Pro</li> </ul>	201	204	201	205	202
UT24	Si nécessaire, vous pouvez également vous reporter au <i>Manuel de référence</i> pour plus de détails sur un point précis.	206	100	206	206	206
UT25	<b>Attention ! Veuillez lire ce chapitre avec attention</b>	207	208	207	208	200
UT26	<b>Configuration nécessaire</b>	205	204	205	205	205
UT27	Pour utiliser OmniPage Pro, vous devez disposer du système suivant :	100	100	100	100	100
UT28	<b>Ordinateur</b>	205	100	202	100	205
UT29	<ul style="list-style-type: none"> <li>• un Macintosh II ou un modèle plus performant</li> <li>• un disque dur avec au moins 4 Mo d'espace disque</li> <li>• 5 Mo de RAM de mémoire totale du système</li> <li>• Système 6.0.7 ou une version plus récente</li> </ul>	100	100	100	100	100
UT30	(Système 7 nécessite l'usage de mémoire virtuelle)	100	100	100	208	100
UT31	Certaines caractéristiques d'OmniPage Pro, comme les bulles d'aide et la publication, ne sont disponibles que sous Système 7.	100	100	100	100	200
UT32	<b>Scanner</b>	205	204	202	205	100
UT33	OmniPage Pro reconnaît la plupart des scanners pleine-page.	100	100	100	100	200
UT34	Vous en trouverez la liste dans l'annexe B.	205	100	205	206	205
UT35	<b>3 - Procédures</b>	205	204	205	205	205
UT36	Ce chapitre décrit la manière d'utiliser OmniPage Pro pour différentes opérations.	202	202	202	202	202
UT37	Il vous sera particulièrement utile si vous savez déjà vous servir d'OmniPage Pro.	207	202	207	207	200
UT38	Pour une introduction à OmniPage Pro,	207	100	202	207	200
UT39	<b>consultez le Manuel condensé</b>	206	100	206	207	202
UT40	d'OmniPage Professional qui contient un guide d'initiation au programme.	207	100	202	202	202
UT41	Les procédures présentées dans ce chapitre vous indiquent comment travailler avec : <ul style="list-style-type: none"> <li>• des documents simples</li> <li>• des articles de revues</li> <li>• des feuilles de calcul</li> </ul>	202	100	201	202	202
UT42	Chaque procédure, présentée sous forme d'organigramme, décrit un type d'opération ou de document, les réglages de numérisation appropriés, les techniques de base pour effectuer l'opération et les conseils appropriés pour tirer le meilleur parti d'OmniPage Pro.	202	100	202	201	202

UT43	Ces procédures vous donneront une idée générale de la manière d'accomplir le travail mais elles ne sont ni complètes ni définitives.	201	100	208	207	202
UT44	Pour des explications détaillées et des instructions sur les réglages décrits dans ce chapitre, consultez	207	207	202	206	205
UT45	le chapitre 4, "Commandes et réglages".	205	207	205	206	205
UT46	Celui-ci est organisé de manière hiérarchique: de gauche à droite de la barre des menus et de haut en bas des commandes de chacun des menus.	201	100	207	201	201
UT47	Par exemple, vous trouverez les descriptions des <i>Options de page</i> de la fenêtre Réglages de numérisation dans la commande <i>Numériser</i> du menu Traitement.	203	203	203	203	203
UT48	<b>Documents simples</b>	205	204	205	205	205
UT49	La procédure utilisée pour reconnaître le texte de documents simples est la méthode de base qui sert à exécuter la plupart des opérations d'OmniPage Pro:	100	100	100	202	207
UT50	ajustez les réglages de numérisation, cliquez sur <i>Numériser</i> pour commencer la numérisation, éditez le fichier texte obtenu dans l'éditeur de texte et enregistrez ce document dans le format souhaité.	100	100	100	100	100
UT51	Les documents simples, comme des lettres ou des documents légaux, sont structurés en colonnes (une seule colonne ou des colonnes multiples, espacées de la même façon).	100	100	100	100	100
UT52	Supposons que vous souhaitiez reconnaître tout le texte d'une page et non pas ses graphiques. *Chargez un document dans votre scanner *Sélectionnez la commande <i>Numériser</i> du menu Traitement. *Ajustez les réglages de numérisation Réglez les options de page : <i>Taille/A4</i> Réglez les options de reconnaissance : <i>Colonnes/Automatiques, Zones/Automatiques</i> Réglez la <i>Luminosité</i> *Cliquez sur <i>Numériser</i> .	100	100	203	203	203
UT53	OmniPage Pro numérise et reconnaît la page. Une fois la reconnaissance terminée, le fichier texte obtenu apparaît dans l'éditeur pour être édité et enregistré.	100	100	100	100	100
UT54	Il y a des numéros de ligne imprimés dans la marge gauche des documents légaux.	203	100	100	100	100
UT55	Si vous voulez conserver le même format que celui du document original, avec les mêmes numéros de ligne, choisissez l'option <i>Colonnes Etroites</i> dans les options de reconnaissance puis enregistrez le fichier en format ASCII. Un retour à la ligne <CR> est ajouté à la fin de chaque ligne. Si vous ne voulez pas reconnaître les numéros de ligne, sélectionnez <i>Interactives</i> pour l'option <i>Zones</i> et délimitez une zone autour du texte à reconnaître .	203	100	100	100	100
UT56	N'oubliez pas de régler la <i>Taille</i> (dans les options de la page) sur <i>Légal US</i> pour les documents légaux.	203	100	208	100	100

UT57	<b>Ouverture d'un fichier image ou de télécopieur</b>	205	204	205	205	205
UT58	Pour reconnaître un fichier image ou de télécopieur :	100	100	200	100	100
UT59	1. Cliquez sur le bouton <i>Image</i> dans la boîte de dialogue Ouvrir. 2. Sélectionnez le dossier contenant le fichier à ouvrir.	100	100	100	100	100
UT60	La liste des fichiers du dossier courant apparaît.	100	100	208	100	100
UT61	3. Sélectionnez le fichier à ouvrir. 4. Cliquez sur <i>Ouvrir</i> .	100	100	100	100	100
UT62	La boîte de dialogue Réglages de reconnaissance apparaît.	100	100	208	100	100
UT63	<b>Luminosité</b>	205	204	205	205	205
UT64	Sélectionnez différents réglages de luminosité pour éclaircir ou assombrir une image numérisée.	100	100	100	100	100
UT65	Ce réglage est nécessaire pour prendre en compte les différences de qualité de papier, d'impression et les caractéristiques des scanners.	100	100	100	208	100
UT66	Il a le même rôle que pour un photocopieur.	203	100	203	208	100
UT67	» <i>Le réglage de luminosité est le meilleur moyen d'améliorer la reconnaissance de texte.</i>	208	208	208	208	100
UT68	Pour régler la luminosité, cliquez sur la flèche de gauche, ou de droite de la jauge.	100	100	100	100	100
UT69	Si la page présente des caractères distincts et contrastés imprimés sur du papier de bonne qualité et non-réfléchissant, réglez la luminosité sur une valeur moyenne.	100	100	208	100	100
UT70	Si la page présente des caractères fins et brisés, réduisez la <i>Luminosité</i> en déplaçant la jauge vers la gauche (Assombrir).	100	100	208	100	100

# ÉLÉMENTS POUR UNE TYPOLOGIE DES UNITÉS TEXTUELLES DE TEXTES TECHNIQUES

## **Résumé**

Ce rapport propose une méthode de découpage d'un texte technique en unités textuelles relevant d'une même intention de l'auteur quant au rôle de l'information transmise.

