



HAL
open science

Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales – 2e partie

Gabriel G. Bès, Pierre-Maurice Fauchère

► To cite this version:

Gabriel G. Bès, Pierre-Maurice Fauchère. Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales – 2e partie. Condenser - Adosa, Clermont-Ferrand, 1982, 3, pp.3-31. hal-01117888

HAL Id: hal-01117888

<https://hal.science/hal-01117888>

Submitted on 18 Feb 2015

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.

Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales – 2^e partie

Gabriel G. Bès et Pierre-Maurice Fauchère
Groupe de recherches sur la condensation de l'information en langue naturelle (CILN)
Université Blaise-Pascal, Clermont II

Condenser, Adosa, Clermont-Ferrand, avril 1982, n° 3, p. 3-31

Résumé

Ce document est la suite de la présentation de la grammaire du système documentaire Vercingétorix I (= V1), publiée dans *Condenser*, Adosa, Clermont-Ferrand, janvier 1981, n° 2, p. 41-89. Après l'examen des relations qui permettent de combiner les contenus significatifs du langage, sont explicités les principes directeurs de la composition des contenus significatifs, dont la connaissance par le documentaliste permet de réduire le coût d'acquisition et d'utilisation du système V1. Les contenus significatifs font par ailleurs l'objet de découpages documentaires, produisant des sous-segments pertinents pour l'indexation. Le § 7 expose les règles syntaxiques du langage V1. Le § 8 commente quelques caractéristiques de V1 : les avantages et limites de la naturalité du langage, les inconvénients supposés d'un langage d'indexation à syntaxe (structures synonymes, difficultés d'utilisation, coût du traitement informatique) et les solutions apportées par V1 sur ces points, et enfin la flexibilité du langage V1.

Deux appendices complètent le texte. Le premier résume les caractéristiques du langage V1, le second donne quelques exemples d'indexation.

Voir aussi

Gabriel G. Bès et Pierre-Maurice Fauchère. « Le système documentaire Vercingétorix I ». *Condenser*, Adosa, Clermont-Ferrand, février 1980, n° 1, p. 57-94. <http://hal.archives-ouvertes.fr/hal-01071423>

Gabriel G. Bès et Pierre-Maurice Fauchère. « Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales (1^{re} partie) ». *Condenser*, Adosa, Clermont-Ferrand, janvier 1981, n° 2, p. 39-89. <http://hal.archives-ouvertes.fr/hal-01117883>

Gabriel G. Bès. « Les structures syntaxiques du langage Vercingétorix I. Problèmes liés à leur naturalité. » *Traitement automatique des langues naturelles et systèmes documentaires*, *Condenser*, Supplément n° 1, Adosa, Clermont-Ferrand, 1984, p. 109-134. <http://hal.archives-ouvertes.fr/hal-01071426>

NB. La version 2 de ce document déposée dans HAL ne diffère de la version 1 que par les liens ci-dessus, qui ont dû être corrigés suite à une défaillance du système ayant conduit au remplacement de certains dépôts par d'autres n'ayant rien à voir.

LE SYSTÈME DOCUMENTAIRE VERCINGÉTORIX I

La description du système Vercingétorix I prend la forme de fiches numérotées selon l'ordre de parution et consacrées à chacun des aspects du système. Ces fiches pourront être ultérieurement reprises selon un plan de classement méthodique.

Les considérations plus théoriques et l'examen de la bibliographie afférente figureront sous une rubrique « Linguistique et langages documentaires ».

LE SYSTÈME DOCUMENTAIRE VERCINGÉTORIX 1

(Textes publiés antérieurement)

Condenser n° 1 (février 1980)

G.G. Bès et P.-M. Fauchère : **Le système documentaire Vercingétorix 1.**

Fiche n°1. — Vercingétorix 1. Définition, objectifs et fondements.

Fiche n°2. — Les informations indexées.

Fiche n°3. — Assistance de l'ordinateur à l'indexation. Principes généraux.

Fiche n°4. — Exemples d'indexation (Série n°1).

Condenser n° 2 (janvier 1981)

G.G. Bès et P.-M. Fauchère : **Le système documentaire Vercingétorix 1.**

Fiche n° 5. — Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales.

6 LES CONTENUS SIGNIFICATIFS DANS LE LANGAGE V1. LEUR UTILISATION DOCUMENTAIRE.

6.1 Principes directeurs de la composition des contenus significatifs.

On a montré, au § 5, comment des contenus significatifs sont déterminés par la mise en rapport des contenus significatifs des items individuels, et ceci au niveau soit du noyau avec coordination interne, par le symbole « et », soit du reste de la phrase documentaire, par la relation de projection, que celle-ci se présente sous la forme de relation d'attribut ou de relation verbale. On a également explicité (cf. aussi § 4) la contribution aux contenus significatifs du langage V1 apportée par les relations entre structures grammaticales, l'anaphorique et le classement en ED et en SED permettant d'introduire aussi bien des déterminants dans les items individuels d'un noyau avec coordination interne que le complément d'objet et le retour en arrière dans le groupe A. Dans ce § 6, on explicitera les principaux critères généraux, sous-jacents aux présentations faites ci-dessus sur des points particuliers; celles-ci sont très largement déductibles de ceux-là, d'où l'intérêt pour le documentaliste de connaître ces principes généraux afin de réduire le coût d'acquisition et d'utilisation du système V1. Ces principes généraux sont présentés par la suite, suivis et/ou précédés des commentaires qui permettent de les justifier ou de les comprendre :

(I) *Si le noyau d'un syntagme de phrase n'est pas un TV, ce syntagme ne peut pas être suivi d'un autre syntagme; autrement dit, la structure LN qui correspond à ce syntagme ne peut pas avoir un complément d'objet, direct ou indirect, ou un complément de régime verbal. Par conséquent, ce noyau, constitué par un item T qui n'est pas un TV ou par une coordination d'items T qui ne sont pas des TV, va correspondre en langue naturelle à un noyau de syntagme nominal.*

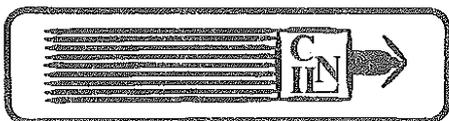
(II) *Si le noyau d'un syntagme de phrase est un TV, la correspondance la plus exacte qu'il aura en langue naturelle est celle d'un nom verbal. Et on sait qu'un nom verbal possède, dans une langue telle que le français, des qualités très utiles pour la documentation:*

(a) *Le nom verbal pose un domaine, mais, à la différence de la forme verbale, n'affirme rien et, en particulier, n'affirme rien sur l'existence de ce domaine. Ainsi, on peut utiliser *indexation automatique* pour un document qui montre qu'une indexation de ce type n'est pas possible, ce qui n'est pas le cas pour l'expression proche *on indexe automatiquement*.*

(b) *A partir de tout infinitif, on peut construire soit un nom verbal reconnu et fixé par la norme linguistique, soit une expression avec le même comportement que le nom verbal. Pour *stocker*, on aura *stockage*, pour *indexer*, *indexation*, pour *financer*, *financement*, pour *aimer*, *amour*, etc., et pour les cas tels que *manger*, on construira *le fait de manger*. Certes, les signifiants de noms verbaux peuvent se construire de manière différente selon les verbes et il se peut qu'il y ait hésitation sur le signifiant fixé par la norme, mais on sait que, d'une part, ceci ne peut constituer une gêne pour le documentaliste (cf. § 2.3) et que, d'autre part, cette variabilité du signifiant n'entraîne aucune conséquence sur le plan documentaire.*

(c) *Le nom verbal réunit en lui certaines caractéristiques du nom et certaines caractéristiques d'une forme verbale. En tant que nom, il peut recevoir les compléments déterminatifs propres à un nom (*l'indexation automatique, l'édition bilingue*) et, en tant que forme verbale, il peut recevoir des compléments déterminatifs paraphrasables par un complément d'objet (*l'indexation des textes*), ou par un complément d'objet indirect ou de régime verbal (*l'accès à une salle*), ou par un sujet (*l'indexation par un documentaliste*), ou par un complément circonstanciel (*la traduction au vingtième siècle*).*

(III) *La vaste gamme de compléments déterminatifs admis par un nom verbal possède de très utiles conséquences pour un langage documentaire tel que V1 : on peut, à peu près, lui accrocher toute sorte de complément, paraphrasable par n'importe quelle fonction (sujet, complément d'objet direct ou indirect, complément de régime verbal, circonstanciel, qualificatif, etc.). Ceci peut être source de non-univocité, c'est-à-dire, de la possibilité d'associer deux signifiés incompatibles à un même complément déterminatif d'un nom verbal; voir, par exemple, le cas bien connu de *le choix du directeur*, interprétable soit comme *le directeur a choisi*, soit comme *on a choisi le directeur*. Pour porter remède à*



Fiche n°5 (décembre 1980)
G. G. Bès et P.-M. Fauchère

Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales

Deuxième partie

SOMMAIRE

Première partie
(parue dans *Condenser* n°2)

- 1 Introduction.
 - 1.1 Terminologie et conventions. 1.1.1 Terminologie et conventions relevant de V1. 1.1.2 Mise au point terminologique sur la langue naturelle.
- 2 L'organisation du lexique.
 - 2.1 Notions. 2.2 Procédés. 2.3 Les supports de radical normalisés. 2.4 Le système de traits.
- 3 Les structures grammaticales ; contraintes combinatoires et relations internes.
 - 3.1 Le syntagme ; le syntagme à noyau simple. 3.2 La combinatoire entre syntagmes ; la phrase documentaire. 3.2.1 Les relations internes de la phrase documentaire constituée de plus d'un syntagme. 3.3 Le groupe d'attributs dans le syntagme de phrase. 3.3.1 Le marqueur de relations *b*. 3.3.2 La spécification de la relation d'attribut. 3.4 Le noyau à coordination interne. 3.5 Récapitulation des relations internes à la phrase documentaire. 3.6 Les constructions documentaires.
- 4 Relations entre structures grammaticales.
 - 4.1 La sous-structure anaphorique. 4.2 Le classement des structures grammaticales. 4.3 Le contenu significatif anaphorique. 4.4 Relations entre structures grammaticales et classement des structures grammaticales : le rôle de l'anaphorique.
- 5 Les contenus significatifs dans le langage V1. Relations constitutives.
 - 5.1 Entités de la phrase documentaire avec contenu significatif. 5.1.1 Le contenu significatif du noyau à coordination interne. 5.1.2 La relation de projection. 5.1.2.1 La relation d'attribut. 5.1.2.1.1 La relation d'attribut sans spécificateur A et avec une seule entité portant sur le noyau. 5.1.2.1.2 La relation d'attribut avec spécificateur A et avec une seule entité portant sur le noyau. 5.1.2.1.3 La relation d'attribut avec ou sans spécificateur A et avec plusieurs entités portant sur le noyau. 5.1.2.2 La relation verbale. 5.2 Le contenu significatif de la construction documentaire.

Deuxième partie

- 6 Les contenus significatifs dans le langage V1. Leur utilisation documentaire.
 - 6.1 Principes directeurs de la composition des contenus significatifs. 6.2 Contenus significatifs des structures grammaticales et découpages documentaires.
 - 7 Les règles syntaxiques des structures grammaticales.
 - 8 Le langage V1 dans le cadre documentaire.
 - 8.1 Avantages et limites de la naturalité. 8.2 Syntaxe et structuration de l'information. 8.3 Flexibilité du langage.
- Appendice I. — Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales (résumé).
- Appendice II. — Mise à jour des exemples d'indexation de la fiche n°4.

cette situation, la grammaire de V1 s'organise autour des principes de base suivants (qui possèdent, par ailleurs, d'autres répercussions heureuses sur le plan du réseau sémantique) :

Le complément d'objet direct ne peut jamais s'exprimer dans le cadre du groupe A.

Si une entrée lexicale fixe un complément verbal de t1 qui correspond à un complément d'objet indirect ou de régime verbal, ce complément ne peut pas s'exprimer dans le cadre du groupe A.

Tous les autres compléments de la phrase, et son sujet, peuvent s'exprimer par la relation d'attribut dans le groupe A.

Le complément d'objet direct ne peut s'exprimer que par le biais de la relation verbale.

(IV) La liberté expressive accordée à la relation d'attribut peut devenir source d'inconvénients. Puisque la machine va reconnaître la relation d'attribut et non les articulations introduites par le documentaliste, le langage V1 risque d'oblitérer des distinctions qui existent en langue naturelle et qui sont importantes sur le plan documentaire. Face à cette question importante, la position adoptée par V1 peut se résumer ainsi :

(a) *Comme hypothèse de travail, on a posé le principe général selon lequel, si une entité A porte sur un noyau (c'est-à-dire si elle est en relation d'attribut avec lui), c'est le contenu significatif de l'item qui est complément A de ce noyau qui va permettre de préciser le contenu spécifique de la relation d'attribut.* Par exemple, si *indexation* est suivi de *documentaliste*, on déduira que *documentaliste* correspond à un item A dont la relation d'attribut est paraphrasable par un sujet ou complément d'agent (*l'indexation par un documentaliste*) ; si *indexation* est suivi d'*automatique*, on déduira que *automatique* possède une fonction qualificative (*indexation automatique*). Les entités A qui, placées après un « b », porteraient sur *documentaliste*, ne devraient rien modifier à ce qu'apporte le contenu significatif de *documentaliste* en ce qui concerne la précision de la relation d'attribut.

