



HAL
open science

Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales - 1re 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 - 1re partie. Condenser - Adosa, Clermont-Ferrand, 1981, 2, pp.41-89. hal-01071424v1

HAL Id: hal-01071424

<https://hal.science/hal-01071424v1>

Submitted on 5 Oct 2014 (v1), last revised 16 Oct 2014 (v2)

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

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

Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales – 1^{re} 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, janvier 1981, n° 2, p. 41-89

Résumé

Ce document présente la grammaire du système documentaire Vercingétorix I (= V1). La fonction de cette grammaire est de spécifier les structures grammaticales schématiques dans lesquelles le documentaliste traduit l'information exprimable en langue naturelle. Les objectifs de fond de cette fiche sont (a) d'explicitier les contraintes auxquelles doivent obéir les structures grammaticales schématiques de V1 et les relations entre structures grammaticales schématiques et (b) d'explicitier les contenus significatifs déterminés par les structures grammaticales schématiques ou qui relèvent des relations entre structures grammaticales schématiques.

Le § 1 introduit la terminologie et les conventions nécessaires à la bonne compréhension de l'ensemble. Le § 2 est consacré à l'organisation du lexique de la grammaire de V1. On y trouvera l'appareil conceptuel permettant d'organiser les aspects de chaque unité lexicale qui sont nécessaires et suffisants au fonctionnement de la syntaxe. Le § 3 présente de manière analytique, et avec un souci d'efficacité pédagogique, les contraintes combinatoires qui règlent les structures grammaticales et les relations internes à celles-ci. Le § 4 traite des relations entre structures grammaticales dans le texte documentaire, via les expressions anaphoriques. Le § 5 est consacré à l'analyse des contenus significatifs de V1 et des relations qui permettent de les combiner : coordination et projection, cette dernière pouvant être attribution ou relation verbale.

Ce document se prolonge par une deuxième partie (§ 6 à 8 et appendices) publiée dans *Condenser*, n° 3, p. 3-31.

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.

Gabriel G. Bès et Pierre-Maurice Fauchère. « [Le langage V1 : structures engendrées par la grammaire et relations entre structures grammaticales \(2^e partie\)](#) ». *Condenser*, Adosa, Clermont-Ferrand, avril 1982, n° 3, p. 3-31.

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.

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)

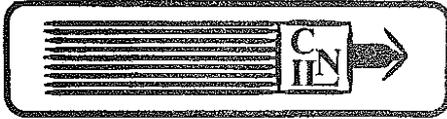
G.G. Bès et P.-M. Fauchère : **Le système documentaire Vercingétorix 1.**

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).



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**

SOMMAIRE

Première partie

- 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

(à paraître dans *Condenser* n°3)

- 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.

1 INTRODUCTION

Dans la fiche n° 2 (« Les informations indexées »), on a distingué les structures grammaticales des autres types d'informations indexées (ceux-ci sont, on s'en souviendra, les structures de transition, créées à des fins stylistiques par une instruction de permutation ; les structures lisibles, produites par la machine, qui sont sous-jacentes aux différents formats de sortie ; le classement de structures, en expressions documentaires (ED) et sous-expressions documentaires (SED), classement établi par le documentaliste en fonction de la pertinence documentaire de l'information véhiculée par une structure grammaticale). Les structures grammaticales sont produites soit par écriture directe du documentaliste, soit par une instruction de commutation. Dans la même fiche n° 2, on a signalé que les structures grammaticales sont composées de deux parties, l'information de base et les articulations, la première assurant l'essentiel de la fonction de compréhension et la deuxième, l'essentiel de la fonction de lisibilité. On distinguera désormais *deux* types d'articulations : les *articulations libres* et les *articulations grammaticales* ; les premières sont entièrement laissées au libre choix du documentaliste, tandis que les secondes sont prévues dans la grammaire comme possibilités de choix à l'intérieur desquelles le documentaliste doit faire porter son élection.

La grammaire de V1 *spécifie* ou *engendre* les structures grammaticales ; on entend par là, d'une manière peu technique, que la grammaire de V1 marque les limites dans lesquelles les structures grammaticales de V1 doivent être construites : elle indiquera, par exemple, comment un item est constitué, selon quelles restrictions les items se combinent entre eux, la liste des notions à partir desquelles on peut construire un item, etc.

Cette spécification des structures grammaticales se fait de trois manières différentes, selon qu'il s'agit de l'information de base, des articulations grammaticales ou des articulations libres.

L'information de base est celle qui est la plus strictement spécifiée par la grammaire : celle-ci offre, certes, un certain nombre de choix, mais le documentaliste doit toujours opérer à l'intérieur de possibilités pré-établies. Si l'information de base a été exprimée de manière adéquate, la machine sera toujours en mesure de l'interpréter, en particulier d'établir les relations entre les items.

Les articulations libres sont celles qui, à l'opposé, sont le plus lâchement spécifiées par la grammaire : celle-ci n'indique que les emplacements de ces articulations ; c'est au documentaliste à remplir ces emplacements de la manière qu'il estime convenable et en puisant dans un ensemble de possibilités non prévues à l'avance par la grammaire. La seule limite en est, par conséquent, le bon sens du documentaliste, guidé par sa connaissance de la fonction des articulations et par l'objectif qu'il veut atteindre : il ne doit pas y avoir interférence avec l'information de base et les articulations grammaticales.

Dans une zone intermédiaire, entre l'information de base et les articulations libres, se situent les articulations grammaticales. Comme les articulations libres, elles ne sont pas indispensables à la reconnaissance par la machine des relations entre les items ; mais, à la différence des articulations libres, elles modifient le contenu significatif (au sens défini ci-dessous) des structures grammaticales et doivent être choisies à l'intérieur des possibilités pré-déterminées par la grammaire.

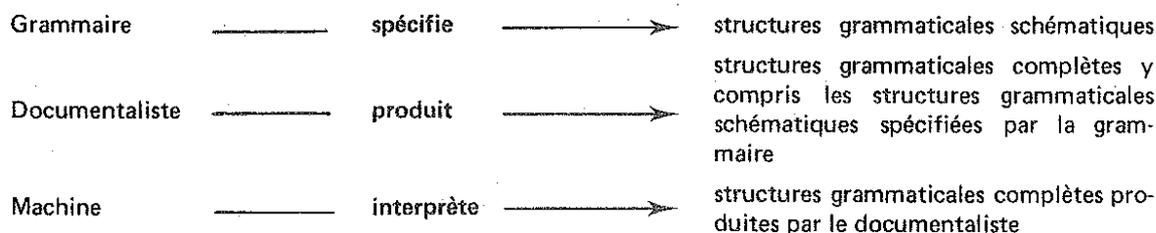
A partir de ce qui vient d'être dit, on peut facilement établir la distinction entre les *structures grammaticales complètes* et les *structures grammaticales schématiques*. Les premières sont les deuxièmes, *plus* le remplissage des articulations libres, décidé par le documentaliste ; les deuxièmes sont constituées par l'information de base, les articulations grammaticales et les emplacements des articulations libres. On a par conséquent :



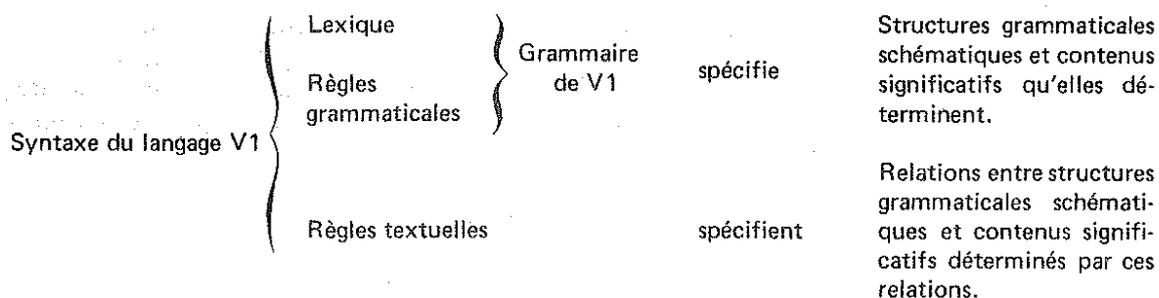
Parallèlement à la distinction entre entités complètes et entités schématiques, on introduira la distinction entre *contenu* et *contenu significatif*, le premier étant associé aux entités complètes et le deuxième aux entités schématiques ou aux relations entre structures grammaticales schématiques. C'est le contenu significatif qui est pertinent pour la recherche documentaire.

La fonction de la grammaire de V1 est de spécifier les structures grammaticales schématiques qui, après remplissage des articulations libres, deviendront les structures grammaticales complètes. Il convient d'insister sur les points suivants : la grammaire de V1 n'est ni une machine de production de structures grammaticales schématiques, ni une machine de reconnaissance de structures grammaticales complètes ou schématiques, ni une machine qui traduit des énoncés produits en langue naturelle en structures grammaticales complètes ou schématiques. C'est le documentaliste qui, connaissant la spécification des structures grammaticales schématiques et, plus particulièrement, leur contenu significatif, traduit une information exprimable en langue naturelle en information exprimée par une structure grammaticale complète. C'est enfin la machine, troisième entité de base du système V1 (cf. fiche n° 2, § 1), qui interprète les structures grammaticales complètes, en opérant notamment sur l'information de base et sur les articulations grammaticales; la machine le fait pour : (1) aider le documentaliste, en particulier lui signaler les erreurs qu'il aura éventuellement commises (cf. fiche n° 3); (2) effectuer les opérations destinées à produire les structures lisibles; (3) effectuer les opérations de recherche documentaire.

On peut représenter schématiquement la fonction de la grammaire de V1 de la manière suivante :



La grammaire de V1 spécifie les structures grammaticales schématiques. Celles-ci s'organisent dans un *texte documentaire* ou ensemble de structures grammaticales associées à un document déterminé. Certaines des relations qui s'établissent entre structures grammaticales dans un texte documentaire sont spécifiées par les *règles textuelles*, qui s'ajoutent ainsi à celles qui, dans la grammaire, spécifient la combinatoire et les relations internes aux structures grammaticales. Les relations entre structures grammaticales déterminent, elles aussi, des contenus significatifs. On a par conséquent :



Les objectifs de fond de cette fiche sont :

- (a) d'expliciter les contraintes auxquelles doivent obéir les structures grammaticales schématiques de V1 et les relations entre structures grammaticales schématiques;
- (b) d'expliciter les contenus significatifs, aussi bien ceux qui sont déterminés par les structures grammaticales schématiques que ceux qui relèvent des relations entre structures grammaticales schématiques.

Le § 2 sera consacré à l'organisation du lexique de la grammaire de V1, mais celui-ci n'y sera pas présenté de manière exhaustive; on y trouvera plutôt l'appareil conceptuel permettant d'organiser les aspects de chaque unité lexicale qui sont nécessaires et suffisants au fonctionnement de la syntaxe. C'est cette non-exhaustivité de la description du lexique qui empêche que la grammaire de V1 soit effectivement utilisable dans son état actuel. Le § 3 présente de manière analytique, et avec un souci d'efficacité pédagogique, les contraintes combinatoires qui règlent les structures grammaticales et les relations internes à celles-ci. Le § 4 traite des relations entre structures grammaticales dans le texte documentaire. Le § 5 et le § 6 sont consacrés aux contenus significatifs de V1. Les exemples, les observations de détail, voire les répétitions pour faciliter l'exposé, ne sont pas ménagés dans ces § 3 à 6, où le souci pédagogique est dominant. En revanche, dans le § 7, sous une forme plus technique et plus formalisée et dans une présentation plus compacte, on trouvera les règles syntaxiques de la grammaire de V1. Le dernier paragraphe (§ 8) amorcera une discussion de V1, sans pourtant aborder les questions de fond sur les rapports de la syntaxe et de la sémantique dans les langages documentaires, ou ceux des langues naturelles avec les langages documentaires. L'appendice n° 1 est un résumé préparé pour faciliter le maniement du langage V1; dans l'appendice n° 2, on trouvera une nouvelle formulation des exemples de la fiche n° 4 en fonction des modifications introduites dans la présente fiche*.