En général les textes techniques ont tendance à être rédigés, comme d'autres textes, en vue d'une lecture de bout en bout. Or la lecture de ces textes est souvent partielle et répétée : c'est une lecture de consultation. On peut penser faciliter le repérage de l'information recherchée par une mise en évidence claire (à l'aide de présentations hypertextes par exemple) des divers types d'unités textuelles qu'on peut y trouver. Nous avons fait une première étude sur un corpus de textes techniques qui a permis d'élaborer une typologie fondée sur une distinction en unités textuelles principales qui fournissent les informations nécessaires au but du texte et des unités secondaires qui ajoutent des informations non directement nécessaires à ce but. Ces unités textuelles de second plan sont elles-mêmes catégorisées selon différentes intentions de l'auteur (commenter, illustrer, résumer, guider la lecture...). Pour une première évaluation nous avons demandé à 5 juges d'analyser un même texte selon cette méthode ce qui a permis de constater un degré d'accord très encourageant.

## **Mots-clés**

*actes de langage ; texte ; premier plan-second plan ; typologies d'unités textuelles, hypertexte.*

## **TYPOLGY FOR TECHNICAL TEXT'S UNITS**

## **Abstract**

A frame is proposed whose aim is to analyse technical texts into text units and to categorize this units according to differents types of information the author is attempting to communicate. This study is based on the speech act theory. Main plan units which provide information focused on the main aim of the text are distinguished from secondary plan units which are additions of different other and not directly necessary materials. On the basis of an empirical analysis of technical texts a categorization of this second plan unit was build in terms of differnts intentions of the writer (to comment, to illustrate, to summarize, to guide lecture...).

A first experiment was carried out in order to measure how well subjects would agree when asked to applied this categorization to a same text. The results are encouraging. The use of such a categorization is suggested in order to structure the display of computerized technical texts specially using hypertext facilities.

## **Keywords**

*speech acts ; text ; main structure - side structure ; text units typology ; hypertext.*

## INTRODUCTION

Nous proposons dans ce rapport une méthode de découpage d'un texte technique en unités textuelles qui rendent compte des intentions du scripteur. Les unités textuelles [Pascual, 91] sont des portions de textes qui comptent 1 ou n propositions et relèvent d'une même intention de l'auteur. En général les textes techniques n'ont pas d'organisation spécifique et ont tendance à être rédigés comme d'autres textes en vue d'une lecture de bout en bout, or la lecture de ces textes est souvent partielle et répétée : c'est une lecture de consultation<sup>1</sup> [Duplan, Jauneau, 86]. Sans organisation particulière, le lecteur doit chercher parmi des informations qui ne le concernent pas, celles qui l'intéressent. On peut penser qu'il est possible de l'aider dans cette activité par une mise en évidence des divers types d'unités textuelles (UT désormais) et ceci en bénéficiant des moyens actuels de type hypertexte par exemple. Pour ce faire il faut connaître la diversité des UT. Notre propos est donc de classer les unités textuelles selon ce qu'à voulu faire l'auteur pour son lecteur. Nous analysons en fait les produits de rédacteurs de textes techniques.

### I. OBJET DE LA RECHERCHE

Plusieurs auteurs indiquent que l'organisation globale d'un texte a un effet sur son traitement [Rouet, 91], [Denis & Denhière, 90]. Il est de ce fait intéressant d'étudier cette organisation. Il existe de nombreuses typologies textuelles, mais elles ne sont guère satisfaisantes car la plupart du temps elles sont trop générales. La tendance actuelle est à l'abandon des typologies englobantes (texte complet) au profit de typologies de parties de textes ; Adam (87) les appelle «séquences». Nous retiendrons la terminologie de l'équipe Virbel qui désigne ces objets par : «unité textuelle».

Il y a plusieurs points de vue pour diviser un texte. Un des points de vue est celui des intentions de l'auteur. La théorie des actes de langage pose que parler une langue, c'est réaliser des actes de langage [Searle, 72]. De même on peut rendre compte d'un texte comme une composition d'actes de langages. Nous reprendrons la démarche de Searle qui consiste à décrire ce que fait le locuteur quand il énonce une phrase avec cette fois, un scripteur et le texte qu'il produit. On recherche quel type d'information le scripteur veut communiquer.

Austin (62) distingue trois composantes fondamentales d'un acte de discours :

- \* le contenu locutoire : c'est le contenu de l'énoncé.
- \* la force (ou valeur) illocutoire : qui exprime le type d'acte en jeu, (ainsi «un conduit démontable par exemple» a une valeur illocutoire d'«illustration»).
- \* l'effet perlocutoire : ce sont les conséquences, les effets que les actes illocutoire ont sur les actions, les pensées, les croyances, etc. des destinataires

Notre typologie des UT s'appuyera sur l'aspect illocutoire des actes effectués par le scripteur. On parlera de valeur illocutoire d'un acte pour exprimer ce que le scripteur a voulu faire.

Que l'acte soit discursif ou réalisé au moyen d'un phénomène de mise en forme matérielle (M.F.M)<sup>2</sup> on considérera qu'il est possible de le caractériser comme un acte de langage à vocation spécifiquement textuelle.

«On peut montrer que toutes les phraséologies associées aux phénomènes de mise en forme matérielle sont des performatifs textuels qui ont une force illocutoire tournée vers le texte lui-même et qui signalent comment les énoncés caractérisés par cette valeur illocutoire doivent être pris en tant qu'objets textuels particuliers dans un texte.» [Pascual, 91]. Une UT peut être

---

<sup>1</sup>On distingue dans la lecture de consultation deux fonctions [Duplan & Jauneau, 86], une fonction de lecture et une fonction de repérage (titres, intertitres, balises...).

<sup>2</sup>procédés type-dispositionnels de mise en forme du texte [Virbel, 92]

constituée de plusieurs actes de langages [Cahours, 89] qui rendent compte de diverses intentions.

## II. POSTULATS POUR UNE TYPOLOGIE DES TEXTES TECHNIQUES

Bien qu'un texte ne soit pas homogène, on fait l'hypothèse qu'il y a un objectif général pour ce texte. On trouve des passages qui relèvent de cet objectif et d'autres qui n'en relèvent pas directement. Nos postulats pour commencer nos travaux de catégorisation sont les suivants :

### II.1 Premier postulat : Le but du scripteur

Chaque texte est rédigé en fonction d'un but que le scripteur s'est assigné (même implicitement). Le but peut être composé si le texte a plusieurs objectifs. «Le *but* représente l'effet spécifique que l'activité langagière est censée produire sur le destinataire ; le but est en quelque sorte un projet de modification du destinataire dans une direction donnée» [Bronckart, 85].

### II.2 Deuxième postulat : L'opposition premier plan / second plan

Chaque texte est constitué d'UT indispensables à l'accomplissement du but du texte, formant en quelque sorte le «squelette» du texte [Combettes, 87] [Weinrich, 73]. Nous les qualifions d'UT de premier plan. Les UT restantes sont de «second plan». Bronckart appelle cette dualité l'opposition «figure» (événements et/ou états mis en évidence) / «fond» (scène) [Bronckart, 85].

### II.3 Troisième postulat : La métaphore

Chaque UT correspond à un acte illocutoire exprimable sous forme d'une métaphore. On parlera donc «d'architecture textuelle» basée sur un ensemble de métaphores [Pascual, 91]. Une métaphore ne fait pas partie du texte, c'est une phrase organisée selon des règles syntaxiques pour décrire une intention de l'auteur qui transparait de son texte. Par exemple : «l'auteur organise son texte en deux parties identifiées pa1 et pa2» «l'auteur numérote pa2» «l'auteur titre» etc. La métaphore dans le cadre de nos recherches est donc une description de la part d'un observateur et non pas une phrase de l'auteur.

Nous faisons notre la définition d'Elsa Pascual (p. 209)

*Une métaphore est une phrase du sous-langage spécialisé relatif à l'architecture textuelle. Elle correspond à la forme discursive d'un phénomène de mise en forme matérielle, exprimant l'intention architecturante sous-jacente à ce phénomène. Une métaphore réalise un acte de discours à vocation textuelle.*

Cependant nous élargissons l'utilisation des métaphores, pour les besoins de notre étude à la caractérisation d'actes de discours n'ayant pas toujours de contrepartie en terme de mise en forme matérielle.

### II.4 Quatrième postulat : Un ensemble fini de métaphores

Il existe un nombre limité de métaphores qui permettent de rendre compte de toutes les UT. Elsa Pascual [Pascual, 91] à proposé une première approche de ce type de langage comprenant trente métaphores décrivant l'architecture d'un texte, auxquelles sont associés un certain nombre d'arguments spécifiant les objets concernés par la métaphore (ensembles d'objets textuels, procédés de MFM etc.). Ceci dit, le langage en question ne prend pas en compte les phénomènes architecturaux complexes (citations, renvois, textes parallèles).

#### *Exemples de métaphore*

\*L'auteur illustre UT0 par Mil

avec Mil ∈ Eil, où Eil = {exemple, analogie, etc.}

\*L'auteur attache UT à id0

Les notations utilisées sont les suivantes :

-  $M_i$  représentent des entités instanciables par des substantifs, pris dans l'ensemble  $E_i$  associé à  $M_i$  ;

-  $idi$  représentent les identificateurs d'objets textuels ;

-  $UT_i$  représentent les unités textuelles .

(Elsa Pascual p. 63)

\*

\* \*

### III. CATÉGORISATION SUR UN CORPUS DE TEXTES TECHNICO REGLEMENTAIRES

#### III.1 Le corpus

Nous avons fait une première étude – de manière empirique – sur un corpus de textes techniques<sup>3</sup> et réglementaires du bâtiment. Les textes du secteur du bâtiment se répartissent pour la plupart dans 3 familles. Les DTU, les normes et les textes réglementaires. Cinq textes ont été choisis de façon à constituer un échantillon relativement représentatif de la documentation du REEF<sup>4</sup> sur un sujet déterminé (les conduits de fumée).

- \* Un document technique unifié (DTU) ;
- \* un arrêté ;
- \* une circulaire interministérielle traitant de l'application de l'arrêté mentionné plus haut ;
- \* un décret ;
- \* une norme<sup>5</sup>.

Nous avons complété la classification en étudiant d'autres textes parmi lesquels sont à signaler deux manuels informatiques :

- \* Mode d'emploi du logiciel «Omnipage Professional»
- \* Guide de prise en main du Macintosh

#### III.2 Méthode utilisée

1° But du texte : Cette première étape nécessite de retenir un «but» pour le texte (premier postulat). La question que l'on se pose est «Quel objectif le scripteur s'est-il fixé pour son texte». Il est souhaitable d'écrire le but du texte quand il n'est pas trivial afin de s'y reporter lors de la seconde étape qui consistera à isoler toutes les UT accessoires quant à l'accomplissement de ce but.

*Exemple simple* Indiquer comment il faut appliquer l'arrêté du 20 juin 1975

*Exemple complexe* Définir les caractéristiques des conduits de fumées composites métalliques et donner les essais destinés à les vérifier ainsi que le mode de désignation et de marquage des conduits.

2° Distinction entre les UT relevant du second plan et celles du premier plan (second postulat). Pour les premiers textes étudiés nous avons :

- passé le texte au scanner
- converti l'image en fichier de traitement de texte à l'aide d'un logiciel de reconnaissance de caractères
- uniformisé le texte à l'écran de l'ordinateur au niveau des polices de caractères, des retraits etc... pour faire abstraction des procédés de mise en forme déjà présent dans les textes (soulignement, encadré...).

---

<sup>3</sup> Cette étude a été menée en collaboration avec le C.S.T.B. (Centre Scientifique et Technique du Bâtiment)

<sup>4</sup> Le REEF est l'ensemble des règles techniques de la construction, environ 15000 pages.

<sup>5</sup>

- Le DTU (document technique unifié) n° 24.2.1 de novembre 1990 (référence AFNOR DTU P 51-202) intitulé "cheminées à foyer ouvert équipées ou non d'un récupérateur de chaleur utilisant exclusivement le bois comme combustible" ainsi que l'erratum et l'additif qui lui sont associés
- L'arrêté du 20 juin 1975 concernant l'équipement et exploitation des installations thermiques en vue de réduire la pollution atmosphérique et d'économiser l'énergie.
- La circulaire interministérielle du 18 décembre 1977 traitant de l'application de l'arrête du 20 juin 1975 concernant l'équipement et l'exploitation des installations thermiques en vue de réduire la pollution atmosphérique et d'économiser l'énergie.
- Le décret n° 74-415 du 13 mai 1974 (interministériel) relatif au contrôle des émissions polluantes dans l'atmosphère et à certaines utilisations de l'énergie thermique.
- La norme NF D 35-303 de Novembre 1989 intitulée Classification des caractéristiques-Méthodes d'essais pour les conduits de fumées composites métalliques rigides démontables ou non.

- imprimé
  - lu séquentiellement et surligné ou entouré les UT de second plan. Il est plus rapide de surligner ou d'entourer les UT de second plan, celle-ci étant moins nombreuses que celles de premier plan.
- Pour cette dernière opération les décisions sont bien sûr empiriques et entachées de subjectivité.

### Exemple

Enregistrement d'un document

Tant que vous n'enregistrez pas le document, il n'existe que dans la mémoire de l'ordinateur

telle une pensée qui s'envole si elle n'est pas notée par écrit.

Choisissez Enregistrer dans le menu Fichier.

Dans la zone de dialogue qui s'affiche, vous pouvez donner un nom au document et définir son emplacement sur le disque dur.

Cliquez sur Enregistrer.

3° Répartition des UT de second plan dans des classes selon l'intention de l'auteur. Il s'agit de regrouper les UT selon des critères explicites et non-ambigus. Nous cherchons à caractériser l'acte illocutoire (et non pas indiquer de quoi elles traitent) [Virbel, 92]. On attribue un code ou un numéro à chaque classe ce qui permet de traiter le document plus rapidement.

### Exemple

Enregistrement d'un document

Tant que vous n'enregistrez pas le document, il n'existe que dans la mémoire de l'ordinateur

telle une pensée qui s'envole si elle n'est pas notée par écrit.

Choisissez Enregistrer dans le menu Fichier.

Dans la zone de dialogue qui s'affiche, vous pouvez donner un nom au document et définir son emplacement sur le disque dur.

Cliquez sur Enregistrer.

O.intra\*

Illustration

\*«O. Intra» signifie Organisateur intra-textuel»

**Remarque** La méthode que nous avons choisie nécessite la prise en compte du contexte pour catégoriser les UT. En dehors du contexte du document il est fortement improbable d'arriver aux mêmes résultats. La même UT pourrait être classée en second plan ou en premier plan dans des documents ayant des buts différents.

## III.3 Résultats

### III.3.1 Résumé de la classification

Nous pouvons ainsi résumer la classification que nous avons établie :

**Univers informationnel**={Premier plan, Second plan}

**Second plan**={Guidage de la compréhension, Guidage de la lecture, Guidage de l'action, Identification}

**Guidage de la compréhension**={Méta-texte, Para-texte}

**Méta-texte**={Composantes structurelles, Programme thématique}

**Para-texte**={Illustration, Synthèse}

**Guidage de la lecture**={Organisateurs intra-textuels, Organisateurs inter-textuels, Adjonctions instructionnelles}

**Guidage de l'action**={Adjonctions interprétatives}

**Identification**={Identité émetteur, Identité édition, Identité Texte, Identité destinataire}

L'«Univers informationnel» est en fait une base textuelle quelconque pour la classification, elle peut se composer : d'une UT, un texte, un corpus de textes (la documentation technique du bâtiment et des travaux publics...). La définition de cette base influence les résultats. Une UT ne sera pas classée de la même manière si elle est analysée dans le cadre de son texte, auquel cas

le but du texte global  $TG^6$  est la référence, ou indépendamment, auquel cas c'est le but du texte partiel  $tp^7$  qui guide la classification.

Nous considérons que les UT de premier plan relèvent toutes de la même intention de l'auteur «atteindre l'objectif qu'il s'est fixé pour son texte». Nous les classons donc toutes dans une seule et même catégorie «UT de premier plan».

De façon plus détaillée la classification se présente ainsi :

#### UT DE PREMIER PLAN

#### UT DE SECOND PLAN

#### GUIDAGE DE LA COMPREHENSION

\***Méta-texte** : UT qui parlent du texte

- Composantes structurelles :

UT qui indiquent l'architecture de TG

- Programme thématique :

UT qui indiquent la substance du texte

\***Para-texte** : UT qui reprennent une partie du texte sous une autre forme

- Illustration :

UT qui exemplifient le discours

- Synthèse :

UT qui reformulent un propos sous une forme différente

#### GUIDAGE DE LA LECTURE

- Organismes intra-textuels :

UT qui permettent le repérage et la navigation à l'intérieur du texte.

- Organismes inter-textuels :

UT qui permettent le repérage et la navigation vers d'autres textes.

- Adjonctions instructionnelles :

UT qui indiquent comment procéder à la lecture

#### GUIDAGE DE L'ACTION

- Adjonctions interprétatives :

UT qui précisent comment procéder aux actions (décrites par le texte)

#### IDENTIFICATION

- Emetteur :

UT qui identifient l'émetteur

- Edition :

UT qui identifient l'édition

- Texte :

UT qui identifient le texte

- Destinataire :

UT qui identifient le destinataire

### III.3.2. Définition des classes

#### Méta-texte

Le méta-texte est composé des UT qui ont pour objet le texte (qui parlent du texte). Le Méta-texte peut être implicite quand le nom du document n'est pas cité ou explicite si le nom du document est cité.

#### Métaphore

#### L'auteur présente TG par Mpr

$Mpr \in Epr$ , où  $Epr = \{\text{composition du texte, articulation, statut de l'information, Sujet, Objet, But, Domaine d'application, Descripteurs (mots clé), Statut du document etc.}\}$

<sup>6</sup>TG texte considéré dans sa globalité

<sup>7</sup>tp texte partiel, peut être un paragraphe, un chapitre, et en général toutes les subdivisions répertoriées.

On peut distinguer :

\*UT exposant la structure du texte, son architecture (composantes structurelles, articulation) [Virbel, 92]

\*UT exposant la substance du texte, son sens (sujet du texte, résumé, domaine d'application, mots-clés, objet...)

Il n'est pas toujours aisé de distinguer les deux aspects du méta-texte, «Composantes structurelles» et «Programme thématique». Ceux-ci sont souvent mêlés, notamment dans les UT qui indiquent le thème de plusieurs parties du texte, comme le font souvent les introductions ou une table des matières. Sont alors présentes des informations concernant l'enchaînement du discours (composantes structurelles) et d'autres indiquant la substance de ses parties. Dans l'exemple qui suit nous avons indiqué en gras l'aspect structurel et en maigre l'aspect substantiel.

#### *Exemple*

*\*Il comprend trois titres : le titre I est consacré, d'une part, à l'équipement et à l'installation des chaufferies, il indique les appareils de réglage des feux et de contrôle dont doit être muni chaque générateur ou chaque installation en fonction de sa puissance. D'autre part, il impose une méthode de calcul des hauteurs de cheminées en vue d'obtenir une bonne diffusion des polluants émis.*

*Le titre II concerne les émissions particulières. Il fixe les indices de noircissement et les indices pondéraux des fumées.*

*Enfin, le titre III est consacré aux dispositions administratives. Il impose notamment l'affichage d'un tableau des périodes de ramonage et la tenue d'un livret de chaufferie.*

- Substance indiquée par mots-clés

#### *Exemples*

*\*conduit de fumée, matériau composite, métal, classification, caractéristique, étanchéité, résistance à la corrosion, résistance mécanique, résistance au choc thermique, essai, marquage*

*\*équipement, qualité de la vie, santé, industrie et recherche*

#### **Para-texte**

Le para-texte correspond aux UT qui reprennent le texte sous une autre forme avec deux buts précis : illustrer ou synthétiser :

\*Illustrer : à l'aide d'analogies – rapprochement de tel problème connu que l'on explique –, d'exemples – supposés induirent convenablement sans explication – [Coltier, 88]. Le scripteur se soucie de donner à son propos un aspect plus concret, il fournit donc une image puisée dans la connaissance supposée partagée qui viendra mettre en lumière ce qu'il veut dire.

#### *Métaphore*

**L'auteur illustre tp par Mil**

avec  $Mil \in Eil$ , où  $Eil = \{\text{exemple, analogie, etc.}\}$

#### *Exemple*

*Tel l'objectif d'une caméra, le finder vous permet de visualiser tous les objets se trouvant sur votre espace de travail (bureau).*

Marqueurs souvent utilisés :

«par exemple», «exemple», «comme», «tel», «ainsi», les parenthèses, les deux points.

\*Synthétiser : le scripteur reprend des informations distantes dans un souci de concision, il récapitule les données principales de son propos sous une forme différente en général : un tableau, une énumération ou de façon discursive.

#### *Métaphore*

**L'auteur récapitule tp à l'aide du procédé Mre**

avec  $Mre \in Ere$ , où  $Ere = \{\text{tableau, figure, énumération, liste}\}$

## Exemples

Le tableau ci-dessous résume les nombres d'éléments de conduits devant subir les essais, la nature des essais applicables aux éléments de conduits et la référence aux paragraphes dans lesquels la méthode des essais est définie.

Nbre d'éléments à tester	Essais	Technique des essais
1	Matériau	7.1
3	Température de contact extérieur	7.4.2
3	Résistance aux chocs extérieurs	7.5.1

Ces données ont déjà été fournies dans le texte, sinon nous les aurions classées dans le texte principal.

## Organisateurs

Les organisateurs sont des UT qui fournissent le repérage, l'aide à la circulation dans le texte ou en direction d'autres textes.

### Métaphore

#### L'auteur indexe TG par Mind<sup>8</sup>

avec Mind  $\in$  Eind, où Eind = {pointeur, sommaire, n° de page, titre, sous-titre, citation, renvoi, mention etc.}

- Les UT ou pointeurs qui permettent le repérage et la circulation dans le texte sont appelés «Organisateurs intra-textuels» : les numéros de page, les titres de parties de texte (On estime que les titres et sous-titres de tp sont des cas particuliers qui jouent un rôle de repérage, les titres et sous-titres de TG ayant plutôt une fonction d'identification.), les sommaires,

- les numéros de parties de texte (chapitre, paragraphes...)

### Exemples

*5.1.1 ; Figure 1 ; Tableau 2 ; art. 1er ; Titre 1er ; a) b) ; 1° 2°*

- Les renvois à une autre partie du même texte [Bronckart, 85] :

— élément (chiffre, astérisque, ...) renvoyant à une note en bas de page, à une autre partie du même texte, à la bibliographie, etc.;

— élément renvoyant à un tableau, à une figure, etc.;

### Exemples

*\*voir plus haut, Cf. p. 10, infra.*

*\*Elle donne également en annexe A (faisant partie intégrante de la norme)[...] en annexe B (ne faisant pas partie intégrante de la norme) les conditions...*

*\* [...] aux paragraphes 5.2.2 et 7.4.3 sans modification du classement d'étanchéité,*

*\* [...] subir l'essai d'étanchéité aux gaz (voir paragraphes 5.1.2.1 et 7.3.1), qui...*

*\*pour l'emploi de ces classes voir NFD35-304*

*\* [...] de son aptitude à résister aux températures du 5.1.1 et à la corrosion du 5.1.3 est fait sur dossier à partir...*

*\*(voir figure 1).*

Les pointeurs sont mis en gras par nous.

*Remarque* Pour délimiter un pointeur on simule son absence en imaginant une proposition qui puisse le remplacer. Par exemple dans la phrase suivante : *pour l'emploi de ces classes voir NFD35-304*, l'absence de pointeur pourrait donner ceci *pour l'emploi de ces classes il convient d'accorder une attention particulière à...*

De même pour *L'essai est réalisé conformément au paragraphe 8.3* qui peut devenir *L'essai est réalisé conformément à la législation...*

Les UT ou pointeurs qui permettent le repérage et la circulation vers d'autres textes sont appelés «Organisateurs inter-textuels» :

<sup>8</sup> indexer au sens informatique du terme soit construire des adresses d'information et en fournir les tables d'adressage

Renvoi à un autre texte (citation ou mention d'un ouvrage). Informations suggérant de quitter le texte.

Explicite, le nom du document pointé est cité :

*Exemples*

*\*Pour une introduction à OmniPage Pro, consultez le Manuel condensé d'OmniPage Professional qui contient un guide d'initiation au programme.*

*\*Vu le décret n° 74-306 du 10 avril 1974 modifiant le décret n° 69 596 du 14 juin 1969 fixant les règles générales de construction des bâtiments d'habitation ;*

Implicite, le nom du document pointé n'est pas précisé :

*Exemple*

*\*les deux éléments d'extrémité sont fixés à la salle d'essais par les dispositifs préconisés par le constructeur dans son catalogue.*

Les pointeurs sont parfois présent dans des UT de premier plan :

*Exemple*

*\*L'essai est réalisé conformément au paragraphe 8.3 de la norme NF C 20-010 de 1986.*

Quelquefois ils le sont dans des UT de second plan :

*Exemple*

*\*Si nécessaire, vous pourrez également vous reporter au Manuel de référence pour plus de détails sur un point précis.*

C'est pourquoi nous suivons volontiers Elsa Pascual qui pose que *les citations et les renvois ne sont pas obligatoirement eux même des objets textuels : ce sont des pointeurs vers des objets textuels.* [Pascual, 91]

### **Adjonctions instructionnelles**

Il est question ici des informations fournies par l'auteur pour procéder à une lecture bénéfique. Elles concernent principalement le but, la procédure ou les pré-requis de la lecture.

#### *Métaphore*

#### **L'auteur guide la lecture par Mguid**

Mguid ∈ Eaa, où Eaa= {commentaire, annotation, glose, précision, explication, interprétation, justification, clarification, etc.} [Pascual, 91]

*Exemple*

*\*Attention ! Veuillez lire ce chapitre avec attention*

*\*Pour des explications détaillées et des instructions sur les réglages décrits dans ce chapitre, consultez le chapitre 4, «Commandes et réglages».*

*\*Ce manuel s'adresse aux débutants en informatique ou à ceux qui ne connaissent pas l'environnement Macintosh. Si vous vous trouvez dans l'un ou l'autre cas, lisez les chapitres dans l'ordre, du premier au dernier.*

*\*Une fois les six premiers chapitres assimilés, vous pourrez vous initier aux applications que vous souhaitez utiliser sur le Macintosh. (les techniques enseignées dans le présent manuel vous seront utiles dans toutes les applications Macintosh.)*

*\*Revenez ensuite à ce manuel pour étudier les chapitres restants. Si nécessaire, vous pourrez également vous reporter au Manuel de référence pour plus de détails sur un point précis.*

*\*Le Manuel de référence contient toutes les informations nécessaires à l'utilisation optimale de votre ordinateur Macintosh. A la différence du présent manuel, qui doit être lu de façon exhaustive, le Manuel de référence doit être utilisé comme un dictionnaire, c'est-à-dire un ouvrage auquel vous vous reportez pour trouver un complément d'information sur un point précis.*

*\*Pour trouver une information donnée, vous pouvez recourir à la table des matières, à l'index et au glossaire du Manuel de référence.*

*\*Si vous souhaitez réviser la Visite guidée, suivez les instructions données à la section «révision de la visite guidée» de la carte de référence.*

*\*Avant de commencer vous devez avoir étudié le chapitre 1*

*\*Passez au chapitre 3 dans lequel vous allez apprendre à installer les applications ou bien faites une pause.*

*\*La lecture de ce manuel requiert une bonne connaissance de base du Macintosh. Si ce n'est pas le cas, reportez vous d'abord au manuel d'utilisation accompagnant l'ordinateur avant de continuer.*

### **Adjonctions interprétatives [Virbel, 92]**

Il est question ici des informations sur le but, la procédure ou les pré-requis des actions à mener sous la forme de commentaires, annotations ou clarifications.

Ces adjonctions informent en général le lecteur sur la procédure, le but ou les pré-requis des actions à mener conformément à la lecture.

#### *Métaphore*

#### **L'auteur conduit l'action du lecteur par Mcdt**

Mcdt  $\in$  Eaa, où Eaa = {commentaire, annotation, glose, précision, explication, interprétation, justification, clarification, etc.} [Pascual, 91]

#### *Exemple*

*\*Si vous utilisez un serveur de fichier, tel qu'AppleShare, vous devez déjà être familiarisé avec cette commande*

*\*Félicitation ! Vous savez vous débrouiller à présent.*

*\*En plus de son rôle décoratif, il est généralement utilisé pour assurer la protection thermique de la poutre en bois.*

*\*Ceci se produit rarement, mais mieux vaut prévenir que guérir*

*\*Nous vous suggérons de prendre l'habitude de protéger les documents confidentiels soit avec un mot de passe général que vous communiquez aux personnes utilisant des documents communs soit avec un mot de passe qui vous est propre pour vos documents personnels*

### **Identification**

L'identification procède d'un étiquetage du texte avec les informations concernant : l'émetteur du texte, le texte, le destinataire, l'édition.

#### *Métaphore*

#### **L'auteur identifie TG avec Mide**

Mide  $\in$  Eide, où Eide = {Nom de l'émetteur, Coordonnées de l'émetteur, Membres du groupe de travail, Nom ou code du groupe de travail, Nom des membres du groupes} U {Date d'édition, Version, Lieu de l'édition, Imprimeur, Support} U {Titres/sous titre du texte, N° ISSN /ISBN, Type de document, Statut du document, N° dans la famille du document, Pays, Dates, Evolution du document, Copyright, Forme de l'ouvrage} U {Nom du destinataire ; Nom du groupe ; Nom d'un sous groupe}

- L'émetteur :

*son nom*

*\*afnor*

*\*cahiers du centre scientifique et technique du bâtiment*

*le nom ou code d'un groupe*

*\*Tuyauteries de fumée UNM 712*

*\*établi par le groupe de coordination des textes techniques*

*le nom des membres d'un groupe*

*\*Le ministre de l'équipement, le ministre de la qualité de la vie, le ministre de la santé et le ministre de l'industrie et de la recherche*

*\*M COC... CSTB*

*M COI... POUJOULAT*

*M COR... CENTRE TECHNIQUE DU BOIS*

*les coordonnées de l'émetteur*

*\*(afnor), tour europe cedex 7 92080 paris la défense*

*\*tél.: (1) 42 91 55 55*

*\*secrétariat du groupe DTU:*

*CSTB, 4, av. du recteur-poincaré, paris 16*

- L'édition qui est la représentation de tout ce qui génère la production du texte au niveau matériel cette fois

**Version**

- \*1er tirage 89-11
- \*S. 88 Mise à jour 1992-2
- \*6e édition

**Imprimeur**

- \*Imprimerie Louis-Jean - 05002 GAP
- \*imb IMPRIMEUR - 70001 VESOUL

**Lieu d'édition**

- \*Fait à Paris,

**Support**

- \*J.O. du ...
- \*cahier 2451

**Date de parution**

- \*(janvier 1982)
- \*du 15 mai 1974

- Le texte est étiqueté par ses attributs les plus courants : titre, sous-titre, date, pays,

**N° ISSN /ISBN**

- \*ISSN 0335-3931
- \*ISBN 2-86891-1 87-0

**Classification dans la famille du document**

- \*NF D 35-303
- \*A 20-06-75
- \*DTU 24.2.1
- \*référence AFNOR DTU P 51-202
- \*C 18-12-77
- \*décret n° 74-415
- \*D 13-05-74

**Copyright**

- \*© CSTB 1990
- \*Toute représentation ou reproduction intégrale ou partielle de cet ouvrage faite sans le consentement du CSTB est illicite. Elle constitue une contrefaçon au sens de la loi du 11 mars 1957.

**Forme de l'ouvrage**

- \*norme française
- \*arrêté
- \*document technique unifié
- \*Circulaire interministérielle
- \*Décret interministériel
- \*Guide, Dictionnaire

**Evolution du document**

- \*(Arrêté) modifié par arrêtés du 7 décembre 1983 et du 10 décembre 1991

- Le destinataire représente la cible ou le «public» auquel le texte est adressé, c'est une représentation sociale. Un scripteur produit un texte en le destinant à un public qui est la représentation sociale qu'il se fait de ses lecteurs potentiels bien qu'il n'ait, physiquement, aucun interlocuteur [Bronckart, 85].

**Identité d'un groupe d'utilisateur**

- \*Ce manuel s'adresse aux débutants en informatique ou à ceux qui ne connaissent pas l'environnement Macintosh®.

*\*Le ministre de la Culture et de l'Environnement, le ministre de l'Équipement et de l'Aménagement du Territoire, le ministre de l'Industrie, du Commerce et de l'Artisanat, le ministre de la Santé et de la Sécurité Sociale à Messieurs les préfets, les chefs des services interdépartementaux de l'industrie et des Mines.*

#### **Identité d'un sous groupe d'utilisateurs**

*\*Notes pour les utilisateurs d'AppleShare.*

**Remarque** Les titres de tp ou du TG figurent en deux classes différentes dans le second plan. En effet la fonction du titre du TG consiste à identifier le document, désigner son contenu et le mettre en valeur [Willemart, 92], c'est pourquoi le titre et sous-titre de l'ouvrage sont mentionnés dans la rubrique «IDENTIFICATION du texte», tandis qu'un titre de tp est classé dans la catégorie des Organisateur intra textuels dans le GUIDAGE DE LA LECTURE. Bien que les titres (de parties de texte) soient dans le second plan, nous reconnaissons qu'ils ont un rôle particulier. Dans les corpus que nous avons étudiés, nous avons constaté qu'ils représentent assez bien le passage qui suit et de ce fait constituent un moyen d'accès à l'information efficace (par l'intermédiaire de sommaires par exemple). Très peu de titres apportent des informations qui ne sont pas mentionnées dans la portion de texte concernée, on peut donc dire que les informations figurant dans les titres sont redondantes pour la grande majorité et c'est pour cette raison que les fonctions «titrer» figurent dans le second plan.

## **IV PREMIERE VALIDATION**

Il nous importe de savoir si les actes de langages ainsi distingués dans cette classification sont perceptibles. Une première expérience a donc consisté à demander à des juges de procéder à un travail de classification d'UT tirées de deux textes, au moyen de la grille d'UT.

### **IV.1 Méthodologie**

#### **IV.1.1 Participant(e)s**

Cinq juges ont participé. Tous ont une formation supérieure au moins égale à Bac +5 et sont des étudiants, utilisateurs de Macintosh. Le groupe compte trois femmes et deux hommes.

#### **IV.1.2 Matériel**

Il comprend la liste des catégories, une consigne expliquant la méthode à mettre en oeuvre, deux textes construits à l'aide d'extraits du manuel d'un logiciel et de la Norme du domaine du bâtiment – texte étudiés plus haut –. Les documents dans leur intégralité étaient également mis à la disposition des juges pour leur permettre de situer les UT dans leurs contextes. Des surligneurs de diverses couleurs leur permettaient de parcourir les diverses étapes de l'expérience.

Les textes sont les suivants :

- Manuel d'Omnipage Professional, logiciel de reconnaissance de caractères pour ordinateur Macintosh.
- La norme NF D 35-303 de Novembre 1989 intitulée Classification des caractéristiques-Méthodes d'essais pour les conduits de fumées composites métalliques rigides démontables ou non.

La liste des catégories comprenait deux niveaux : premier plan et second plan.

#### **IV.1.3 Déroulement de l'expérience**

Les participant(e)s prennent tous/toutes part individuellement à une séance expérimentale. En début d'expérience la consigne écrite et la classification est donnée. La consigne indique la tâche à effectuer. L'expérience se déroule en deux phases. Dans la première les participants prennent connaissance de la consigne, de la classification et des documents d'où proviennent les UT à

classer. L'expérimentateur informe les sujets qu'ils n'auront plus accès aux documents d'où proviennent les UT (pour ne pas être influencé par les éventuels titres données aux UT dans ces documents). L'expérimentateur présente le premier texte au sujet. L'ordre des textes est contrebalancé. Le sujet effectue une première lecture en distinguant à l'aide d'un surligneur le second et le premier plan puis il indique, au cours d'un second passage le type d'UT en entourant plus finement les phrases, propositions ou groupes de mots et en reportant sur son texte le numéro de la catégorie qu'il a choisie. Si aucune catégorie ne convient il peut proposer un autre type d'acte en le spécifiant sur son texte. Quand sur la classification, certaines catégories avaient des termes de leur définition soulignés, les sujets devaient choisir l'un de ses termes ou en proposer un qui convienne mieux.

*Exemple* Illustration : exemple (aide à inductions à partir desquels le lecteur doit comprendre sans explication) ; analogie (comme tel problème connu que l'on explique)

#### IV.1.4 Recueil et traitement des données

Deux types de données sont recueillies. Les temps d'exécution et les classifications des UT fournies par les participants. Nous ne comparons pas les résultats avec une grille théorique mais nous essayons de mesurer la proportion d'accord inter-juges tant en ce qui concerne la distinction des UT de second plan que leur répartition dans des classes.

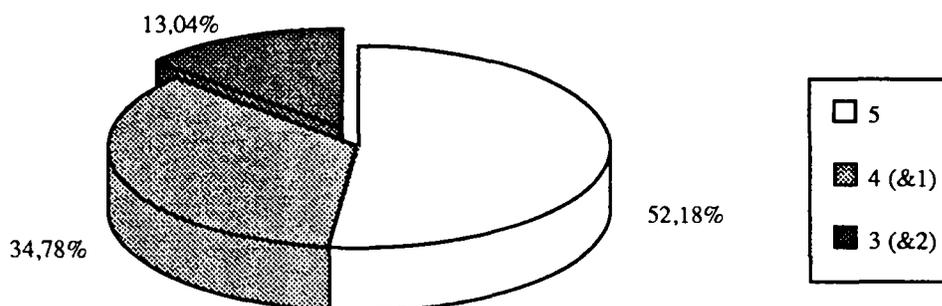
## IV.2 Résultats

### IV.2.1 Le temps

Les participants consacrent en moyenne 80 minutes pour effectuer la classification des deux textes. Ce temps ne comprend ni la lecture de la consigne ou la première lecture de la classification, ni la consultation préalable des documents proposée au début de l'expérience. Les UT provenant du manuel d'Omnipage sont classées en cinquante minutes (2 pages) tandis que celles de la norme ne nécessitent que trente minutes (1 page). Les sujets consacrent moins de temps à la première phase de l'expérience, (environ 30%) qui consiste à différencier premier plan et second plan, qu'à la seconde, l'identification des UT.

### IV.2.2 Distinction premier plan second plan

Distinction 1er plan/2nd plan sur les deux textes en % d'UT. Nombre de juges d'accord.



Pour 52,18% des UT l'accord est total entre les juges (5 juges d'accord) pour la séparation entre premier et second plan. Pour 34,78% des UT un juge n'est pas d'accord (4 juges d'accord). Ainsi une grande majorité des UT (87%) sont classées par au moins 4 juges sur cinq dans le même plan. C'est un résultat satisfaisant compte tenu que les juges faisait ce type de tâche pour la première fois.

### *Norme du bâtiment*

1er/2e Plan	Nbre UT	% d'UT
5	11	61,1
4&1	3	16,7
3&2	4	22,2
Totaux	18	100

Les UT de la norme sont distinguées avec un accord de tout les juges pour 61,1% des UT. 77,8% des UT sont distinguées par au moins quatre juges sur cinq (ligne 5 et 4&1).

### *Manuel informatique*

1er/2e Plan	Nbre d'UT	% d'UT
5	25	49,02
4&1	21	41,18
3&2	5	9,8
Totaux	51	100

90,2% des UT sont distinguées avec l'accord d'au moins 4 juges sur 5.

#### **IV.2.2 Différenciation entre les UT**

Accord	% d'UT	Nbre UT	Cumul
5	24,64	17	17
4.1	34,78	24	41
3.2	7,25	5	46
3.1.1	17,39	12	58
2.2.1	7,25	5	63
2.1.1.1	7,25	5	68
1.1.1.1.1	1,45	1	69
Totaux	100	69	

Le tableau reprend les UT des deux textes qui en comptaient 70. L'UT 19 (il s'agit d'un tableau) n'a pas été prise en compte dans les calculs, les sujets ayant découpé cette UT de manière très diverses, tantôt par colonnes, tantôt par lignes, ou encore comme un seul item. Les juges ont tous affecté la même catégorie à 17 UT. Une seule UT a donné lieu à un désaccord total (ligne 1.1.1.1.1). Si on cumule les deux premières lignes du tableau on s'aperçoit que près de 60% des UT sont classées dans la même catégorie par au moins 4 juges sur 5. Les principaux désaccords entre les juges opposant au moins deux d'entre eux aux deux autres (catégorie 3.2 et 2.2.1) concernent 10 UT (soit 14,5%). Il convient de distinguer les scores à l'intérieur des classes.

## Evaluation de l'accord inter-juge par catégorie

	Catégorie	% homogénéité/ confusion
211	Identité Texte	75
208	Adjonctions Interprétatives	60
207	Adjonctions Instructionnelles	45
206	Organisateur Inter-textuel	73,33
205	Organisateur Intra-textuel	66,25
203	Illustration	68
202	Programme Thématique	65
201	Composantes Structurelles	60
100	1er Plan	88

Ce tableau reprend les principaux résultats de l'annexe "Patterns de confusion". Les pourcentages indiqués correspondent au nombre de citations d'une classe pour une UT, sur les citations d'autres classes.

Voici quelques analyses qualitatives de désaccord entre les juges.

La catégorie «Identité texte» est affectée d'un bon score du point de vue de l'homogénéité. Toutefois l'UT 2 a provoqué un désaccord entre les juges – 3 juges catégories 211, «Identité texte» et deux juges 209 «Identité émetteur» – :

### UT 2

*homologuée  
par décision du Directeur Général de l'afnor  
le 20 octobre 1989  
pour prendre effet le 20 novembre 1989.*

Les résultats obtenus pour la catégorie «Adjonctions Interprétatives» sont peu représentatifs car ils ne concernent que deux UT. L'une d'entre elle, l'UT 25 a d'ailleurs fait l'objet d'un désaccord entre 208 «Adjonctions Interprétatives» (deux juges), 207 «Adjonctions Instructionnelles» (deux juges) et 200 (second plan indéterminé) :

### UT 25

*Attention ! Veuillez lire ce chapitre avec attention*

Les juges ont hésité entre un guidage de la lecture et un guidage de l'action.

Les organisateurs (206 et 205) sont bien reconnus bien qu'un des juges ait attribué à tous les titres la catégorie 204 «Synthèse» contre 205 «Organisateurs intra-textuels» pour les autres juges, considérant que les titres étaient une reformulation d'un propos du texte sous une forme différente.

Trois UT occasionnent un désaccord concernant la catégorie 203 :

### UT 13 en gras

*Exemple :  
un conduit démontable dont la température...*

Trois juges considèrent cette UT comme un titre (Organisateur intra textuel), les deux autres ne l'ont pas distinguée de l'exemple lui même et l'on classé dans «Illustration».

## UT 52

*graphiques. Supposons que vous souhaitez reconnaître tout le texte d'une page et non pas ses*

- \*Chargez un document dans votre scanner*
- \*Sélectionnez la commande Numériser du menu Traitement.*
- \*Ajustez les réglages de numérisation*
- Réglez les options de page :*
  - Taille/A4*
- Réglez les options de reconnaissance :*
  - Colonnes/Automatiques, Zones/Automatiques*
- Réglez la Luminosité*
- \*Cliquez sur Numériser.*

Le désaccord pour l'UT 52 porte davantage sur la distinction premier plan, second plan deux juges ayant classé cette UT en premier plan (100) et les trois autres en second plan (203 illustration).

## UT 66 en gras

*Ce réglage est nécessaire pour prendre en compte les différences de qualité de papier, d'impression et les caractéristiques des scanners.*

*Il a le même rôle que pour un photocopieur.*

*Le réglage de luminosité est le meilleur moyen d'améliorer la reconnaissance de texte.*

Deux juges ont classé cette UT en «illustration», deux autres en premier plan et le dernier juge en guidage de l'action «Adjonctions Interprétatives»

Trois désaccords sont constatés pour la classe 202, «programme thématique». Les trois sont relatif à la distinction premier plan, second plan. Deux des juges ont classé ces trois UT en premier plan.

L'UT 6 a été classée par un juge en 208 : guidage de l'action «adjonction interprétatives». Les UT 5 et 8 ont été catégorisées par 3 juges en 202 «programme thématique» et par les deux autres en UT de premier plan.

## UT 5

*La présente norme définit les caractéristiques fonctionnelles des conduits de fumées composites métalliques démontables ou non ainsi que les essais destinés à les vérifier, et le mode de désignation et de marquage.*

## UT 8

*La présente norme sert de base aux normes d'application de tous les conduits de fumées composites métalliques rigides démontables ou non quels que soient le métal, le type de foyer industriel ou domestique, le type de combustible ou la nature de la construction.*

## UT 6

*Certaines caractéristiques sont exigibles pour tous les conduits de fumées quel que soit le type de foyer ou de structure d'immeuble.  
Ces caractéristiques sont définies dans des classes ;*

Les trois UT suivantes ont été classifiées en «composantes structurelles», un juge a attribué la catégorie Programme thématique aux UT 22 et 23.

## UT 46

*Celui-ci (le chapitre) est organisé de manière hiérarchique : de gauche à droite de la barre des menus et de haut en bas des commandes de chacun des menus.*

## UT 22

*Il comprend quatre sections :*

## UT 23

- Configuration nécessaire*
- Contenu du coffret*
- Installation du logiciel*
- Démarrage d'OmniPage Pro*

Une seule UT a obtenu un classement différent de chaque juge :

**UT 43 en gras**

*Chaque procédure, présentée sous forme d'organigramme, décrit un type d'opération ou de document, les réglages de numérisation appropriés, les techniques de base pour effectuer l'opération et les conseils appropriés pour tirer le meilleur parti d'OmniPage Pro.*

*Ces procédures vous donneront une idée générale de la manière d'accomplir le travail mais elles ne sont ni complètes ni définitives.*

Un juge a classifié en premier plan, un autre en 208 guidage de l'action «Adjonctions interprétatives», un troisième en 207 guidage de la lecture «Adjonctions instructionnelles», un quatrième en 202 «programme thématique» et enfin le dernier en 201 «composantes structurelles». On peut postuler que la rédaction de ce type d'UT composée d'informations disparates est à éviter. Presque chaque mot rend compte d'une intention différente.

### IV.3 Conclusion

Cette étude confirme la présence non négligeable dans les textes techniques d'unités textuelles secondaires correspondant à diverses intentions de l'auteur complémentaires de l'intention lié à l'objectif principal du texte.

Nous nous étions proposé d'organiser la distinction de ces unités textuelles de second plan en une classification qui permit de rendre compte d'un texte technique.

Nous avons vérifié que la classification que nous avons obtenue empiriquement permettait cet objectif sur un corpus de textes, certes encore relativement restreint. Dans ce cadre des juges réussissent, sans apprentissage, à appliquer la classification avec un pourcentage d'accord satisfaisant. Un apprentissage améliorerait certainement cet accord.

Si cette classification doit être encore affinée et validée grâce à un corpus plus conséquent il est d'ores et déjà possible d'envisager son utilisation dans la mise en forme matérielle des documents techniques.

On peut faire l'hypothèse qu'une rapide distinction par le lecteur des diverses unités selon les intentions de l'auteur, est un facteur de facilitation de la consultation des textes. Les procédés typo-dispositionnels, dont ceux qu'autorisent les documents électroniques, doivent permettre de supporter cette distinction. Souvent employés, ils ne le sont pas toujours de manière cohérente [Combettes,86].

Une prochaine étape de la recherche pourrait être l'étude de procédés de mise en forme matériel visant à améliorer la rédaction et la présentation des textes techniques et principalement ,l'agencement des divers types d'unités textuelles de second plan en vue de faciliter leur reconnaissance.



## BIBLIOGRAPHIE

- ADAM J.-M. (1987). Types de séquences textuelles élémentaires. *Pratiques* n°56. déc 1987. p. 54-79.
- AUSTIN J.-L. (1962). *Quand dire, c'est faire*. Ed (française) du Seuil, 1970. 1962 Ed. originale
- BRONCKART J.-P. (1985). *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*. Ed. Delachaux & Niestlé. Paris, 1985. ISBN 2.603.00539.1. 176 pages.
- CAHOUR B. (1989). *Niveau de compétence des interlocuteurs et répartition de l'initiative dans des dialogues de consultation*. Rapport de recherche n° 1124. INRIA. Rocquencourt. déc 1989. ISSN. 0249-6399. 40 pages.
- COLTIER D. (1988). Introduction et gestion des exemples dans les textes à thèse. *Pratiques* n°58. juin 1988. p.23-40.
- COMBETTES B. (1987). Types de texte et faits de langue. *Pratiques* n°56. déc 1987. p. 5-16.
- COMBETTES B. (1986). Le texte explicatif. *Pratiques* n°51. sept. 1986. p. 23-38.
- DENIS M. & DENHIERE G. (1990). - Comprehension and recall of spatial descriptions. *C. P. C., European Bulletin of Cognitive Psychology* Vol. n°2. 1990. p. 115-143.
- DUPLAN P. & JAUNEAU, R. (1986). *Maquette et mise en page*. Edition l'usine nouvelle. Paris, 1986. ISBN 2-281-31045-0. 310 pages.
- CARCIA-DEBANC C. (1989). Le tri de textes : Modes d'emploi. *Pratiques* n°62. juin. 1989.
- PASCUAL E. (1991). *Représentation de l'architecture textuelle et génération de texte*. Thèse. Institut de recherche en informatique de Toulouse. Université de Toulouse.
- ROUET, J.-F. (1991). *Compréhension de textes didactiques par des lecteurs inexpérimentés dans des situations d'interaction sujet-ordinateur*. Thèse. Université de Poitiers.
- SEARLE J.-R. (1972). *Les actes de langage..* Essai de philosophie du langage. Collection savoir. Hermann. Paris, 1972. 261 pages.
- VIRBEL J. (1992). *Formalisation d'une classe de relations structurelles de textes*. CNED 92' : Colloque National sur l'écrit et le document . Nancy, 6-7 juillet 1992. IRIT, Université Paul Sabatier Toulouse. Actes, p. 192-199.
- WEINRICH H. (1973) . *Le temps*. Editions du seuil. Paris, 1973. ISBN 2-02-002044-0. 333 pages.
- WILLEMART D. (1992). Frontispices et pages de titre sous la révolution in *Les cahiers des sciences de l'information et de la communication*.n°2. DEA. Rhône-Alpes. Univ. Grenoble 2,3 Lyon 2,3 E.N.S.S.I.B. fév. 1992. p. 87-94.

## **ANNEXES**

N° UT	Réponses des juges					Patterns de confusion	
						1er/2e plan	Pattern de confusion
UT1	211	211	211	211	211	5	5
UT4	211	211	211	211	200	5	4.1
UT2	211	211	211	209	209	5	3.2
UT3	211	211	211	204	100	4&1	3.1.1
UT67	208	208	208	208	100	4&1	4.1
UT25	208	208	207	207	200	5	2.2.1
UT37	207	207	207	202	200	5	3.1.1
UT38	207	207	202	200	100	4&1	2.1.1.1
UT44	207	207	206	205	202	5	2.1.1.1
UT43	207	208	202	201	100	4&1	1.1.1.1.1
UT7	206	206	206	206	206	5	5
UT24	206	206	206	206	100	4&1	4.1
UT39	206	206	207	202	100	4&1	2.1.1.1
UT9	205	205	205	205	204	5	4.1
UT18	205	205	205	205	204	5	4.1
UT20	205	205	205	205	204	5	4.1
UT26	205	205	205	205	204	5	4.1
UT35	205	205	205	205	204	5	4.1
UT48	205	205	205	205	204	5	4.1
UT57	205	205	205	205	204	5	4.1
UT63	205	205	205	205	204	5	4.1
UT11	205	205	205	206	100	4&1	3.1.1
UT17	205	205	205	204	100	4&1	3.1.1
UT34	205	205	205	206	100	4&1	3.1.1
UT45	205	205	205	207	206	5	3.1.1
UT13	205	205	205	203	203	5	3.2
UT15	205	205	100	100	204	3&2	2.2.1
UT28	205	205	100	100	202	3&2	2.2.1
UT32	205	205	204	202	100	4&1	2.1.1.1
UT14	203	203	203	203	203	5	5
UT47	203	203	203	203	203	5	5
UT52	203	203	203	100	100	3&2	3.2
UT66	203	203	100	100	208	3&2	2.2.1
UT36	202	202	202	202	202	5	5
UT21	202	202	202	202	201	5	4.1
UT5	202	202	202	100	100	3&2	3.2
UT8	202	202	202	100	100	3&2	3.2
UT40	202	202	202	207	100	4&1	3.1.1
UT41	202	202	202	201	100	4&1	3.1.1
UT42	202	202	202	201	100	5	3.1.1
UT6	202	202	100	100	208	3&2	2.2.1
UT22	201	201	201	201	202	5	4.1
UT46	201	201	201	207	100	4&1	3.1.1
UT23	201	201	205	204	202	5	2.1.1.1
UT10	100	100	100	100	100	5	5
UT12	100	100	100	100	100	5	5
UT16	100	100	100	100	100	5	5
UT27	100	100	100	100	100	5	5
UT29	100	100	100	100	100	5	5
UT50	100	100	100	100	100	5	5
UT51	100	100	100	100	100	5	5
UT53	100	100	100	100	100	5	5
UT59	100	100	100	100	100	5	5

Patterns de confusion

UT61	100	100	100	100	100	5	5
UT64	100	100	100	100	100	5	5
UT68	100	100	100	100	100	5	5
UT30	100	100	100	100	208	4&1	4.1
UT31	100	100	100	100	200	4&1	4.1
UT33	100	100	100	100	200	4&1	4.1
UT54	100	100	100	100	203	4&1	4.1
UT55	100	100	100	100	203	4&1	4.1
UT58	100	100	100	100	200	4&1	4.1
UT60	100	100	100	100	208	4&1	4.1
UT62	100	100	100	100	208	4&1	4.1
UT65	100	100	100	100	208	4&1	4.1
UT69	100	100	100	100	208	4&1	4.1
UT70	100	100	100	100	208	4&1	4.1
UT49	100	100	100	207	202	3&2	3.1.1
UT56	100	100	100	208	203	3&2	3.1.1



---

Unité de Recherche INRIA Rocquencourt  
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique  
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)

Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)

Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)

Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

---

EDITEUR

INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