(b) *Le spécificateur A est l'instrument formel pour préciser de manière contrôlée les distinctions que l'on souhaite introduire dans la relation d'attribut.* V1, dans sa version actuelle, en introduit deux, mais rien n'empêche d'en ajouter d'autres, si les exigences documentaires le justifient.

(V) *Les items qui sont noyau et/ou complément d'un noyau assurent les contraintes sémantiques de sélection* (ce principe est en rapport étroit avec celui qui précède).

On sait que, pour des raisons « sémantico-pragmatiques », toute forme verbale ne peut être suivie de n'importe quel complément d'objet et que tout nom ne peut être suivi de n'importe quel complément déterminatif : *on n'indexe pas l'agriculture* et il est difficile de concevoir une *discothèque syntaxique*. Il existe donc des contraintes sémantiques entre un noyau et les compléments qu'il admet — ce sont les contraintes sémantiques de sélection qui, encore une fois, jouent un rôle important dans le réseau sémantique. *Dans le cas le plus général, le documentaliste n'a pas à se soucier des contraintes sémantiques de sélection* : il n'aura pas à se demander s'il peut écrire *indexation t1 de l'agriculture t0*, simplement parce que dans son entrée — les documents qu'il traite — on n'en parle pas.

Mais, parfois, certaines entrées lexicales peuvent fixer les caractéristiques sémantiques des contenus significatifs qui peuvent se mettre en relation les uns avec les autres ; or, dans tous les cas, ces caractéristiques sont fixées pour l'item qui est noyau et/ou l'item qui est son complément verbal ou son complément A. Si ces items possèdent les caractéristiques exigées, tout le reste du contenu significatif du syntagme (syntagme A ou syntagme de phrase) ou du bloc (bloc A ou bloc de syntagmes de phrase) auquel appartient chacun des items concernés par les contraintes de sélection est indifférent pour le fonctionnement des contraintes de sélection. Exemple :

Supposons qu'*indexation* exige comme complément verbal un item qui possède le trait [+ document]. On pourra écrire *indexation t1 d'un livre t0*, *indexation t1 d'une thèse t0*, *indexation t1 d'un périodique t0*, etc., *livre*, *thèse* et *périodique* possédant, parmi d'autres notions, le trait [+document]. Or, *livre* (ou *thèse*, ou *périodique*) admet un groupe A, mais la contrainte de sélection ne tient pas compte de l'organisation des contenus significatifs dans ce groupe A : il faut et il suffit qu'elle soit satisfaite par le seul item complément.

(VI) *La répétition des emboîtements — que ce soit dans le groupe A ou dans la phrase documentaire — peut s'étendre indéfiniment, la seule restriction en est le contenu même du document qu'on veut indexer. Cela permet d'ajouter des précisions aux précisions.*

Par exemple, un noyau de syntagme de phrase étant donné, on peut le préciser par un item A qui porte sur lui, mais cet item A, suivi du symbole « b », peut à son tour être précisé par un autre item A, et ainsi de suite. De même, le complément d'objet direct, représenté par un syntagme de phrase X qui porte sur un syntagme de phrase Y, peut avoir un complément d'objet direct lui-même. C'est, par exemple, le cas de *financement t1 du traitement t1 de l'indexation t1*, où *indexation t1* correspond à un complément d'objet de traitement. Ces emboîtements « en cascade » — aussi bien dans le groupe A qu'entre syntagmes de phrase — obéissent à une contrainte très stricte (elle est une conséquence des contraintes combinatoires, cf. § 3) : *si B porte sur A, on peut introduire à la suite de B une autre entité qui porte sur B mais qui ne peut en aucun cas porter directement sur A.*

Exemple : *Catalogue t0 d'un centre de documentation a3 spécialisé en s médecine a1* ne peut jamais correspondre à *catalogue spécialisé en médecine d'un centre de documentation*. De même *traduction t1 de l'indexation t1 automatique a1* ne peut jamais correspondre à *traduction automatique de l'indexation*.

(VII) *Les possibilités du langage V1 ne s'épuisent pas dans les contenus significatifs spécifiés par la grammaire de V1. Celle-ci détermine les structures grammaticales de V1, mais le langage V1 incorpore la spécification de relations entre structures grammaticales, qui déterminent d'autres contenus significatifs qui viennent compléter ceux qui sont déterminés strictement par la grammaire.*

L'anaphorique et le classement en ED ou SED, présentés au § 4, permettent l'introduction dans une structure grammaticale de : (1) un complément déterminatif portant sur l'un seulement des items qui composent un noyau avec coordination interne, cf. § 5.1.1 *in fine*; (2) le complément d'objet direct, ou, pour les entrées lexicales qui l'exigent, le complément d'objet indirect ou de régime verbal, *dans le groupe A*, cf. § 5.1.2.1.1 *in fine*; (3) le « retour en arrière » dans ce même groupe A, cf. § 5.1.2.1.3 *in fine*.

L'utilisation de l'anaphorique exige l'existence d'une identité référentielle entre structures grammaticales différentes. En revanche, l'instruction de commutation (cf. fiche n° 2, p. 3-4), qui est essentiellement une commodité d'écriture pour le documentaliste, peut être utilisée si l'on souhaite produire deux ou plus de deux structures grammaticales *qui soient partiellement identiques et partiellement différentes au niveau du langage, mais sans qu'il y ait une identité référentielle entre elles*. Exemples :

- (1 a) Indexation t1 automatique a1 par un centre de documentation a3.
 1 2 3
- (1 b) Traduction t1 automatique a1 par un centre de documentation a3.
 1 2 3
- (2 a) Édition t1 en URSS a1 de livres t0 techniques a1.
 1 2 3 4
- (2 b) Édition t1 aux USA a1 de livres t0 techniques a1.
 1 2 3 4

Dans le cas de la première paire, pour obtenir (1 b), on utilisera, à la suite de (1 a), l'instruction de commutation suivante :

& 1 = traduction t1.

et dans le deuxième cas, pour obtenir (2 b), celle qui suit, qui opère sur (2 a) :

& 2 = aux USA a1.

Il est clair que (1 a) et (1 b) d'une part, (2 a) et (2 b) d'autre part, sont des paires de structures grammaticales qui ont des éléments en commun et des éléments différents, l'identité et la différence se situant au niveau de l'expression par le langage et ne portant pas sur le monde référentiel : on ne dit pas que *l'indexation* est considérée *avec* ou *en même temps* que *la traduction*, constituant une entité complexe au niveau du référent, ni qu'il s'agit référentiellement du même processus d'édition en URSS et aux USA ; si on avait voulu exprimer ces derniers contenus — le deuxième très improbable sur le plan

référentiel — on sait qu'on aurait dû respectivement écrire :

- (1) Indexation t1 et traduction t1 automatiques a1 par un centre de documentation a3.
- (2) Édition t1 en URSS a1 et aux USA a1 de livres t0 techniques a1.

Les principes I à VI valent aussi pour l'organisation des constructions documentaires ; ce sont des principes qui relèvent strictement de la grammaire de V1. Ils conditionnent les correspondances suivantes entre contenus significatifs des constructions documentaires de V1 et structures LN :

Si la phrase documentaire est constituée d'un seul syntagme, elle correspondra à un syntagme nominal, avec le noyau du syntagme de phrase correspondant à un noyau nominal et les entités A qui portent sur lui correspondant à des compléments déterminatifs juxtaposés.

Si la phrase documentaire est constituée de deux ou plusieurs syntagmes de phrase, son contenu significatif s'organisera comme celui d'un bloc de syntagmes de phrase.

Un bloc de syntagmes de phrase correspond à un syntagme nominal qui possède un complément déterminatif paraphrasable par un complément d'objet direct (cas général) ou un complément d'objet indirect ou de régime verbal (cas particulier) qui sont, dans l'un et l'autre cas, en rapport avec une forme verbale. Cette forme verbale, en langue naturelle, correspond au noyau du premier syntagme de phrase, à gauche du bloc ou de la phrase documentaire. Les entités A qui portent sur ce même noyau et qui n'ont pas de spécificateur A vont correspondre à des compléments déterminatifs, chacun étant paraphrasable par un sujet, ou par un complément à fonction qualificative, ou par un complément circonstanciel ou, si l'entrée lexicale du noyau ne les exige pas, par un complément d'objet indirect ou par un complément de régime verbal, mais jamais par un complément d'objet direct, ni, si l'entrée lexicale les exige, par un complément d'objet indirect ou de régime verbal. Les entités A qui portent sur ce même noyau et qui possèdent un spécificateur A vont correspondre à des compléments déterminatifs dont la relation avec le noyau nominal est paraphrasable par le signifié du spécificateur A.

6. 2 Contenus significatifs des structures grammaticales et découpages documentaires

La question de la mise en rapport des contenus ponctuels qui apparaissent dans une phrase est inhérente à tout langage qui incorpore des relations syntaxiques : celles-ci sont, intrinsèquement, des moyens pour combiner des entités simples afin d'obtenir une structure plus complexe, le contenu associé à cette structure plus complexe n'étant pas équivalent à la simple addition ou à la juxtaposition des contenus des entités séparées. Les pages précédentes ont été consacrées à la présentation du système de relations incorporé à la grammaire de V1 et aux contenus significatifs qu'elles déterminent. On remarquera que le mécanisme formel de constitution des contenus significatifs possède certaines propriétés générales.

Dans tous les cas, on a une ou plusieurs entités X qui sont mises en rapport entre elles (la coordination dans le noyau) ou qui portent sur une entité Y (composition par la relation de projection). Comme résultat, on crée une autre entité Z. Cette entité Z va porter sur une autre entité Z' pour créer une entité Z'', et ainsi de suite jusqu'à épuisement de la structure grammaticale. On remarquera qu'entre la ou les entités X et l'entité Y, il n'y a pas de chevauchement de contenu, que le contenu de Z n'est pas la juxtaposition des contenus de X et de Y, mais qu'il possède des caractères propres, déterminés par ce que sont les entités X et Y et le type de mise en rapport permettant de produire Z. Le procédé de composition par projection est formellement uniforme : il opère par embloquements successifs à partir des entités de droite, et ceci aussi bien dans le groupe A que dans la phrase documentaire et la construction documentaire. Ces caractéristiques de la grammaire de V1 permettent de spécifier des contenus significatifs à différents niveaux de complexité, pour des parties et des sous-parties des structures grammaticales — les *contenus significatifs intermédiaires* — mais aussi pour les structures grammaticales dans leur ensemble.

Il faut souligner que ce système de relations *n'implique pas* :

- (a) que le documentaliste doit produire mentalement ou écrire effectivement les

structures grammaticales suivant l'ordre de droite à gauche des emboîtements qui conditionnent les contenus significatifs : il ne sera contrôlé, sur le plan des structures grammaticales, que sur les caractéristiques du produit créé — les structures grammaticales constituées — et non sur la manière de le concevoir. La grammaire *spécifie* les structures grammaticales, le documentaliste doit *produire* les structures grammaticales. Il les produit en interaction avec la machine et, si pour l'étape finale de l'introduction en machine de l'information indexée, il doit s'accommoder de ce processus interactif — qui a, d'ailleurs, été conçu pour lui rendre cette opération naturelle et commode, il peut organiser comme il l'entend les étapes préalables.

(b) que la machine doit interpréter les structures grammaticales suivant un balayage de droite à gauche. On n'abordera pas dans cette fiche la question des procédés informatiques de reconnaissance des structures grammaticales de V1 ; on se limitera à remarquer que les structures grammaticales de V1 possèdent toute l'information nécessaire pour que la machine puisse reconnaître les relations qui déterminent les contenus significatifs ; le procédé de lecture de gauche à droite, ou son contraire ou un procédé mixte, relève d'une question d'optimisation informatique.

(c) que les contenus significatifs déterminés par la grammaire de V1 épuisent tous les contenus d'une structure grammaticale qui sont pertinents sur le plan documentaire.

Si les points (a) et (b) sont déductibles des distinctions déjà établies sur les rôles du documentaliste, de la grammaire et d'une machine qui interprète les structures grammaticales produites par le documentaliste (cf. § 1), le point (c) mérite qu'on s'y attarde.

Les structures grammaticales suivantes :

- (1) Financement t1 de la traduction t1 des textes t0 anciens a1.
- (2) Gestion t1 de l'indexation t1 automatique a1 des documents t0 techniques a1.
- (3) Microfilms t0 d'un centre de documentation a3 b automatisé a4 en Inde a1 sur s l'agriculture a1.
- (4) Livres t0 imprimés a1 b par des maisons d'édition a3 b spécialisées a4.

possèdent les emboîtements et les contenus significatifs qui sont déterminés par la grammaire de V1. Mais il est légitime d'observer que chacune de ces structures possède d'autres possibilités d'organisation du contenu significatif, qui peuvent être pertinentes sur le plan documentaire et qui ne coïncident pas avec les contenus significatifs. C'est par exemple le cas des groupements d'items présentés ci-dessous ; ils relèvent des exemples précédents (1) à (4), et déterminent des contenus significatifs correspondant aux structures LN placées à leur droite :

(1) Financement t1 de la traduction t1 Traduction t1 des textes t0	Financement de la traduction Traduction des textes
(2) Gestion t1 de l'indexation t1 Gestion t1 de l'indexation t1 automatique a1 Indexation t1 automatique t1	Gestion de l'indexation Gestion de l'indexation automatique Indexation automatique
(3) Microfilms t0 d'un centre de documentation a3 Centre de documentation t3 automatisé a4 Centre de documentation t3 sur s l'agriculture a1 Centre de documentation t3 en Inde a1 sur s l'agriculture a1	Microfilms d'un centre de documentation Centre de documentation automatisé Centre de documentation sur l'agriculture Centre de documentation en Inde sur l'agriculture
(4) Livres t0 imprimés a4 b par des maisons d'édition a3	Livres imprimés par des maisons d'édition

Les contenus significatifs ainsi déterminés possèdent les caractéristiques suivantes :

- (1) ils ne coïncident pas avec les contenus significatifs spécifiés par la grammaire ;
- (2) ils sont le résultat d'une mise en rapport des contenus significatifs des items individuels ;
- (3) ils ne possèdent pas les qualités requises pour que les propriétés du mécanisme formel de

composition des contenus significatifs restent valables (les contenus partiels peuvent chevaucher entre eux, à la différence des contenus significatifs des entités qui portent les unes sur les autres); (4) ils s'intègrent dans les contenus significatifs des structures grammaticales; (5) ils ne se construisent pas de manière hasardeuse ou arbitraire. On les appellera *contenus significatifs découpés* et on appellera *découpages documentaires* les groupements d'items qui les déterminent. Si le besoin se fait sentir, on nommera *contenu significatif grammatical* ce que jusqu'ici on a nommé *contenu significatif*.

Les découpages documentaires relèvent des structures schématiques : c'est une autre manière de mettre en rapport les entités de ces structures schématiques, « autre » par rapport à celle qui est déterminée par la grammaire de V1. Mais les entités qui composent les découpages documentaires sont les mêmes que celles qui composent les entités syntaxiques du langage qui, elles, sont associées aux contenus significatifs grammaticaux. Par ailleurs, les relations entre les entités des découpages documentaires sont, elles aussi, tirées de celles qui existent dans les structures grammaticales. C'est pour ces raisons que les contenus significatifs découpés s'intègrent dans les contenus significatifs grammaticaux et ne se construisent pas de manière hasardeuse. On observera, en effet, que les découpages suivants détermineraient des contenus significatifs *non intégrables* dans les contenus significatifs grammaticaux des structures grammaticales respectives; par conséquent, ces découpages *ne sont pas* des découpages documentaires possibles des structures grammaticales (1) à (4) et doivent être exclus par le système :

(1) Financement t1 des textes t0	Financement des textes
Financement t1 ancien a1	Financement ancien
Traduction t1 ancienne a1	Traduction ancienne
(2) Gestion t1 automatique a1	Gestion automatique
Indexation t1 technique a1	Indexation technique
Documents t0 automatiques a1	Documents automatiques
Gestion t0 technique a1	Gestion technique
(3) Microfilms t0 automatisés a1	Microfilms automatisés
L'agriculture t0 en Inde a1	L'agriculture en Inde
(4) Livres t0 spécialisés a4	Livres spécialisés
Maisons d'édition t0 imprimées a4	Maisons d'édition imprimées

Les découpages documentaires jouent un rôle très important sur le plan documentaire : on peut en effet prévoir qu'il sera tout à fait exceptionnel qu'une question documentaire coïncide exactement avec une structure grammaticale appartenant à l'information indexée. Le cas le plus général sera celui de coïncidences partielles, qui pourront exister soit entre les questions documentaires et les contenus significatifs grammaticaux intermédiaires, qui sont associés aux entités syntaxiques du langage, soit entre les questions documentaires et les contenus significatifs découpés. Ces coïncidences entre questions documentaires et groupements d'items, que ceux-ci soient des découpages documentaires ou des entités syntaxiques du langage, pourront être exploitées par la machine, dans la phase de recherche documentaire, pour la spécification ou la généralisation des questions documentaires, en liaison avec le fonctionnement du réseau sémantique. Mais le documentaliste n'a à se soucier d'indiquer formellement ni les contenus significatifs grammaticaux intermédiaires, ni les découpages documentaires : c'est à la machine à les repérer automatiquement.

Le rôle des relations syntaxiques sur le plan documentaire peut ainsi être entièrement compris : ces relations permettent de créer des structures complexes, aux contenus composites, qui, en tant que tels, représentent le contenu des documents, mais elles permettent aussi de définir les limites dans lesquelles peuvent se créer les découpages documentaires, dont la raison d'être n'est pas d'exprimer le contenu d'un document, mais de déterminer les possibilités de coïncidence partielle entre une question documentaire et un aspect de ce contenu. Les relations internes aux structures grammaticales, les entités syntaxiques déterminées par ces relations et les contenus significatifs grammaticaux associés à ces entités syntaxiques fonctionnent ainsi comme un système de référence — une espèce de système de coordonnées — par rapport auquel les questions documentaires peuvent découper des zones partielles dont la spécification sera faite en fonction du système de référence. Il est clair que, sans ce système, il serait impossible de préciser, à partir d'une question documentaire déterminée et pour un document donné, quels découpages documentaires sont possibles.

7 LES RÈGLES SYNTAXIQUES DES STRUCTURES GRAMMATICALES

Il est possible de présenter de manière compacte les règles syntaxiques qui permettent de spécifier les structures grammaticales de V1; on utilisera une notation empruntée à la grammaire générative. Les règles 1 à 14 sont des règles de réécriture indépendantes du contexte; les règles I à IV sont des transformations exprimées de manière très peu technique.

Conventions sur les symboles

→	« se réécrit comme » (le symbole qui apparaît à gauche d'une « → » doit être remplacé par le ou les symboles qui apparaissent à droite).
{ }	choix obligatoire d'un parmi les éléments internes, soit disposés en colonnes, soit séparés par une virgule.
()	choix optionnel de l'élément interne ou d'un parmi les éléments internes.
[]	symbole complexe constitué par les éléments qui sont à l'intérieur des crochets; le symbole à gauche de « : » possède les traits indiqués à droite.
#	symbole limite.
Str. Gram.	structure grammaticale.
PD	phrase documentaire.
CD	construction documentaire.
Synt. Phrase	syntagme de phrase.
Nc	noyau à coordination interne.
NVc	noyau verbal à coordination interne.
A	symbole catégoriel des items A.
T	symbole catégoriel des items T.
TV	symbole catégoriel des items verbaux.
TR	symbole catégoriel des termes repris.
RD	symbole catégoriel des radicaux discriminants.

Règles syntaxiques 1 à 14

1	# Str. Gram. # → { PD } { CD }
2	CD → PD Prédictat r PD
3	PD → Synt. Phrase
4	Synt. Phrase → { T (Groupe A) { TV (Groupe A) (Synt. Phrase) Nc (Groupe A) NVc (Groupe A) (Synt. Phrase) }
5	Nc → T et T (et Nc')
6	Nc' → T (et Nc')
7	NVc → TV et TV (et NVc')
8	NVc' → TV (et NVc')
9	Groupe A → A ((b) groupe A)
10	Prédictat → { avec l'objectif d(e), au moyen d(e), en relation avec, en coopération avec, influençant, provoquant, entraînant }
11	T → { { TR { [RD : + t0] } } { t0, t3 } } TV
12	TV → { TR { [RD : + t1] } } { t1, +t1, -t1, t2, t3 }
13	A → { TR { a1, +a1, -a1, a2, a3 } RD { a1, +a1, -a1, a2, a3, a4 } } }
14	TR → Forme anaphorique (Référence) Terme repris (Développement de l'anaphorique)

Règles syntaxiques I à IV

I Les symboles [RD : +t0], [RD : +t1] et RD dans une suite x engendrée par les règles 1 à 14 doivent être remplacés par un radical discriminant du lexique, précédé, le cas échéant, par le signifiant du spécificateur de notion. Les radicaux discriminants qui se substituent à ces symboles doivent être marqués dans le lexique comme admettant le procédé qui est indiqué immédiatement à droite de ces symboles dans la suite x. Dans le cas du remplacement des symboles [RD : +t0] et [RD : +t1], les radicaux discriminants qui se substituent à eux doivent, en plus, posséder dans le lexique respectivement les traits [+t0] et [+t1] (les traits sont acquis automatiquement par les entrées lexicales en fonction de leur capacité d'admettre respectivement les procédés t0 et t1).

II Le signifiant d'un spécificateur A peut être introduit comme premier élément d'une sous-structure avec un item A. Le spécificateur A « fonction » exige que cet item A soit un item a1 ou +a1. Le spécificateur A « domaine » exige que l'item A soit complément A d'un noyau dont le ou les items portent les traits [+document] et/ou [+domaine].

III La préposition de complément verbal doit être introduite comme premier élément d'une sous-structure possédant un item T qui est complément verbal d'un item T dont l'entrée lexicale exige cette préposition.

IV @ (=articulation libre) peut être introduit dans les contextes qui suivent :

- (a) immédiatement avant et immédiatement après un radical discriminant ;
- (b) immédiatement avant un signifiant de spécificateur de notion ;
- (c) immédiatement avant une forme anaphorique, pourvu que celle-ci ne soit pas précédée d'un signifiant de spécificateur A ou d'une préposition de complément verbal.

Par application des règles 1 à 14, on obtient les suites :

- (1) # [RD : +t0] t0 RD a1 #
- (2) # [RD : +t1] t1 RD a1 [RD : +t0] t0 RD a1 RD a1 #
- (3) # [RD : +t1] t3 RD a1, en coopération avec r forme anaphorique TR t0 #
- (4) # [RD : +t1] t1 RD a1 [RD : +t0] t0 #

Le rôle de la règle I est de substituer un radical discriminant particulier du lexique aux symboles [RD : +t0], [RD : +t1] et RD ; à la suite de son application, les suites précédentes vont, par exemple, devenir :

- (1) # BIBLIOTHEQ t0 MEX a1 #
- (2) # INDEXA t1 AUTOMATIQ a1 TEXTE t0 TECHNI a1 TRAD a1 #
- (3) # Centre DOCU t3 AUTOMATIQ a1, en coopération avec r forme anaphorique TR t0 #
- (4) # ACCE t1 LIBRE a1 BIBLIOTHEQ t0 #

La règle II introduit le signifiant du spécificateur A ; étant donné ses contraintes, elle ne peut s'appliquer qu'à (2'), pour donner :

- (2^{II}) # INDEXA t1 AUTOMATIQ a1 TEXTE t0 TECHNI a1 sur s TRAD a1 #

La règle III permet d'introduire la préposition exigée par le noyau verbal ; (4^I) deviendra :

- (4^{III}) # ACCE t1 LIBRE a1 à BIBLIOTHEQ t0 #

Par application de la règle IV, on introduit les emplacements réservés aux articulations libres

- (1^{IV}) # BIBLIOTHEQ @ t0 @ MEX @ a1 #
- (2^{IV}) # INDEXA @ t1 AUTOMATIQ @ a1 @ TEXTE @ t0 TECHNI @ a1 sur s @ TRAD @ a1 #
- (3^{IV}) # Centre @ DOCU @ t3 AUTOMATIQ @ a1, en coopération avec r forme anaphorique TR t0 #
- (4^{IV}) # ACCE @ t1 LIBRE a1 à @ BIBLIOTHEQ @ t0 #

Les structures grammaticales précédentes vont correspondre aux structures grammaticales complètes qui suivent (on suppose qu'un terme repris particulier s'est substitué à TR) :

- (1) Bibliothèques t0 au Mexique a1.
- (2) Indexation t1 automatique a1 des textes t0 techniques a1 sur s la traduction a1.
- (3) Centre de documentation t3 automatique a1 en collaboration avec r cette bibliothèque t0.
- (4) Accès t1 libre a1 à la bibliothèque t0.

8 LE LANGAGE V1 DANS LE CADRE DOCUMENTAIRE

La conception du langage de V1, qu'il s'agisse des structures engendrées par la grammaire au sens strict ou des relations que le langage permet d'établir entre ces structures, a été dirigée par le souci de satisfaire le mieux possible les conditions d'adéquation externes que constituent les exigences de l'indexation et de la recherche documentaire automatisée; c'est en fonction de celles-ci que le langage doit être jugé. Or ces exigences entrent en conflit les unes avec les autres. Il est souhaitable d'opter pour un langage dont l'acquisition et le maniement soient aussi aisés que possible pour le documentaliste; on est donc orienté vers la recherche de la « naturalité », c'est-à-dire d'un rapport étroit entre les structures du langage documentaire et celles de la langue naturelle. Mais il faut éviter les inconvénients de l'usage de la langue naturelle dans la recherche documentaire (synonymie, polysémie et ambiguïtés de toute sorte, redondance) et la formalisation doit être suffisante pour permettre le traitement par la machine. Pour répondre aux besoins précis de l'utilisateur et réduire le bruit documentaire, on est tenté d'affiner les procédés d'indexation, mais on risque à l'inverse de provoquer du silence. Si l'on multiplie les moyens d'expression très spécialisés, on améliore sans doute l'adaptation du langage à un domaine particulier, mais on en rend l'apprentissage plus difficile, le traitement par l'ordinateur plus lourd et plus coûteux, la transposition à d'autres domaines malaisée. Et ainsi de suite. Les facteurs qui entrent en jeu sont trop complexes et leur interaction n'a pas été suffisamment étudiée pour qu'il soit possible de dire sur un plan théorique où se situe dans tel ou tel cas le point d'équilibre; et l'on en est réduit bien souvent aux tâtonnements et aux leçons de l'expérience.

V1 représente donc une tentative et une voie de recherche, parmi d'autres possibles, pour organiser de manière optimale les relations entre les trois entités qui interviennent dans la mise en œuvre du système documentaire et qui permettent le transfert de l'information du document vers l'utilisateur final : la machine, le langage et le documentaliste (qui, dans la perspective où se place V1, reste un intermédiaire indispensable pour une représentation riche et correcte du contenu des documents, cf. fiche n° 1). On commentera ci-dessous un certain nombre de caractéristiques de V1 qui dépendent des choix faits au niveau du langage et on essaiera de montrer comment elles situent V1 par rapport aux exigences externes qui s'imposent à lui.

8.1 Avantages et limites de la naturalité

Puisque V1 devait être un système de recherche documentaire *automatisée*, les informations indexées devaient nécessairement avoir les propriétés requises pour pouvoir être traitées de manière univoque par un ordinateur, mais on voulait également exploiter l'idée qu'un langage documentaire construit en fonction de certaines structures des langues naturelles pouvait être appris plus rapidement et manié plus commodément qu'un langage entièrement artificiel.

La manifestation la plus immédiate de la naturalité de V1 se trouve sans doute dans l'apparence même des expressions documentaires, qui sont directement lisibles par l'être humain. Cette lisibilité directe facilite le dialogue entre l'homme et la machine et améliore les conditions de travail du documentaliste, mais il ne peut s'agir d'une naturalité totale car la *fonction de lisibilité* (par l'être humain) doit se concilier avec la *fonction de compréhension* (par la machine), qui exige un certain degré de formalisation (cf. fiche n° 2, § 2. 2). Cette problématique de V1 fait penser à celle qui a présidé à la conception des systèmes de caractères pour la reconnaissance optique OCR/A et OCR/B de l'ISO (Organisation internationale de normalisation); ces systèmes de caractères doivent être à la fois lisibles pour un être humain, sans apprentissage spécial, et reconnaissables par les appareils de lecture optique. Dans ce cas comme dans celui de Vercingétorix, on cherche à satisfaire *simultanément* (par la fusion intime de l'information de base et des articulations en ce qui concerne V1) — et non parallèlement — les exigences minimales de l'être humain et

celles de la machine, sans pour autant que les solutions adoptées présentent nécessairement les qualités optimales par rapport à chacune de ces entités prise séparément.

Si l'on peut penser que le langage de V1 est, malgré une certaine complexité, plus rapide à acquérir et plus facile à utiliser qu'un langage complètement artificiel, ce n'est pas seulement parce que la fonction de lisibilité est satisfaite et que les unités du lexique, comme dans un simple thésaurus, renvoient à des entités du monde référentiel qui sont désignées ou peuvent être désignées de la même façon dans la langue naturelle, mais, plus profondément, parce que la naturalité existe aussi dans l'organisation même des structures grammaticales. Les relations syntaxiques ne sont pas des relations abstraites empruntées à la logique par exemple; il ne s'agit pas non plus de l'interprétation des rôles que peuvent jouer les différentes entités les unes par rapport aux autres en grandes fonctions sémantiques, plus ou moins arbitrairement découpées, telles que agent, lieu, temps, moyen, etc., qui risquent de présenter souvent des chevauchements. Ces relations sont la transposition dans V1 de certaines des relations les plus simples et les plus spontanément utilisées par le documentaliste dans sa propre langue. C'est dans la connaissance intuitive que l'indexeur a de sa langue naturelle, langue qui sert de référence à la naturalité de V1, qu'il peut trouver les critères qui lui permettront de déterminer quels types de relations doivent exister dans les structures grammaticales pour exprimer le contenu d'un document ou d'une question documentaire; ainsi, c'est par référence à la langue naturelle qu'il pourra savoir si la relation entre un item t1 et une notion qui doit le déterminer doit être rendue par un complément verbal ou par un item A. Les critères que fournit, pour l'indexation avec V1, la langue naturelle sont clairs, accessibles par l'intuition sans qu'il soit besoin de définitions complexes, indépendants des connaissances, de l'expérience ou de la personnalité de l'indexeur et partageables par plusieurs indexeurs. Les procédés admis par chaque notion sont déductibles du mot de rappel, qui est lui-même déterminé par règle générale: le documentaliste n'a donc pas à les mémoriser.

Mais cette recherche de la naturalité connaît nécessairement des limites, principalement dans deux directions:

a) Le langage documentaire doit jouer le rôle de filtre par rapport à la masse d'informations qui existe dans le document à indexer; il n'est pas utile et il n'est pas même souhaitable que le langage documentaire puisse véhiculer tous les types d'information, toutes les nuances que portent les phrases en langue naturelle; beaucoup de ces informations et de ces nuances n'ont pas d'intérêt documentaire. Pour aider le documentaliste dans son travail de sélection de l'information, le langage documentaire a une fonction de simplification à remplir, à laquelle doit correspondre une simplification des structures du langage par rapport à celles de la langue naturelle.

b) La langue naturelle est porteuse d'ambiguïtés au niveau lexical et au niveau syntaxique. Certaines de ces ambiguïtés ne sont pas acceptables sur le plan documentaire et doivent être éliminées. Au niveau syntaxique, l'ambiguïté est réduite par la formalisation et la simplification du langage, au niveau lexical par l'utilisation d'un vocabulaire contrôlé. En fait, il faut distinguer le cas où les tournures de la langue naturelle sont *multivoques* (la tournure peut être interprétée de deux ou plusieurs manières différentes, mais il est exclu que dans une situation d'énonciation particulière les deux possibilités soient vraies simultanément) et le cas où elles sont simplement *ambiguës* (la langue ne choisit pas entre deux possibilités qui peuvent être vraies séparément ou simultanément). Dans le premier cas, V1 doit éliminer la multivocité de la langue naturelle (c'est le cas de la coordination entre adjectifs en langue naturelle, cf. § 5.1.2.1.3, p. 47); dans le second cas, V1 peut conserver l'ambiguïté de la langue naturelle ou éventuellement la lever (par exemple, pour les items t1, le documentaliste ne *doit* pas choisir entre les traits [+ action] et [+ effet], mais il *peut* le faire s'il est possible et utile de trancher).

On pourrait penser que la « naturalité » de V1 (c'est-à-dire sa liaison étroite avec une langue naturelle, la langue de référence étant en l'occurrence le français) est un obstacle majeur à la réalisation de systèmes multilingues tels qu'ils se révèlent souvent souhaitables aujourd'hui. En fait, il n'est pas sûr qu'il en soit ainsi. D'abord, la langue de référence

peut être une autre langue que le français sans que les principes essentiels de V1 soient altérés (on va retrouver dans d'autres langues des éléments tels que les emboîtements et la composition de signifiés complexes à partir d'éléments plus simples, la transitivité, la formation d'une famille de mots par dérivation à partir d'un même radical, etc.). Une étude est d'ailleurs actuellement en cours sur les modifications à apporter à une version de V1 dont la langue de référence serait l'anglais (par exemple, la règle fixant la place de l'attribut, au sens de V1, ne pourra être la même que dans la version française). D'autre part, il est possible de concevoir des règles de passage entre des versions de V1 adaptées au même domaine scientifique et technique, mais s'appuyant sur des langues de référence différentes. Dans un premier temps, cette possibilité de passer d'une langue à une autre ne concernera que les entités schématiques; elle ne permettra pas de traduire les structures lisibles dans une autre langue, mais elle autorisera une recherche documentaire réellement translinguistique, les questions de l'utilisateur pouvant être posées au système dans une version linguistique de V1 différente de celle qui a servi à indexer les documents.

8.2 Syntaxe et structuration de l'information

Il n'est pas nécessaire d'insister longuement sur les avantages qui ont fait le succès actuel des langages documentaires du type thésaurus : unités lexicales (« descripteurs ») directement lisibles (une certaine dose, donc, de « naturalité »); principes de fonctionnement simples à comprendre; maniement accessible à des documentalistes de niveau moyen; facilité du traitement automatique des données indexées et de la gestion automatisée du thésaurus lui-même; etc. Les limites de ce type de langages ne sont pas moins évidentes : fausses combinaisons de descripteurs au moment de l'interrogation (d'où le développement de la pré-coordination dans les thésaurus); impossibilité d'employer efficacement des descripteurs de signification vague et générale, correspondant à des termes qui ne peuvent prendre une utilité documentaire véritable qu'en combinaison avec d'autres*; impossibilité de répondre à l'apparition de concepts nouveaux ou aux besoins non prévus au départ autrement que par la post-coordination, avec les risques qu'elle comporte, et à condition encore que les éléments constitutifs de ces concepts figurent parmi les descripteurs (et cela est vrai, même si ces concepts nouveaux se trouvent avoir une expression linguistique parallèle à celle d'autres concepts qui ont été traités en pré-coordination). Comme on peut le constater, ces limites tiennent essentiellement au fait que les unités d'indexation que forment les descripteurs sont totalement autonomes les unes par rapport aux autres. Des groupes de chercheurs ont donc été conduits, à plusieurs reprises, à élaborer des langages plus évolués, comportant des relations syntaxiques. Mais il faut reconnaître qu'il y a actuellement une désaffection certaine pour ces langages évolués. Comme le dit Jacques Chaumier (*Les langages documentaires*, Paris, Entreprise Moderne d'Édition, 1978, p. 24), « on peut remarquer que sur la dizaine de langages documentaires de ce type décrits par M. Coyaud dans son ouvrage paru en 1966, il n'y en a plus un seul qui soit utilisé de nos jours. La difficulté d'utilisation de ces langages et leur coût de mise en œuvre les ont écartés de la scène documentaire ». De nouvelles recherches dans ce sens doivent tenir compte des échecs passés — échecs qui ne font pas disparaître, d'ailleurs, l'intérêt intellectuel de certaines constructions — et ne se justifient que si l'on apporte des remèdes aux principaux risques des langages à syntaxe. Le système V1 tient compte de cette situation et présente à cet égard un certain nombre de caractères positifs.

Un premier risque, sans doute inévitable dans les langages à syntaxe, car les relations ne peuvent guère être strictement exclusives les unes des autres, est qu'une idée donnée peut souvent s'exprimer de deux ou plusieurs manières différentes. On réintroduit donc par

* Dans le *Thésaurus Sciences de l'information* du CNRS de tels descripteurs existent en nombre assez important : ainsi *accroissement*, *adaptation*, *élaboration*, etc.; on constate sans surprise que ces descripteurs ne sont presque pas utilisés. Certains d'entre eux sont réunis explicitement dans une facette intitulée *Facteur commun*, qui est ainsi définie : « ensemble de qualificatifs et de modificateurs, à associer aux notions principales », ce qui revient à reconnaître, d'une certaine manière, le besoin d'une syntaxe.

le biais de la syntaxe la synonymie que l'on essaie de maîtriser au niveau du vocabulaire et cela constitue un facteur de silence. Cette pluralité d'expression existe également dans V1 ; on ne souhaitait d'ailleurs pas l'éliminer complètement, car elle va dans le sens de la naturalité ; l'indexeur dispose donc parfois, comme dans sa langue naturelle, de plus d'une manière de formuler ce qu'il a à dire (bien évidemment, cette possibilité est beaucoup plus limitée qu'en langue naturelle : il fallait arbitrer entre les deux exigences de naturalité et de simplicité). C'est une des fonctions du *réseau sémantique** que de faire jouer a posteriori entre des structures grammaticales formellement différentes, mais dont le contenu significatif est équivalent et qui doivent être associées sur le plan documentaire**, les règles de passage qui permettent de réduire le risque de silence. Le réseau sémantique comportera donc des règles de passage capables de reconnaître l'équivalence entre les paires suivantes :

- (1a) Machine t0 à s photocopier a1.
- (1b) Photocopieuse t2.
- (2a) Entrée t1 gratuite a1.
- (2b) Gratuité t0 de l'entrée a1.
- (3a) Périodiques t0 scientifiques a1 en accès a1 b libre a1.
- (3b) Accès t1 libre a1 aux périodiques t0 scientifiques a1.

En revanche, il devra éviter d'associer les paires suivantes :

- (4a) Bibliothèques t0 en médecine a1.
- (4b) La médecine t0 dans les bibliothèques a1.
- (5a) Centre de documentation t3 sur s l'architecture a1.
- (5b) Architecture t0 d'un centre de documentation a3.

Les règles constitutives du réseau sémantique prennent en compte aussi bien les items et les traits qui, dans le lexique, sont associés aux notions dont ils sont composés que les relations établies entre les items par les règles syntaxiques. L'ensemble des règles qui constituent le réseau ne peut être maîtrisé que parce que le langage documentaire présente une grande simplification par rapport aux facteurs très complexes qui interviennent dans les langues naturelles.

Un deuxième risque, mentionné par J. Chaumier, est la difficulté d'utilisation ; la complexité d'un langage à syntaxe est évidemment plus grande que celle d'un thésaurus. Ne rend-elle pas indispensable le recours à un personnel nettement plus qualifié, alourdissant ainsi dans des proportions importantes le coût de l'indexation ? La réponse de V1 est ici la naturalité, qui joue un rôle positif pour l'acquisition du langage et pour son maniement, comme on l'a précédemment montré. Les documentalistes-indexeurs peuvent trouver dans leur langue maternelle, dont tout homme possède une connaissance intuitive, une bonne part des critères nécessaires à une utilisation correcte de V1. Quant aux chercheurs, ingénieurs et techniciens qui consacrent une partie de leur temps à l'analyse et à l'indexation de documents de leur spécialité, ils peuvent acquérir le langage V1 sans être découragés par un investissement en temps trop considérable.

La complexité du langage est d'ailleurs limitée par la recherche systématique de l'économie des moyens d'expression. Plutôt que de multiplier les procédés d'expression très spécialisés et d'un rendement problématique, on a opté pour des procédés présentant une certaine généralité, capables parfois de remplir plusieurs fonctions dans le langage et utilisables à des fins variées, en fonction du domaine d'application ou des objectifs poursuivis.

* Le réseau sémantique n'a fait l'objet, jusqu'à présent, que d'études partielles portant sur son organisation générale et sur quelques modèles de règles. Comme le réseau sémantique est une pièce essentielle du système V1, la possibilité de faire fonctionner efficacement le système est conditionnée par la mise en place effective du réseau et sa capacité de remplir les fonctions principales qui lui ont été assignées, fonctions dont on donne quelques exemples par la suite.

** Lorsque le contenu significatif de plusieurs structures grammaticales formellement différentes est équivalent et que le système, par le biais du réseau sémantique, doit renvoyer directement et automatiquement de l'une à l'autre, on dit qu'il y a *équivalence absolue* entre ces structures grammaticales. L'*équivalence absolue* s'étend aux cas où il existe de légères nuances de sens entre les structures LN qui correspondent aux structures grammaticales de V1, lorsque ces nuances n'ont pas été jugées significatives sur le plan documentaire (cf. par exemple les structures (3a) et (3b) ci-dessous).

L'anaphorique est un exemple typique de tels procédés. Il joue, comme on l'a vu, un rôle double dans le langage : d'une part, en donnant le moyen de porter remède à certaines contraintes de la grammaire qui sont gênantes pour l'indexation, il permet de résoudre par une solution générale plusieurs problèmes différents, en évitant de compliquer la grammaire pour résoudre séparément ces contraintes; d'autre part, il crée des liens entre expressions documentaires représentant le contenu d'un même document, en notant les identités référentielles que l'indexeur constate d'une ED à une autre, ce qui favorise la structuration de l'indexation des documents et permet une recherche approfondie, tout en simplifiant l'écriture des structures grammaticales où apparaît l'anaphorique. Les procédés, au sens particulier de V1 (c'est-à-dire les procédés T et A qui ont pour fonction la catégorisation d'une notion en items), représentent eux aussi une solution qu'on peut qualifier de générale : ils sont en nombre très limité; ils ont un rôle à la fois syntaxique (en liaison avec la position relative de l'item dans la structure grammaticale) et sémantique (addition ou non de traits sémantiques à une notion); pour répondre aux besoins de tel ou tel domaine d'application, on peut créer de nouveaux procédés T et A, porteurs d'un contenu sémantique particulier et fonctionnant de la même manière que les procédés déjà existants, ou supprimer, au contraire, certains de ceux qui existent dans la version actuelle, et cela sans remettre en cause l'élément caractéristique du langage V1 qui est, ici, l'existence de procédés T et A parallèles, capables de remplir simultanément une fonction syntaxique et une fonction sémantique.

Il faut noter enfin que les langages à syntaxe dont J. Chaumier constate l'abandon n'ont pu bénéficier du progrès de l'informatique durant la dernière décennie et notamment des possibilités de dialogue interactif entre l'homme et la machine. Grâce à ces possibilités, dont le système Vercingétorix 1 doit tirer parti, la machine peut soulager la mémoire et l'attention de l'indexeur, en lui fournissant des informations, en le guidant et en le contrôlant (cf. fiche n° 3); elle peut, inversement, effectuer automatiquement certaines opérations sous le contrôle de l'indexeur. Ce dernier peut concentrer l'essentiel de son effort sur une bonne compréhension du document et sur la recherche des critères pertinents pour fixer les choix offerts à tous les niveaux par le langage V1. Pour l'indexation des questions, le réseau sémantique devrait faciliter de même la tâche du documentaliste : il lui présentera des ED dont le contenu significatif n'est pas strictement équivalent à celui de la question de départ (qui ne sont pas en situation d'*équivalence absolue*, au sens indiqué plus haut, et qui ne sont donc pas automatiquement prises en compte par la machine), mais qui ont un contenu significatif très voisin, avec des nuances qui peuvent n'être pas pertinentes pour un document donné ou pour un utilisateur final donné (on parle dans ce cas d'*équivalence relative*). Si le documentaliste veut modifier la question de départ, la machine devrait lui proposer les diverses possibilités dans le sens d'une généralisation, d'une spécification ou d'un déplacement de la question.

Un troisième risque des langages à syntaxe est la difficulté, la lenteur et, par conséquent, le coût du traitement informatique. Comment assurer de manière commode le stockage des indexations et la comparaison des questions posées par les utilisateurs avec le contenu de la mémoire documentaire? Ces problèmes, qui sont résolus d'une manière satisfaisante et très simple dans son principe pour les langages qui s'appuient sur des thésaurus (par la constitution de fichiers inversés et l'utilisation d'une logique de recherche reposant sur les opérateurs booléens), sont beaucoup plus complexes et délicats quand on se trouve devant un langage à syntaxe. Or, une propriété intéressante des structures grammaticales de V1 est que les items d'une PD ou les constituants d'une construction documentaire peuvent être caractérisés de manière non ambiguë par la place qu'ils occupent dans la série des emboîtements qui forment la PD ou dans la suite des constituants de la construction documentaire. Par exemple, si l'on prend les deux PD suivantes, où l'on trouve le même item *formation t1* :

- a) Formation t1 professionnelle a1 du personnel des bibliothèques t3.
- b) Contrôle t1 du financement t1 de la formation t1 des utilisateurs t3 de bases de données t0.

chacune des occurrences de cet item se trouve dans une certaine *situation syntaxique* par rapport aux autres items de la même PD. Cette situation syntaxique dépend de la réponse

à des questions telles que : l'item T a-t-il ou non des entités A qui portent sur lui et, si oui, combien? ; l'item T est-il complément verbal d'un autre item? ; si oui, cet autre item est-il lui-même un complément verbal? ; etc. Malgré la complexité des combinaisons possibles, la situation syntaxique peut être notée de manière purement formelle (c'est-à-dire indépendamment de la valeur sémantique des items et, par conséquent, du contenu significatif de la PD) à partir des relations syntaxiques entre les éléments d'une structure grammaticale (cf. § 3) ; ce travail est effectué automatiquement par l'ordinateur. Il s'agit d'une propriété extrêmement importante pour la recherche automatique, car elle permet à la machine de créer un fichier inversé par notions, items et situations syntaxiques d'items, fichier à partir duquel il est possible d'accéder aux structures grammaticales et, par elles, aux ED pertinentes sur le plan documentaire. A partir de ce fichier, le travail d'appariement entre représentation des questions documentaires et représentation des documents peut se faire dans des conditions, somme toute, assez comparables à ce qui se passe dans les systèmes reposant sur des thésaurus traditionnels.

On peut donc dire qu'à l'égard de chacun des principaux inconvénients que présentent souvent les langages à relations syntaxiques, V1 apporte des réponses positives. De ce fait, on ne paie pas d'un prix trop élevé les avantages que procurent les relations syntaxiques : précision et fiabilité de l'indexation, richesse d'expression des nuances et des idées complexes. En outre, grâce à ces relations syntaxiques et aux relations entre structures grammaticales, on obtient des indexations plus explicites qu'avec un simple thésaurus, en ce sens que, lorsque le document est représenté par des descripteurs isolés, on ne sait pas toujours pourquoi l'indexeur a été conduit à choisir ou au contraire à éliminer tel ou tel descripteur ; sa décision s'explique par une certaine vision du contenu du document, qu'il n'est pas toujours facile de reconstituer à partir de descripteurs isolés — et on peut penser que l'indexeur lui-même, dans certains cas, ne s'est pas formulé de manière parfaitement nette les raisons de son choix. Au contraire, lorsque les unités du lexique sont liées par des relations syntaxiques et s'intègrent dans des ED qui représentent pleinement les diverses idées que l'indexeur a reconnues dans le contenu d'un document, les choix deviennent beaucoup plus explicites, et le rapport étroit qui existe entre l'image mentale de l'information à sélectionner — image exprimée, ou exprimable à tout moment, en langue naturelle, qui guide l'indexation — et la suite des ED est favorisé par la naturalité de V1. Ce caractère plus explicite de l'indexation avec V1 va dans le sens d'une meilleure cohérence ou uniformité de l'analyse et de l'indexation, entre plusieurs indexeurs travaillant en parallèle ou pour le même indexeur à des moments différents. Il s'agit justement là d'une des fonctions importantes qu'on assigne aux langages documentaires.

Un autre point mérite d'être pris en considération : celui du niveau d'indexation et de la pertinence plus ou moins grande de chaque unité d'indexation par rapport au contenu du document. La situation de V1 et celle d'un langage à thésaurus sont assez différentes à cet égard. Dans V1, chaque ED est, en principe, par elle-même et directement (conformément à la définition même des ED et à la fonction du classement des structures grammaticales, cf. § 4. 2), d'un niveau pertinent pour le document qu'elle sert à représenter, c'est-à-dire ni trop générale ni trop spécifique ; tandis que les descripteurs d'un thésaurus, pris isolément, ne sont pas, bien souvent, pertinents pour le document auquel ils sont appliqués : ils ne trouvent une certaine pertinence (mais de manière aléatoire et peu fiable) qu'associés les uns aux autres dans les opérations de recherche. L'exemple suivant permettra d'illustrer ce fait : soit un document qui intéresse « l'automatisation des bibliothèques spécialisées en sociologie aux États-Unis » ; il pourra être indexé avec les descripteurs *automatisation, bibliothèque, sociologie, États-Unis* ; le document a évidemment un contenu trop particulier pour répondre à une question qui porterait sur l'automatisation, sur les bibliothèques, sur la sociologie ou sur les États-Unis, *en général*. Si l'on regroupe les descripteurs deux par deux, on trouve des situations différentes selon les cas : (a) le couple *automatisation / sociologie* est peu vraisemblable, il a donc peu de chances d'apparaître ; (b) le document ne peut guère répondre à des questions sur la « sociologie aux États-Unis » ou sur l'« automatisation aux États-Unis » ; (c) bien que trop spécifique, le document pourra, à la rigueur, être de quelque utilité pour une recherche intéressant les

couples *bibliothèque / États-Unis*, *bibliothèque / sociologie*, *bibliothèque / automatisation* (mais il faut remarquer que les deux derniers couples peuvent traduire aussi des questions sur la « sociologie des bibliothèques » ou les « bibliothèques sur l'automatisation »). C'est pour certaines combinaisons de *trois* descripteurs que le document commence vraiment à être pertinent. La simple juxtaposition des descripteurs rend donc compte de la présence des concepts correspondants dans le contenu du document, sans dire si ces descripteurs sont pertinents isolément ou à l'intérieur de quelles combinaisons ils sont pertinents. Avec un langage documentaire du type thésaurus, on doit certes rechercher quel est le bon niveau d'indexation *quand on peut choisir entre plusieurs descripteurs en relation hiérarchique*, mais on ne peut aller plus loin; dans V1, au contraire, la détermination du bon niveau d'indexation est une des composantes essentielles de l'utilisation du langage. Il faut ajouter qu'on peut, avec V1, exprimer sans risque de fausses combinaisons autant d'idées différentes qu'on juge utile d'en associer au contenu d'un document, selon le degré de précision à atteindre et compte tenu des contraintes d'encombrement, alors qu'avec des descripteurs juxtaposés les risques de fausses combinaisons augmentent avec le nombre des descripteurs employés*.

L'organisation déterminée par les relations syntaxiques, c'est-à-dire celle des structures grammaticales (PD et constructions documentaires), se prolonge par une structuration plus large qui repose sur l'emploi de l'anaphorique. On rencontre ainsi au total trois niveaux de structuration de l'indexation dans V1 qui sont, du plus étroit au plus large :

- le niveau des structures grammaticales, où s'organisent les relations entre les notions;
- le niveau de l'ED, où s'établit le rapport de pertinence avec le document;
- le niveau constitué par un ensemble d'ED associées au contenu d'un même document, lorsqu'un rapport d'identité référentielle est constaté entre des entités représentées dans les ED.

Les deux premiers niveaux se confondent lorsqu'il y a coïncidence entre une structure grammaticale et une ED. Dans le cas contraire, les différentes structures grammaticales, constituant autant de SED, doivent obligatoirement être reliées par l'anaphorique, pour que puisse s'établir sans aucune ambiguïté, au niveau de l'ED, le rapport de pertinence avec le document qui est, comme on l'a vu, un élément important de V1. V1 met en outre à la disposition de l'indexeur – c'est le troisième niveau – la possibilité d'établir des relations entre ED et par conséquent de structurer tout ou partie de l'indexation d'un document. Il peut essayer de rendre compte, de cette manière, de la structuration qu'il reconnaît dans le document lui-même et donner ainsi de celui-ci une image plus complète et plus fidèle. Les entités qui figurent dans la première ED (ou dans l'ED qui se trouve à la tête d'un groupe d'ED reliées par l'anaphorique) ont toute chance, lorsqu'elles sont rappelées dans la suite de l'indexation à l'aide de l'anaphorique, d'être celles qui constituent le ou les thèmes centraux autour desquels s'organise le contenu du document. Il faut noter toutefois que cette structuration, qui n'est assurée que par l'anaphorique (et non par d'autres éléments tels que l'ordre des ED) et qui s'appuie donc sur la reconnaissance de rapports d'identité référentielle, ne peut traduire la structuration du document lui-même que dans la mesure où elle repose sur ce type de rapports; en revanche, une organisation générale fondée sur d'autres types de rapports logiques ou hiérarchiques ne peut s'exprimer, dans l'état actuel du système Vercingétorix, par des relations entre ED. Le troisième niveau de structuration du langage reste donc facultatif, en ce sens que les ED ne seront reliées que dans la mesure où des rapports d'identité référentielle sont constatés; et, si de tels rapports existent, ils pourront déterminer plusieurs groupes d'ED ou ne pas concerner certaines ED, qui resteront isolées.

* On sait que certains systèmes à thésaurus essaient de pallier les inconvénients mentionnés en distinguant les descripteurs essentiels et les descripteurs secondaires ou en introduisant une pondération ou des liens. Ces diverses méthodes compliquent le fonctionnement de ces systèmes sans apporter tous les avantages d'un véritable langage à relations syntaxiques.

On peut dire, pour conclure sur ce point, que V1 amène le documentaliste à ne pas se contenter d'une lecture superficielle du document; il ne s'agit pas d'une « cueillette de descripteurs ». Certes, l'opération intellectuelle extrêmement complexe qu'est l'analyse et la sélection de l'information ne relève pas de la problématique de V1, mais de celle de la contraction de texte (cf. *Condenser*, n° 1, « La condensation de l'information », p. 5) et le langage documentaire n'intervient, à proprement parler, que lorsque l'indexeur s'est déjà formé une représentation, mentale ou explicitement formulée, du contenu du document et de l'information à retenir. Cependant, le langage V1 fournit un cadre précis pour cette opération et conduit l'indexeur à se poser des questions, sur le niveau de pertinence des ED ou l'existence de rapports d'identité référentielle par exemple, qui constituent autant d'éléments d'une grille d'analyse possible. Sans doute cette lecture approfondie augmente-t-elle le temps nécessaire à l'indexation de chaque document; elle détermine donc un surcoût. Mais celui-ci paraît acceptable chaque fois que l'on désire que l'information mémorisée soit non seulement plus précise et plus riche, mais aussi plus fiable et de meilleure qualité; et il peut probablement être compensé en partie par le fait qu'il est possible d'obtenir des sous-produits divers de l'analyse documentaire, soit directement sous forme de résumés en langue naturelle rédigés par l'indexeur, soit après traitement en machine sous forme de produits documentaires variés, dérivés des structures lisibles, tels que bulletin de résumés, index permuté, diffusion sur profil (cf. fiche n° 2, § 3).

8.3 Flexibilité du langage

Le langage V1 possède des caractéristiques qui permettent au système dans son ensemble de présenter les qualités de souplesse et d'adaptation que l'on souhaite de plus en plus trouver dans les systèmes documentaires.

Nous avons déjà vu que la naturalité n'était pas nécessairement un obstacle à l'existence de versions multilingues ou translinguistiques de V1 (p. 61). On a vu aussi que l'organisation du langage autour de procédés d'expression suffisamment généraux (et non de procédés *ad hoc*) facilitait l'adaptation du système à des domaines d'application différents du domaine expérimental d'origine (p. 63). On vient enfin de rappeler qu'on pouvait demander au système V1, à côté de la recherche documentaire en conversationnel à partir de questions précises posées par un utilisateur donné (recherche rétrospective), des sous-produits variés dérivés des structures lisibles; cela est rendu possible à la fois par la richesse de l'information que peuvent véhiculer les ED et par la naturalité. Ce sont ces mêmes facteurs — richesse de l'information et naturalité — qui favorisent la compatibilité du langage V1 avec d'autres langages documentaires. Le lexique de V1, appliqué aux sciences et techniques de l'information et de la documentation, est, à titre expérimental, élaboré à partir du *Thesaurus Sciences de l'information* du CNRS. On peut donc étudier avec précision sur cet exemple les relations qui existent entre les indexations réalisées avec des descripteurs juxtaposés et les expressions documentaires de V1 et avoir une idée des conditions dans lesquelles une conversion automatique des unes aux autres serait envisageable. Sans entrer dans les détails, on peut dire qu'un certain niveau de compatibilité peut être obtenu dans les conditions suivantes :

a) il faut que le lexique du langage de départ et celui du langage d'arrivée possèdent en commun une même organisation du monde référentiel, notamment quant à l'extension des concepts, notions ou entités représentés par des descripteurs, d'une part, par des items ou groupements d'items, d'autre part; il faut, par exemple, que les groupements d'items *formation t1 professionnelle a1, échange t1 de documents t0, index t0 permuté a4*, renvoient aux mêmes réalités du monde référentiel que, respectivement, les descripteurs *formation professionnelle, échange (de) document, index permuté*;

b) il faut que la conversion se fasse du langage le plus riche vers le langage le moins riche, soit de V1 vers le thésaurus (avec, évidemment, une déperdition plus ou moins importante d'information); il y a dans les expressions documentaires de V1 toute l'information nécessaire pour former les suites de descripteurs; l'inverse n'est pas vrai.

Compte tenu de ces exigences, on peut imaginer diverses possibilités. On peut souhaiter, par exemple, avoir accès dans le cadre de V1 à un stock de références et

d'indexations mémorisées antérieurement sur la base d'un thésaurus de type classique. Il n'est pas possible dans ce cas de convertir automatiquement les suites de descripteurs en expressions documentaires et de les mettre dans le même fichier que les indexations directement réalisées avec le langage V1 (cf. condition (b) ci-dessus : les descripteurs ne possèdent en général pas, isolément, le rapport de pertinence avec le document exigé d'une ED; et la machine n'a pas l'information nécessaire pour combiner correctement les descripteurs). En revanche, les questions documentaires exprimées dans le langage V1 peuvent être traduites par l'ordinateur en suites de descripteurs, à partir des découpages documentaires dont il a été question p. 57, et servir à interroger un fonds distinct. Le réseau sémantique peut intervenir pour autoriser des conversions assez complexes à partir des traits sémantiques présents dans les items (ex. : *élaboration t1 automatique a1* (d'un index, par exemple) → *traitement automatisé*). Tout porte à croire que la compatibilité serait possible aussi avec d'autres types de langages que les thésaurus, par exemple avec le langage de PRECIS.

S'il est important qu'un langage documentaire puisse être adapté à des problèmes et à des besoins variés, il l'est sans doute davantage encore que, dans une application particulière, il présente une souplesse suffisante pour suivre l'évolution rapide des sciences et des techniques ou pour permettre à l'utilisateur, selon les résultats qu'il souhaite obtenir pour une recherche donnée, d'arbitrer commodément et efficacement entre les exigences contradictoires de pertinence et d'exhaustivité.

Sur le premier point — flexibilité du langage documentaire dans le temps —, V1 apporte des réponses originales. Assurément, le souci de réduire les facteurs de bruit et de silence qu'apporte l'usage de la langue naturelle pour représenter le contenu des documents a conduit à adopter, comme pour la plupart des langages documentaires, un vocabulaire strictement contrôlé : V1 repose sur une liste fermée de notions et de traits sémantiques, comme il comporte des listes fermées d'éléments syntactico-sémantiques tels que les procédés, les spécificateurs d'attribut et les prédicats des constructions documentaires. Les listes de notions et de traits sémantiques peuvent être établies par les spécialistes du domaine considéré selon des méthodes similaires à celles qui sont employées pour la construction de thésaurus; elles correspondent, si elles sont bien faites, à un certain consensus chez les utilisateurs potentiels à un moment donné et garantissent un niveau minimal d'homogénéité lexicale des indexations. Mais cette rigidité est partiellement compensée par des facteurs qui jouent dans le sens de la souplesse et peuvent faciliter l'adaptation à l'évolution des sciences et des techniques. Il ne s'agit pas ici de la faculté de mise à jour du lexique qui existe évidemment comme pour un thésaurus, avec le risque de créer des distorsions plus ou moins grandes entre les fichiers antérieurs ou postérieurs aux modifications*, mais de facteurs qui exercent leur rôle dans un état donné du langage et qui tiennent à certaines des caractéristiques essentielles de V1. Parmi ces facteurs, on notera spécialement les suivants :

- a) l'existence de relations syntaxiques offre, par la combinaison des items, des possibilités considérables d'exprimer des concepts ou des idées non prévus à l'avance;
- b) certains procédés, comme t2 ou t3 dans la version actuelle, permettent de rendre une idée avant même que la langue ait créé un mot pour la représenter;
- c) la fonction de la particule *et* et du noyau à coordination interne est précisément d'autoriser, au moment de l'indexation, la création de notions complexes, non déterminées à priori;
- d) il est prévu une possibilité qui n'a pas encore été évoquée et sur laquelle des indications seront données dans une fiche sur les articulations : l'indexeur pourra insérer dans les articulations non seulement les éléments de liaison qui les constituent normalement, mais des termes de sens plein, étrangers au lexique de V1 (noms propres, termes très spécifiques, néologismes, etc.); ces termes seront éventuellement associés à des traits

* Dans le cas de V1, le réseau sémantique peut d'ailleurs incorporer des règles d'équivalence absolue ou relative pour améliorer le passage d'un fichier à l'autre et limiter ces distorsions.

sémantiques qui permettront de les situer dans le réseau sémantique; la machine pourra, sur des ED qu'elle aura préalablement sélectionnées et analysées à partir de l'information de base, opérer une recherche complémentaire en tenant compte de ces termes non contrôlés.

La possibilité d'arbitrer entre les exigences de pertinence et d'exhaustivité existe à plusieurs niveaux dans V1. Un bon exemple est celui du spécificateur d'attribut; l'introduction de distinctions fines dans le langage permet d'accroître la pertinence, mais elle risque de générer du silence et donc de nuire à l'exhaustivité; selon les besoins, on fera donc intervenir ou non, au moment de l'interrogation, la recherche des spécificateurs d'attribut. Certains critères, en principe clairs, peuvent être difficiles à manier dans la pratique et être interprétés de manière un peu différente par les indexeurs; si l'on recherche moins la pertinence que l'exhaustivité, le réseau sémantique, par des règles d'équivalence relative, fournira le moyen d'effacer les distinctions établies par l'usage de ces critères.

On espère avoir réussi à montrer que V1 prend en compte les conditions réelles et concrètes de la recherche documentaire automatisée et cherche à présenter sur chacun des points traités une réponse appropriée, mais il convient de souligner en terminant que l'élaboration de V1 offre aussi un intérêt théorique: elle fournit un cadre permettant de mieux étudier les rapports entre les langues naturelles et les langages documentaires et de mieux connaître les facteurs qui interagissent dans les systèmes de représentation de l'information.

Appendice I

LE LANGAGE V1 : STRUCTURES ENGENDRÉES PAR LA GRAMMAIRE ET RELATIONS ENTRE STRUCTURES GRAMMATICALES

(résumé *)

(A) Les *structures grammaticales* sont les phrases documentaires et les constructions documentaires; elles comportent trois types d'information: (a) l'information de base (items); (b) les articulations grammaticales (spécificateurs de notion et spécificateurs A, prépositions de complément verbal); (c) les articulations libres, fixées par le documentaliste. Les *structures grammaticales schématiques*, constituées de (a), de (b) et des emplacements réservés pour (c), possèdent un *contenu significatif*. Les *structures grammaticales complètes*, constituées de (a), (b) et (c), possèdent un contenu. Les contenus significatifs du langage V1 comprennent les contenus significatifs des structures grammaticales schématiques et les contenus significatifs déterminés par les relations entre structures grammaticales. La grammaire de V1, composée d'un lexique et des règles syntaxiques grammaticales, engendre les structures grammaticales schématiques (§ 1, p. 2-3).

* Sont indiqués entre parenthèses les passages correspondants du texte. Les numéros des pages renvoient à la pagination en italiques de la fiche n°5 (*Condenser* n°2 pour les pages 1 à 49, n°3 pour les pages 51 et suivantes).

(B) La grammaire de V1. Le lexique (§ 2)

Le *lexique* est constitué d'*entrées lexicales* et de *renvois* (§ 2, p. 8 et 14).

(1) Une *entrée lexicale*

(a) est caractérisée par une *notion* dont le contenu est précisé par le *mot de rappel*, qui renvoie à une *famille restreinte de mots* et, optionnellement, par des précisions sur le contenu. Le mot de rappel identifie l'entrée lexicale (§ 2.1);

(b) fixe de manière contraignante : les procédés admis par chaque notion (§ 2.2); le radical discriminant et, éventuellement, les variations dans le radical discriminant (§ 2.3); la ou les prépositions exigées par le complément verbal d'un t1, si ce complément ne correspond pas en LN à un complément d'objet direct (§ 2.2, p. 11);

(c) indique de manière non contraignante les supports de radical normalisés (§ 2.3) et, dans t3, les signifiants du spécificateur de notion « personne », et, s'ils diffèrent des formules générales, ceux du spécificateur de notion « institution » (§ 2.3, p. 14);

(d) précise les traits sémantiques, de sélection et particuliers (§ 2.4).

(2) *Incorporation des entrées lexicales au lexique de V1*

Une *famille restreinte de mots* (§ 2, p. 8 sq.) est composée des mots et expressions LN avec un signifié invariant, suivi des éléments de flexion et des indicateurs de genre et de nombre; elle admet les suffixes (a) de nominalisation, (b) « machine ou outil pour faire quelque chose », (c) « personne qui fait quelque chose, qui est spécialiste en quelque chose ou dont l'activité est caractéristique d'une institution déterminée », (d) « le fait ou la qualité d'être quelque chose ».

Si, dans la famille restreinte de mots, il y a un verbe et pas de nom sans suffixe, le verbe est incorporé à l'infinitif comme mot de rappel d'une entrée lexicale. S'il y a un nom sans suffixe, et pas de verbe, on incorpore le nom sans suffixe. S'il y a un verbe et un nom sans suffixe, on incorpore les deux dans deux entrées lexicales distinctes. S'il y a un adjectif ou un adjectif et un nom avec le suffixe « le fait ou la qualité d'être x », x étant l'adjectif, et s'il n'y a ni verbe ni nom sans suffixe, on incorpore l'adjectif (les noms sans suffixe dont le contenu est paraphrasable par celui d'un nom avec suffixe sont assimilés à ceux-ci; certaines expressions en langue naturelle sont assimilées à des mots).

(3) *Procédés* (§ 2.2)

Procédés T : t0 ou t1, t2, t3; A : a1, a2, a3, a4.

TV : t1, t2; t3, si la notion respective admet t1 (le mot de rappel est un verbe).

(4) *Admission de procédés* (§ 2.2):

Mot de rappel	Complément verbal correspond à :	Trait sémantique de la notion	Procédés admis
verbe	obj. direct		t1 t2 t3 a1 a2 a3 a4
verbe	obj. ind. ou régime verbal		t1 t2 t3 a1 a2 a3
nom ou adjectif		[+ Inst.] et/ou t0 [+ AST] t0	t3 a1 a3 a1
nom ou adjectif		t0	a1

(5) *Procédés T* (en italiques, on trouvera la formule générale qu'il faut employer en l'absence d'un nom avec le suffixe adéquat en LN.)

t0 (§ 2, p. 10) : correspond en LN au mot de rappel, si celui-ci (cas le plus général) est un nom ou une expression assimilée; aucun trait sémantique n'est apporté par le procédé dans ce cas. Si le mot de rappel est un adjectif, t0 correspond en LN à un nom avec suffixe « la qualité d'être x », x étant l'adjectif du mot de rappel; le signifié du suffixe est le trait sémantique apporté.

t1 (§ 2, p. 10-11) : correspond en LN à un nom verbal ou à un verbe à l'infinitif ou à une expression assimilée ou à *le fait de x*, *x* étant l'infinitif du mot de rappel. Une des trois configurations suivantes représente les traits sémantiques apportés par t1 (entre parenthèses, la notation correspondante) :

$$\begin{array}{ccc} \left[\begin{array}{c} + \text{ action} \\ \pm \text{ effet} \end{array} \right] & (t1) \text{ ou} & \left[\begin{array}{c} - \text{ action} \\ + \text{ effet} \end{array} \right] & (-t1) \text{ ou} & \left[\begin{array}{c} + \text{ action} \\ - \text{ effet} \end{array} \right] & (+t1) \\ \text{I} & & \text{II} & & \text{III} \end{array}$$

Le complément verbal admis par t1 correspond, dans la plupart des cas, à un complément d'objet direct. Dans quelques cas particuliers, il correspond à un complément d'objet indirect ou de régime verbal; dans de tels cas, l'entrée lexicale indique la ou les prépositions de complément verbal exigées.

t2 (§ 2, p. 11) : correspond en LN à un nom avec le suffixe « machine qui sert à quelque chose » ou à *machine à x*, *x* étant l'infinitif du mot de rappel.

t3 (§ 2, p. 11) : quatre cas peuvent se présenter; la notion admet

- **t0** [+ institution] : correspond en LN à un nom avec suffixe « personne dont l'activité est caractéristique d'une institution déterminée » ou à *personnel de x* ou à *personnel spécialisé de x*, *x* étant le nom de l'institution du mot de rappel;

- **t0** [+ AST] : correspond en LN à un nom avec suffixe « personne qui est spécialiste en quelque chose » ou à *spécialiste en x*, *x* étant le nom du mot de rappel;

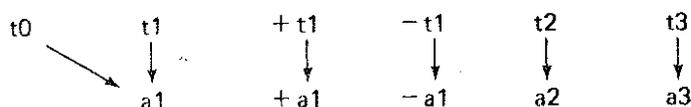
- **t1** et reçoit le spécificateur de notion « personne » : correspond en LN à un nom avec suffixe ou à un nom dont le signifié est « personne qui fait quelque chose » ou à *personne qui x* ou à *personnel de x* ou à *personnel spécialisé de x*, *x* étant une forme verbale qui incorpore le radical discriminant;

- **t1** et reçoit le spécificateur de notion « institution » : correspond en LN à *centre de x* ou *service de x* ou *organisme de x*, *x* étant une forme verbale qui incorpore le mot de rappel.

Le choix du spécificateur de notion « personne » ou « institution » doit être effectué par le documentaliste. Là où le spécificateur « institution » exige un signifiant particularisé, celui-ci sera indiqué dans l'entrée lexicale correspondante.

(6) Procédés A (§ 2, p. 12)

a1, **a2**, **a3** introduisent le trait sémantique de « propre ou relatif à quelque chose, en relation avec quelque chose »; ils correspondent en LN soit à un adjectif, soit à un syntagme nominal ou prépositionnel dans la fonction de complément déterminatif. Conservation de traits sémantiques :



a4 introduit le trait sémantique de « ayant subi l'action de *x* » ou « produit par l'action de *x* », *x* étant un infinitif qui admet un complément d'objet direct; il correspond en LN au participe passé. (Pour exprimer l'idée « qui a été soumis à l'action de *x*, mais qui n'est pas le produit de cette action », ajouter *original a1* à la suite A à laquelle appartient l'item a4.)

(C) La grammaire de V1. Les règles syntaxiques grammaticales. Contenus significatifs déterminés par les structures grammaticales.

(1) *Item non anaphorique avec ses articulations* (§ 1.1.1, p. 5)

Items non-anaphoriques

(Signifiant du spécificateur A)	(artic. libre)	(Signifiant du spécificateur de notion)	(artic. libre)	Radical discriminant	(artic. libre)	Indicateur de procédé
(Préposition de compl. verbal)	(artic. libre)	(Signifiant du spécificateur de notion)	(artic. libre)	Radical discriminant	(artic. libre)	Indicateur de procédé

Le contenu significatif de l'item non anaphorique est déterminé par sa notion, les traits sémantiques de son indicateur de procédé et, le cas échéant, son spécificateur de notion. Voir ci-dessous (D) l'item anaphorique et son contenu significatif. Le contenu significatif de l'item, avec ou sans anaphorique, est l'unité de contenu significatif de la phrase documentaire (§ 5.1, p. 41).

(2) *Le syntagme de phrase* est constitué d'un noyau et d'un groupe A (§ 3.1).

(a) *Le noyau* peut être (§ 3.1)

- *isolé* : non suivi du groupe A ;
- *simple* : un seul item T ;
- *à coordination interne* (§ 3.4) : deux ou plus de deux items T reliés par *et*.

Les entités du langage portent sur le noyau en entier et sur chacun des items qui le composent : si un complément verbal le suit, le noyau à coordination interne doit être constitué d'items TV. Pour le reste, même combinatoire et participation aux mêmes relations que le noyau simple. Mise en rapport des contenus significatifs des items individuels comme opération préalable à la composition (§ 5.1.1). Création d'une entité nouvelle de caractère complexe. L'anaphorique permet d'introduire des déterminations à droite qui portent sur chacun des items.

(b) *Le groupe A* est constitué de tous les items A qui suivent le noyau dans un syntagme de phrase (§ 3.3).

Le marqueur b doit être utilisé à la suite immédiate d'au moins un item A et doit précéder au moins un item A. Un groupe A peut incorporer plusieurs *b* (§ 3.3.1).

Noyau A : item A qui précède immédiatement *b* (§ 3.3.1).

Suite A : item(s) A qui, sans être précédé(s) par *b*, suit (suivent) le noyau A ou de syntagme de phrase. Chaque item A de la suite A est *complément A* du noyau (§ 3.3.1).

Le syntagme A est composé du noyau A, du marqueur *b* et de la suite A (§ 3.3.1, p. 24).

Deux ou plus de deux syntagmes A déterminent la constitution d'un *bloc A*. Le dernier syntagme A, à droite du syntagme de phrase, tout bloc A, et chacun des items A d'une suite A qui n'est pas un noyau A, *portent (relation d'attribut)* sur le noyau (noyau A ou de syntagme de phrase) qui le(s) précède immédiatement (§ 3.3.1, p. 25).

Spécificateurs A : articulation grammaticale d'un item A qui précise la relation d'attribut (§ 3.3.2 et § 5.1.2.1.2).

	Signifié :	Signifiants :	Contrainte :
(1)	« domaine »	spécialisé(e)(s) en, spécialisé(e)(s) dans, portant sur, sur	L'item A est complément A d'un noyau avec [+ document] et/ou [+ domaine]
(2)	« fonction »	pour, à	L'item A est un item a1 ou + a1 dont la notion correspondante admet t1

La *relation d'attribut* (§ 5.1.2.1) entre une entité A (item A, syntagme A, bloc A) et un noyau A correspond à celle qui existe en LN entre un complément déterminatif et un noyau nominal ou entre un complément déterminatif et un noyau adjectival; la relation d'attribut entre une entité A et un noyau de syntagme de phrase correspond à celle qui existe en LN entre un complément déterminatif et un noyau nominal. Le complément déterminatif ne doit pas être paraphrasable par un complément d'objet direct ou par un complément qui correspond en V1 à un complément verbal. Pour tourner cette restriction, il faut avoir recours à l'anaphorique. Si le spécificateur A est utilisé, la relation d'attribut doit être paraphrasable par le signifié du spécificateur A (§ 5.1.2.1.2). Si plusieurs entités A portent sur un même noyau, chacune a, avec ce noyau, la même relation que si elle était seule et, avec les autres, une relation qui correspond, en LN, à celle de la juxtaposition (§ 5.1.2.1.3). Dans ce cas, le noyau possède une invariance référentielle : toutes les entités A doivent porter sur un même référent exprimé par ce noyau. Après un syntagme A ou un bloc A qui porte sur un noyau, il est interdit d'introduire une autre entité A qui porte sur ce même noyau. Pour tourner cette restriction, il faut avoir recours à l'anaphorique.

(3) La *phrase documentaire* (= PD) est constituée par un ou plusieurs syntagmes de phrase. Un syntagme de phrase dont le noyau est constitué d'un item TV, ou de plusieurs items TV reliés par *et*, est susceptible d'être suivi d'un autre syntagme de phrase (§ 3.2 et 3.4). Le noyau d'un syntagme de phrase est complément verbal du noyau du syntagme de phrase qui le précède (§ 3.2.1, p. 22). Deux ou plus de deux syntagmes de phrase placés à la suite du premier syntagme de phrase à gauche d'une phrase documentaire déterminent la constitution d'un *bloc de syntagmes de phrase*. Le dernier syntagme de phrase à droite de la phrase documentaire, et tout bloc de syntagmes de phrase, *porte (relation verbale)* sur le syntagme de phrase qui le précède immédiatement (§ 3.2.1, p. 22).

La *relation verbale* (§ 5.1.2.2) entre un syntagme de phrase, ou un bloc de syntagmes de phrase, et un syntagme de phrase correspond, dans le cas le plus général, à celle qui existe en LN entre un complément d'objet direct et un infinitif ou une forme verbale personnelle ou entre un complément déterminatif, paraphrasable par un complément d'objet direct, et son noyau nominal. Dans certains cas, fixés dans le lexique, la relation verbale correspond à celle qui existe en LN entre le complément d'objet indirect ou de régime verbal et un infinitif ou une forme verbale personnelle ou entre un complément déterminatif, paraphrasable par un complément d'objet indirect ou de régime verbal, et son noyau nominal.

(4) Une *construction documentaire* (CD) est constituée de deux phrases documentaires reliées par un prédicat (§ 3.6). Notation : phrase documentaire, signifiant du prédicat r phrase documentaire. Le contenu significatif (§ 5.2) de la phrase documentaire de droite porte sur le contenu significatif de la phrase documentaire de gauche. Prédicats incorporés :

Signifié	Signifiant	Radical discriminant repris par l'anaphorique
« but »	avec l'objectif d(e)	objectif
« instrument »	au moyen d(e)	moyen
« relation »	en relation avec	relation
« coopération »	en coopération avec	coopération
« influence »	influençant, provoquant, entraînant	influence

(D) Relations entre structures grammaticales. L'anaphorique. Contenus significatifs déterminés par les relations entre structures grammaticales.

(1) *Item anaphorique avec ses articulations* (§ 1.1.1 et 4.1)

Items anaphoriques

$\left\{ \begin{array}{l} \text{Signifiant du} \\ \text{spécificateur A} \\ \text{(artic. libre)} \end{array} \right\}$	Forme	(Référence)	Terme	(Développement	Indicateur
	anaphorique		repris	de l'anaphorique)	de procédé
Préposition de	Forme	(Référence)	Terme	(Développement	Indicateur
compl. verbal	anaphorique		repris	de l'anaphorique)	de procédé

La préposition de complément verbal et le spécificateur A ont le même comportement que dans les sous-structures grammaticales sans anaphorique. Sur le développement de l'anaphorique, cf. fiche n° 2, § 2.1.1.3.

Forme anaphorique : ce, cet, cette, ces (§ 4.1, p. 32).

Référence : précise le renvoi en excluant l'ambiguïté; se note entre parenthèses (cf. fiche n° 4, p. 6).

Terme repris est le signifiant — expression ou mot — qui calque le signifiant d'un antécédent, non prédicatif ou prédicatif (§ 4.1, p. 32).

L'*antécédent non-prédicatif* d'un terme repris est le mot ou l'expression qui apparaît dans une sous-structure grammaticale d'une structure grammaticale qui précède celle où apparaît le terme repris ou dans la PD de gauche d'une CD, le terme repris apparaissant dans la PD de droite, et qui est le seul à posséder les caractéristiques suivantes : (a) le signifiant de cet antécédent incorpore le même radical discriminant que celui qui est incorporé dans le terme repris, *ou bien* (b) le signifiant de cet antécédent est le signifiant d'un spécificateur de notion et il est identique à celui du terme repris; le terme repris peut, dans ce cas, être suivi ou non de la même expression que celle qui suit l'antécédent; *et* (c) les traits sémantiques introduits par l'indicateur de procédé de l'item de la sous-structure grammaticale où apparaît l'antécédent sont les mêmes que les traits sémantiques introduits par l'indicateur de procédé de l'item où apparaît le terme repris (§ 4.1, p. 32). Chacune des sous-structures grammaticales du noyau avec coordination interne peut contenir un antécédent qui aura le même comportement par rapport au(x) terme(s) repris respectif(s) que tout autre antécédent (§ 4.1, p. 33).

L'*antécédent prédicatif* d'un terme repris apparaît dans le prédicat d'une structure grammaticale qui précède celle où apparaît le terme repris; le radical discriminant de ce prédicat est incorporé dans le terme repris (§ 4.1, p. 33).

Le *contenu significatif de l'antécédent* est le signifié de l'antécédent, prédicatif ou non-prédicatif. Le signifié d'un antécédent *non-prédicatif* est constitué par la notion de l'item qui se situe dans la sous-structure grammaticale où apparaît l'antécédent et par les traits sémantiques que son indicateur de procédé introduit, et en plus, si l'item est un t3 ou un a3, par le signifié du spécificateur de notion. Le signifié d'un antécédent *prédicatif* est la relation établie par le prédicat entre les deux phrases documentaires (§ 4.1, p. 33).

L'*indicateur de procédé* (§ 4.1, p. 33), placé immédiatement à la suite du terme repris, est celui qui est exigé par la syntaxe de la structure grammaticale où apparaît l'item anaphorique; l'indicateur de procédé placé après l'antécédent est celui qui est exigé par la syntaxe de la structure grammaticale où apparaît l'antécédent.

(2) *Classement des structures grammaticales* (cf. fiche n° 2, § 2.3 et fiche n° 5, § 4.2), en expressions documentaires (ED) ou en sous-expressions documentaires (SED), selon que le contenu significatif véhiculé par la structure grammaticale est, par lui-même, fortement pertinent ou faiblement pertinent, par rapport à un document donné. C'est au niveau de l'ED entière que « s'établit le contact » entre le document et l'indexation; l'ED est l'unité qui exprime le contenu d'un document. L'utilisation conjointe du classement de structures et de l'anaphorique permet de choisir le niveau d'indexation le plus pertinent.

(3) *Contenu significatif anaphorique (ou contenu significatif de l'item anaphorique, cf. § 4.3).*

Contenu significatif du terme repris x : contenu significatif anaphorique de l'item où apparaît le terme repris x.

Terme repris x : item anaphorique où apparaît le terme repris x.

Un antécédent peut être calqué par un ou plusieurs termes repris (les définitions, modalités et contraintes qui suivent portent toujours sur des termes repris qui calquent un seul et même antécédent).

Terme repris final : dernier terme repris qui apparaît soit dans l'ED où apparaît l'antécédent, soit dans une ED où apparaît un terme repris initial.

Terme repris initial : premier terme repris qui, dans une ED autre que celle où apparaît l'antécédent, calque cet antécédent.

Déterminations à droite d'un item d'un noyau x : contenu significatif de toutes les entités A – item(s) A, syntagme A ou bloc A – qui, dans la phrase documentaire où apparaît le noyau x, portent sur lui, et/ou contenu significatif du syntagme de phrase ou du bloc de syntagmes de phrase qui, dans la phrase documentaire où apparaît le noyau, porte sur le syntagme de phrase où apparaît le noyau x.

Modalités de constitution des contenus significatifs des termes repris :

Le contenu significatif

(I) *des termes repris qui suivent dans une ED soit l'antécédent, soit un terme repris initial* est le contenu significatif du terme repris qui précède immédiatement, avec, s'il y en a, ses déterminations à droite, ou, à défaut, le contenu significatif de l'antécédent, avec, s'il y en a, ses déterminations à droite.

(II) *du terme repris initial non suivi d'une référence renvoyant à un terme repris* est soit le contenu significatif du terme repris final de l'ED où apparaît l'antécédent, avec, s'il y en a, ses déterminations à droite, soit, à défaut du terme repris final, le contenu significatif de l'antécédent, avec, s'il y en a, ses déterminations à droite.

(III) *du terme repris initial suivi d'une référence renvoyant à un terme repris final précédent ou à un terme repris initial précédent non suivi d'autres termes repris* est le contenu significatif du terme repris auquel il renvoie avec, s'il y en a, ses déterminations à droite.

Contraintes d'utilisation de l'anaphorique (§ 4.3, p. 36) :

(I) Si plusieurs SED s'intègrent à une même ED, la dernière SED doit incorporer tous les contenus significatifs des SED précédentes.

(II) Un terme repris et sa forme anaphorique qui apparaissent dans une ED ne peuvent incorporer un contenu significatif anaphorique auparavant introduit dans une autre ED que si le contenu significatif anaphorique ainsi créé et le contenu significatif auparavant introduit, expriment le même référent.

(4) *Rôle de l'anaphorique (§ 4.4)*

(a) Réintroduire dans le contenu significatif de l'item anaphorique et, par conséquent, dans celui de la structure grammaticale à laquelle il appartient, un contenu significatif antérieurement exprimé;

(b) Marquer l'existence d'une *identité référentielle* entre l'entité à laquelle se réfère le contenu significatif de l'item anaphorique et l'entité à laquelle se réfère le contenu précédemment exprimé et réintroduit par le moyen de l'anaphore.

Le rôle (a) s'exerce entre SED appartenant à la même ED; les rôles (a) et (b) entre ED différentes. Le rôle (a) permet de tourner certaines contraintes syntaxiques de V1 : introduction d'une détermination qui porte sur un des items qui constitue le noyau avec coordination interne (§ 5.1.1, p. 42), d'une entité A qui correspond au complément d'objet direct ou indirect ou de régime verbal dans l'expression LN qui correspond au groupe A (§ 5.1.2.1.1, p. 45), d'une entité A qui remonte en arrière, portant sur un noyau plus à gauche que celui qui précède immédiatement cette entité A (§ 5.1.2.1.3, p. 47).

Appendice II

MISE A JOUR DES EXEMPLES D'INDEXATION DE LA FICHE N° 4

Les exemples de la fiche n° 4 ont été reformulés ci-dessous dans leur zone (4) (*Indexation avec V1 : écriture par le documentaliste*) en fonction, notamment, des précisions et des modifications introduites dans la fiche n° 5 sur le langage V1 ; dans quelques cas on a introduit des modifications de détail relevant du contenu des documents indexés.

Exemple 1 :

1. Recherche scientifique t0 en Inde a1, en bibliothéconomie a1 et science de l'information a1.
2. Organismes t0 pour cette recherche scientifique a1.
1 2
3. &1 = projets t0.
4. Financement t1 de cette recherche scientifique t0.
1 2
5. &1 = diffusion t1.

Exemple 2 :

1. Centre de documentation t3 et filmothèque t0 au Royaume Uni a1 spécialisés en s films a1 b éducatifs a1.

Exemple 3 :

1. Bibliothèque t0 universitaire a1 en Allemagne (république fédérale) a1.
2. Historique t0 de cette bibliothèque a1.
3. Statistiques t0 et rapport t0 d'activité a1 de cette bibliothèque a1 portant sur s les acquisitions a1 b de cette bibliothèque a1.
4. Ces statistiques t0 et ce rapport t0 portant sur s les catalogues a1 b de cette bibliothèque a1.
1 2 3
- 4 5-9. &3 = les fonds a1 b, le budget a1 b, le prêt a1 b dans, la reprographie a1 b dans, la formation a1 b professionnelle a1 dans.
- 10.1 Utilisation t1 de cette bibliothèque t0. 2 Ces statistiques t0 et ce rapport t0 portant sur s cette utilisation a1.
- 11.1 Gestion t1 de cette bibliothèque t0. 2 Ces statistiques t0 et ce rapport t0 portant sur s cette gestion a1.
12. Section t0 médicale a1 de cette bibliothèque a1.

Exemple 4 :

1. Recherche t1 documentaire a1 automatisée a4 par les bibliothèques a1 b universitaires a1 en France a1.
2. Accès t1 de ces bibliothèques a1 en mode conversationnel a1 aux bases de données t0.
3. Formation t1 professionnelle a1 du personnel t0 de ces bibliothèques a1 affecté à cette recherche a1.

Exemple 5 :

1. Service de microreproduction t3 par microfilm a1 d'une bibliothèque a1 b nationale a1 au Canada a1, avec l'objectif de r la conservation t1 et la microédition t1.

Exemple 6 :

1. Évaluation t1 de l'indexation t1 par indicateur de rôle a1.
2. Conversion t1 automatique a1 en index a1 b permuté a4 de cette indexation t1.
3. Cette évaluation t1, au moyen de r cette conversion t1.

Exemple 7 :

1. Centre de documentation t3 en France a1 sur s l'urbanisme a1.
2. Automatisation t1 de ce centre t3, influençant r ce centre t3.

1	2	3	4
---	---	---	---
3. & 4 = les utilisateurs t3 de ce centre t3.
4. Activités t0 documentaires a1 de ce centre a3.
5. Comportement t0 du personnel a1 b de ce centre a3.
6. Enquête t1 auprès de ces utilisateurs t3.

Exemple 8 :

1. Banque de données t0 terminologiques a1 multilingues a1.
2. Cette banque t0, avec l'objectif de r l'assistance t1 automatique a1 au traducteur t3.
3. Données t0 terminologiques a1 stockées dans cette banque a1.
4. Recherche t1 en mode conversationnel a1 de ces données t0.
5. Interrogation t1 en mode conversationnel a1 de cette banque t0.
- 6.1 Traitement t1 automatique a1 des textes t0. 2 Élaboration t1 des glossaires t0. 3 Interrogation t1 par ce traitement a1 b pour s cette élaboration a1 de cette banque t0 (4,2,3).

1	2	3	4
---	---	---	---

Exemple 9 :

1. Bibliothèques t0 universitaires a1 d'Allemagne (république fédérale) a1, en coopération

1	2	3	4
---	---	---	---

 avec r les bibliothèques t0 de Belgique a1.

5	6
---	---
2. & 6 = des Pays-Bas a1.
- 3.1 Échange t1 de documents t0. 2 Cette (1) coopération t1 internationale a1 par cet

1	2
---	---

 échange a1 et par prêt a1 b interbibliothèque a1.

3	4	5
---	---	---
4. &1 cette (2) coopération t1.

Exemple 10 :

- 1.1 Bibliothèques t0 spécialisées a1 b en éducation a1. 2 Ces bibliothèques t0 en Alle-

1	2
---	---

 magne (R.D.A.) a1 dépendant des organismes de l'enseignement a3 b supérieur a1. 3 & 3-4 = dépen-

3	4
---	---

 dant d'autres organismes a1 b spécialisés a4 b en éducation a1.
2. Ces bibliothèques [spécialisées allemandes] t0, en coopération avec r les organismes d'in-

1	2
---	---

 formation t3 et les organismes de documentation t3 spécialisés en s éducation a1.

3	4	5
---	---	---
3. & 3 5 = les bibliothèques t0 publiques a1.
- 4.1 Ces (1) bibliothèques [spécialisées allemandes] t0, en coopération avec r les bibliothèques
 t0 et les organismes d'éducation t3 en URSS a1.
5. Coopération t1 internationale a1 de ces (1) bibliothèques [spécialisées allemandes] a1.
6. Formation t1 professionnelle a1 et politique a1 des personnels t0 de ces (1) biblio-
 thèques [spécialisées allemandes] a1.