Dans la suite de ce § 1 on introduira la terminologie et les conventions nécessaires à la bonne compréhension des § 2 à 6.

1.1 Terminologie et conventions

Comme dans les fiches précédentes, la terminologie et les conventions de notation sur des points particuliers seront présentées au fur et à mesure de leur apparition. Par ailleurs, on continuera à se servir aussi de la terminologie couramment admise dans les écrits de ce type — si cette terminologie, en ce qui concerne son utilisation dans ce travail, ne prête pas trop à l'équivoque — et cela en ayant conscience qu'on est loin de disposer de définitions entièrement satisfaisantes, comme par exemple pour des termes tels que *mot*, *signifié*, *langue naturelle*, etc.

Dans les deux paragraphes qui suivent, on groupera, en revanche, des précisions utiles aussi bien sur la terminologie et les conventions de caractère général qui relèvent de V1 (§ 1.1.1) que sur la terminologie portant sur la langue naturelle (§ 1.1.2).

1.1.1 Terminologie et conventions relevant de V1

La distinction entre entités schématiques et entités complètes, introduite à propos des structures grammaticales, s'applique également à d'autres entités du langage V1 (ces entités seront introduites et/ou développées au § 3), à savoir : phrase documentaire, construction documentaire, syntagme de phrase, bloc de syntagme de phrase, syntagme A, bloc A, suite A. Dans tous ces cas, on retrouve le schéma déjà introduit à propos des structures grammaticales. Cela implique que, par exemple, une entité comme le syntagme de phrase peut être soit un syntagme de phrase complet, soit un syntagme de phrase schématique.

* Nous remercions A. E. Hurtado, qui a bien voulu discuter avec nous un certain nombre d'aspects de la grammaire de V1.

L'astérisique, placé à la suite d'une expression quelconque, indique que l'expression en question est déviante et, par conséquent, non admise par la grammaire de V1. Les parenthèses, utilisées dans un contexte où on évoque des caractéristiques spécifiques de la grammaire de V1, indiquent que les expressions qu'elles enferment sont optionnelles.

1.1.2 Mise au point terminologique sur la langue naturelle

La liste terminologique qui suit ne vise qu'à fixer une nomenclature minimale qu'on utilisera par la suite, et ceci de la manière la plus simple et la plus expéditive qui soit : en présentant une étiquette terminologique suivie de quelques exemples de structures en langue naturelle (désormais langue naturelle = LN) et en ajoutant, parfois, quelques commentaires ou précisions. Ainsi, on ne définira pas, à proprement parler, ce qu'est une *racine verbale*, mais on indiquera simplement :

RACINE VERBALE : *Indexation* ; *Indexer* ; Il *indexe* souvent.

Cette mise au point terminologique concerne la catégorisation des mots et des constructions et/ou relations syntaxiques.

Catégories de mots

NOM : *porte* ; une *table* ; les *étagères* ; la *traduction*.

NOM VERBAL : la *traduction* ; l'*indexation*.

SUFFIXÉ DE NOMINALISATION : *traduction* ; *vernissage* ; *financement*.

RACINE VERBALE : *indexation* ; *indexer* ; il *indexe* souvent ; *indexé*.

FORME VERBALE : *indexation* ; *indexer* ; il *indexe* souvent ; *indexé* (mot où est incorporée une forme verbale).

VERBE OU FORME VERBALE À UN MODE PERSONNEL : il *indexe* souvent ; il *a indexé* ; il *doit avoir indexé* (racine verbale suivie de la terminaison personnelle, ou auxiliaire suivi d'un participe passé, ou modal avec terminaison personnelle suivi d'un infinitif).

INFINITIF : *traduire* ; *avoir traduit*.

ADJECTIF : *bon* ; *automatique* ; livre *facile* à lire ; collection *imprimée* en offset (Les participes passés qui n'apparaissent pas dans une forme verbale à un mode personnel et/ou avec un infinitif, sont assimilés à des adjectifs).

Constructions et relations syntaxiques

SYNTAGME NOMINAL :

<i>les</i>	<i>filles</i>	<i>joyeuses</i> ;
<i>un</i>	<i>enfant</i>	<i>malade</i> ;
<i>des</i>	<i>livres</i>	<i>faciles à lire</i> ;
<i>ton</i>	<i>couteau</i>	<i>en acier</i> ;
<i>les</i>	<i>travailleurs</i>	<i>du Maroc</i> ;
<i>les</i>	<i>documents</i>	<i>traduits par Michel</i> ;
<i>la</i>	<i>traduction</i>	<i>automatique</i> ;
<i>la</i>	<i>traduction</i>	<i>des ouvrages scientifiques</i> ;
<i>l'</i>	<i>avertissement</i>	<i>au lecteur</i> ;
<i>le</i>	<i>manuscrit</i>	<i>en impression</i> ;
<i>le</i>	<i>développement</i>	<i>de l'institution</i> .

Schéma : Pré-déterminant

Noyau nominal

Complément déterminatif

SYNTAGME PRÉPOSITIONNEL :

<i>par</i>	<i>la traduction automatique</i> ;
<i>pour</i>	<i>le développement de l'institution</i> .

Schéma : Préposition

Syntaxme nominal

SYNTAGME ADJECTIVAL :

<i>facile</i>	<i>à lire</i> ;
<i>apte</i>	<i>à être imprimé</i> ;

loppées de manière systématique. On remarquera, cependant, que ce § 2 porte sur l'*organisation* du lexique; on n'y trouvera pas, par conséquent, le *lexique* de V1 lui-même dans une présentation achevée et exhaustive, mais seulement les principes de son organisation.

Le *Thésaurus Sciences de l'Information* (Paris, CNRS, 1977) a été utilisé comme point de référence pour la confection du lexique : il fournit les limites du champ référentiel considéré, l'équilibre de ses différentes parties, le recensement des notions utiles et leur situation les unes par rapport aux autres. On a adopté donc, par principe, à ce stade de l'élaboration de V1, la vision qu'il donne de ce champ référentiel, sans que cela signifie pour autant une adhésion sans réserves aux options globales ou de détail de cet outil documentaire.

Le lexique de V1 est constitué d'un nombre fini d'*entrées lexicales* et de *renvois*. Les entrées lexicales sont organisées en fonction d'un certain nombre de concepts : (a) les *notions*, leur correspondance avec les signifiés en langue naturelle et leur notation par un *radical discriminant*; (b) les *procédés* avec leur notation, les *indicateurs de procédé*; (c) les *supports de radical normalisés*, pour noter les notions suivies et/ou précédées d'une articulation déterminée et les *variations du radical discriminant*; (d) le *système de traits, sémantiques, de sélection et particuliers*. Chaque entrée lexicale doit se spécifier en fonction des concepts qu'on vient d'énumérer. La spécification de chaque entrée particulière et le système général de concepts qui est sous-jacent à chaque spécification particulière doivent être soigneusement distingués. On présentera par la suite chacun des concepts généraux dont on a fait état ci-dessus et la répercussion de ce concept dans l'organisation de chaque entrée particulière.

Les renvois seront traités dans le même paragraphe que les supports de radical normalisés et les variations du radical discriminant.

2.1 Notions

Chaque entrée lexicale est constituée, parmi d'autres éléments, par une *notion*. Pour cerner ce qu'est une notion, on introduira au préalable les concepts de *famille de mots* et de *famille restreinte de mots*, concepts qui relèvent de la langue naturelle.

Chacun des groupes (1) et (2), qui suivent, constitue une *famille de mots* :

(1) aimer, amour, aimant, aimé, aimés, (ils) aiment, etc.

(2) photographe, photographie, photographier, (il) photographiait, photographié, etc.

Les mots de chaque famille possèdent un contenu permanent qui apparaît dans chacun des mots de la famille et qu'on doit retenir en faisant abstraction des catégories grammaticales (ce n'est ni le verbe *aimer* ni le nom verbal *amour*, mais ce qu'il y a de commun aux deux), des éléments de dérivation (ce n'est ni *photographie* ni *photographe*, mais ce qu'ils ont en commun), des accidents grammaticaux et des éléments de flexion (ce n'est ni *amour*, ni *amours*, ni *aimait*, ni *aimer*, ni *aimé*, mais ce qu'il y a de commun et de permanent à toutes ces formes).

Une *famille restreinte de mots* est un type particulier de famille de mots. Plus précisément, on dira qu'une famille restreinte de mots en langue naturelle est constituée par tous les mots de la langue qui possèdent un radical fixe et/ou des variations de radical régulières ou codifiées dans le dictionnaire, associés à un signifié invariant et pouvant être suivis par des indicateurs de genre et de nombre, par tous les éléments de flexion verbale et par certains suffixes : (a) de nominalisation dans les noms verbaux (p. ex., dans *indexation*), (b) dont le signifié est « machine ou outil pour faire quelque chose » (p. ex., dans *photocomposeuse*), (c) dont le signifié est « personne qui fait quelque chose, qui est spécialiste en quelque chose ou dont l'activité est caractéristique d'une institution déterminée » (p. ex., dans *indexeur*, *chimiste*, *bibliothécaire*), (d) dont le signifié est « le fait ou la qualité d'être x », x étant l'adjectif de la famille de mots (p. ex. dans *gratuité*). Un lexème suivi d'un suffixe dont le signifié est autre que ceux indiqués dans (a) à (d) est traité comme un radical unique. Grâce à cette caractérisation des familles de mots, si *indexer* et

indexeur sont dans une même famille, *normaliser* — avec un suffixe autre que ceux qu'on vient de signaler — et *norme* appartiennent à des familles différentes; à remarquer également que *Français*, *français* (langue française) et *français* (de la France), ne possédant pas un signifié invariant, n'appartiennent pas à une même famille de mots. Par extension, on considère qu'appartient aussi à une famille restreinte de mots, tout mot ou syntagme qui, sans présenter le radical fixe et/ou des variations de radical régulières ou codifiées dans le dictionnaire, est associé à un contenu qui coïncide avec celui de ce radical suivi des suffixes admis dans la famille restreinte de mots. Par exemple, *machine à écrire* appartient à une même famille que *dactylographe*. Par extension toujours, pour la constitution des familles restreintes de mots, on traite certaines expressions en langue naturelle (les syntagmes du type *recherche scientifique*, *page de couverture* ou *mise en page*) comme étant des mots.

Il est maintenant possible de revenir au concept de *notion*. Le contenu de chaque notion ne sera pas défini dans le lexique de V1 de manière explicite, en fonction, par exemple, d'un système complexe d'entités sémantiques ou par des traductions particulières. Chaque entrée lexicale, en ce qui concerne la notion, ne comporte que, obligatoirement, un *mot de rappel* en langue naturelle et, optionnellement, des précisions sur le contenu de la notion, précisions qui seront apportées par des exemples ou des constructions types.

Le mot de rappel, en langue naturelle, sert à faire état de la famille restreinte de mots en langue naturelle dont le signifié invariant est identique au contenu de la notion. Il identifie l'entrée lexicale et est à la base de son classement alphabétique. Ce mot de rappel en langue naturelle est soit un verbe à l'infinitif, soit un nom, soit un adjectif. Le choix de chacune de ces formes n'a pas été fait au hasard, mais en fonction de certains principes généraux, résumés par la suite (dans la formulation de ces principes, le nom sans suffixe mais dont le signifié est équivalent à celui d'un nom avec suffixe — par exemple *interprète* — est assimilé à un nom avec suffixe) :

Familles restreintes de mots constituées par :

- (1) Verbe, nom(s) avec suffixe, (adjectif)
- (2) Nom sans suffixe, (nom(s) avec suffixe), (adjectif)
- (3) Verbe, nom sans suffixe, (nom(s) avec suffixe), (adjectif)
- (4) (nom avec le suffixe « le fait ou la qualité d'être x », x étant l'adjectif), adjectif

Mot(s) de rappel incorporé(s) :

verbe à l'infinitif
 nom sans suffixe
 verbe à l'infinitif et nom sans suffixe, dans deux entrées lexicales distinctes
 adjectif

Exemples :

- | | |
|--|--|
| (1) photographe, photographie, photographe, photographique, photographié
interpréter, interprétation, interprète
traduire, traducteur, traduction, traduit | photographe

interpréter
traduire |
| (2) bibliothèque, bibliothécaire
livre
syntaxe, syntaxique
page de couverture | bibliothèque
livre
syntaxe
page de couverture |
| (3) financer, finances, financement, financier, financé
documenter, document, documentation, documentaire
indexer, indexation, indexeur, indexé, index | financer, finances
documenter, document

indexer, index |
| (4) Bilinguisme, bilingue
gratuité, gratuit | bilingue
gratuit |

On peut visualiser une entrée lexicale d'un lexique imprimé comme étant composée d'une série de cases, placées en lignes et en colonnes. Dans la première case de cette représentation graphique du lexique, à l'intersection de la première ligne avec la première

colonne, est écrit le mot de rappel. Les premières lettres du mot de rappel ou certaines d'entre elles, notées en capitales, coïncident, dans le cas le plus général, avec le radical discriminant, c'est-à-dire avec la suite de caractères alpha-numériques nécessaires et suffisants pour que la machine reconnaisse la notion (cf. fiche n° 2, § 2.2). Dans la case en-dessous de celle occupée par le mot de rappel, on incorpore éventuellement les précisions sur le contenu (cf. tableaux, à la fin de ce § 2).

2.2 Procédés

Les notions du lexique sont catégorisées en fonction de leur aptitude à acquérir un ou plusieurs *procédés*, parmi une liste finie et très restreinte. Les procédés ont une double fonction : (a) indiquer la relation syntaxique d'une notion dans une structure grammaticale; (b) ajouter, dans certains cas, des traits sémantiques à cette notion. Il existe deux grandes catégories de procédés : T et A. Les procédés T sont acquis par les items qui sont noyau d'un syntagme de phrase; les procédés A sont acquis par les items du groupe A, qui, en ligne générale, possèdent une fonction d'attribut par rapport au noyau (cf. § 3).

La catégorie T comprend : t0 ou t1, t2, t3. La catégorie TV comprend : t1, t2; t3, lorsque la notion correspondante admet t1 (le mot de rappel est un verbe; cf. ci-dessous).

Une notion suivie de :

- t0 correspond, en langue naturelle, au nom du mot de rappel (sans un des suffixes admis par la famille restreinte de mots et qui ne possède pas le signifié de « celui qui fait quelque chose ») ou, si le mot de rappel est un adjectif, à un nom dont le suffixe est « la qualité d'être x ». Ex. : *document t0*, *bibliothèque t0*, *gratuité t0*. A remarquer qu'une notion suivie de t0 peut correspondre, en langue naturelle, aux signifiés de « machine » ou « personne », sans la spécification de leur fonction; ex. : *machine t0*, *agent t0*, *personnel t0*. Les syntagmes nominaux étant traités comme des mots dans la constitution des familles restreintes de mots, une notion suivie de t0 peut correspondre aussi à un syntagme nominal; ex. : *recherche scientifique t0*.

- t1 correspond, en langue naturelle, à un nom verbal, ou à un verbe à l'infinitif, ou à un syntagme nominal dont le noyau nominal est un nom verbal, ou à un syntagme figé du type *mise en page*, ou à un syntagme nominal du type *le fait de x*, x étant un verbe à l'infinitif. Le procédé t1 ajoute les traits sémantiques qui suivent au contenu de la notion :

+ action	ou	- action	ou	+ action
± effet		+ effet		- effet
I		II		III

La configuration I doit se lire « action avec ou sans effet »; la configuration II, « effet sans action » et la configuration III, « action sans effet ». Sauf indication contraire, par *procédé t1* on comprend t1 avec une quelconque des trois configurations. Les structures LN qui suivent illustrent ces trois possibilités :

- (1) La photographie est un art.
La traduction en France n'est pas encouragée.
- (2) La photographie placée sur ton bureau m'attriste.
Cette traduction anglaise de Molière est mauvaise.
- (3) On va t'apprendre à photographier.
Traduire est toujours difficile.

La configuration I est la plus largement utilisée; elle exprime l'indissociation de la cause et de l'effet. Si, dans l'interprétation d'un document, il y a un doute sur la configuration à utiliser, on doit avoir recours à la configuration I car, exprimant toujours l'action, elle couvre les deux possibilités d'avoir ou non l'effet. C'est celle qui est comprise par la machine chaque fois qu'elle trouve un t1 suivant une notion. En revanche, la configuration II, par la suite, sera notée par le symbole « - » précédant t1, et la configuration III,

par le symbole « + ». On souligne de cette manière, dans ces deux configurations, l'absence ou la présence actuelle du trait fondamental *action*. Exemples :

- (1) La photographie t1 à l'étranger a1.
La traduction t1 en France a1.
- (2) Les photographies - t1 d'un musée a1.
Les traductions - t1 d'une bibliothèque a1.
- (3) Photographier + t1 des estampes t0.
Traduire + t1 des auteurs t0 anglais a1.

Dans la plupart des cas, la racine du nom verbal ou le verbe à l'infinitif qui correspond à une notion suivie de t1 admettent un complément d'objet direct ; c'est le cas de tous les exemples qui précèdent immédiatement. Dans certains cas particuliers, ils admettent soit un complément d'objet indirect, soit un complément de régime verbal : dans ces cas particuliers, chaque entrée lexicale indique, dans la case correspondant au procédé, la préposition de complément verbal exigée par la notion suivie de t1 ou les prépositions exigées, parmi lesquelles le documentaliste doit en choisir une.

- t2 correspond, en langue naturelle, soit à un nom avec un suffixe dont le signifié est « machine qui sert à » (par exemple, *agrandisseur t2, photocomposeuse t2*), soit à un syntagme nominal du type *machine à x*, x étant un verbe à l'infinitif (par exemple, *machine à traduire t2*).

- t3 correspond à plusieurs possibilités en langue naturelle. Si la notion a admis un t0 et si elle possède le trait sémantique [+ institution], elle va correspondre en langue naturelle, lorsqu'elle est suivie de t3, à un nom avec un suffixe dont le signifié est « personne dont l'activité est caractéristique d'une institution déterminée » (par exemple, *bibliothécaire t3, libraire t3*) ou à l'expression *personnel* ou *personnel spécialisé de x*, x étant le nom de l'institution. Si la notion a admis un t0 et si elle possède le trait sémantique [+ AST] (AST = art, science, technique), elle va correspondre en langue naturelle, lorsqu'elle est suivie de t3, à un nom avec un suffixe dont le signifié est « personne qui est spécialiste en quelque chose » (*mathématicien t3, chimiste t3*) ou à l'expression *spécialiste en x*, x étant le nom d'une discipline. On incorpore ici les suggestions de Dorota Norman, dans *Relations sémantiques et syntaxiques dans un langage documentaire, Thesaurus Sciences de l'information*, Mémoire de maîtrise, Université de Clermont II, 1980, p. 61-62.

Si la notion suivie par t3 a admis un t1, elle va correspondre, en langue naturelle :

- soit à un nom avec un suffixe dont le signifié est « personne qui fait quelque chose » (par exemple, *indexeur t3, traducteur t3*) ou à un nom dont le signifié est « personne qui fait quelque chose » (par exemple, *interprète t3*) ou à un syntagme nominal du type *personne qui x* (par exemple, *personne qui stocke t3*) ou *personnel de x* ou *personnel spécialisé de x*, x étant une forme verbale qui incorpore le radical discriminant.

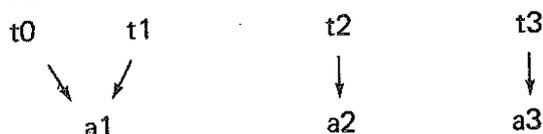
- soit à un syntagme nominal du type *centre* ou *service* ou *organisme de x*, x étant une forme verbale qui incorpore le radical discriminant (par exemple, *centre de documentation t3, service de traduction t3*).

Le choix entre *personne* (sous une des trois formes envisagées, suffixe, nom ou syntagme nominal) et *institution* (*centre, service, organisme*) doit être effectué par le documentaliste, qui écrira dans l'articulation grammaticale de l'item soit *centre, service* ou *organisme de x*, x étant en général un nom verbal (par exemple, *centre de, dans centre de documentation t3*), soit le suffixe, le nom ou le syntagme nominal qui correspondent à *personne* (par exemple, *mathématicien t3, interprète t3, personnel spécialisé de bibliothèque t3, traducteur t2*). On appellera spécificateurs de notion ces articulations grammaticales qui doivent obligatoirement accompagner une notion et qui précisent son contenu. Le signifiant du spécificateur « institution » est une des trois formes indiquées, parmi lesquelles le documentaliste doit obligatoirement choisir ; ces formes, étant générales et fixées à l'avance pour tout le lexique, ne seront pas rappelées dans chaque entrée lexicale ; en revanche, là où le spécificateur de notion « institution » demande un signifiant particularisé (par exemple, *maison d'édition t3*), celui-ci sera indiqué dans l'entrée lexicale

correspondante, dans la deuxième case de la colonne du procédé t3 (cf. ci-dessous). Le suffixe qui correspond en langue naturelle au spécificateur « personne » sera traité dans chaque entrée lexicale comme un support de radical, le signifiant du spécificateur « personne » étant l'absence d'un signifiant « institution » (cf. ci-dessous).

La catégorie A comprend : a1, a2, a3, a4.

Les procédés a1, a2, a3 introduisent le même trait sémantique, très large, de « propre ou relatif à quelque chose, en relation avec quelque chose ». La relation entre les procédés T et les procédés A s'établit comme suit :



a1 est le procédé qui, tout en faisant acquérir la fonction d'attribut à la notion, conserve les traits sémantiques propres aux procédés t0 ou t1 ; a2 fait de même avec t2, et a3 avec t3. Les notions suivies de ces procédés correspondent, en langue naturelle, soit à un adjectif, soit à un syntagme nominal ou prépositionnel dont le noyau nominal correspond au contenu de la notion. Exemples :

- Document t0 graphique a1.
- Document t0 d'une bibliothèque a1.
- Schéma t0 d'une photocomposeuse a2.
- Thèse t0 d'un documentaliste a3.
- Microfiches t0 d'un centre de documentation a3.

Comme les traits sémantiques introduits par le procédé t1 se retrouvent dans a1, le même choix qui existait en t1 entre trois configurations de traits sémantiques existe aussi pour les a1 qui admettent un t1, avec une notation analogue. Exemples (les structures grammaticales de V1, à gauche, correspondent aux structures LN, à droite) :

- | | |
|--------------------------------------|-----------------------------|
| Manuel t0 de traduction a1. | Manuel de traduction |
| Collection t0 de photographies - a1. | Collection de photographies |
| Documents t0 à photocopier + a1. | Documents à photocopier |

Le procédé a4 introduit le trait sémantique « ayant subi l'action de x », x étant un infinitif qui admet un complément d'objet direct. Une notion suivie du procédé a4 correspond, en langue naturelle, à un participe passé. Ex. : *ouvrage t0 acquis a4* correspond, en langue naturelle, à *ouvrage acquis*. L'item a4 de V1 ne conserve cependant pas l'ambiguïté de certains participes passés en langue naturelle. On sait, en effet, qu'une expression du type *document photocopié* peut s'interpréter soit comme *document produit par l'action de photocopier*, soit comme *document original qui a été reproduit par la photocopie*. Or l'item a4, en V1, doit toujours être interprété comme « produit par l'action de x », c'est-à-dire conformément à la première des deux interprétations précédentes. Pour exprimer le deuxième cas, on utilisera simultanément la notion *original* et l'item a4 ; on aura, par conséquent, les deux possibilités suivantes :

- (1) Document t0 photocopié a4.
- (2) Document t0 original a1 photocopié a4.

Il est à remarquer qu'aucune notion ne peut acquérir tous les procédés et que les procédés admis par chaque notion sont déductibles à partir du mot de rappel. La détermination des procédés que chaque notion peut acquérir se fait de la manière suivante :

- (1) Le mot de rappel est-il un verbe ?
 - si *oui*, la notion admet un t1 ;
 - si *non* (le mot de rappel est un nom ou un adjectif), la notion admet un t0.
- (2) La notion admet t0 :
 - la notion possède-t-elle le trait [+ Institution] ou [+ AST] ?
 - si *oui*, la notion admet aussi t3, a1 et a3,
 - si *non*, la notion admet aussi a1.

(3) La notion admet t1 :

Le complément verbal correspond en LN à un complément d'objet direct ?

si *oui*, la notion admet aussi t2, t3, a1, a2, a3, a4 ;

si *non*, la notion admet aussi t2, t3, a1, a2, a3.

Bien que les procédés acquis par chaque notion soient entièrement déductibles par règle à partir du mot de rappel, on les notera dans les entrées lexicales de la manière suivante : les 8 cases qui suivent la première, dans la première ligne, correspondent à chacun des procédés ; on écrira dans chaque case le procédé, si celui-ci peut être acquis par la notion ; l'écriture se fera en caractères gras pour les procédés t0 et t1. La présence du symbole t1 indique que la notion admet le procédé t1 avec les trois configurations possibles de traits sémantiques (t1, + t1, - t1) ; si, dans une entrée particulière, on voulait signaler une restriction, on ne noterait que la configuration admise par la notion respective (+ t1, - t1) ; une convention analogue vaut pour le procédé a1. Dans la case correspondant à t1 on notera aussi la ou les prépositions exigées par la forme verbale, si celle-ci ne se construit pas avec le complément d'objet direct (cf. tableaux ci-dessous).

2.3 Les supports de radical normalisés

Si le mot de rappel est un nom, le support de radical que le documentaliste doit écrire lorsqu'il veut exprimer un item t0 coïncide avec le mot de rappel, aux marques de nombre et de genre près ; la même situation se présente pour les supports de radical qui doivent être utilisés pour exprimer des items a1 dont le mot de rappel est un adjectif. Dans tous les autres cas, les supports de radical utilisés pour noter une notion présentent des modifications par rapport au mot de rappel ; autrement dit, les articulations placées après le radical discriminant et avant l'indicateur de procédé ne coïncident pas avec la terminaison du mot de rappel. Exemples : *indexation t1, indexeur t3, indexé a4*, ne coïncident pas avec le mot de rappel *indexer*.

Le documentaliste, à partir de sa langue naturelle, qu'il connaît, sait quelle est l'articulation qui convient pour chaque radical discriminant : il écrira ainsi spontanément, par exemple, *documentaliste t3* et non *documenteur t3*, ou *indexeur t3* et non *indexiste t3*. Cette écriture est importante pour l'obtention des structures lisibles, sinon pour l'information de base : on sait, en effet, que pour interpréter l'information de base, il suffit à la machine de reconnaître le radical discriminant et l'indicateur de procédé (cf. fiche n° 2, § 2.2). En revanche, si le documentaliste se trompait dans le choix, par exemple, d'une forme de suffixe, la structure lisible serait très altérée.

Malgré cette liberté laissée au documentaliste, les entrées lexicales de V1, pour lui faciliter la tâche, prévoient, à titre indicatif, la mention des supports de radical normalisés : ce sont les signifiants consacrés par la norme ou l'usage. Ils seront écrits sur la deuxième ligne et dans la colonne qui correspond à chaque procédé. Pour les procédés T, on n'inclura que les supports de radical avec les terminaisons suffixales, qui peuvent changer dans chaque entrée lexicale, et quand elles ont été consacrées par la norme. Pour les cas où aucune forme consacrée par l'usage n'existe, on laissera un blanc, qui doit être interprété comme une invitation à utiliser soit l'infinitif pour le procédé t1, soit le syntagme type du procédé en question : *le fait de x*, pour le procédé t1 ; *machine à x*, pour le procédé t2 ; *personne qui x*, *personnel (spécialisé)*, pour le procédé t3, x étant, dans les trois cas, une forme verbale. Dans le cas de t3, on indiquera aussi le mot ou l'expression particulière qui peut être utilisé comme signifiant du spécificateur « institution » (par exemple *maison d'édition t3, imprimerie t3*). Pour le procédé a1, on indiquera, quand il existe, l'adjectif le plus courant (par exemple, *photographique*, pour l'entrée lexicale dont le mot de rappel est *photographier*), restant bien entendu que le documentaliste est libre soit de l'utiliser, soit d'utiliser un syntagme prépositionnel (par exemple, *photographique a1* ou, si le contexte l'exige, *de la photographie a1*). Pour le procédé a4, on indiquera le participe passé qui lui correspond. Les cases a2 et a3 resteront libres : on emploie les supports de radical de, respectivement, t2 et t3, avec les articulations nécessaires (cf. tableaux ci-dessous).

On sait que l'information de base est exprimée par le radical discriminant et par l'indicateur de procédé. Dans le cas de t3, le contenu du spécificateur de notion (« institution » ou « personne »), qui est une articulation grammaticale, est incorporé au contenu significatif des entités schématiques (cf. § 1). Pour rendre possible le traitement par la machine des informations indexées, la notation de ces informations (information de base et contenu des spécificateurs de notion) est entièrement contraignante : le documentaliste ne peut modifier ni le radical discriminant ni les indicateurs de procédé. Dans le cas du spécificateur de notion de t3, s'il souhaite exprimer « institution », son choix doit porter obligatoirement sur une des trois formes générales prévues (cf. ci-dessus) ; ou bien, sur la forme ou les formes particulières incorporées dans chaque entrée lexicale ; si, dans un item t3, la machine ne retrouvait aucune de ces formes, elle l'interpréterait comme possédant le spécificateur de notion « personne ». Celui-ci est donc noté de manière stricte par le fait négatif de l'absence du spécificateur « institution ». C'est pour cette raison que le signifiant particulier qui doit être introduit par le documentaliste pour exprimer le spécificateur de notion « personne » dans t3 peut être assimilé à un support de radical.

Les supports de radical ne sont pas contraignants. Est-ce qu'on doit dire *photocopieur* ou *photocopieuse* ? Dans un cas analogue, le lexique de V1 peut soit prévoir les deux signifiants comme alternative possible, soit un seul. Dans tous les cas, le documentaliste est habilité à introduire le signifiant de son choix, prévu ou non dans l'entrée lexicale, à condition de respecter le radical discriminant. Si, par distraction, devant un indicateur de procédé, il écrivait un mot LN sans radical discriminant, il serait averti par la machine. Dans un stade ultérieur du système V1, on peut prévoir que, dans ce même cadre d'assistance de l'ordinateur à l'indexation (cf. fiche n° 3), la machine avertisse le documentaliste si celui-ci altérerait un support de radical prévu par le lexique ou en choisissait un autre.

Les supports de radical introduits dans les entrées lexicales doivent être distingués des variations des radicaux discriminants, qui sont introduites dans les mêmes cases que les supports de radical et dans la même structure LN. On se souviendra, en effet, que la notion de « famille restreinte de mots » admettait des expressions LN avec des signifiants entièrement différents, pourvu qu'ils soient associés à une invariance de contenu. Par exemple, *machine à écrire* est dans une famille restreinte de mots avec *dactylographier* et non avec *écrire*. Or, dans tous les cas où en langue naturelle, dans une famille restreinte de mots, un contenu invariant est associé à plusieurs signifiants différents, *le radical discriminant présentera des variations*. Ces formes particulières qu'il peut acquérir sont notées en lettres capitales dans la colonne correspondant à chaque procédé et elles ont le même caractère contraignant que le radical discriminant noté dans la première case de la première colonne. Lorsque le radical discriminant noté dans cette première case n'est pas invariant (c'est-à-dire dans les cas où, dans une entrée lexicale, existent des variations dans le radical discriminant), ce fait est indiqué dans la première case, en notant, suivi d'un astérisque, l'indicateur de procédé dans la colonne duquel apparaît la variation du radical discriminant.

A partir de ce qui vient d'être dit, on peut introduire le deuxième type d'entrées du lexique : ce sont les *renvois*, exprimés en structures LN. Un renvoi est constitué de : (a) une expression LN qui possède le signifié invariant d'une famille restreinte de mots associé à un signifiant qui contient une variation du radical discriminant ; (b) l'indication du mot de rappel qui identifie l'entrée lexicale où on peut trouver le moyen d'exprimer le signifié de l'expression LN de (a). On utilisera les deux points (« : ») entre a et b. Exemples :

écrire à la machine : dactylographier.
machine à écrire : dactylographier.

2.4 Le système de traits

Il a été remarqué ci-dessus (cf. § 2.1) que le lexique de V1 ne comporte pas une description formalisée du contenu de chaque notion : c'est le mot de rappel et, éventuellement, les précisions sur le contenu de la notion qui permettent de délimiter ce contenu par le

biais des structures LN correspondantes. Or, en parallèle avec cette manière de procéder, le lexique de V1 prévoit un vocabulaire général de traits sémantiques, à partir duquel on peut décrire partiellement, et en fonction d'objectifs précis, le contenu de chaque entrée lexicale. Les traits sémantiques possédés par chaque notion sont notés dans l'entrée lexicale correspondante et, à la différence du mot de rappel et des précisions sur le contenu, la machine peut opérer sur eux. Par exemple, les notions dont le mot de rappel est *affiche*, *estampe*, *encyclopédie*, *nomenclature*, *livre*, *thèse*, etc. possèdent toutes le trait sémantique [+ document].

Deux principes de base règlent l'incorporation des traits sémantiques : (1) les traits sémantiques incorporés à chaque entrée lexicale sont déductibles du contenu de la notion ; cela implique que, en connaissant le vocabulaire général de traits sémantiques et le contenu de chaque notion particulière, un documentaliste peut prévoir les traits sémantiques assignés à chaque notion ; (2) en principe, les traits sémantiques sont incorporés dans chaque entrée lexicale, et le documentaliste n'a pas à les noter. On déduit de ceci que si le documentaliste doit connaître la raison d'être et les implications, pour le système d'ensemble, des traits sémantiques, leur maniement ne doit pas constituer pour lui une charge supplémentaire, car, en règle générale, il ne doit pas s'en occuper.

La fonction des traits sémantiques est de rendre possible le fonctionnement du réseau sémantique et, par ce biais, de permettre les généralisations et les spécifications de questions dans la recherche documentaire. Plus particulièrement, les traits sémantiques vont être utilisés notamment pour : (1) mettre en relation les notions entre elles ; (2) mettre en relation les notions avec les traits sémantiques introduits par les indicateurs de procédé ; (3) établir des relations entre structures grammaticales différentes, notamment d'équivalence ; (4) exprimer les contraintes de sélection.

(1) Relation entre notions

Chaque entrée lexicale est constituée d'une notion. Or, il est évident qu'entre *film* et *filmer*, il existe une relation qui est absente entre *agriculture* et *télétype*. Les traits sémantiques sont utilisés pour exprimer les relations qui intéressent la recherche documentaire. Sans prétendre ici à l'exhaustivité, plusieurs cas sont à prévoir :

(a) Un même contenu dans une famille de mots au sens large, en langue naturelle, apparaît en deux entrées lexicales différentes, en vertu des contraintes imposées par la notion de « famille restreinte de mots » et/ou par les principes qui déterminent l'inclusion des entrées lexicales. Par exemple, *norme* et *normaliser*, *film* et *filmer* (cf. 2.1). Dans chaque paire, les mots cités appartiennent à des entrées lexicales différentes ; un trait sémantique commun doit permettre de les mettre en relation.

(b) Il existe des emboîtements entre les contenus de notions, par exemple entre *document*, *ouvrage*, *livre* et *opuscule* ; les traits sémantiques doivent les spécifier.

(c) Des notions, sans correspondre à une famille de mots au sens large en langue naturelle, possèdent des éléments de contenu identique, qui doivent être exprimés par des traits sémantiques. Exemple : *bibliothèque*, *discothèque*, *photothèque*.

(2) Relation entre notions et traits sémantiques des procédés

Les traits sémantiques introduits par les procédés (notamment t2 et t3) et/ou par les spécificateurs de notion (« institution » ou « personne ») peuvent se retrouver partiellement dans des notions du lexique. Le lexique doit en effet incorporer une notion *machine*, avec un trait [+ machine] qui réapparaît, sous une autre forme, dans « machine qui sert à », de t2 ; de même, le lexique incorpore *personne*, *agent*, *organisme* et d'autres, qui réapparaissent, avec des modifications, dans le procédé t3 et/ou dans ses spécificateurs de notion.

(3) Relations, notamment d'équivalence, entre structures grammaticales

Le système V1, pour satisfaire la caractéristique de naturalité des structures grammaticales (c'est-à-dire leur « ressemblance » avec la langue naturelle), permet d'exprimer un même contenu, ou quasiment un même contenu, par des structures grammaticales différentes, entre lesquelles il faut, pour éviter le silence documentaire, établir des relations d'équivalence. Or, ces relations d'équivalence peuvent être formulées en tenant compte, parmi d'autres facteurs, des traits sémantiques des notions.

Exemple (cf. aussi § 8.2) : (1) *Bibliothèque t0 en médecine a1* est équivalent à (1') *Bibliothèque t0 spécialisée en s médecine a1* ou à (1'') *Bibliothèque t0 spécialisée a1 b en médecine a1*. Ce sont trois structures grammaticales que le documentaliste peut produire pour exprimer un même contenu. En revanche, (2) *Bibliothèque t0 avec microfiches a1* n'est pas équivalent à (2') *Bibliothèque t0 spécialisée en s microfiches a1* ou à (2'') *Bibliothèque t0 spécialisée a1 b en microfiches a1*. Comme « en » dans (1) et « avec » dans (2) appartiennent aux articulations, dont la machine ne tient pas compte, pour exprimer l'équivalence du premier cas, et non du deuxième, la règle correspondante – qui doit pouvoir être introduite en machine – doit opérer sur le trait sémantique [+ AST] (AST = art, science ou technique) de *médecine*, qui est absent de *microfiche*.

(4) Contraintes de sélection

En termes généraux, on a une contrainte de sélection entre deux items dont l'un est un complément A ou un complément verbal de l'autre, si chaque item ou les deux doivent posséder un trait sémantique particulier. Par exemple, la notion *abonner*, dans le lexique de V1, est utilisée pour exprimer des structures LN telles que *abonnement à une revue*, *à une base de données* et non pas *abonnement à l'électricité*. L'item T, complément verbal de *abonner à*, doit donc porter le trait sémantique [+ UD] (UD = unité documentaire ; on désigne par là tout ce qui concerne le domaine des sciences de la documentation et de l'information en tant que telles).

Les contraintes de sélection sont soit incorporées dans les entrées lexicales sous la forme de *traits de sélection* (c'est le cas de l'exemple précédent dans l'entrée de *abonner*), soit incorporées dans la formulation de règles concernant les spécificateurs A (cf. 5.1.2.1.2).

On n'indiquera pas de manière exhaustive dans cette fiche le système de notation des traits sémantiques et des traits de sélection dans le lexique de V1. On remarquera, cependant, que l'organisation graphique des entrées lexicales est prévue pour incorporer sans trop de difficultés ces informations. La première colonne, à gauche, accueille les informations propres à la notion, valables pour tout procédé, sauf indication contraire notée au moyen de l'indicateur de procédé suivi d'un astérisque. Chacune des autres colonnes est destinée à un procédé différent. On a présenté ci-dessus (cf. 2.2 et 2.3) les informations contenues dans les deux premières lignes de cases, concernant, respectivement, les procédés admis, d'une part, et les supports de radical normalisés, avec les variations du radical discriminant, d'autre part. La troisième ligne de cases est destinée aux traits sémantiques : dans la première colonne, s'ils concernent la notion dans toutes ses manifestations, et dans chacune des autres colonnes, s'ils concernent les procédés respectifs. La quatrième ligne est destinée aux traits de sélection qui, eux, ne peuvent apparaître que dans les colonnes 2 à 9 ; la cinquième ligne enfin est destinée aux *traits particuliers*.

Par *trait particulier* on entend, sur un plan général, toute entorse aux principes généraux de fonctionnement du lexique qu'il est nécessaire d'indiquer de manière *ad hoc* dans chaque entrée lexicale. Un trait particulier est un échec d'une généralisation. Soit, comme exemple, le problème posé par *bande dessinée*, descripteur incorporé au *Thésaurus Sciences de l'information*, sur lequel opère le système V1. Le même *Thésaurus* incorpore aussi *bande magnétique* et *bande perforée*. Il est clair que *bande*, dans *bande dessinée*, n'est pas la même notion que dans *bande magnétique* ou *bande perforée*. Dans ces deux derniers cas, *bande* est paraphrasable par « ruban destiné à l'enregistrement de l'information ou servant de support à de l'information enregistrée ». Cette notion de « bande » apparaît, par ailleurs, dans *bande cinématographique*, *bande de magnétophone*, *bande d'ordinateur*, *bande vidéo*. Ces expressions, absentes du *Thésaurus*, peuvent être facilement formées par V1, grâce à ses relations syntaxiques. Dans *bande dessinée*, en revanche, *bande* n'a pas le signifié de « ruban... » et, par conséquent, *bande dessinée* ne correspond pas à *bande t0 dessinée a4*. Il est donc nécessaire d'ouvrir deux entrées lexicales, l'une pour *bande*, l'autre pour *bande dessinée*. Il faut aussi introduire un trait particulier dans *bande*, celui qui doit spécifier l'interdiction d'écrire *bande t0 dessinée a4* comme structure de V1 qui corresponde à *bande dessinée*.

t0 t1 t2 t3 a1 a2 a3 a4

MOT de rappel [avec le radical discriminant en capitales]	←							→						
	procédés admis													
Indications sur les variations de radical discriminant	Prépos. compl. verb.													
	← VARIATIONS DES RADICAUX DISCRIMINANTS →													
Précisions sur le contenu	←							→						
	supports de radical normalisés													
	←							→						
	traits sémantiques													
	←							→						
	traits de sélection													
	←							→						
	traits particuliers													

Schéma d'une entrée lexicale

ACCEDer		t1	t2	t3	a1	a2	a3	
- à une salle ; - à la documentation - accès de fièvre *		accès						
BIBLIOTHEQue	t0			t3	a1		a3	
				bibliothécaire				
CHIMie	t0			t3	a1		a3	
				chimiste	chimique			
DACTylographier * t2		t1	t2	t3	a1	a2	a3	a4
		dactylographie	machine à ECR	dactylographe	dactylographique			dactylographié
EDIter		t1	t2	t3	a1	a2	a3	a4
		édition		éditeur maison d'édition				édité

GRATuit	t0				a1				
non payant; injustifié *	gratuité								
IMPRIMER * t1	t1	t2	t3		a1	a2	a3	a4	
	IMPRESSiON		imprimeur imprimerie					imprimé	
INDEXER * t1	t1	t2	t3		a1	a2	a3	a4	
	INDEXATION		indexeur					indexé	
PHOTOCOMposer	t1	t2	t3		a1	a2	a3	a4	
	photocom- position	photocom- poseuse						photocomposé	
TRADuire	t1	t2	t3		a1	a2	a3	a4	
	traduction		traducteur					traduit	

Exemples d'entrées lexicales (lignes 1 et 2)

3 LES STRUCTURES GRAMMATICALES ; CONTRAINTES COMBINATOIRES ET RELATIONS INTERNES

Les structures grammaticales suivantes :

(1) Gestion t1 de l'indexation t1 automatique a1 des documents t0 techniques a1 traduits a4.
(2) Cette gestion t1, en relation avec r le financement t1 d'un centre de documentation t3.
sont, respectivement, une phrase documentaire (PD) et une construction documentaire (CD). Toutes deux sont des structures complexes, constituées de structures plus simples : une construction documentaire est constituée de deux phrases documentaires mises en rapport par un prédicat ; la phrase documentaire est constituée de *syntagmes de phrase*. La structure grammaticale (1) ci-dessus est une phrase documentaire constituée de trois syntagmes de phrase :

gestion t1
indexation t1 automatique a1
documents t0 techniques a1 traduits a4

On exposera par la suite les principes qui règlent la combinatoire interne du syntagme de phrase et de la phrase documentaire et les relations que cette combinatoire détermine, pour terminer (§ 3.6) avec la construction documentaire. On désignera désormais le syntagme de phrase par l'expression abrégée *syntagme*, mais on aura recours à l'expression complète — *syntagme de phrase* — lorsqu'on voudra insister sur le fait qu'il s'agit bien de ce type de syntagme et non du syntagme d'attributs, qu'on introduit ci-dessous au § 3.3.1.

3.1 Le syntagme ; le syntagme à noyau simple

Un syntagme est constitué d'un *noyau* et, optionnellement, d'un *groupe d'attributs*, désormais *groupe A*. Le noyau est soit *simple* soit avec *coordination interne*. Le premier type étant de loin le plus largement utilisé, on n'introduira le deuxième qu'à la fin de ce § 3 (cf. 3.4) et en fonction de ce qui aura été dit sur le syntagme simple.

Un *noyau simple* est constitué d'un seul item avec, le cas échéant, ses articulations ; cet item doit être un T (t0 ou t1 ou t2 ou t3). Les formules des syntagmes à noyau simple sont donc les suivantes :

[x] T
[x] T groupe A

On dira, en première approximation, que le groupe A est constitué de l'item A ou des items A avec, le cas échéant, leurs articulations, qui, dans un syntagme, suit ou suivent le noyau, que celui-ci soit simple ou à coordination interne. Les trois syntagmes de l'exemple précédent sont donc constitués de la manière suivante (on dira du premier des syntagmes qui suivent que son noyau est *isolé*, c'est-à-dire non suivi du groupe A) :

gestion t1
noyau
indexation t1 automatique a1
noyau groupe A
documents t0 techniques a1 traduits a4
noyau groupe A

Dans le reste de ce § 3, on ne parlera que des *items*, sans introduire l'explicitation « avec, le cas échéant, ses articulations », qui, sauf avis contraire, doit être sous-entendue.

3.2 La combinatoire entre syntagmes ; la phrase documentaire

Une *phrase documentaire* est constituée d'un ou plusieurs *syntagmes de phrase*. La contrainte qui règle la combinatoire interne des phrases documentaires est la suivante : un syntagme de phrase dont le noyau est constitué d'un item TV est susceptible d'être suivi d'un autre syntagme de phrase. Il faut remarquer qu'il n'y a aucune limite numérique à l'adjonction d'un syntagme à la suite qui le précède, composée elle-même d'un ou plusieurs syntagmes.

Exemples : les expressions suivantes sont des phrases documentaires (remarquer que les deux premières sont aussi des syntagmes : ce sont des phrases documentaires constituées d'un seul syntagme) :

- (1) Bibliothèque t0.
- (2) Bibliothèque t0 universitaire a1 au Mexique a1.
- (3) Traitement t1 automatique a1 de l'indexation t1.
- (4) Gestion t1 par un centre de documentation a3 du stockage t1 des documents t0.
- (5) Gestion t1 automatisée a4 par un centre de documentation a3 de la reproduction t1 par microfilm a1 des documents t0 anciens a1.
- (6) Financement t1 du traitement t1 automatique a1 de l'indexation t1.

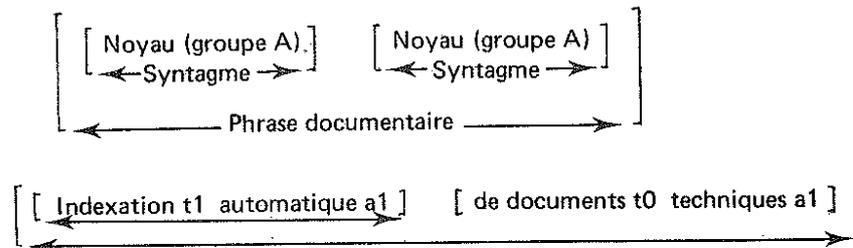
En revanche, les expressions suivantes ne sont pas des phrases documentaires. Dans tous les cas, on n'y a pas respecté les contraintes sur la combinatoire des syntagmes : dans les deux premiers cas, parce qu'un complément attributif, à lui seul, ne constitue pas un syntagme et, à fortiori, ne constitue pas une phrase documentaire ; dans les deux cas suivants, parce qu'un syntagme en suit un autre dont le noyau est un item qui n'est pas un TV :

- (1) Automatique a1 par microfilm a1 *.
- (2) Universitaire a1 en France a1 *.
- (3) Documents t0 anciens a1 reprographie t1 *.
- (4) Discothèque t0 conservation t1 des documents t0 *.

3.2.1 Les relations internes de la phrase documentaire constituée de plus d'un syntagme

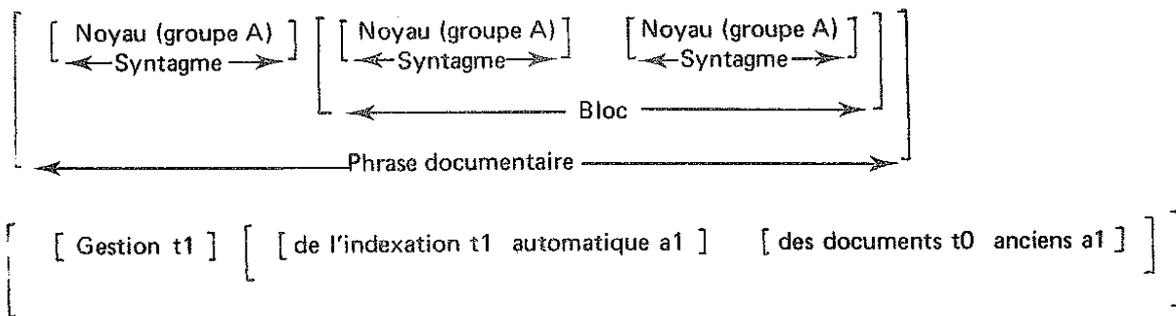
Une phrase documentaire, si elle est constituée d'un seul syntagme, aura les relations internes propres au syntagme, lesquelles seront abordées au cours de la présentation du groupe A. Si la phrase documentaire est constituée de deux ou de plus de deux syntagmes, les relations entre ceux-ci seront celles de structures à emboîtements hiérarchiques (ces emboîtements sont indiqués dans les exemples qui suivent par des parenthésisations fermées par le bas. Cette notation n'a d'autre objectif que pédagogique : elle n'apparaît pas dans les informations indexées, la machine pouvant retrouver automatiquement les emboîtements ainsi notés).

Supposons que les syntagmes qui composent une phrase documentaire sont au nombre de deux ; on aura la structure emboîtée suivante (une formule générale sera suivie d'une expression particulière) :



Les items à l'intérieur de chaque syntagme entretiennent entre eux des relations significatives. On doit en effet comprendre que c'est l'indexation qui est automatique et que ce sont les documents techniques qu'on indexe automatiquement. Les relations ne s'organisent donc pas entre les items constitutifs d'un syntagme, qui seraient mis isolément en rapport avec les éléments constitutifs de l'autre syntagme, mais entre un syntagme en tant que tel et l'autre syntagme (on abordera ci-dessous au § 3.3 les relations à l'intérieur du syntagme de phrase). Le syntagme de droite et celui qui le précède s'emboîtent ainsi dans la structure plus large qu'est la phrase documentaire. On dira que le syntagme de droite *porte* sur le syntagme qui le précède. On dira par ailleurs que le noyau du syntagme de droite est *complément verbal* du noyau du syntagme précédent.

Si la phrase documentaire est constituée de trois syntagmes, on aura la structure emboîtée suivante :



On remarquera que le syntagme le plus à droite porte sur le syntagme qui le précède. Ensemble, ils constituent un *bloc*, qui porte sur le syntagme précédent, au même sens que le syntagme de droite porte sur le syntagme qui précède; en effet, *indexation t1 automatique a1 des documents t0 anciens a1* est mis en rapport *comme un tout* avec *gestion t1*. Par ailleurs, le noyau de chaque syntagme, à l'exception du syntagme le plus à gauche de la phrase documentaire, est le complément verbal du noyau du syntagme qui le précède immédiatement : *documents t0* est le complément verbal de *indexation t1* et *indexation t1* est le complément verbal de *gestion t1*.

On sait que la contrainte qui règle la combinatoire interne de la phrase documentaire n'établit aucune limite au nombre de syntagmes d'une phrase documentaire. Or, quel que soit ce nombre, pourvu qu'il soit égal ou supérieur à deux, on retrouvera les mêmes relations internes : (a) le dernier syntagme à droite de la phrase documentaire et tout bloc de syntagmes portent sur le syntagme immédiatement précédent; (b) le noyau d'un syntagme est le complément verbal du noyau du syntagme précédent.

Dans tous les cas, les syntagmes s'emboîtent entre eux pour former les blocs de syntagmes et la phrase documentaire. Un bloc de syntagmes est constitué soit de deux syntagmes, soit d'un syntagme et d'un autre bloc, interne, de syntagmes. Tout bloc de syntagmes incorpore le dernier syntagme à droite de la phrase documentaire. Le bloc de syntagmes constitué par le dernier syntagme à droite de la phrase documentaire et par le syntagme qui le précède est le bloc d'ordre 1. Le bloc d'ordre 1 qui, comme tout bloc de syntagmes, porte sur le syntagme qui le précède, constitue avec celui-ci le bloc d'ordre 2; celui-ci, avec le syntagme qui le précède (et sur lequel il porte) formera le bloc d'ordre 3, et ainsi successivement, jusqu'à la formation du bloc d'ordre *n*. Celui-ci comprendra tous les syntagmes de la phrase documentaire à l'exception d'un seul, le premier syntagme à gauche, sur lequel il porte. On notera, dans ce contexte, que la phrase documentaire présente la structure emboîtée des blocs de syntagmes mais, à la différence de ceux-ci, elle ne porte sur aucun syntagme.

On aura les possibilités d'emboîtements suivantes dans les phrases documentaires de deux ou plus de deux syntagmes (le symbole « ... » sera utilisé pour indiquer un nombre indéfini d'expressions; pour améliorer la lisibilité, on a développé en langage courant ce qu'expriment les formules entre crochets) :

$$(1) \left[\left[\text{Syntagme} \right] \left[\text{Syntagme} \right] \right]$$

← Phrase documentaire →

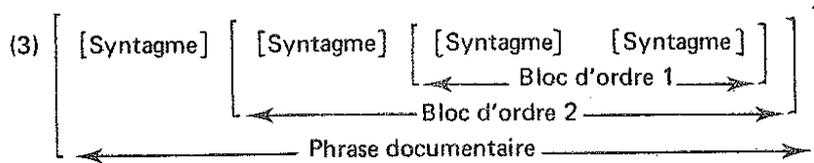
Phrase documentaire constituée de deux syntagmes.

$$(2) \left[\left[\text{Syntagme} \right] \left[\left[\text{Syntagme} \right] \left[\text{Syntagme} \right] \right] \right]$$

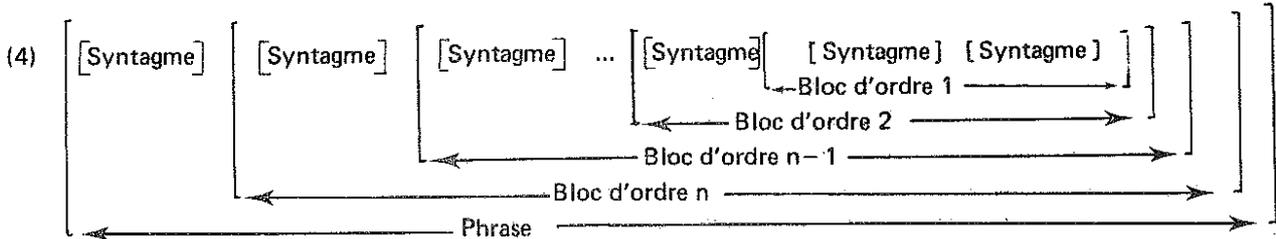
← Phrase documentaire →

← Bloc d'ordre 1 →

Phrase documentaire constituée d'un syntagme et d'un bloc d'ordre 1; le bloc d'ordre 1 est constitué de deux syntagmes.

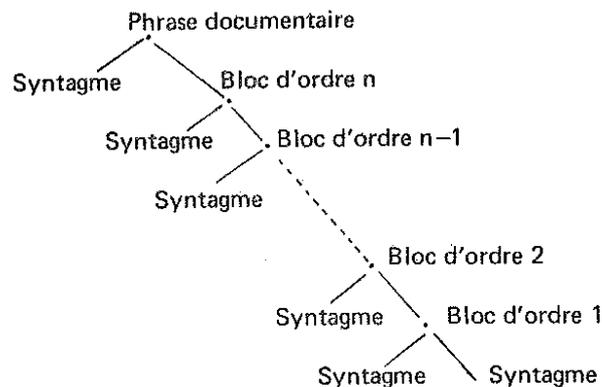


Phrase documentaire constituée d'un syntagme et d'un bloc d'ordre 2; le bloc d'ordre 2 est constitué d'un syntagme et d'un bloc d'ordre 1.



Phrase documentaire constituée d'un syntagme et d'un bloc d'ordre n; le bloc d'ordre n est constitué d'un syntagme et d'un bloc d'ordre n-1. Il existe un nombre quelconque de blocs intermédiaires entre le bloc d'ordre n-1 et le bloc d'ordre 2.

Les emboîtements dans la phrase documentaire peuvent aussi être représentés, et de manière équivalente, par la structure arborescente suivante (seul le cas (4) est ici représenté sous forme d'arbre) :



3.3 Le groupe d'attributs dans le syntagme de phrase

Au § 3.1 on a caractérisé le groupe A comme étant constitué de tous les items A qui suivent le noyau dans un syntagme. Cette caractérisation reste valable avec des modifications : le groupe A possède des possibilités d'organisation interne qu'on présentera par la suite. Ces possibilités reposent sur l'utilisation du symbole « b » et sur la spécification de la relation d'attribut.

3.3.1 Le marqueur de relations *b*

Le symbole « b » note un marqueur de relations : sa fonction est d'indiquer des relations internes au groupe A. Un groupe A est constitué de tous les items A qui suivent le noyau. Si le symbole « b » n'apparaît pas parmi eux, chacun des items A porte sur le noyau du syntagme; on dira aussi que chacun des items A est complément d'attribut (désormais, complément A) du noyau. C'est la situation attestée dans l'exemple qui suit, et dans tous ceux qui ont précédé dans ce § 3 :

Indexation t1 automatique a1 par un centre de documentation a3 .

Il est à remarquer que *ce n'est pas* le groupe A en tant que tel qui porte sur le noyau.

En effet, *automatique a1* et *par un centre de documentation a3* ne se combinent nullement entre eux pour former un contenu composite, qui serait à mettre en rapport avec *indexation t1* ; c'est bien *automatique a1* et *par un centre de documentation a3* qui portent individuellement sur le noyau.

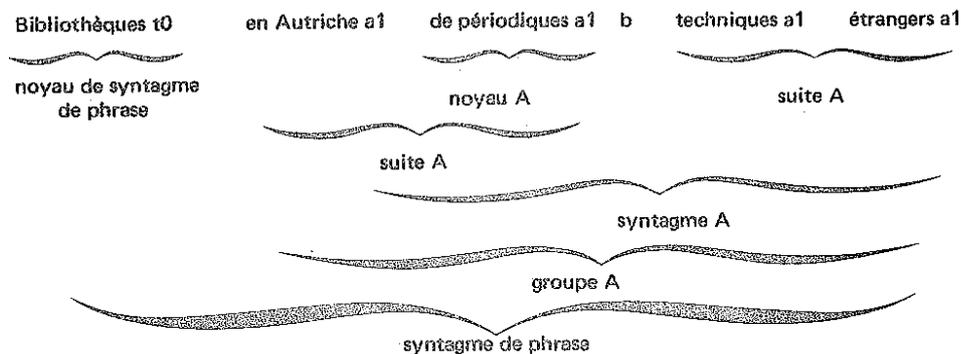
En revanche, si l'on souhaite organiser autrement les rapports entre les items A et le noyau, on peut utiliser *b* : il sera placé à la suite d'un item A et il indiquera que chaque item A qui suit (sans être précédé lui-même d'un autre *b*) porte sur l'item A qui précède immédiatement *b*. On dira aussi que chacun des items A, qui porte sur un item A est complément A de celui-ci. Par exemple, dans :

Bibliothèques *t0* en Autriche *a1* de périodiques *a1* *b* techniques *a1* étrangers *a1*.

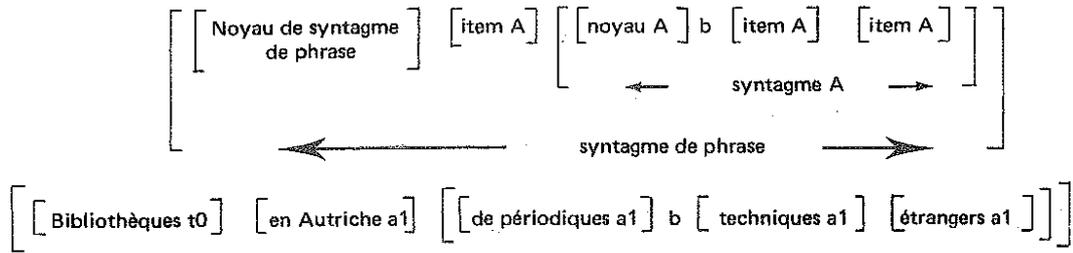
techniques a1 porte sur *périodiques a1* et en est complément A, de même que *étrangers a1*, mais aucun d'eux ne porte sur *bibliothèques t0* : ce sont les périodiques qui sont techniques et étrangers, et non les bibliothèques.

La contrainte combinatoire qui règle l'utilisation du marqueur *b* est la suivante : *b* doit être utilisé à la suite immédiate d'au moins un item A et doit précéder au moins un item A.

L'emploi du marqueur *b* entraîne la formation d'un syntagme d'attribut (désormais, syntagme A), constitué d'un noyau de syntagme d'attribut (désormais, noyau A), du marqueur *b* et d'une suite d'items A (désormais, suite A). Le noyau A est constitué par l'item A qui précède immédiatement le marqueur *b* ; la suite A est constituée de tous les items A qui, sans être précédés par un autre symbole « *b* », suivent le noyau A. On dira aussi que les items A qui, sans être précédés par un marqueur *b*, suivent le noyau de syntagme de phrase, constituent une suite A. Chacun des items A d'une suite A est complément A du noyau A ou du noyau de syntagme de phrase qui précède la suite A. Dans l'exemple qui suit, *en Autriche a1* est complément A de *bibliothèques t0*, de même que *périodiques a1* ; *techniques a1* est complément A de *périodiques a1*, de même que *étrangers a1* :

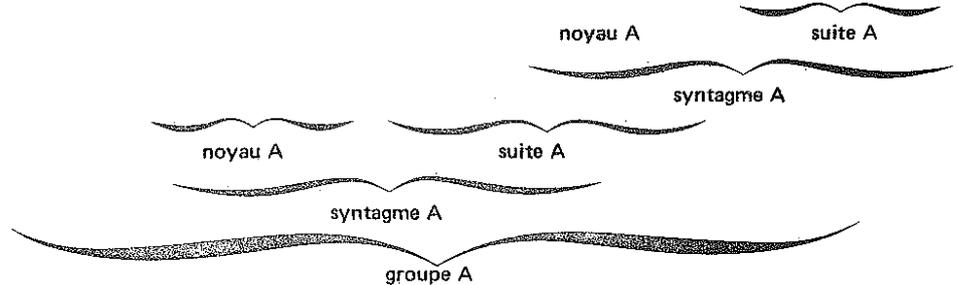


On remarquera que si *techniques a1* et *étrangers a1* portent individuellement sur *périodiques a1*, c'est le syntagme *périodiques a1 b techniques a1 étrangers a1* qui porte sur *bibliothèques t0*, au même titre que *en Autriche a1*. A l'instar du syntagme de phrase, mais à la différence de la suite A et du groupe A, le dernier syntagme A à droite d'un groupe A porte en tant que tel sur une autre entité (dans ce cas, le noyau du syntagme de phrase) : les items qui composent ce syntagme A se combinent en effet entre eux pour construire un contenu composite, lequel sera mis en rapport avec le contenu de l'entité sur laquelle porte le syntagme A. Il est clair que l'on retrouve dans le cadre du syntagme de phrase le principe d'organisation en structures emboîtées déjà rencontré dans la phrase ; l'exemple précédent sera donc représenté ainsi :



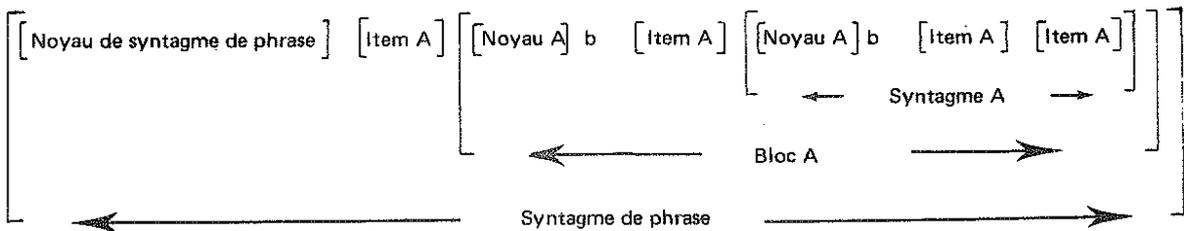
Aucune restriction sur le nombre des *b* dans un groupe A n'a été introduite dans la contrainte combinatoire qui règle son utilisation. Il est donc possible d'en utiliser plusieurs, « en cascade » dans un même groupe A, chaque *b* étant suivi et précédé d'un ou plusieurs items A. L'introduction de chaque *b* entraînera la constitution d'un nouveau syntagme A, avec, comme noyau A, l'item A qui précède *b*, et, comme suite A, le ou les items A qui, sans être précédés d'un autre *b*, suivent *b*; chacun des items A de la suite A sera complément A du noyau qui précède. Dans l'exemple qui suit, *publique a1* est complément A de *université a1*, de même que *Inde a1*; *centrale a1* est complément A de *bibliothèque a1*, de même que *université a1*; *automatique a1* est complément A de *centre de documentation t3*, de même que *bibliothèque a1* :

Centre de documentation t3 automatique a1 d'une bibliothèque a1 b centrale a1 d'une université a1 b publique a1 en Inde a1

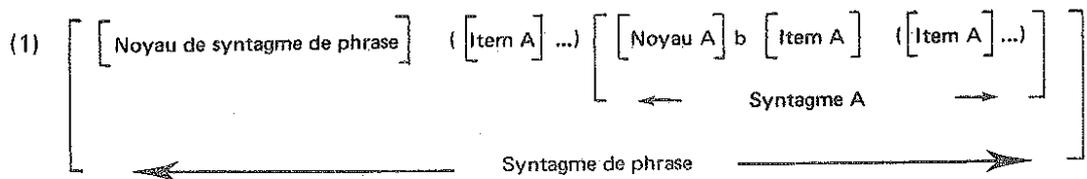


Par ailleurs, l'adjonction d'un *b* à un groupe A qui en possède déjà au moins un provoque la formation d'un bloc de syntagmes A, ou bloc A, à l'intérieur duquel certaines entités portent sur d'autres entités, dans les conditions précisées ci-dessous.

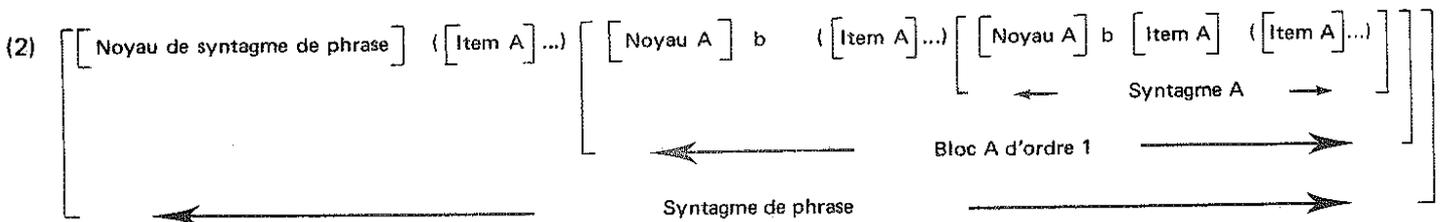
Le dernier syntagme A, à droite du groupe A, soit *université a1 b publique a1 en Inde a1* porte sur le noyau A qui précède, soit *bibliothèque a1*. Sur ce même noyau A porte l'item *centrale a1*. Il faut remarquer que le dernier syntagme A, à droite, porte sur le noyau A qui précède et non sur le syntagme A (à la différence du dernier syntagme de phrase qui porte, lui, sur le syntagme de phrase qui précède). Le noyau A *bibliothèque a1*, le *b* qui le suit, l'item *centrale a1* qui porte sur lui et le syntagme A *université A1 b publique a1 en Inde a1*, qui porte également sur *bibliothèque a1*, constituent un bloc A; ce bloc A (*bibliothèque a1 b centrale a1 d'une université a1 b publique a1 en Inde a1*) porte sur le noyau du syntagme de phrase – *centre de documentation t3* – de même que *automatique a1*. Le syntagme de phrase de l'exemple antérieur présente l'organisation suivante des structures emboîtées :



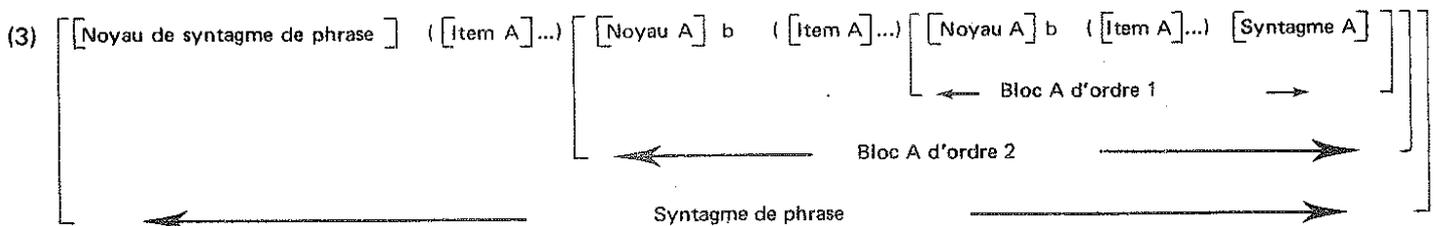
Sur un plan général, on trouvera dans tout groupe A les mêmes relations internes : (a) le dernier syntagme A, à droite du syntagme de phrase, et tout bloc A porte sur le noyau qui le précède immédiatement, que ce noyau soit noyau A ou noyau de syntagme de phrase; (b) chacun des items A d'une suite A qui n'est pas un noyau A porte sur le noyau qui le précède immédiatement, que ce noyau soit noyau A ou noyau de syntagme de phrase; (c) chacun des items A d'une suite A est complément A du noyau qui précède la suite A, que ce noyau soit noyau A ou noyau de syntagme de phrase. Par ailleurs, les blocs A s'emboîtent en structures de plus en plus larges. Tout bloc A incorpore le dernier syntagme A, à droite du syntagme de phrase. Le bloc A constitué par le dernier syntagme A, à droite du syntagme de phrase, par le noyau A du syntagme A précédent et par les items de la suite A qui portent individuellement sur celui-ci, est le bloc A d'ordre 1. Si le noyau sur lequel porte le bloc A d'ordre 1 est un noyau A, le bloc A d'ordre 1 va constituer avec celui-ci et avec les autres items A, s'il y en a, qui portent individuellement sur ce même noyau A, le bloc A d'ordre 2. Le bloc A d'ordre 2 formera avec le noyau A qui le précède et sur lequel il porte, et avec les items A, s'il y en a, qui portent individuellement sur ce noyau A, le bloc A d'ordre 3. Ainsi, on obtiendra successivement des blocs A de plus en plus larges, jusqu'à la formation du bloc A d'ordre n, le bloc A le plus large : celui-ci portera sur le noyau du syntagme de phrase et il constituera, avec les autres items A, s'il y en a, qui portent sur ce noyau de syntagme de phrase, le groupe A. On aura les possibilités d'emboîtements suivantes dans les syntagmes de phrase avec un ou plus d'un marqueur *b* :



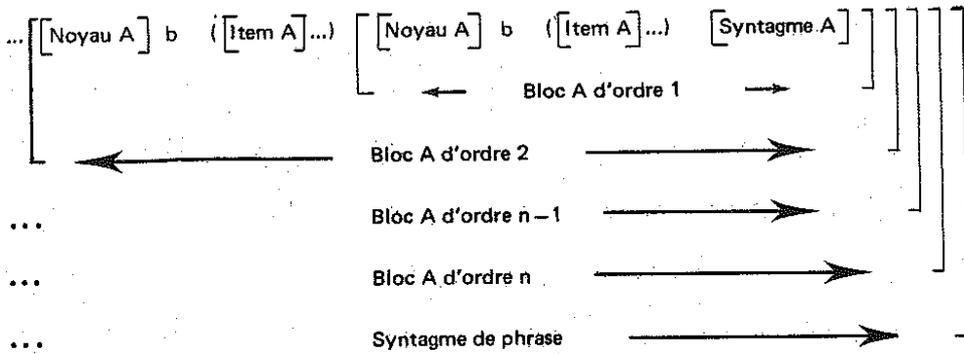
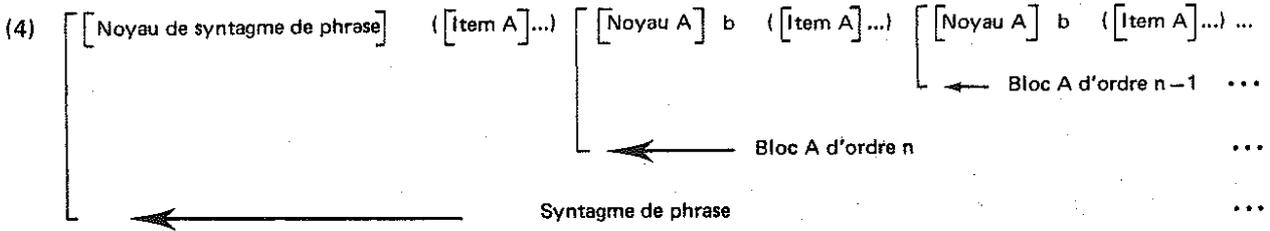
Syntagme de phrase constitué d'un noyau de syntagme de phrase, d'un item A ou, optionnellement, de plusieurs, qui portent sur lui, et d'un syntagme A, qui porte également sur lui. Le syntagme A est constitué d'un noyau A, du marqueur *b* et d'un ou, optionnellement plusieurs items A, qui portent sur ce noyau A.



Syntagme de phrase constitué d'un noyau de syntagme de phrase, d'un item A ou, optionnellement, de plusieurs, qui portent sur lui, et d'un bloc A d'ordre 1, qui porte également sur lui. Le bloc A d'ordre 1 est constitué d'un noyau A, du marqueur *b*, d'un item A ou, optionnellement, de plusieurs et d'un syntagme A, qui porte également sur lui; sur le syntagme A, cf. (1).

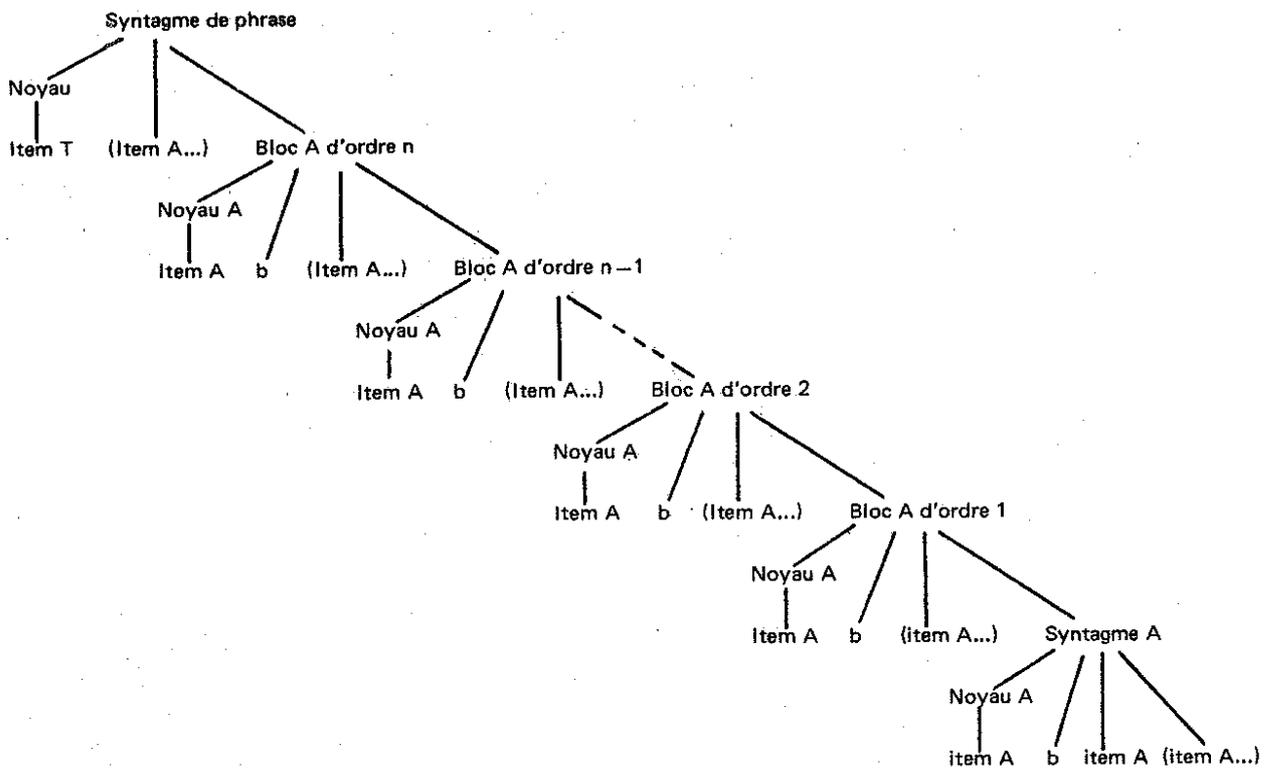


Syntagme de phrase constitué d'un noyau de syntagme de phrase, d'un item A ou, optionnellement, de plusieurs, qui portent sur lui, et d'un bloc A d'ordre 2, qui porte également sur lui. Le bloc A d'ordre 2 est constitué d'un noyau A, du marqueur *b*, d'un item A ou, optionnellement, de plusieurs et d'un bloc A d'ordre 1, qui porte également sur lui; sur le bloc A d'ordre 1, cf. (2).



Syntagme de phrase constitué d'un noyau de syntagme de phrase, d'un item A ou, optionnellement, de plusieurs, qui portent sur lui, et d'un bloc A d'ordre n, qui porte également sur lui. Le bloc A d'ordre n est constitué d'un noyau A, du marqueur b, d'un item A ou, optionnellement, de plusieurs, et d'un bloc A d'ordre n-1, qui porte également sur lui. Il existe un nombre quelconque de blocs intermédiaires entre le bloc A d'ordre n-1 et le bloc A d'ordre 2; sur ce bloc d'ordre 2, cf. (3).

Les emboîtements dans le syntagme de phrase peuvent aussi être représentés, et de manière équivalente, par la structure arborescente suivante (seul le cas (4) est ici représenté sous forme d'arbre) :



3.3.2 La spécification de la relation d'attribut

La syntaxe de V1 ne spécifie que très faiblement le contenu du rapport établi entre chaque item A, syntagme A ou bloc A, et le noyau sur lequel il porte, que ce noyau soit noyau A ou de syntagme de phrase (cf. ci-dessous § 5.1.2.1). C'est ainsi que dans les deux expressions suivantes

Reproduction t1 par microfilm a1
Centre de documentation t3 en France a1

les sous-structures *par microfilm a1* et *en France a1*, que l'on peut considérer comme possédant respectivement un sens instrumental et un sens locatif, ne sont traitées, au niveau de l'information de base, qu'en fonction de la seule relation d'attribut, établie entre les items dont les supports de radical sont *microfilm* et *France* et les noyaux respectifs. Cette situation, qui a des avantages certains de simplicité et de généralisation, ne va pas, dans quelques cas, sans difficultés (on réservera la discussion plus circonstanciée de ces questions pour la fiche sur le réseau sémantique de V1 et pour un travail ultérieur sur les contraintes et conditions d'adéquation d'un langage documentaire à relations syntaxiques). On peut souhaiter, par conséquent, distinguer une relation d'attribut de toutes les autres, en particulierisant son contenu. Pour ce faire, la grammaire de V1 prévoit la possibilité de spécifier la relation d'attribut au moyen des *spécificateurs d'attribut*, introduits dans les articulations grammaticales d'un item A.

Un spécificateur d'attribut (désormais, spécificateur A) possède un signifié, auquel sont associés un ou plusieurs signifiants. Par exemple, V1 incorpore un spécificateur A, dont le signifié est « contenu » et dont les signifiants sont *spécialisé(e)(s) en, spécialisé(e)(s) dans, portant sur, sur*. Le documentaliste choisit, dans cette liste fermée de signifiants possibles, celui qu'il souhaite incorporer à la structure grammaticale qu'il veut produire. Le signifiant du spécificateur A sera suivi du symbole « s » et placé en tête de l'articulation d'un item A (cf. § 1.1.1). Exemples :

Document t0 portant sur s l'indexation a1
Bibliothèques t0 spécialisées en s médecine a1
Films t0 sur s l'URSS a1

L'utilisation d'un spécificateur A ne crée pas un item nouveau et n'introduit pas une nouvelle relation : la sous-structure grammaticale *portant sur s l'indexation a1* est considérée comme formée d'un seul item, dont le support de radical est *indexation*, item qui sera, si nécessaire, numéroté par la machine en tant que tel (cf. fiche n°3, p. 1) ; cet item A continue à porter sur *document t0* et en est complément A ; il possède toutes les possibilités combinatoires des items A, en particulier par rapport au marqueur *b*. En effet, un item A précédé d'un spécificateur A peut apparaître soit après *b*, dans la suite A, soit avant *b*, comme noyau A, soit avant et après. Exemples :

Diagramme t0 d'un document a1 b portant sur s l'indexation a1
Document t0 portant sur s l'indexation a1 b automatique a1
Diagramme t0 d'un document a1 b portant sur s l'indexation a1 b automatique a1

L'utilisation des spécificateurs A obéit à des contraintes sévères ; elles portent sur les traits que doivent posséder soit l'item A ainsi spécifié, soit le noyau sur lequel porte cet item A, soit l'un et l'autre. Ces contraintes seront présentées au § 5.1.2.1.2.

3.4 Le noyau à coordination interne

Le noyau d'un syntagme peut être soit simple soit avec coordination interne (cf. § 3.1) ; le premier type a été seul décrit et illustré jusqu'ici. Un noyau à coordination interne est constitué de deux ou plus de deux items T reliés par la particule *et*. Un noyau à coordination interne forme un bloc conceptuel à partir des items qui le composent (cf. § 5.1.1) ; de ce fait, les entités qui portent sur lui (les entités A — item A, syntagme A, bloc A — ou

le syntagme de phrase ou le bloc de syntagmes de phrase) doivent porter sur le noyau à coordination interne en entier et en même temps sur chacun des items qui le composent. Exemples :

- (1) Documents t0 et livres t0 et estampes t0.
- (2) Indexation t1 et condensation t1.
- (3) Films t0 et microfilms t0 d'un centre de documentation a3.
- (4) Bibliothèques t0 et centres de documentation t3 en URSS a1.
- (5) Indexation t1 et condensation t1 automatiques a1 des textes t0 anglais a1.
- (6) Stockage t1 des livres t0 et documents t0 anciens a1 sur s les mathématiques a1.

Les exemples (1) et (2) illustrent des phrases documentaires constituées d'un syntagme à noyau *isolé*. