



HAL
open science

MADO : un langage de contrôle de données documentaires; France Logiciel No de dépôt 6-89-03-004-00

Jean-Claude Poupa

► **To cite this version:**

Jean-Claude Poupa. MADO : un langage de contrôle de données documentaires; France Logiciel No de dépôt 6-89-03-004-00. 72 p., 1988. <hal-01914392>

HAL Id: hal-01914392

<https://hal.science/hal-01914392v1>

Submitted on 8 Jun 2020

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.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

**INSTITUT NATIONAL
DE LA RECHERCHE AGRONOMIQUE**

Station d'Economie et Sociologie rurales de Rennes

MADO

Un langage de contrôle de données documentaires

Mai 1988

jean-claude poupa

INRA

Station d'Economie et Sociologie rurales
65, rue de Saint-Brieuc
35042 RENNES cedex

RESUME

MODELISER l'information documentaire, NORMALISER la représentation de cette information sous une forme indépendante de l'environnement puis concrètement IMPLANTER ce modèle sur toutes les machines d'un réseau, tels furent les objectifs d'un projet documentaire et informatique aujourd'hui opérationnel. Le résultat est une méthode et sa traduction informatique, désignées sous le nom de MADO (Module d'Analyse de Documents). Le logiciel associé est à la fois sur les stations de travail des documentalistes du Département d'Economie et Sociologie Rurales et sur le serveur de l'INRA. Il peut être porté immédiatement sur tous les systèmes d'exploitation supportant la norme ISO du langage Pascal. Le produit donne la possibilité de constituer une «réserve centrale» d'informations documentaires codées et normalisées, matière première utilisable pour le développement d'applications : bibliographies thématiques, plaquettes de présentation de publications, exportation de références vers des serveurs...

MOTS-CLES :

Sciences de l'information ;
Système d'information ;
Méthode d'analyse informatique ;
Linguistique ;
Langage documentaire ;
Ergonomie cognitive ;
Analyse de l'activité ;
Analyse de la tâche.

MADO
UN LANGAGE DE CONTROLE
DE DONNEES DOCUMENTAIRES

SOMMAIRE

Introduction	7
1. EXPERTISER LE BESOIN DOCUMENTAIRE ET CADRER UNE SOLUTION	9
1.1. Domaine et définitions	10
1.2. Le choix des variables.....	11
Encadré n° 1 :	12
Les variables du projet ESR	
1.3. L'égalité de deux variables documentaires.....	13
1.4. L'approche informatique traditionnelle.....	14
1.5. L'approche linguistique.....	15
1.6. La mise en oeuvre de l'approche linguistique.....	16
2. MODELISER L'INFORMATION DOCUMENTAIRE	17
2.1. Description syntaxique générale :	18
un premier niveau de structuration	
2.1.1. Exemple	18
2.1.2. Définition du langage à partir d'une grammaire formelle	18
a. Alphabet reconnu par MADO.....	19
b. Structure syntaxique générale.....	19
d'une référence documentaire	
2.1.3. Les contrôles globaux dans la phase	20
de reconnaissance syntaxique	
a. Le paramétrage des noms de drapeau et de variables.....	20
b. Les restrictions	21
2.2. Le second niveau de structuration : le type syntaxique.....	22
2.2.1. Exemples	22
2.2.2. Choix des types syntaxiques	22
a. Calcul de la fonction d'état	22
b. Les types de propriétés contrôlées.....	23
2.3. La construction syntaxique des variables.....	25
Encadré n° 2 :	26
Description syntaxique des variables de l'application ESR	
2.4. Les propriétés intrinsèques des variables.....	27
2.5. Les propriétés relatives aux références.....	28
2.6. Les contrôles d'intégrité de référence paramétrables.....	30
2.7. Les contrôles d'intégrité de référence non paramétrables.....	32
2.8. Les contrôles d'intégrité de notice.....	33

3. NORMALISER LA REPRESENTATION DE L'INFORMATION DOCUMENTAIRE.....	35
3.1. Le prétraitement de texte.....	36
3.2. Le prétraitement d'une référence.....	37
3.3. Les conventions dactylographiques.....	38
3.4. Le codage des laboratoires.....	39
3.5. La génération de textes prédéfinis.....	40
3.6. Les conversions des lettres sous la forme majuscule.....	41
3.7. Les conventions spécifiques.....	42
3.8. La gestion des variables communes.....	43
3.9. La syntaxe de la forme normale.....	43
Encadré n° 3 :.....	45
Description syntaxique de la forme normalisée des variables de l'application ESR	
4. PROPOSER UN PRODUIT INFORMATIQUE ADAPTABLE AUX CONTEXTES DOCUMENTAIRES.....	47
4.1. La traduction informatique de MADO.....	48
4.2. L'utilisation de MADO.....	49
4.3. Le paramétrage de MADO par le relais du document de paramètres.....	50
4.3.1. Les noms des variables et la matrice de contrôle.....	50
4.3.2. La liste des organismes.....	50
4.3.3. Les messages d'erreur.....	51
4.3.4. Les codes pays.....	51
4.3.5. Les codes langues.....	51
4.3.6. Le premier plan de classement : RESAGRI.....	51
4.3.7. Le second plan de classement : AGRIS.....	51
4.4. Les paramètres initialisés par programme.....	52
Tableau 1	
4.5. Le module de saisie-correction.....	53
4.6. Les contrôles d'appartenance des mots-clé aux lexiques.....	54
4.7. Le prototype de développement associé AIDA.....	55
CONCLUSION	57
ANNEXE 1 : Le document de paramètres MADO.....	59

INTRODUCTION

Les sciences de la documentation sont depuis longtemps confrontées aux problèmes de gestion de l'information et de communication, terrain sur lequel se sont développées les technologies informatiques. Les progrès les plus récents offrent des moyens d'archivage à la hauteur des besoins documentaires, des réseaux de communication nationaux et internationaux qui rendent immédiatement repérable et accessible tout document, des techniques d'édition qui répondent à des exigences professionnelles diversifiées.

Paradoxalement, l'informatique documentaire telle qu'elle est utilisée pour les besoins quotidiens des équipes scientifiques, reste souvent artisanale : les fichiers locaux des laboratoires ne communiquent pas toujours avec les bases de données centrales et un même document peut être décrit plusieurs fois pour des applications non communicantes. Pratiquement, le mode de représentation logique des données peut dépendre simultanément du matériel, du logiciel et de l'application.

MODELISER l'information documentaire, NORMALISER la représentation de cette information sous une forme indépendante des équipements informatiques, puis concrètement IMPLANTER ce modèle sur toutes les machines d'un réseau, tels furent les objectifs d'un projet aujourd'hui opérationnel. Le résultat est une méthode et sa traduction informatique pour une application. Le logiciel associé, MADO (Module d'Analyse de DOcuments), est utilisé à la fois sur les micro-ordinateurs des documentalistes du Département d'Economie et Sociologie rurales et sur le serveur de l'INRA. Respectant la norme ISO du langage Pascal, il est portable sur tout système d'exploitation.

La version actuelle de MADO communique avec le logiciel documentaire TEXTO, utilisé en amont pour saisir les références puis les corriger, et, en aval, pour administrer ces références contrôlées et normalisées (1). Un second projet informatique se donne comme objectif de fournir aux documentalistes un environnement de travail plus convivial sur micro-ordinateur. Le produit résultant, EDIBASE, prend en charge "l'intendance informatique" et propose un outil spécialisé dans la saisie-corrrection de références documentaires.

(2)

(1) M.A. Farget. Proposition pour un manuel utilisateurs texto-mado, INRA, Grenoble. Décembre 1987.

(2) M. Meunier ; P. Richard
EDIBASE, Saisie de références documentaires normalisées. Manuel d'utilisation. Version 1.00. INRA, Montpellier, Mars 1988

Le résultat actuel pour le Département d'Economie et Sociologie rurales est la disponibilité d'une "réserve centrale" d'informations documentaires contrôlées et normalisées, matière première utilisable pour le développement d'applications : bibliographies thématiques, plaquettes de présentation de publications, exportation de références vers des serveurs...

1. EXPERTISER LE BESOIN DOCUMENTAIRE ET CADRER UNE SOLUTION

C'est dans le cadre du projet documentaire du Département d'Economie et de Sociologies Rurales, animé par M.A. Farget (1), que s'est progressivement exprimée une demande informatique, autour de l'objectif initial dit de "transparence documentaire", visant à disposer au niveau du Département d'une base de données commune regroupant les fichiers locaux des services de documentation des stations. Dans le contexte technologique de démarrage de ce projet (1983), il fut décidé d'intégrer cette base commune dans la banque de données RESAGRI. L'application de cette décision a nécessité de décentraliser les fonctions de saisie-contrôle des données documentaires au niveau des laboratoires, depuis des terminaux reliés au serveur central de l'INRA, avec des outils logiciels développés dans le contexte du traitement par lots.

Le fait d'utiliser une chaîne de contrôle non conçue pour être mise en oeuvre en libre-service, combiné à des aléas de télé-transmission imputables au réseau commuté, risquait de décourager totalement une population motivée de documentalistes. Le projet informatique décrit dans ce rapport (dit "projet ESR") est né dans ce contexte. A l'objectif initial de transparence documentaire s'est ajouté celui de la valorisation des publications du Département.

(1) M.A. Farget, le projet documentaire du Département d'Economie et Sociologie rurales : une étape et des résultats. INRA, Grenoble, Avril 1988

1.1. DOMAINE ET DEFINITIONS

Les informations recherchées sont extraites de documents, généralement sous la forme écrite : articles, ouvrage, actes de congrès, rapports, annuaires... Mais de plus en plus d'autres formes sur d'autres supports sont proposées : base de données et bases de connaissances, images, films, documents sonores... Le terme "document" sera donc utilisé dans sa signification la plus étendue.

L'objectif est de caractériser ces documents pour en évaluer le contenu et les localiser. Le domaine traité se restreint à la description, sous forme d'un texte structuré, d'entités documentaires définies à plusieurs niveaux : articles d'une revue et revue ; ouvrage et collection ; chapitre d'une monographie et monographie ; base de données et chapitres de la base... Les descriptions élémentaires sont dites "références documentaires". Les références reliées entre elles constituent une "notice bibliographique".

1.2. LE CHOIX DES VARIABLES

Une étape préliminaire consiste à définir sans ambiguïté la nature des informations recherchées par l'utilisateur, dans un langage commun aux documentalistes et à ces usagers. La méthode cartésienne est de rigueur : il faut identifier clairement les informations retenues, les décomposer en leurs parties constitutives, isoler chaque entité puis classer ces entités. Au terme d'un processus de décomposition et de synthèse, piloté par les fonctionnalités recherchées, l'analyse aboutit à proposer un ensemble structuré de variables, chaque variable étant dûment définie quant au contenu documentaire. La sélection de ces variables est une étape fondamentale, qui délimite d'emblée le champ des applications réalisables

Les choix effectués pour cette application s'appuient sur les acquis des services de documentation de l'INRA, et sur les résultats des travaux des documentalistes du Département d'Economie et Sociologie rurales. La liste ci-dessous (voir encadre numéro 1) fournit les noms choisis dans le cadre de cette application et précise la signification documentaire des variables.

ENCADRE NUMERO 1
LES VARIABLES DU PROJET ESR

ND	: numéro de document ou clé d'accès
TD	: type de document
NI	: niveau d'indexation
IN	: indicateur bibliographique
LO	: localisation du document ou cote
SU	: support du document
DA	: date de publication
AU	: liste nominative des auteurs du document
AF	: organisme d'affiliation du premier auteur ou des auteurs
A1	: organisme coresponsable de la production du document
A2	: organisme coresponsable de la production du document
A3	: organisme coresponsable de la production du document
DI	: nature du diplôme pour un document universitaire
TI	: titre principal du document
LI	: complément du titre principal
TF	: titre traduit en français
TG	: titre traduit en anglais
AG	: adresse d'un congrès
DG	: date d'un congrès
NE	: numéro d'édition
IB	: international standard book number
ED	: nom de l'éditeur
AE	: adresse de l'éditeur
PG1	: pagination
PG2	: complément sur la pagination
PG3	: numéro de volume
PG4	: numéro de chapitre
LA	: liste des codes langues du texte
LR	: liste des codes langues des résumés
IL	: nombre d'illustrations
RB	: nombre de références bibliographiques
CA	: cartes
TB	: nombre de tables
CI	: public visé ou cible
AD	: adresse de l'organisme assurant la diffusion du document
R1	: plan de classement RESAGRI
R2	: plan de classement AGRIS
MC1	: mots clés principaux du lexique RESADEC
MC2	: mots clés secondaires du lexique RESADEC
MC3	: mots clés géographiques
MC4	: dates
MC5	: noms cités
MC6	: candidats descripteurs RESAGRI
MC7	: mots clés principaux du lexique TECAGRI
MC8	: mots clés secondaires du lexique TECAGRI
MC9	: mots clés principaux du lexique AGROVOC
MC10	: mots clés secondaires du lexique AGROVOC
MC11	: indexation locale
MC12	: indexation locale
RS	: résumé
SO	: titre d'une publication en série
CP	: code pays pour une publication en série
IS	: international standard serial number
VOL	: numéro de volume d'une publication en série
NUM	: numéro d'une publication en série
NS	: titre d'un numéro spécial d'une publication en série
L1	: variable locale
L2	: variable locale
L3	: variable locale
L4	: variable locale
L5	: variable locale
L6	: variable locale
L7	: variable locale
L8	: variable locale
L9	: variable locale
L10	: plan de classement Département
AUG	: liste nominative de responsables du niveau collectif d'un document
AFG	: organisme d'affiliation du premier responsable de niveau collectif
A1G	: organisme coresponsable du niveau collectif d'un document
A2G	: organisme coresponsable du niveau collectif d'un document
A3G	: organisme coresponsable du niveau collectif d'un document
TIG	: titre du niveau collectif d'un document
PGG	: pagination du niveau collectif d'un document

1.3. L'EGALITE DE DEUX VARIABLES DOCUMENTAIRES

Dans les bases de données documentaires, les variables sont des chaînes de caractères ayant une signification précise pour les documentalistes : identification de l'auteur, d'un laboratoire, titre, indexation, résumé... Pour une machine, ces chaînes sont des combinaisons aléatoires des caractères de l'alphabet. D'un point de vue strictement mathématique nous énoncerons le fait fondamental suivant :

“deux chaînes de caractères sont égales si et seulement si elles contiennent le même nombre de caractères et les caractères de même rang sont égaux”.

Compte tenu des degrés de liberté en matière d'écriture (majuscules ou minuscules, signes diacritiques des alphabets latins, orthographe, grammaire...), des données ayant des contenus sémantiques identiques du point de vue documentaire sont différentes pour une machine qui teste l'égalité mathématique.

Il en résulte des difficultés évidentes, qui se traduisent concrètement par un silence à l'interrogation. Pour y remédier sont proposées des procédures de manipulation de chaînes dont la richesse et la souplesse sont une qualité essentielle des systèmes de gestion de bases de données documentaires. En l'absence de contrôles rigoureux en amont, il faut imaginer une multitude de formes syntaxiques probables, voire imaginer la nature des erreurs possibles. Dans un tel contexte, conduire une recherche documentaire exige un savoir-faire spécifique.

1.4. L'APPROCHE INFORMATIQUE TRADITIONNELLE

Pour résoudre ce problème ont été établies des règles strictes, consignées dans des manuels de catalogages. Dans le cadre d'une informatique scientifique centralisée, des bordereaux de codage étaient proposés. La codification d'une référence documentaire restait alors une affaire de spécialistes : les contrôles et le formatage étaient effectués par des personnels spécialisés, puis les bordereaux transmis vers un atelier de saisie, la machine n'intervenant qu'en aval pour refuser ou valider le contenu de chaque bordereau. Tout bordereau valide était ensuite transmis vers les bases de données.

Au fur et à mesure que se développaient de telles applications, les usagers demandaient des informations plus précises et plus complètes. Cela se traduisait par l'ajout de variables ou d'informations dans certaines variables, de conventions nouvelles, de listes d'exception, etc... Les bordereaux sont devenus d'une complexité déroutante, et leur réalisation l'affaire d'un personnel hautement qualifié et expérimenté. Les améliorations liées à l'évolution des techniques déplaçaient le problème sans le résoudre.

Ces expériences, parfois difficiles, ont été à l'origine de situations de rejets tant du côté des usagers que des documentalistes. Elles ont induit une vision de l'informatique et de ses contraintes souvent négative.

1.5. L'APPROCHE LINGUISTIQUE

L'approche linguistique s'appuie sur les acquis des méthodes dites de linguistique formelle, nées des théories de Chomsky. Le principe est de définir la syntaxe des variables, et plus globalement de la référence documentaire, à l'aide d'une grammaire formelle. De cette modélisation mathématique se déduisent, à l'aide de théorèmes, des algorithmes qui pourront être traduits dans des langages informatiques.

Ces méthodes se sont développées à partir des années 60 et sont traditionnellement utilisées pour construire les compilateurs. Le rôle de l'analyste est de trouver des formes syntaxiques qui englobent les contraintes effectives répertoriées, puis de les réduire à une forme normale. L'ensemble de ces formes, dites variables linguistiques ou alphabet non terminal, va constituer le langage. La forme normale servira de base pour représenter symboliquement les objets documentaires élémentaires.

Construire ce langage implique un long travail d'observation du contenu des tâches, pour proposer une modélisation du domaine à informatiser, qui devra simultanément prendre en compte le métier des documentalistes et s'abstraire du contexte technologique. L'objectif est en fin de compte de sous-traiter à la machine un maximum d'opérations automatisables, de telle sorte que les documentalistes puissent "faire de la documentation".

1.6. LA MISE EN OEUVRE DE L'APPROCHE LINGUISTIQUE

Les analyses conduites sous la responsabilité du service documentaire central de l'INRA ont fourni le matériau de base, à partir des analyses et des expériences relatives à la participation de l'INRA à la base nationale RESAGRI, ainsi qu'à la base internationale AGRIS. La collaboration active de documentalistes de stations a permis de déboucher sur une première structure générale, proposée en assemblée générale et validée. Sa traduction informatique a fait l'objet d'une première note de travail fin 1985, avec des propositions précises quant aux choix des formes syntaxiques en regard des variables retenues.

Ce premier schéma s'est concrétisé dans une version complète, laquelle a fait l'objet d'une formalisation théorique (avril 1986) puis d'une expérimentation dans les laboratoires (juin 1986) (1). L'existence d'une forte dynamique de groupe dans le cadre d'une organisation structurée, combinée avec la disponibilité d'un outil opérationnel, a débouché sur une évaluation et des propositions. Le résultat est, au terme d'un long processus interactif de dialogue entre informatique et documentation, un produit adapté dont les fonctionnalités dépassent largement les demandes exprimées dans les cahiers des charges initiaux.

(1) M.A. Farget ; J.C. Poupa. Le projet documentaire du Département, situation au 1.9.86
INRA, Grenoble, 1986.

2. MODELISER L'INFORMATION DOCUMENTAIRE

Un modèle de données est une construction mathématique sur la base d'énoncés qui, progressivement, limitent le champ de définition de la structure. Cette construction s'appuie au départ sur une structure algébrique très générale, le monoïde, pour s'acheminer vers un langage formel de plus en plus contraint : structuration générale du texte en références et variables, définition de types syntaxiques puis structuration des variables... Des propriétés sémantiques, exprimées sous forme de prédicats, sont affectées aux types, puis aux variables construites avec ces types, puis aux références à travers les relations entre variables. D'autres propriétés sont déduites de la lecture d'un document de paramètres. Enfin sont définies des relations entre références.

L'objectif de ce chapitre est de présenter une méthode, appliquée sur un exemple réel, qui aboutit à une modélisation des données documentaires. Il faudra ensuite examiner si les références documentaires du monde réel appartiennent ou pas au modèle, c'est-à-dire satisfont la totalité des règles répertoriées.

Bien que cette vérification puisse s'effectuer au terme d'un long raisonnement strictement mathématique, l'intérêt de la méthode est de déboucher sur des algorithmes et une traduction informatique opérationnelle.

2.1. DESCRIPTION SYNTAXIQUE GENERALE : UN PREMIER NIVEAU DE STRUCTURATION

2.1.1. Exemple

Pour illustrer l'architecture globale, une référence documentaire ressemble à la page d'un cahier d'écolier. La marge gauche contient éventuellement des annotations : les noms de variables. La partie droite contient le texte : le contenu des variables. La première ligne de la page contient impérativement un nom ou numéro de page : le nom ou le numéro de la référence. Les autres lignes contiennent les informations documentaires relatives aux variables désignées dans la marge.

Au niveau de l'algorithme de reconnaissance, la marge est matérialisée par le caractère "deux-points". La première ligne est une variable spécifique annonçant une nouvelle référence, dite variable "drapeau". La référence est lue dans son intégralité, pratiquement jusqu'à rencontre d'un autre drapeau ou d'une marque de fin. Le nom du drapeau et les noms de variables sont des données externes, propres à l'application et gérées par l'administrateur du projet documentaire.

L'exemple ci-dessus est une référence documentaire, au sens de cette définition. Le drapeau se nomme ici "référence" et les trois variables "Titre", "Date" et "Index".

Référence	: 1
Titre	: Popi
Date	: Juillet 87
Index	: Léo
	: Popi
	: Petit ours brun

Cette syntaxe très générale est en fait un format d'édition fréquemment utilisé par les logiciels documentaires.

Il est possible de construire des grammaires formelles qui génèrent de telles formes, et d'en déduire les algorithmes de reconnaissance. Une grammaire formelle se définit sur un alphabet, pratiquement les symboles graphiques des langues écrites (lettres majuscules, minuscules, caractères accentués, chiffres arabes, ponctuations et divers), à l'aide de règles dites de production.

2.1.2. Définition du langage à partir d'une grammaire formelle

Du point de vue mathématique un langage est une structure algébrique. Plus précisément, c'est une partie d'un monoïde non commutatif défini sur un ensemble A de symboles, dit "alphabet". Les éléments de cet alphabet sont les symboles graphiques usuels des langues occidentales. La loi de composition interne est l'opérateur dit de "concaténation", qui à partir des éléments de l'alphabet génère des "mots". L'élément neutre de ce monoïde est dit "mot vide".

Une grammaire formelle est un système de génération de langage, à partir d'un alphabet et d'un ensemble de règles construites avec l'opérateur de concaténation, et dites "règles de production". La description de l'alphabet se restreint à l'énumération des caractères retenus. La description des règles de production utilise une méthode standardisée de définition de la syntaxe des langages de programmation, dite "forme de Backus et Naur". Ce cadre conceptuel permet de décrire syntaxiquement des ensembles de références documentaires, et à l'intérieur de ces références, de décrire des formes syntaxiques pour chacune des variables constitutives.

a) *l'alphabet reconnu par MADO*

Les caractères autorisés sont regroupés ci-dessous en sous-ensembles. Dans les énumérations, les caractères sont placés entre accolades et séparés par des blancs. Les accolades n'appartiennent pas à l'alphabet mado.

```

MAJUSCULE = {A B C D E F G H I J K L M N O P Q R S T U V W X Y Z}
MINUSCULE = {a b c d e f g h i j k l m n o p q r s t u v w x y z , à â ä è é ê ë ì
              ï ð ò ù û ü ç}
CHIFFRE    = {0 1 2 3 4 5 6 7 8 9}
PONCTUATION = { , . ; : ? ! ( ) [ ] }
SPECIAL    = { + - * / ' " % & }
MARQUE DE CONTINUITE = { _ }

```

Les caractères "BLANC", "MARQUE DE FIN DE LIGNE", "MARQUE DE FIN DE FICHER" sont des caractères non imprimables appartenant à l'alphabet. Tous les autres caractères sont ignorés.

b). *Structure syntaxique générale d'une référence documentaire*

Les conventions adoptées pour décrire ce langage sont les suivantes :

- le symbole "::<=" signifie "est défini par"
- les chevrons ouvrant "<" et fermant ">" encadrent le nom d'un composant syntaxique ;
- les crochets ouvrant "[" et fermant "]" encadrent un composant facultatif qui ne peut pas être répété ;
- les accolades ouvrantes "{" et fermantes "}" encadrent un composant facultatif qui peut être répété plusieurs fois ;
- le symbole "|" signifie "ou bien".

Le composant syntaxique <caractère> désigne un caractère de l'alphabet différent des marques de fin de ligne, de fin de fichier, de continuité. Les règles de production exprimées sous la forme de Backus et Naur sont regroupées ci-dessous.

<référence>	::= <référence>{<référence>}<fin de fichier>
<référence>	::= <drapeau>{<variable>}
<drapeau>	::= <nom de drapeau><séparateur><ligne>
<nom de drapeau>	::= <caractère>{<caractère>}
<séparateur>	::= :
<ligne>	::= {<caractère>}[<marque de continuité>]<fin de ligne>
<variable>	::= {<nom de variable>{<séparateur><texte>}}
<texte>	::= <ligne>{<marge><séparateur><ligne>}
<marge>	::= {<blanc>}

MADO reconnaît actuellement, compte-tenu d'autres contraintes syntaxiques et sémantiques prises en compte, un sous-ensemble du langage généré par cette grammaire.

2.1.3. Les contrôles globaux dans la phase de reconnaissance syntaxique

La grammaire générale qui structure le texte soumis à analyse autorise une relative liberté syntaxique. L'intérêt de cette grammaire est de définir un format d'entrée aussi souple que possible, de telle sorte que puissent être reconnues des références en provenance de systèmes quelconques, orientés documentaires ou bureautiques. Mais l'objet de cette construction formelle est avant tout pratique. Les algorithmes associés vont donc restreindre d'emblée le champ théorique de définition sous forme de règles générales qui restructurent la référence avant de la soumettre aux procédures de contrôles.

C'est dans cette phase que sont localisées au niveau informatique un certain nombre d'erreurs dites fatales dans la mesure où le traitement de la référence est abandonné.

a) Le paramétrage des noms de drapeau et de variables

Les noms de drapeau et de variables sont ceux choisis dans l'application documentaire. Ils sont lus dans un document de paramètres, géré par l'administrateur du projet documentaire, et convertis sous la forme majuscule. L'algorithme d'analyse syntaxique va en outre vérifier l'appartenance des noms effectivement reconnus. Ce contrôle d'appartenance s'exprime à travers les règles fondamentales suivantes :

F1 : le nom du drapeau, converti sous la forme majuscule, est toujours égal à celui lu dans le document de paramètres.

F2 : Tout nom de variable, converti sous la forme majuscule, appartient à l'ensemble des noms de variables lus dans le document de paramètres.

Dans la pratique l'occurrence de ces erreurs est rare et survient lors de premières expérimentations ; lecture d'un texte quelconque qui a priori ne commence pas par un nom de drapeau ; choix de noms de variables non autorisés ; absence de marge ; lecture de code exécutable, etc.

b) Les restrictions

La définition syntaxique de plusieurs composants autorise des combinaisons quelconques de caractères, la longueur de la chaîne résultante n'étant pas bornée. Une telle définition n'a souvent qu'un intérêt théorique.

Dans la pratique, il a été convenu de restreindre la longueur d'une ligne physique à LL caractères (LL = 80), et la longueur du nom de variable à LNTX caractères (LNTX = 4), l'optique "saisie au kilomètre" s'avérant finalement sans intérêt. Ce choix a conduit à introduire dans les définitions syntaxiques le composant <marque de continuité>, pour signifier qu'il y a coupure du texte en fin de ligne du fait de cette convention. Une ligne physique est donc constituée d'une marge, d'un délimiteur de marge, d'une zone utile contenant un texte et éventuellement d'une marque de continuité. La longueur du texte est donc limitée à LLMADO caractères, avec $LLMADO = LL - LNTX - 2$, valeur fréquemment utilisée dans les contrôles sémantiques associés aux variables.

La définition syntaxique d'une référence n'introduit aucune limite finie. Le choix retenu dans cette version a consisté à se limiter à un maximum de LR lignes (LR = 100).

Ces choix se traduisent dans les règles fondamentales suivantes :

F3 : Le composant syntaxique <ligne> contient au plus LLMADO éléments de type <caractère>.

F4 : Le composant syntaxique <référence> contient au plus LR éléments de type <ligne>.

2.2. LE SECOND NIVEAU DE STRUCTURATION : LE TYPE SYNTAXIQUE

2.2.1 Exemple

Avec les seules règles associées à la grammaire de définition précédente, le contenu d'une variable est un texte libre. Pour les besoins d'une description documentaire sont définies d'autres contraintes : une variable d'indexation est une liste de mots-clés séparés par des points-virgules ; un mot-clé est une chaîne de caractères alpha-numériques de longueur bornée ; une date est une chaîne ayant la forme "AAA/MM/JJ", AAAA désignant l'année, MM le mois, JJ le jour ; etc... Les définitions précises de ces structures construites sur un texte libre constituent les "types syntaxiques", ou "variables syntaxiques", qui seront dits "type date", "type mot-clé"... Avec ces conventions, l'exemple précédent s'écrit désormais :

Référence	: 1
Titre	: Popi
Date	: 1987/07
Index	: Léo ; Popi ; Petit ours brun

2.2.2. Choix des types syntaxiques

La sélection de structures syntaxiques adaptées s'effectue sur la base d'une observation rigoureuse des pratiques documentaires, qu'elles soient codifiées ou qu'elles traduisent un savoir-faire associé au métier. Elle nécessite une collaboration étroite avec des "experts documentalistes", dont le rôle est simultanément de traduire le besoin effectif et de vérifier que la formalisation proposée est l'expression rigoureuse de ce besoin. Cette formalisation est soumise à une contrainte forte de non ambiguïté syntaxique, toutes les règles devant dans les faits s'interpréter de façon claire et unique.

Les types retenus pour la version MAD0 du projet documentaire du Secteur des Sciences Sociales sont regroupés en quatre catégories :

- les types élémentaires, qui se dérivent en un élément de l'alphabet,
- les types de base qui se dérivent en une suite de types élémentaires,
- les types spécifiques construits avec les types de base,
- les types listes.

a) Les types élémentaires

Ils traduisent en fait l'appartenance d'un caractère à une partie de l'alphabet. Le formalisme utilisé définit ces types à partir des règles de production. Par exemple, le type chiffre est produit par la règle:

<chiffre> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

Les types élémentaires pratiquement utilisés sont répertoriés ci-dessous, sans que soient fournies les règles de production qui se déduisent immédiatement des commentaires associés.

<caractère>	est un caractère quelconque de l'alphabet ;
<majuscule>	est une lettre majuscule ;
<minuscule>	est une lettre minuscule, voyelles accentuées et ç inclus ;
<lettre>	est une lettre majuscule ou minuscule ;
<chiffre>	est un chiffre arabe ;
<alphanumérique>	est une lettre ou un chiffre ;
<ponctuation>	est un caractère de ponctuation, parenthèses et crochets inclus ;
<caractère d'indexation>	est une lettre, un tiret, une apostrophe, ou un blanc
< caractère de pagination>	est un chiffre, un point ou la lettre "p" ;
<joker>	désigne les caractères "*" ou "&" ;
<vide>	désigne l'élément neutre du monoïde, ou mot vide

b) Les types de base

Ils sont définis à partir des règles de production suivantes :

<texte>	::= {<caractère>}
<nombre>	::= {<chiffre>}
<mot clé>	::= {<caractère d'indexation>}
<pagination>	::= {<caractère de pagination>}
<code pays>	::= <lettre><lettre><lettre>
<code langue>	::= <lettre><lettre><lettre>
<plan de classement>	::= <lettre>[<lettre>][>lettre>][<lettre>]
<initiales>	::= <majuscule>.<[majuscule]>.[<majuscule>.]
<sigle>	::= {<majuscule>}

On remarquera que le type <texte>, qui introduit un maximum de degrés de liberté, contient tous les autres types.

c) Les types spécifiques

Les constructions adoptées sont actuellement les suivantes :

<adresse>	::= <texte> (<code pays>)
<organisme>	::= [<sigle>;]<texte>;[<texte>]<adresse>
<date>	::= <nombre> [/ <nombre>] [/ <nombre>]
<auteur>	::= <texte> [, <initiales>] [(<texte>)]
<cote>	::= <texte> - <texte> I <nombre> - <texte>

d) Les types <listes>

Les règles ci-dessous illustrent la notion de liste telle qu'elle est pratiquement utilisée dans les applications documentaires.

<indexation>	<mot clé>{;<mot clé>}
<dates>	<date>{;<date>}
<langues>	<code langue>{;<code langue>}
<plans>	<plan de classement>{;<plan de classement>}
<cotes>	<cote>{;<cote>}

Ces constructions peuvent s'explicitier à l'aide d'une règle plus générale, dite "métarègle", ayant la forme suivante :

$$\langle \text{liste} \rangle ::= \langle \text{élément} \rangle [; \langle \text{élément} \rangle]$$

Cette "métarègle" signifie que la métavariable syntaxique <élément> peut prendre les valeurs <mot clé>, <date>, <code langue>, <plan de classement> ou <cote> pour générer des types dits <liste>.

2.3. LA CONSTRUCTION SYNTAXIQUE DES VARIABLES

Les variables contiennent des données documentaires, qu'il s'agit de structurer à l'aide des types syntaxiques, souvent appelés composants, précédemment définis.

Le choix de ces règles de construction est une étape fondamentale, qui détermine les degrés de liberté pour chacune des variables résultantes : toutes les constructions pourraient être de type <texte> et cela reviendrait à ne rien contrôler. A l'opposé une forme syntaxique trop contraignante peut compliquer démesurément la tâche.

Dans la pratique, la méthode retenue a consisté à prendre pour cible le format normalisé exigé par la base de données internationale AGRIS, puis à rechercher les solutions qui accordent une liberté syntaxique maximale dans le cadre des contraintes imposées. A l'issue de plusieurs itérations, les règles de construction se sont stabilisées. Elles sont traduites dans la fiche de description syntaxique des variables (encadré numéro 2)

ENCADRE N° 2

DESCRIPTION SYNTAXIQUE DES VARIABLES DE L'APPLICATION ESR

ND	: <nombre>{<nombre>
TD	: <caractère>
NI	: <caractère>
IN	: <caractère>{<caractère>}
LO	: <cote>{;<cote>}
SU	: <caractère>
DA	: <date>{;<date>}
AU	: <auteur>{;<auteur>} Anonymel<joker>
AF	: <nombre> <organisme>
A1	: <nombre> <organisme>
A2	: <nombre> <organisme>
A3	: <nombre> <organisme>
DI	: <texte>
TI	: <texte>
LI	: <texte>
TF	: <texte>
TA	: <texte>
TG	: <texte>
AG	: <adresse>
DG	: <date>{;<date>}
NE	: <nombre>
IB	: <nombre>-<nombre>-<nombre>-<alphanumérique>
ED	: <texte>
AE	: <adresse>
PG1	: <pagination> <joker>IN<texte>In<texte>
PG2	: <nombre>
PG3	: <nombre>
PG4	: <nombre>
LA	: <code langue>{;<code langue>}
LR	: <code langue>{;<code langue>}
IL	: <nombre> <joker>
RB	: <nombre> <joker>
CA	: <texte>
TB	: <nombre> <joker>
CI	: <caractère>
AD	: <adresse> <nombre>
R1	: <plan de classement>{;<plan de classement>}
R2	: <plan de classement>{;<plan de classement>}
MC1	: <mot clé>{;<mot clé>}
MC2	: <mot clé>{;<mot clé>}
MC3	: <mot clé>{;<mot clé>}
MC4	: <mot clé>{;<mot clé>}
MC5	: <mot clé>{;<mot clé>}
MC6	: <mot clé>{;<mot clé>}
MC7	: <mot clé>{;<mot clé>}
MC8	: <mot clé>{;<mot clé>}
MC9	: [*]<mot clé>{;[?]<mot clé>}
MC10	: <mot clé>{;<mot clé>}
MC11	: <mot clé>{;<mot clé>}
MC12	: <mot clé>{;<mot clé>}
RS	: <texte> <joker>
SO	: <texte>
CP	: <code pays> (<code pays>)
IS	: <nombre>-<nombre>-<alphanumérique>
VOL	: <nombre>
NUM	: <nombre>
NS	: <texte>
L1	: <texte>
L2	: <texte>
L3	: <texte>
L4	: <texte>
L5	: <texte>
L6	: <texte>
L7	: <texte>
L8	: <texte>
L9	: <texte>
L10	: <nombre>
AUG	: <auteur>{;<auteur>} Anonymel<joker>
AFG	: <nombre> <organisme>
A1G	: <nombre> <organisme>
A2G	: <nombre> <organisme>
A3G	: <nombre> <organisme>
TIG	: <texte>
PGG	: <pagination> <joker>IN<texte>In<texte>

2.4. LES PROPRIETES INTRINSEQUES DES VARIABLES

Les définitions syntaxiques ne donnent que des règles de construction, indépendamment du contenu sémantique de l'information documentaire. Certaines conventions sémantiques peuvent se traduire aisément sous forme de prédicats mathématiques : appartenance à un ensemble, proposition conditionnelle, relations arithmétiques et/ou booléennes. D'autres s'avèrent plus difficiles à traduire, ou requièrent des ressources informatiques plus importantes pour être évaluées. D'autres enfin, ne sont pas formalisées sous forme d'algorithmes et leur application reste sous l'entière responsabilité des documentalistes.

Les propriétés sémantiques des variables sont décrites ci-dessous. Cette liste n'est pas close et peut évoluer au fur et à mesure que sont formalisés de nouveaux contrôles. Les règles retenues sont affectées d'un numéro d'ordre, précédé de la lettre V, pour rappeler qu'elles sont des attributs de Variables, même si définies sur des composants syntaxiques communs à plusieurs variables. La numérotation suit l'ordre des variables. Les valeurs initiales paramétrées sont fournies dans le tableau 1 du paragraphe 4.4.

- V1 : Les deux premiers chiffres contenus dans ND forment un nombre appartenant à l'ensemble CD des codes documentalistes.
- V2 : La forme majuscule du caractère contenu dans TD appartient à l'ensemble CODETD des codes "type de document".
- V3 : La forme majuscule du caractère contenu dans NI appartient à l'ensemble CODENI des codes "niveau d'indexation".
- V4 : La liste contenue dans LO est limitée à LLI1 éléments de type <cote>.
- V5 : Tout élément de type <cote> est une chaîne d'au plus LCOTE caractères.
- V6 : Si un élément de type <cote> revêt la forme <nombre>-<texte>, alors ce nombre appartient à l'ensemble NPO des numéros prédéfinis d'organismes.
- V7 : Si un élément de type <cote> revêt la forme <texte>-<texte>, le premier composant <texte> est un préfixe de cote appartenant à la liste des cotes d'organismes désignés par un numéro prédéfini.
- V8 : La forme majuscule du caractère contenu dans SU appartient à l'ensemble SUPP des codes "support de document".
- V9 : La liste contenue dans DA est limitée à LLI3 éléments de type <date>.
- V10 : Le nombre associé au premier composant optionnel d'un élément de type <date> est obligatoirement formé de 4 chiffres.
- V11 : Le nombre associé au second composant optionnel d'un élément de type <date> est obligatoirement formé de 2 chiffres.
- V12 : Le nombre associé au troisième composant optionnel d'un élément de type <date> est obligatoirement formé de 2 chiffres.
- V13 : La liste contenue dans DG est limitée à LLI4 éléments de type <date>.
- V14 : La liste contenue dans AU est limitée à LLI2 éléments de type <auteur>.
- V15 : La liste contenue dans AUG est limitée à LLI2 éléments de type <auteur>.
- V16 : Tout élément de type <auteur> est une chaîne d'au plus LLMADO caractères.
- V17 : Tout élément de type <auteur> contient au moins une majuscule.
- V18 : Tout élément de type <auteur> contient au moins une minuscule.
- V19 : Les caractères de l'ensemble CINA des "caractères interdits dans un nom d'auteur" ne peuvent pas être présents dans le premier composant <texte> d'un élément de type <auteur>.
- V20 : Si le contenu des variables dites d'affiliation (AF, A1, A2, A3, AFG, AIG, A2G, A3G) est de type <nombre>, ce nombre appartient à l'ensemble NPO des numéros prédéfinis d'organismes.
- V21 : Si une variable appartient à l'ensemble CPP, alors la propriété de parité des parenthèses est vérifiée sur le texte contenu dans la variable.
- V22 : Si une variable appartient à l'ensemble CPC, alors la propriété de parité des crochets est vérifiée sur le texte contenu dans la variable.
- V23 : Si une variable appartient à l'ensemble VD1, elle contient une seule chaîne d'au plus LLMADO caractères.
- V24 : Tout caractère appartenant à l'ensemble CIA des "caractères interdits dans une adresse" ne peut pas être présent dans un composant de type <adresse>.
- V25 : La forme majuscule du code associé au type <code pays> appartient à l'ensemble PAYS des "codes pays".

- V26 : Les nombres associés au composant <pagination> des variables PG1 et PGG contiennent au plus TAILLE chiffres.
- V27 : Les composants <nombre> des variables NE, PG3 et VOL sont formés d'au plus 2 chiffres consécutifs.
- V28 : Les composants <nombre> des variables PG2, PG4 et L10 sont formés d'au plus 4 chiffres consécutifs.
- V29 : La liste contenue dans LA est limitée à LLI5 éléments de type <code langue>.
- V30 : La forme majuscule du code associé au type <code langue> appartient à l'ensemble LANGUE des "codes langues".
- V31 : La liste contenue dans LR est limitée à LLI6 éléments de type <code langue>.
- V32 : La forme majuscule du caractère contenu dans CI appartient à l'ensemble DEGRE des «degré» d'intérêt.
- V33 : Si le contenu de la variable AD est de type <nombre>, le nombre associé appartient à l'ensemble NPO des "numéros prédéfinis d'organismes".
- V34 : La liste contenue dans R1 est limitée à LLI7 éléments de type <plan de classement>.
- V35 : Le code constitué avec la forme majuscule du composant <plan de classement> de la variable R1 est, après suppression des lettres "O" terminales, composé d'au plus 3 caractères et appartient à l'ensemble PLAN1.
- V36 : La liste contenue dans R2 est limitée à LLI8 éléments de type <plan de classement>.
- V37 : Le code constitué avec la forme majuscule du composant <plan de classement> de la variable R2, est composé de 3 caractères et appartient à l'ensemble PLAN2.
- V38 : La liste contenue dans MC1 est limitée à LMC1 éléments de type <mot clé>.
- V39 : La liste contenue dans MC2 est limitée à LMC2 éléments de type <mot clé>.
- V40 : La liste contenue dans MC3 est limitée à LMC3 éléments de type <mot clé>.
- V41 : La liste contenue dans MC4 est limitée à LMC4 éléments de type <date>.
- V42 : La liste contenue dans MC5 est limitée à LMC5 éléments de type <mot clé>.
- V43 : La liste contenue dans MC6 est limitée à LMC6 éléments de type <mot clé>.
- V44 : La liste contenue dans MC7 est limitée à LMC7 éléments de type <mot clé>.
- V45 : La liste contenue dans MC8 est limitée à LMC8 éléments de type <mot clé>.
- V46 : La liste contenue dans MC9 est limitée à LMC9 éléments de type <mot clé>, éventuellement préfixés par les caractères "*" ou "?".
- V47 : La liste contenue dans MC10 est limitée à LMC10 éléments de type <mot clé>.
- V48 : La liste contenue dans MC11 est limitée à LMC11 éléments de type <mot clé>.
- V49 : La liste contenue dans MC12 est limitée à LMC12 éléments de type <mot clé>.
- V50 : Tout élément de type <mot clé> est une chaîne de longueur limitée à LCOTE caractères.
- V51 : Les formes majuscules des caractères contenus dans IN appartiennent à l'ensemble CODINDI des codes "indicateur bibliographique".

2.5. LES PROPRIETES RELATIVES AUX REFERENCES

Les valeurs de certaines variables induisent l'existence de propriétés propres à la référence. La méthode utilisée consiste à définir des attributs, ou variables secondaires indiquant l'existence de caractéristiques sémantiques particulières. Les définitions de ces attributs sont précisées dans l'encadré ci-dessous, sous forme de définitions numérotées de A1 à A8.

- A1 : *L'attribut PERE est une variable logique qui vaut "vrai" si le nombre contenu dans ND est divisible par 100, "faux" sinon.*
- A2 : *L'attribut TYPED est une variable caractère qui reçoit comme valeur la forme majuscule du caractère contenu dans TD.*
- A3 : *L'attribut NIVEAU est une variable caractère qui reçoit comme valeur la forme majuscule du caractère contenu dans NI.*
- A4 : *L'attribut NPERE est une variable caractère égale à NIVEAU si PERE vaut "vrai", invariante sinon.*
- A5 : *L'attribut INDIBI est une table indicée par les 26 lettres de l'alphabet. Chaque élément de la table est une variable logique qui vaut "vrai" si la lettre associée est présente dans IN, "faux" sinon.*
- A6 : *L'attribut ANONYME vaut "vrai" si la variable AU est absente ou si AU contient le composant <joker> ou la mention "Anonyme", "faux" sinon.*
- A7 : *L'attribut FRANCAIS vaut "vrai" si le premier élément de la liste de la variable LA a pour valeur "FRE", "faux" sinon.*
- A8 : *L'attribut ANGLAIS vaut "vrai" si le premier élément de la liste de la variable LA a pour valeur "ENG", "faux" sinon.*
- A9 : *L'attribut PAGINE vaut "vrai" si la variable PGI contient un composant syntaxique de type <pagination>, "faux" sinon.*

L'attribut PERE situe une référence à l'intérieur d'une notice bibliographique. En fait cette notion permet de gérer potentiellement des arborescences, ici limitées à deux niveaux, un "père" pouvant, vu les conventions, avoir 99 "fils". Il est utilisé d'une part dans le paramétrage des contrôles d'intégrité de référence, d'autre part, pour reconnaître une notice et effectuer les contrôles d'intégrité de notice.

Les attributs TYPEB (pour type bibliographique), NIVEAU (pour niveau d'indexation) et INDIBI (pour indicateur bibliographique), ont une signification documentaire précise et sont fondamentaux pour piloter les contrôles d'intégrité de référence. L'ajout d'indicateurs bibliographiques permet l'ajout de nouveaux contrôles pour les besoins d'une application, à la demande du responsable de l'application documentaire.

L'attribut NPERE (pour niveau d'indexation du document "père") précise le niveau d'indexation de la référence contenant les données communes de la notice. Il est "hérité" pour les "fils" dans une notice, et non défini si l'attribut PERE de la première référence d'un ensemble vaut "faux".

2.6. LES CONTROLES D'INTEGRITE DE REFERENCE PARAMETRABLES

Un certain nombre de contrôles s'appliquent à la référence dans son ensemble et sont relatifs principalement à des obligations-interdictions de présence de variables selon des contextes calculés à l'aide d'une fonction d'état.

La procédure de pilotage de cette famille de contrôles consiste à définir une matrice M dite de contrôle, à NBVDOC lignes et NBETAT colonnes, NBVDOC désignant le nombre de variables et NBETAT le nombre de valeurs possibles de la fonction d'état. Cette matrice M est un paramètre externe géré par l'administrateur du projet. Si le terme $M(i,j)$ prend la valeur k, cela signifie que la variable numéro i d'une référence d'état j possède la propriété k.

a) Calcul de la fonction d'état

Sur les trois attributs NPERE, TYPEB et NIVEAU est définie une fonction dite d'état qui permet de classer les références dans des sous-ensembles ou familles ayant une signification documentaire précise : article de revue, ouvrage vu dans son ensemble, partie d'un ouvrage, logiciel, film, etc... Cette fonction est calculée comme suit :

Si NIVEAU = 'L'	alors état = 1;
Si NIVEAU = 'A' et TYPEB = 'J'	alors état = 2;
Si NIVEAU = 'M' et TYPEB = 'B'	alors état = 3;
Si NIVEAU = 'A' et TYPEB = 'B' et NPERE = 'M'	alors état = 4;
Si NIVEAU = 'C' et TYPEB = 'B'	alors état = 5;
Si NIVEAU = 'A' et TYPEB = 'B' et NPERE = 'C'	alors état = 6;
Si NIVEAU = 'M' et TYPEB = 'G'	alors état = 7;
Si NIVEAU = 'M' et TYPEB = 'F'	alors état = 8;

La fonction d'état n'est pas définie si l'une de ces conditions n'est pas remplie et les contrôles d'intégrité de référence ne seront pas effectués. D'autres conditions peuvent s'ajouter pour générer d'autres valeurs de cette fonction d'état.

b) Les types de propriétés contrôlées

Cinq codes de contrôles sont actuellement autorisés dans la définition de la matrice de contrôle M mais cette définition peut être étendue pour d'autres besoins. Ces codes ont la signification suivante :

0	: la variable est facultative pour toute référence ;
1	: la variable est obligatoire pour toute référence ;
2	: la variable est interdite pour toute référence ;
3	: la variable est obligatoire si PERE vaut vrai, ignorée sinon ;
4	: la variable est facultative si PERE vaut vrai, ignorée sinon ;
5	: les contrôles de propriétés syntaxiques et sémantiques de la variable sont occultés.

Les valeurs de ces contrôles peuvent toutefois être occultés par des règles de priorités supérieures, définies au niveau des contrôles d'intégrité de référence non paramétrables.

2.7. LES CONTROLES D'INTEGRITE DE REFERENCE NON PARAMETRABLES

Ces propriétés sont des règles documentaires qu'il s'agit de formaliser sous forme de prédicats logiques. Elles sont numérotées de 1 à 34, cette liste étant évolutive. Les 4 premières règles s'appliquent en priorité au cas où elles seraient en contradiction avec celles déduites du contenu de la matrice M de contrôle.

Pour exprimer ces règles, on utilise les attributs définis par les règles A1 à A9 et on se donne des fonctions booléennes ainsi définies :

Si X est le nom d'une variable documentaire,

- la fonction P(X) vaut "vrai" si X est présente, "faux" sinon ;
- la fonction A(X) vaut "vrai" si X est absente, "faux" sinon ;

- R1 : $P(ND) = \text{vrai}$.
 R2 : $P(TD) = \text{vrai}$.
 R3 : $P(IN) = \text{vrai}$.
 R4 : $SI (P(A1) \text{ et } A(AF)) \text{ ALORS } (P(AU) \text{ OU } A(AU))$.
 R5 : $(A(AU) \text{ ET } A(A1)) = \text{faux}$.
 R6 : $SI P(AF) \text{ ALORS } P(AU)$.
 R7 : $SI P(A3) \text{ ALORS } P(A2)$.
 R8 : $SI P(A2) \text{ ALORS } P(A1)$.
 R9 : $SI (\text{état} = 5) \text{ ALORS } SI P(AFG) \text{ ALORS } P(AUG)$.
 R10 : $SI (\text{état} = 5) \text{ ALORS } SI P(A3G) \text{ ALORS } P(A2G)$.
 R11 : $SI (\text{état} = 5) \text{ ALORS } SI P(A2G) \text{ ALORS } P(A1G)$.
 R12 : $SIPERE = \text{"vrai"} \text{ ALORS } SI (P(TG) \text{ OU } P(DG) \text{ OU } P(AG)) \text{ ALORS } (P(TG) \text{ ET } P(DG) \text{ ET } P(AG))$.
 R13 : $SI PAGINE = \text{"faux"} \text{ ALORS } P(PG2)$.
 R14 : $SI (P(ED) \text{ OU } P(AE)) \text{ ALORS } (P(ED) \text{ ET } P(AE))$.
 R15 : $SI (P(ED) \text{ OU } P(AE)) \text{ ALORS } (\text{NON INDIBI('V')})$.
 R16 : $SI (TD = \text{"B"}) \text{ ALORS } ((P(ED) \text{ ET } P(AE)) \text{ OU } ((P(AF) \text{ OU } P(A1)) \text{ ET } P(AD)))$.
 R17 : $SI P(MC2) \text{ ALORS } P(MC1)$.
 R18 : $SI P(DI) \text{ ALORS } P(MC1)$.
 R19 : $SI P(DI) \text{ ALORS } (P(AF) \text{ OU } P(A1))$.
 R20 : $SI P(R2) \text{ ALORS } (P(TA) \text{ ET } P(LR) \text{ ET } P(R1) \text{ ET } P(MC9))$.
 R21 : $SI \text{INDIBI('K')} \text{ ALORS } ((P(TG) \text{ ET } P(DG) \text{ ET } P(AG)))$.
 R22 : $SI \text{INDIBI('U')} \text{ ALORS } (P(AF) \text{ ET } P(A1) \text{ ET } P(DI) \text{ ET } P(AD))$.
 R23 : $SI \text{INDIBI('V')} \text{ ALORS } (P(AF) \text{ OU } (P(A1)) \text{ ET } P(AD))$.
 R24 : $SI \text{INDIBI('Z')} \text{ ALORS } P(RB)$.
 R25 : $SI \text{INDIBI('P')} \text{ ALORS } (P(AF) \text{ ET } P(LI0))$.
 R26 : $SI (\text{NON } (\text{INDIBI('U')} \text{ OU } \text{INDIBI('V')})) \text{ ET } (\text{état} = 3) \text{ ET } ((A(AF) \text{ ET } A(A1)) \text{ OU } A(AD)) \text{ ALORS } (P(ED) \text{ ET } P(AE))$.
 R27 : $SI (\text{ETAT} = 4) \text{ alors } PERE = \text{"faux"}$
 R28 : $SI (\text{ETAT} = 3) \text{ alors } PERE = \text{"vrai"}$.
 R29 : $SI (\text{ETAT} = 6) \text{ alors } PERE = \text{"faux"}$
 R30 : $SI (\text{ETAT} = 5) \text{ alors } PERE = \text{"vrai"}$
 R31 : $SI (\text{ETAT} > 2) \text{ et } \text{ANONYME} \text{ alors } A(AU)$.
 R33 : $SI \text{ le premier code de la variable LA n'est pas "FRE"} \text{ ALORS } P(TF)$.
 R34 : $SI \text{ le premier code de la variable LA n'est pas "ENG"} \text{ ALORS } P(TA)$.

2.8. LES CONTROLES D'INTEGRITE DE NOTICE

Les notices sont des ensembles de références reliées par des relations hiérarchiques. Dans cette application, un seul niveau de filiation est autorisé.

Le numéro de notice est le nombre constitué avec les 6 premiers chiffres de la variable ND. Le numéro de bordereau est constitué avec les 2 derniers chiffres. Les propriétés contrôlées référencent les attributs définis de A1 à A9 et sont actuellement les suivantes :

- N1 : SI une référence est la première d'une notice ALORS PERE vaut "vrai"*
- N2 : Le numéro de notice est invariant dans les références d'une notice.*
- N3 : Les numéros de bordereaux d'une notice sont rangés dans l'ordre croissant.*
- N4 : SI (NIVEAU = 'L') pour la première référence d'une notice, ALORS (NIVEAU = 'L') pour toutes les références de la notice.*
- N5 : SI ((NIVEAU = 'M') OU (NIVEAU = 'A') OU (NIVEAU = 'C')) pour la première référence d'une notice, ALORS (NIVEAU = A) pour toutes les autres références de la notice.*

3. NORMALISER LA REPRESENTATION DE L'INFORMATION DOCUMENTAIRE

Un modèle de données documentaires a sa raison d'être par rapport à un ensemble de fonctionnalités, relevant du domaine documentaire, en réponse à des besoins scientifiques d'équipes ou institutionnels. Les applications se situent simultanément en amont des projets scientifiques, à travers la recherche de documents, et en aval, par le biais d'une valorisation sous des formes et sur des supports diversifiés. C'est en regard de ces familles d'applications que s'imposent d'autres contraintes pratiques : des conventions non ambiguës de formatage doivent être adoptées pour que les tests d'égalité traduisent la signification documentaire des questions au cours d'une interrogation ; les conventions dactylographiques usuelles doivent être respectées lorsqu'il s'agit d'éditer des produits diffusés ; etc...

L'objectif de cette étape de normalisation est de définir sur des références documentaires contrôlées une liste de règles systématiques, qui satisfont à un maximum de contraintes, liées aux pratiques et conventions professionnelles. L'application de ces règles aboutit à une représentation standardisée des références documentaires, dite **FORME NORMALE**.

3.1. LE PRETRAITEMENT DU TEXTE.

Avant d'effectuer la reconnaissance syntaxique et les contrôles sémantiques associés, une tâche préliminaire consiste à acquérir un à un les caractères qui composent le texte, et pour chacun d'eux, examiner la relation d'appartenance à l'alphabet retenu dans la définition du langage. Les caractères n'appartenant pas à cet alphabet sont ignorés.

Au cours de cette phase sont effectués des prétraitements, dont l'objet est de faciliter l'application des conventions dactylographiques usuelles : continuités de lignes ; espaces devant et derrière les caractères non alphanumériques. Ces prétraitements sont formalisés dans les conventions suivantes :

- C1 : Les caractères qui n'appartiennent pas à l'ensemble ALPHABET sont ignorés*
- C2 : le caractère "MARQUE DE FIN DE LIGNE" non suivi du caractère "MARQUE DE CONTINUITE" est remplacé par le caractère "BLANC".*
- C3 : Toute suite de caractères "BLANC" située derrière un caractère de l'ensemble AGORA est ignorée.
L'ensemble AGORA contient le caractère "BLANC" et les caractères de l'ensemble suivants : { . , ; ; ? ! } { [- + * / & }.*

La convention C1 autorise pratiquement la lecture de n'importe quel code en entrée, le texte reconnu étant formé par juxtaposition des caractères autorisés : les caractères parasites sont ainsi filtrés (une séquence telle "chapitre n° 1) devient "chapitre n 1).

La convention C2 permet d'uniformiser la notion dactylographique "d'espace séparateur" entre deux mots.

La convention C3 supprime les blancs en fait derrière les ponctuations et quelques autres caractères non alphanumériques. Le caractère "BLANC" appartenant à l'ensemble AGORA, toute "zone blanche" est réduite à un seul "blanc séparateur" ou au "mot vide".

3.2. LE PRETRAITEMENT D'UNE REFERENCE

La définition syntaxique d'une référence, et les contrôles associés (paragraphe 2.1.) n'introduisent aucune contrainte quant à l'ordre des variables à l'intérieur d'une référence. En outre, aucune règle n'interdit la répétition de variables. L'algorithme de reconnaissance générale de la référence réalise une restructuration en appliquant les conventions suivantes :

- C4 : les variables sont rangées selon l'ordre prédéfini dans le document de paramètres géré par l'administrateur du projet, le drapeau restant au début de la référence.*
- C5 : Si une variable est répétée dans une référence, son contenu définitif est constitué par juxtaposition des occurrences de cette variable, en respectant l'ordre de ces occurrences.*

La convention C5 peut entraîner des interprétations imprévues : le fait par exemple d'oublier des drapeaux crée une référence unique dans laquelle toutes les variables de même nom sont regroupées ; si cette fusion n'active pas la règle F4, l'algorithme de contrôle va diagnostiquer des suites d'erreurs a priori incompréhensibles.

3.3. LES CONVENTIONS DACTYLOGRAPHIQUES.

Ces conventions gèrent les espaces séparateurs dans un texte, y compris auprès des caractères de ponctuation. Elles font références à des ensembles de caractères paramétrables. Ces règles ne sont pas encore stabilisées en l'absence d'une normalisation.

- C6 : Les blancs sont supprimés en début de variable
- C7 : Le blanc en fin de variable est supprimé
- C8 : Un composant <ligne> ne contenant que des blancs est supprimé
- C9 : Les variables vides sont supprimées.
- C10 : Les blancs sont supprimés devant et derrière les caractères de l'ensemble
PASDEBLANC = { - + * / & }.
- C11 : Un blanc est ajouté derrière les caractères de l'ensemble
PONCTU = { . , ; : ! ?)] } sauf dans la liste des cas suivants :
- le caractère suivant appartient à cet ensemble ;
- le caractère suivant est un blanc ;
- le caractère courant est un point ou une virgule et les caractères précédents et suivants sont des chiffres (en clair c' est un nombre décimal);
- le caractère courant est le point-virgule dans un type <liste> ;
- la variable en cours est de type <auteur>.
- C12 : Un blanc est ajouté devant les caractères de l'ensemble
BDEVANT = { : ; ! ? ({ } } sauf dans la liste des cas suivants :
- le caractère précédent appartient à cet ensemble ;
- le caractère précédent est un blanc ;
- le caractère courant est le premier caractère d'une variable ;
- le caractère courant est le point-virgule dans un type <liste> ;
- C13 : Le caractère "et commercial" (&) n'est pas transmis dans l'édition de la forme normale.
- C14 : Le composant <marque de continuité> est ajouté dans la forme normale lorsque le composant <ligne> à transmettre n'est pas le dernier de la variable.

3.4. LE CODAGE DES LABORATOIRES.

Les auteurs de publications scientifiques appartiennent généralement à des laboratoires dans des Instituts de Recherches, ou à des services dans des organismes professionnels. Les informations relatives à ces laboratoires ou services doivent être reproduites fidèlement, tout écart aussi infinitésimal soit-il, se traduisant par un silence anormal à l'interrogation et une "coquille" à l'édition. Le silence à l'interrogation est inacceptable pour des applications dont le but est par exemple de dresser la liste exhaustive des publications d'un laboratoire.

Pour pallier à cet inconvénient, et éviter aux documentalistes une saisie répétitive pour une variable complexe, un numéro prédéfini fut affecté aux laboratoires et services connus. La présence d'un code numérique déclenche la recherche dans le document de paramètres, géré par l'administrateur du projet documentaire. Les informations lues sont transcrites sans que soient appliquées les conventions dactylographiques précédentes.

Ce mécanisme est activé pour les variables dites d'affiliation des auteurs (AF, A1, A2, A3, AFG, A1G, A2G, A3G), la localisation ou cote du document (LO) et l'adresse de diffusion d'un document (AD). Le code numérique est le contenu du composant <nombre> présent dans les définitions syntaxiques des variables.

- L1 : Si les variables AF, A1, A2, A3, AFG, A1G, A2G, A3G contiennent la forme syntaxique <nombre>, ce nombre est remplacé par l'intitulé de l'organisme lu dans le document de paramètres.*
- L2 : Si la variable LO contient la forme syntaxique <nombre>, ce nombre est remplacé par la cote de l'organisme lue dans le document de paramètres.*
- L3 : Si la variable AD contient la forme syntaxique <nombre>, ce nombre est remplacé par l'adresse de l'organisme lue dans le document de paramètres.*

3.5. LA GENERATION DE TEXTES PREDEFINIS.

Les valeurs de certaines variables peuvent ne pas être définies, et cette information doit être précisée pour ne pas être confondue avec une absence de variable. La convention retenue consiste alors à utiliser le composant syntaxique <joker>, et à le remplacer par un texte prédéfini associé à la famille de variables.

Elle est appliquée dans les cas suivants :

- C15 : Le composant <joker> des variables AU et AUG est remplacé par le texte "Anonyme".*
- C16 : Le composant <joker> des variables PG1 et PGG est remplacé par le texte "non paginé".*
- C17 : Le composant <joker> des variables IL, RB et TB est remplacé par le caractère "*".*
- C18 : Le composant <joker> de la variable RS est remplacé par le texte "non résumé".*

Dans d'autres cas le formatage est invariant et peut se déduire du contexte. Les conventions ci-dessous s'appliquent aux variables de pagination, dans l'ordre pour C20 à C22

- C19 : Si les variable PG1 ou PGG contiennent les formes syntaxiques "N<texte>" ou "n<texte>, ces composants sont remplacés par le texte "non paginé".*
- C20 : Dans un composant <pagination>, les caractère "p" et"." sont ignorés.*
- C21 : Si le tiret est présent dans un composant <pagination>, le texte de ce composant est préfixé par la chaîne "pp".*
- C22 : Si le tiret n'est pas présent dans un composant <pagination>, le texte de ce composant est postfixé par la chaîne "p."*

3.6. LES CONVERSIONS DES LETTRES SOUS LA FORME MAJUSCULE.

Les lettres minuscules sont converties en majuscules, avec suppression des signes diacritiques, les autres caractères étant inchangés, dans les cas suivants :

- C23 : les noms de drapeau et de variables sont convertis en majuscules.*
- C24 : le composant <code pays> est converti sous la forme majuscule.*
- C25 : le composant <code langue> est converti sous la forme majuscule.*
- C26 : le composant <plan de classement> est converti sous la forme majuscule.*
- C27 : le composant <mot clé> est converti sous la forme majuscule.*
- C28 : les composants <caractère> des variables TD, NI, SU, CI, IN sont convertis en majuscule.*
- C29 : si les variables IB et IS contiennent une lettre, elle est convertie sous la forme majuscule.*

3.7. LES CONVENTIONS SPECIFIQUES

Cette rubrique permet de regrouper dans des ensembles des variables pour lesquelles existe une convention commune de formatage.

Ainsi une convention documentaire interdit la présence d'un point final dans un titre, ce qui se traduit avec la règle suivante :

C30 : le point final est supprimé dans les variables de l'ensemble SPF.

3.8. LA GESTION DES VARIABLES COMMUNES

La notion de notice a été définie pour éviter de répéter la saisie de variables communes dans un ensemble de références. Les applications situées en aval sont chargées de reconstituer la hiérarchie pour éventuellement gérer convenablement ces variables communes. Elles peuvent toutefois être répétées mais l'algorithme de formatage ne sait les transmettre que si l'attribut PERE est vrai. Plus précisément la règle est la suivante :

C31 : si le code de contrôle associé à la variable vaut 3 ou 4, la variable est transmise seulement si l'attribut PERE de la référence vaut "vrai".

ENCADRE NUMERO 3

DESCRIPTION SYNTAXIQUE DE LA FORME NORMALISEE
DES VARIABLES DE L'APPLICATION ESR

ND	: <nombre>
TD	: <majuscule>
NI	: <majuscule>
LO	: <cote>{;<cote>}
SU	: <majuscule>
DA	: <date>{;<date>}
AU	: <auteur>{;<auteur>}:Anonyme
AF	: <organisme>
A1	: <organisme>
A2	: <organisme>
A3	: <organisme>
DI	: <texte>
TI	: <texte>
LI	: <texte>
TF	: <texte>
TA	: <texte>
TG	: <texte>
AG	: <adresse>
DG	: <date>{;<date>}
NE	: <nombre>
IB	: <nombre>-<nombre>-<nombre>-<alphanumérique>
ED	: <texte>
AE	: <adresse>
PG1	: <pagination>:non paginé
PG2	: <nombre>
PG3	: <nombre>
PG4	: <nombre>
LA	: <code langue>{;<code langue>}
LR	: <code langue>{;<code langue>}
IL	: <nombre> *
RB	: <nombre> *
CA	: <texte>
TB	: <nombre> *
CI	: <majuscule>
AD	: <adresse>
R1	: <plan de classement>{;<plan de classement>}
R2	: <plan de classement>{;>plan de classement}
MC1	: <mot clé>{;<mot clé>}
MC2	: <mot clé>{;<mot clé>}
MC3	: <mot clé>{;<mot clé>}
MC4	: <mot clé>{;<mot clé>}
MC5	: <mot clé>{;<mot clé>}
MC6	: <mot clé>{;<mot clé>}
MC7	: <mot clé>{;<mot clé>}
MC8	: <mot clé>{;<mot clé>}
MC9	: <mot clé>{;<mot clé>}
MC10	: <mot clé>{;<mot clé>}
MC11	: <mot clé>{;<mot clé>}
MC12	: <mot clé>{;<mot clé>}
RS	: <texte>
SO	: <texte>
IS	: <nombre>-<nombre><alphanumérique>
VOL	: <nombre>
NUM	: <nombre>
NS	: <texte>
L1	: <texte>
L2	: <texte>
L3	: <texte>
L4	: <texte>
L5	: <texte>
IN	: <caractère>{<caractère>}
CP	: <code pays>
L6	: <texte>
L7	: <texte>
L8	: <texte>
L9	: <texte>
L10	: <nombre>
AUG	: <auteur>{<auteur>}:anonyme
AFG	: <organisme>
A1G	: <organisme>
A2G	: <organisme>
A3G	: <organisme>
TIG	: <texte>
PGG	: <pagination>:non paginé

4. PROPOSER UN PRODUIT INFORMATIQUE ADAPTABLE AUX CONTEXTES DOCUMENTAIRES

Sur le plan informatique, MADO est une traduction fidèle de la description faite dans ce rapport. C'est une construction modulaire, paramétrable par un administrateur de projet, très structurée, prévue pour s'adapter à d'autres contextes. Il est a priori exclu de modifier le noyau du système mais il est toujours possible d'ajouter des règles, voire de nouvelles constructions syntaxiques, dans le cadre de cette méthodologie et sous réserve évidemment de disposer d'équipes de développement. Ce chapitre décrit cet environnement informatique.

4.1. LA TRADUCTION INFORMATIQUE DE MAD0

La version actuelle respecte fidèlement la norme ISO du langage PASCAL, et même plus précisément le sous-ensemble de cette norme disponible dans la version dite "TURBO PASCAL" de Borland. Seule exception incontournable à cette règle, quelques lignes, dûment isolées, réglant les communications avec les systèmes d'exploitation hôtes en matière de gestion de fichiers. Le prototype fut mis au point avec le compilateur de Digital Research sur le système CP/M 86, puis porté sans problème sur le système central Multics pour la mise au point, et enfin exploité sous MSDOS. Une version NOS/VE devrait prochainement être installée sur le serveur du Centre Interuniversitaire de Calcul. L'installation sous UNIX ou OS/2 ne nécessitera qu'une compilation d'un code déjà validé par trois compilateurs.

A l'heure actuelle la version du Département d'Economie et Sociologie Rurales peut donc s'exécuter sur toute machine. On notera toutefois qu'en période de charge normale du système central Multics, la version micro sur machine compatible IBM XT s'avère être la plus rapide. En période de faible utilisation de Multics, les temps d'exécution sont deux fois plus rapides par rapport à la version micro. L'ordre de grandeur de ces temps est voisin de la seconde pour le millier de caractères.

4.2. L'UTILISATION DE MADO

Dans le cadre du projet documentaire du Département d'Economie et de Sociologie rurales, MADO est utilisé en relation avec le logiciel de gestion de références documentaires TEXTO, également disponible sous MULTICS, MSDOS et NOS/VE. Les références à analyser sont des éditions issues de TEXTO. Les références contrôlées sont transmises dans une base TEXTO à l'aide de la commande dite "d'ajout piloté". Le couple (MADO, TEXTO) est depuis plus d'un an opérationnel et a actuellement traité plusieurs milliers de références documentaires.

Un aspect important qui conditionne l'acceptabilité du produit est la qualité du diagnostic d'erreur, exprimé à travers des messages. Ces messages sont gérés dans le document de paramètres par l'administrateur du projet documentaire, qui dispose ainsi d'un moyen pour adapter le contenu de ces messages à des publics d'utilisateurs aux motivations et compétences diversifiées.

MADO acquiert des caractères à l'aide d'une primitive unique référencée dans le module de lecture. Dans cette application, cette primitive lit ses caractères dans le fichier texte édité par TEXTO.

Les références contrôlées et normalisées sont transmises vers une procédure de recopie, qui édite actuellement deux fichiers. Le premier est la forme normale issue de MADO, rangée dans un fichier texte et pouvant d'ailleurs être relue puisque définie comme une partie du langage reconnu par MADO. Le second formate les variables pour répondre aux exigences de la commande TEXTO dite "ajout piloté". Cette seconde forme est sur le plan linguistique ambiguë, ce qui peut se traduire dans la pratique par une confusion entre le nom de la variable et son contenu.

L'identification des références est éditée à l'écran avec l'indication "je prends" ou "j'exclus". Dans le second cas les messages d'erreurs sont édités. Une copie de ces éditions destinées à l'opérateur est recopiée dans un fichier dit de compte-rendu.

4.3. LE PARAMETRAGE DE MADO PAR LE RELAIS DU DOCUMENT DE PARAMETRES.

Au logiciel MADO est associé un fichier manipulable sous un éditeur de texte dit "document de paramètres". Ce fichier est géré et contrôlé par le responsable de l'application documentaire.

Les lignes pour lesquelles le premier caractère non blanc est "\$" sont des commentaires. Les données sont en format libre. Un document de paramètres de l'application ESR est fourni en annexe à titre d'exemple.

4.3.1. LES NOMS DES VARIABLES ET LA MATRICE DE CONTROLE

Le premier ensemble de paramètres regroupe la liste des noms de variables, et en regard de ces noms les codes de contrôle.

Les numéros d'ordre des lignes sont les numéros des variables (1 à NBVDOC avec NBVDOC = 75) et les numéros des colonnes correspondant aux valeurs de la fonction d'état (1 à NBETAT avec NBETAT = 8).

La dernière ligne désigne toujours le drapeau et les codes associés sont lus mais ignorés car occultés par une priorité d'ordre supérieur.

L'ordre initial des variables est laissé à l'initiative du responsable de l'application, mais une fois choisi, cet ordre doit resté invariant durant le cycle de vie de l'application. Des variables peuvent s'ajouter au-delà de l'avant-dernière ligne, et des états introduits à droite de la dernière colonne, sous réserve d'avoir auparavant intégré dans MADO le traitement de ces nouvelles variables ainsi que l'algorithme de calcul de la fonction d'état. Ces dernières opérations ne peuvent être envisagées que sous contrôle informatique rigoureux, pour élargir une application par exemple : MADO reconnaîtra alors les références des niveaux antérieurs.

MADO vérifie au moment du chargement des paramètres certaines relations de cohérence : ainsi les états 1 (obligatoire pour toute référence) et 3 (interdite pour référence fille) pour une même variable dans une arborescence sont contradictoires.

4.3.2. LA LISTE DES ORGANISMES

Le document de paramètres contient ensuite une liste d'organismes repérés par des numéros d'organismes (sur la première ligne). Pour chaque organisme sont fournies les valeurs des variables de types :

<organisme> sur 4 lignes,
<cote de l'organisme> sur 1 ligne,
<adresse> sur 3 lignes.

Si un numéro d'organisme est répété, les valeurs retenues par MADO seront les dernières lues. La version actuelle reconnaît des numéros compris entre 1 et NBLABO, avec NBLABO = 99. La liste se termine par l'occurrence du caractère "*" à la place d'un numéro d'organisme. L'ensemble des codes numériques d'organismes présents dans le document de paramètres représente la valeur de NPO (numéro prédéfini d'organisme) référencée dans la description des contrôles sémantiques.

4.3.3. LES MESSAGES D'ERREUR

Le paramètre suivant est la liste des messages d'erreurs, terminée par le caractère "*" sur une ligne. Les éléments de cette liste sont des lignes avec en tête un numéro d'erreur suivi d'un libellé d'erreur destiné aux usagers de l'application. Les numéros d'erreurs varient de 0 à MAXERR, avec MAXERR = 150.

La liste des erreurs est initialisée par le responsable de l'application documentaire, toute erreur non répertoriée provoquant l'édition du message par défaut numéro 0. Si un numéro d'erreur est répété, c'est la dernière valeur lue qui est retenue.

4.3.4. LES CODES PAYS

La liste des code pays est actuellement limitée à un maximum de NBPAYS (NBPAYS = 170). La liste se termine avec le code fictif "****".

MADO recherchant ces codes par balayage séquentiel de la table des codes pays, l'exécution des contrôles sera plus rapide si les codes fréquemment utilisés sont placés en début de liste.

4.3.5. LES CODES LANGUES

La convention est identique à celle des codes pays, avec, cette fois, un maximum de 50 codes langues (NBLANGUE = 50)

4.3.6. LE PREMIER PLAN DE CLASSEMENT : RESAGRI

A chaque lettre de l'alphabet est associée une liste qui recevra les codes plan commençant par cette lettre. Chaque liste est limitée à un maximum de 70 codes (NBCODOLA = 70) et sera balayée séquentiellement par MADO. Les codes de moins de 3 lettres sont complétés à droite par ajout de la lettre "0". L'occurrence du code "****" signifie que le plan est lu.

4.3.7. LE SECOND PLAN DE CLASSEMENT : AGRIS

La convention est identique à celle des codes pays, avec un maximum de NBPLAN2 codes de 3 caractères. (NBPLAN2 = 200)

4.4. LES PARAMETRES INITIALISES PAR PROGRAMME

Les règles formalisées dans les chapitres précédents font référence à des variables de type numérique, ensembles de variables, ensembles de caractères. Ces variables sont initialisées dans une procédure globale activée à l'appel de MAD0, et, dans certains cas à l'intérieur des procédures de contrôle associées à la reconnaissance et au traitement des types syntaxiques.

TABLEAU 1

LES PARAMETRES INITIALISES PAR PROGRAMME

NOM	SIGNIFICATION	VALEUR INITIALE	TYPE
CD	Codes documentalistes	[1..30]	ensemble de nombres
CODETD	Codes "type de document"	{B,J,G,F}	ens. de caractères
CODENI	Codes "niveau"	{M,L,A,C}	ens. de caractères
LL I1	Longueur de la liste de cotes	10	nombre
LCOTE	Longueur de la chaîne formant la cote	74	nombre
SUPP	Codes "support"	{M}	ens. de caractères
LL I3	Longueur liste dates de publication	2	nombre
LL I4	Longueur liste dates congrès	2	nombre
LL I2	Longueur liste auteur	20	nombre
CINA	Caractères interdits dans un nom d'auteur	{+ * / : ? ! [] }	ens. de caractères
CPP	Contrôle parité des parenthèses	{AU.. AG, AE, AD, RS, SO, AUG...TIG}	ens. de variables
CPC	Contrôle parité des crochets	{AF..TG, RS,SO,AFG...TIG}	ens. de variables
VD1	Codage de la variable sur une ligne	{ND...NI, SU, NE, IB, ED, PG1...PG4, IL.. CI, R1, R2, IS, VOL, NUM, L5, IN, CP PGG}	ens. de variables
SPF	Suppression point final	{TL, TF, TA, TG, SO, NS, TIG}	ens. de variables
CIA	Caractères interdits dans une adresse	{., ? ! ; : + * / }	ens. de caractères
TAILLE	Dimension des nombres dans un composant pagination	4	nombre
LL I5	Longueur liste codes langues du texte	5	nombre
LL I6	Longueur liste codes langues du résumé	3	nombre
LL I7	Longueur liste plans de classement 1	1	nombre
LL I8	Longueur liste plans de classement 2	1	nombre
LMC1	Longueur liste mots clés 1	7	nombre
LMC2	Longueur liste mots clés 2	17	nombre
LMC3	Longueur liste mots clés 2	16	nombre
LMC4	Longueur liste mots clés 2	3	nombre
LMC5	Longueur liste mots clés 2	12	nombre
LMC6	Longueur liste mots clés 2	3	nombre
LMC7	Longueur liste mots clés 2	7	nombre
LMC8	Longueur liste mots clés 2	4	nombre
LMC9	Longueur liste mots clés 2	15	nombre
LMC10	Longueur liste mots clés 2	14	nombre
LMC11	Longueur liste mots clés 2	20	nombre
LMC12	Longueur liste mots clés 12	20	nombre
LMOTCLE	Longueur maximale du mot clé	74	nombre
CODINDI	Codes indicateurs bibliographiques	{KUVZP}	ens. de caractères

4.5. LE MODULE DE SAISIE-CORRECTION

L'interface avec un module de saisie-correction interne est prévue. MADDO transmet à une procédure de correction la référence courante et la liste des erreurs. Cette procédure est chargée de dialoguer avec l'utilisateur. Elle retourne au programme appelant une référence corrigée, qui pourra à nouveau être contrôlée. Le dialogue peut se poursuivre, sauf abandon de l'un des partenaires. Ce module n'a pas été développé et le rôle de la procédure associée se limite à l'édition des messages d'erreurs et à la transmission d'un ordre d'abandon.

Une référence vide pouvant être vue comme une référence à corriger, des fonctionnalités de saisie-correction pourraient s'intégrer via ce point d'ancrage.

Un éclatement des contrôles à ce niveau est une solution pour assurer une plus grande interactivité.

4.6. LES CONTROLES D'APPARTENANCE DES MOTS-CLES AUX LEXIQUES.

Les mots-clés sont des chaînes souvent composées de plusieurs mots, appartenant à des ensembles, ou lexiques, dont la dimension se mesure en plusieurs milliers d'éléments, synonymies incluses. Les variables d'indexation sont des listes de mots clés définis sur de tels lexiques, un lexique étant généralement associé à chacune des variables.

Ces contrôles d'appartenance lexicale ne sont pas décrits au chapitre 2, dans la mesure où ils ne sont pas fournis avec la version initiale de MADO. Les algorithmes sont cependant disponibles et utilisés par l'administrateur du projet documentaire, dans le cadre d'un prototype de développement nommé AIDA (Administration Informatisée des Références Analysées), présenté dans le paragraphe suivant. Si les temps de contrôle sont relativement instantanés, les temps de chargement des lexiques peuvent devenir prohibitifs : sur une machine compatible IBM XT, il faut compter environ une minute pour charger une dizaine de milliers de mots clés.

4.7. LE PROTOTYPE DE DEVELOPPEMENT ASSOCIE : AIDA

Les procédures associées à MADO contrôlent et normalisent les références documentaires. La phase de contrôle revient à définir une référence documentaire comme un objet mathématique structuré, doté de propriétés. La phase de normalisation réalise des transformations sur cet objet.

Le module AIDA (Administration Informatisée des Références Analysées) gère ces objets normalisés, en régénérant les structures de données associées et les valeurs des attributs de références. Il est ensuite possible de transformer à nouveau ces références pour répondre à des exigences particulières de formatage ou d'édition, ou d'effectuer d'autres contrôles. La génération du code pour communiquer avec la chaîne documentaire INRA s'effectue actuellement dans une version adaptée de ce prototype. Des contrôles d'appartenance lexicale sont également expérimentés avec succès sur le lexique RESAGRI.

Cet environnement informatique est utilisable pour développer d'autres applications : indexation automatique sur le contenu d'autres variables, production du format documentaire ISO, etc . Des procédures de structuration de lexiques de mots-clés sont disponibles pour optimiser les dimensions de tables dans les algorithmes de contrôles utilisés par AIDA, lesquels pourraient ensuite être intégrés dans une version de MADO s'exécutant sur des machines plus performantes.

Le programme ALEX (analyse lexicale) effectue des dénombrements en fonction de la première lettre du mot-clé. Le programme VELEX (vérification lexicale) classe les mots en fonction de leur longueur et propose ensuite de vérifier l'appartenance de mots saisis au clavier au lexique analysé. Ces procédures sont réservées aux informaticiens pour des développements ultérieurs.

CONCLUSION

Né dans la dynamique du projet documentaire du Département d'Economie et Sociologie rurales, MADO est un produit informatique opérationnel qui matérialise une méthodologie de construction d'un système de contrôle et de normalisation de descriptions de documents, quelle que soit la forme de ces documents (publications, images, logiciels, bases de données, bases de connaissances, systèmes experts, etc.)

La construction de ce système s'est appuyée d'abord sur l'existence d'analyses du système d'information documentaire de l'INRA, et, plus particulièrement, d'un cahier des charges élaboré dans le cadre d'un appel d'offres pour réaliser une nouvelle chaîne de contrôle des données documentaires de l'Institut.

Elle s'est ensuite poursuivie par un long processus interactif entre informatique et documentation pour ajuster progressivement la représentation formalisée des tâches documentaires à la réalité du métier au sein du Département et des unités.

Pour les documentalistes des stations, MADO, seul ou dans l'environnement EDIBASE, est l'outil logiciel qui prend en charge une multitude de contrôles et d'opérations de formatage, pour homogénéiser la représentation des informations. Il permet de développer des applications de station et de répondre aux demandes du Département dans de bonnes conditions.

Au niveau du Département, MADO restitue au responsable du projet documentaire des descriptions contrôlées et normalisées, qui sont ensuite administrées dans une base de données et distribuées dans des applications.

L'outil de développement, AIDA, autorise des contrôles supplémentaires (appartenance des mots clés aux lexiques) et assure les interfaces avec le système d'information documentaire de l'INRA, lequel alimente les bases de données RESAGRI et AGRIS.

Pour l'équipe de conception et de développement, l'intérêt principal de MADO est son pouvoir d'adaptation à tout environnement informatique et documentaire. La version du Département d'Economie et Sociologie rurales est, avant tout, une démonstration de la faisabilité d'une solution méthodologique, dont le résultat est l'élaboration d'un langage. L'existence de ce langage permet de dialoguer de façon non ambiguë et d'enrichir progressivement les procédures de contrôle et de normalisation, voire d'intégrer des modules interactifs pour d'autres fonctionnalités : contrôle de thésaurus, aide à l'indexation, contrôle de la forme syntaxique d'un résumé, etc.

Cette intégration devient possible dès lors que les progrès technologiques offrent des machines plus puissantes avec des systèmes d'exploitation plus performants et un environnement logiciel mieux adapté aux métiers.

Elle nécessite cependant d'autres développements, dans le contexte d'une architecture générale très modulaire, et dépasse les objectifs d'un projet documentaire propre à un Département de recherche.

Ce langage est-il généralisable au niveau d'organismes de recherches, à celui d'autres services documentaires dans d'autres institutions, voire de façon plus générale à l'administration de données textuelles ?

Dans la mesure où il produit une représentation logique des données textuelles indépendante tant des configurations matérielles et logicielles que des traitements, cette hypothèse paraît réaliste. Les bases de données textuelles perdraient ainsi une certaine spécificité liées aux difficultés de "modélisation des données" (1), et s'intégreraient pleinement dans les systèmes d'informations scientifiques.

(1) F. Velez. La prise en compte de documents structurés dans un SGBD : aspects, modèles, langage et architecture du système, in M. Adiba et alii (coord.)
Nouvelles perspectives des bases de données. Edition Eyrolles, 1986.

Annexe 1 : Le document de paramètres MADO

\$ Document du 15/10/87

\$ M.A. Farget J.C. Poupa

\$ Mis à jour le 3 décembre 1987 (contrôle agris) J.C. Poupa

§ TABLE DE CONTROLE

nd	1 1 1 1 1 1 1 1
td	1 1 1 1 1 1 1 1
ni	1 1 1 1 1 1 1 1
lo	3 3 3 3 3 3 0 0
su	0 4 4 4 4 4 0 0
da	3 3 3 3 3 3 0 0
au	1 1 1 0 0 0 0 0
af	0 0 0 0 0 0 0 0
a1	0 0 0 0 0 0 0 0
a2	0 0 0 0 0 0 0 0
a3	0 0 0 0 0 0 0 0
di	0 0 0 0 0 0 0 0
ti	1 1 1 1 1 1 0 0
li	0 0 0 0 0 0 0 0
tf	0 0 0 0 0 0 0 0
ta	0 0 0 0 0 0 0 0
tg	0 4 4 4 0 0 0 0
ag	0 4 4 4 0 0 0 0
dg	0 4 4 4 0 0 0 0
ne	0 2 4 4 4 4 0 0
ib	0 0 4 4 4 4 0 0
ed	0 0 4 4 4 4 0 0
ae	0 0 4 4 4 4 0 0
pg1	1 1 1 1 1 1 0 0
pg2	0 0 0 0 0 0 0 0
pg3	0 0 0 0 0 0 0 0
pg4	0 0 0 0 0 0 0 0
la	0 1 1 1 1 1 0 0
lr	0 0 0 0 0 0 0 0
il	0 0 0 0 0 0 0 0
rb	0 0 0 0 0 0 0 0
ca	0 0 0 0 0 0 0 0
tb	0 0 0 0 0 0 0 0
ci	0 1 1 1 1 1 0 0
ad	0 0 0 0 0 0 0 0
r1	0 1 1 1 0 1 0 0
r2	0 0 0 0 0 0 0 0
mc1	0 1 1 1 0 1 0 0
mc2	0 0 0 0 0 0 0 0
mc3	0 0 0 0 0 0 0 0
mc4	0 0 0 0 0 0 0 0
mc5	0 0 0 0 0 0 0 0
mc6	0 0 0 0 0 0 0 0
mc7	0 0 0 0 0 0 0 0
mc8	0 0 0 0 0 0 0 0
mc9	0 0 0 0 0 0 0 0
mc10	0 0 0 0 0 0 0 0

mc11	00000000
mc12	00000000
rs	01110100
so	43444422
is	44444422
vol	44444422
num	43444422
ns	44444422
11	00000000
12	00000000
13	00000000
14	00000000
15	00000000
in	00000000
cp	43444422
16	00000000
17	00000000
18	00000000
19	00000000
110	00000000
aug	02224400
afg	02224400
a1g	02224400
a2g	02224400
a3g	02224400
tig	02223300
pgg	02223300
ref	11111111

\$ LISTE DES ORGANISMES

1

INRA
Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie rurales ENSSAA
Dijon (FRA)
INRA-ESR-DIJ
Station d'Economie et Sociologie Rurales
ENSSAA 26, bd du Docteur Petitjean
21100 DIJON (FRA)

2

INRA
Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Grenoble (FRA)
INRA-ESR-GRE
Laboratoire de Recherches Economiques INRA-IREP
Université des Sciences Sociales
BP 47 x 38040 GRENOBLE (FRA)

3

INRA
Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Thiverval-Grignon (FRA)
INRA-ESR-GRI
Station d'Economie et Sociologie Rurales
INA Paris-Grignon
78850 THIVERVAL GRIGNON (FRA)

4

INRA
Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Lille (FRA)
INRA-ESR-LIL
Laboratoire de Recherches Economiques sur les IAA de Lille
BP 39
59650 VILLENEUVE D'ASCQ Cedex (FRA)

5

INRA
Institut National de Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Montpellier (FRA)
INRA-ESR-MON
Station d'Economie et Sociologie Rurales
ENSA 9 place Viala
34060 MONTPELLIER Cedex (FRA)

6

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Nancy (FRA)

INRA-ESR-NCY

Laboratoire d'Economie Forestière et Agricole
14, rue Girardet
54042 NANCY Cedex (FRA)

7

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Orléans (FRA)

INRA-ESR-ORL

Laboratoire d'Economie et Sociologie Rurales
CRF Ardon
45160 OLIVET (FRA)

9

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Paris (FRA)

Station d'Economie et Sociologie Rurales

6, passage Tenaille

75014 PARIS (FRA)

10

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Rennes (FRA)

INRA-ESR-REN

Station d'Economie et Sociologie Rurales

65, rue de Saint-Brieuc

35042 RENNES Cedex (FRA)

11

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Rungis (FRA)

INRA-ESR-RUN

Laboratoire de Recherches Economiques et Sociales

3, rue du Caducée BP 333

94153 RUNGIS Cedex (FRA)

12

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Toulouse (FRA)

INRA-ESR-AUZ

Station d'Economie et Sociologie Rurales
CRA Auzeville BP 27
31326 CASTANET-TOLOSAN (FRA)

13

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales des Antilles-Guyane
Pointe-à-Pitre (FRA)

INRA-ESR-ANT

Laboratoire d'Economie et Sociologie Rurales du CRA des Antilles-
Guyane
BP 1232
94184 POINTE-A-PITRE

14

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Toulouse (FRA)

INRA-ESR-TOU

Station d'Economie et Sociologie Rurales
Université des Sciences Sociales
Place Anatole France
31042 TOULOUSE Cedex (FRA)

15

INRA

Institut National de la Recherche Agronomique
CRZV. Laboratoire d'Economie de l'Elevage
Theix (FRA)

INRA-ESR-THE

Laboratoire d'Economie de l'Elevage
CRZV Theix
63122 CEYRAT (FRA)

16

INRA

Institut National de la Recherche Agronomique
ESR. Station d'Economie et Sociologie Rurales
Versailles (FRA)

INRA-ESR-VER

Laboratoire d'Economie et Sociologie Rurales
ENSH-ENSP 4, place Hardy
78000 VERSAILLES (FRA)

17

INRA

Ecole Nationale Supérieure Agronomique de Rennes

Chaire d'Economie Rurale

Rennes (FRA)

INRA-ESR-REN

Ecole Nationale Supérieure Agronomique de Rennes

65, rue de Saint-Brieuc

35042 RENNES Cedex (FRA)

18

ENSSAA

Ecole Nationale Supérieure des Sciences Agronomiques Appliquées

Chaire d'Economie Rurale

Dijon (FRA)

INRA-ESR-DIJ

Ecole Nationale Supérieure des Sciences Agronomiques Appliquées

26, Bd du Docteur Petitjean

21100 DIJON (FRA)

\$ MESSAGES D'ERREURS

- 0 : Il y a une erreur mais je ne sais pas la décrire
- 1 : Cette variable est interdite dans ce contexte
- 2 : Cette variable est obligatoire et je ne la trouve pas
- 3 : Il me faut 8 chiffres, pas un de plus, pas un de moins
- 4 : Il me faut un caractère et un seul
- 5 : Les seuls types de document autorisés sont B, J, G, F pour l'instant
- 6 : Les seuls niveaux autorisés sont L, M ou A et je n'en vois pas d'autre
- 7 : Vous fermez une parenthèse ou un crochet non ouvert
- 8 : Vous avez oublié de fermer la parenthèse ou le crochet
- 9 : Il a été convenu de coder cette variable sur une seule ligne
- 10 : Un article dans cette liste dépasse la longueur ici permise
- 11 : Je cherche un code pays. Où est-il donc ?
- 12 : Ce code pays ne figure pas sur mes tablettes
- 13 : Ce code langue m'est inconnu
- 14 : L'existence de ce plan de classement ne m'a pas été signalée
- 15 : Je ne connais pas ce numéro de laboratoire
- 16 : La longueur totale de la cote dépasse le maximum autorisé (74)
- 17 : Cette liste contient trop d'éléments
- 18 : La longueur d'un composant auteur dépasse la limite permise (74)
- 19 : Une seule virgule est autorisée pour séparer le nom du prénom
- 20 : 3 initiales au plus sont autorisées donc 3 points et pas un de plus
- 21 : Une variable du type organisme est composée d'au moins 2 éléments
- 22 : Une variable du type organisme contient au plus 4 éléments
- 23 : Un sigle est composé de lettres capitales consécutives. J'en autorise 12.
- 24 : J'essaie de lire un numéro de labo et je ne reconnais pas un chiffre
- 25 : En fait, je ne connais pas d'autre support que M
- 26 : Les codes autorisés sont S, T ou V
- 27 : La date s'écrit AAAA, ou AAAA/MM, ou AAAA/MM/JJ
- 28 : Les seuls caractères autorisés sont des chiffres
- 29 : Sont admis parmi les chiffres, le tiret, le point et la lettre p
- 30 : Votre ISBN n'a pas la même syntaxe que celle qui m'a été décrite
- 31 : Votre ISSN n'a pas la même syntaxe que celle qui m'a été décrite
- 34 : Je ne trouve pas ce code AGRIS dans la table qui m'a été transmise
- 35 : Une liste de cotes contient au moins 10 éléments
- 36 : Une liste d'auteurs n'excède pas 10 éléments
- 37 : Une seule date de publication est permise
- 38 : Pour un congrès, 2 dates au plus sont autorisées
- 39 : C'est trop : nous nous limitons à 5 codes langues pour le texte
- 40 : C'est trop : nous nous sommes limités à 3 codes langues pour le texte
- 41 : Pour l'instant, je n'accepte qu'un seul plan de classement resagri
- 42 : Pour l'instant, je n'accepte qu'un seul plan de classement agris
- 43 : Plus de 7 mots-clés principaux, c'est trop !
- 44 : Plus de 17 mots-clés secondaires, c'est vraiment trop !
- 45 : Plus de 16 mots clés géographiques, je dois refuser !
- 46 : Je n'accepte pas plus de 3 mots-clés
- 47 : Plus de 12 noms, je ne peux pas prendre !

- 48 : Il m'est interdit d'autoriser plus de 3 candidats descripteurs
49 : Plus de 7 mots clés, ce n'est pas autorisé (tecagri)
50 : Plus de 4 mots clés, ce n'est pas autorisé (tecagri)
51 : Il faut se limiter à 15 mots clés au plus (agrovoc)
52 : Il faut se limiter à 14 mots clés au plus (agrovoc)
53 : Plus de 20 mots clés pour la première base locale, c'est trop !
54 : Plus de 20 mots clés pour la seconde base locale, c'est trop !
56 : Si A3 est présente alors A2 devrait aussi être présente
57 : Si A2 est présente alors A1 devrait aussi être présente
58 : Les trois variables de congrès sont indissociables
59 : Les variables ED et AE sont indissociables
60 : Cette variable est superflue pour un bordereau fils
61 : Si MCS est présente alors MCP devrait être aussi présente
62 : Cette variable est obligatoire pour un bordereau père
63 : Objet documentaire non identifié : j'ai du capituler en cours d'analyse
64 : Le type de document est toujours obligatoire
65 : Le niveau est toujours obligatoire
66 : Le numéro de document est toujours obligatoire
67 : Les variables AD, ainsi que AF ou A1 sont ici obligatoires
68 : Je n'ai pas rencontré l'ombre d'une variable MADO
69 : Plus de 100 lignes après normalisation : MADO doit capituler
70 : J'aperçois un caractère indésirable dans cette adresse (.,:;!/?+*/)
71 : Je ne reconnais pas votre code de documentaliste (2 premiers chiffres)
72 : Les caractères non numériques à droite sont ignorés
73 : La longueur de cette ligne est limitée à 74 caractères
74 : Cette variable devrait être codée sur 3 lignes au plus
75 : Si le plan R2 est présent, le plan R1 devrait aussi être des nôtres
76 : Les variables AU et A1 ne peuvent pas être simultanément absentes
77 : La variable DI est présente alors que AF et A1 sont absentes
78 : La variable AF est présente alors que AU est absente ?
79 : Le code pays n'est pas précédé du nom de la ville
80 : Il y a un point virgule en début de variable
81 : Je recherche un code pays à la fin d'une variable organisme
82 : Mon document père est incorrect ou absent
83 : Cette variable n'est pas présente dans mon document père
84 : Il a été convenu de respecter l'ordre filial
85 : La première variable de la première référence est impérativement REF
86 : Ce code localisation n'est pas autorisé (nos labos sont codés de 1 à 13)
87 : Pour un document non paginé, il faut me fournir cette variable)
88 : J'ai le plan de classement AGRIS (R2). Il me faudrait aussi le plan R1
89 : J'exige la présence de la variable ND mais une seule fois
90 : J'exige la présence de la variable TD mais une seule fois
91 : J'exige la présence de la variable NI mais une seule fois
92 : N'auriez-vous pas fortuitement oublié la variable REF ?
93 : Je capitule après lecture de 100 lignes et je croque la référence
94 : Les seuls indicateurs autorisés sont, à ma connaissance : K, P, U, V, Z
95 : L'indicateur "universitaire" rend obligatoire AF ou A1, AD, DI
96 : L'indicateur "congrès" rend obligatoire la présence de TG, DG, AG.
97 : L'indicateur "bibliographie" rend obligatoire la présence de RB
98 : L'indicateur "non conventionnel" rend obligatoire AF ou A1, AD.
99 : La présence de ED et AE est ici obligatoire

- 100 : Il y un caractère interdit pour désigner un auteur
101 : Ce numéro de laboratoire est incorrect
102 : Il m'est interdit de transcrire l'adresse de cet organisme externe
103 : L'orthographe d'un nom propre nécessite la présence d'une majuscule
104 : Un nom d'auteur contient au moins une lettre majuscule
105 : C'est surprenant de ne pas trouver de majuscule à gauche du code pays
106 : C'est surprenant de ne pas trouver de minuscule à gauche du code pays
107 : Une initiale d'auteur est une lettre majuscule suivie d'un point
108 : Une initiale est toujours suivie d'un point, y compris la dernière
109 : Un tiret (-) doit séparer le préfixe de labo et la cote locale
110 : J'attends un nombre seul, ou suivi de p. , ou précédé de pp (revue)
111 : Une monographie de niveau A, descendante d'un niveau M, est une fille
112 : Pour une monographie, un document de niveau M est toujours père
113 : Je n'accepte que des nombres d'au plus 3 chiffres séparés par le tiret (-)
114 : La variable AF est interdite si le document est anonyme
115 : A ma connaissance, ce préfixe de cote n'est pas autorisé
116 : Une variable organisme contient au moins une lettre
117 : Une variable organisme contient au moins une lettre majuscule
118 : Une variable organisme sans lettre minuscule, ce n'est pas joli
119 : J'attends ici un code pays de 3 lettres, éventuellement entre parenthèses
120 : Seules sont éditées les 30 premières erreurs
121 : A défaut de cote locale, le tiret reste obligatoire derrière le code labo
122 : Je vois un caractère indésirable dans cette variable (.,:;!+*/
123 : Je n'ai pas trouvé la ville à gauche du code pays
124 : La première référence d'un dossier est impérativement un père
125 : La descendance d'un niveau L est nécessairement d'un niveau L
126 : La descendance d'un niveau M est toujours de niveau A
127 : La descendance d'un niveau C est obligatoirement d'un niveau A
128 : Un document de niveau C est nécessairement un père
129 : Une monographie de niveau A, descendante d'un niveau C, est une fille
130 : Pour les publications du Département, la variable AF est obligatoire
131 : Le point-virgule est indésirable et perturbateur derrière un code pays
132 : Il y a après cette variable au moins un nom de variable qui m'est inconnu
134 : La variable AFG est présente alors que AUG est absente
135 : Si A3G est présente alors A2G devrait aussi être présente
136 : Si A2G est présente alors A1G devrait aussi être présente
137 : Les caractères d'indexation sont les lettres, le tiret et l'apostrophe
140 : La variable L10 est absente pour une publication du département
141 : L'indicateur V est interdit si ED et AE sont présents
142 : Il me faudrait ED et AE ou à défaut AD et (AF ou A1)
143 : Il y a au moins une ligne codée sur 80 caractères ou plus derrière moi
144 : Pour un texte de langue anglaise, la traduction du titre est superflue
145 : Pour un texte de langue française, la traduction du titre est superflue
146 : Pour AGRIS (R2 présent), il me faut la variable langue du résumé (LR)
147 : Pour AGRIS (R2 présent), il me faut la variable d'indexation MC9
148 : Pour AGRIS, si la langue du texte (LA) n'est pas l'anglais, il me faut TA
149 : Pour AGRIS, si la langue du texte (LA) n'est pas française, il me faut TF

\$ CODES PAYS

FRA
 AFG ZAF ALB DEU DDR AND AGO SAU
 ARG AUS AUT BHS
 BHR BGD BRB BEL BEN BTN BYS BUR BOL
 BWA BRA BGR
 BDI CMR CAN CAF CHL CHN CYP COL COM
 COG KOR PRK CRI
 CIV CUB DNK DJI DOM EGY SLV ARE ECU
 ESP USA ETH FJI
 FIN GAB GMB GHA GRC GRD GTM GIN GNB
 GNQ GUY HTI
 HVO HND HUN IND IDN IRQ IRN IRL ISL ISR
 ITA JAM JPN
 JOR KEN KHM KWT LAO LSO LBN LBR LBY
 LIE LUX MDG MYS
 MWI MDV MLI MLT MAR MUS MRT MEX
 MCO MNG MOZ NAM NRU
 NPL NIC NER NGA NOR NCL NZL OMN UGA
 PAK PAN PCZ PNG PRY NLD PER PHL POL
 PRT QAT RHO ROM GBR RWA ASM SEN
 SYC SLE SGP SOM SDN LKA SWE CHE SUR
 SWZ SYR TZA TCD
 CSK THA TGO TON TTO TUN TUR UKR SUN
 URY VAT VEN VNM
 YEM YMD YUG ZAR ZMB CPV

\$ CODES LANGUES

FRE
 AFR ALB GER ENG ARA BEL BUL CHI KOR
 CRO DAN SPA
 EST FIN FLE FRE GRE HEB HIN HUN IND
 ICE ITA JAP
 LAT LAV LIT MAC DUT NOR PER POL POR
 RUM RUS SER
 SLO SLV SWE CZE TUR UKR MUL

\$ PLAN DE CLASSEMENT R1

AAB AEA AEB AEC AED AEE AEF AEG AEH
 AEJ AEK AEL AEM AEN AEP AGA AGB AGC
 BA BB BC BDA BDB BE
 CA CB CC CD CE CF CGA CGB CGC
 DA DBA DBB DBC DBD DBE DCA DCB DCC
 DCD DCE DCF
 EA EBA EBB ECA ECB ECC ECD ECE ECF ED
 EEA EEB EEC EED EFA EFB EFC EFD EFE
 EFF EFG EFH EGA EGB EGC EHA EHB EHC
 EHD EJA EJB EJC EJD EJE EJF EJG EJH
 FA FBA FBB FBC FBD FBE FBF FBG FBH
 FBJ FBK FCA FCB FCC FCD FCE FCF FCG
 FCH FCJ FD FE FFA FFB FFC FFD FGA FGB
 FGC FGD FGE
 GA GB GCA GCB GCC GCD GCE GCF GDA
 GDB GDC GDD GDE GDF GDG GDH GEA
 GEB
 GEB GED GEE GEF GFA GFB GFC GFD GFE
 GFF GFG
 HAA HAB HAC HAD HAE HEF HAG HBA
 HBB HBC HBD HBE HBF HCA HCB HCC
 HCD HCE
 HEC HED HEE HFA HFB HFC HFD HFE HFF
 HFG HFH HFJ HGA HGB HGC HHA HHB HHC
 HJA HJB
 HJC HKA HKB HKC HLA HLB HLC HLD HLE
 HLF HLG HLH
 JA JB JCA JCB JCC JCD JCE JCF J

\$PLAN DE CLASSEMENT R2

A01 A50
B10 B50
C10 C20 C30
D10 D50
E10 E11 E12 E13 E14 E16 E20 E21 E40 E50 E51
E70 E71 E72 E73 E 80
F01 F02 F03 F04 F05 F06 F07 F08 F30 F40 F50
F60 F61 F62 F63 F70
H01 H10 H20 H50 H60
J10 J11 J12 J13 14 J15
K01 K10 K11 K50 K70
L01 L02 L10 L20 L40 L50 L51 L52 L53 L60 L70
L72 L73 L74
M01 M11 M12 40
N01 N02 N10 N20
P01 P05 P06 P07 P10 11 P30 P31 P32 P33 P34
P35 P36 P40
Q01 Q02 Q03 Q04 Q05 Q51 Q52 Q53 Q54 Q55
Q60 Q70 Q80
S01 S20 S30 S40
T01
U10 U30 U40

Ce document a été composé à BUROSCOPE
80, avenue des Buttes de Coësmes
35700 RENNES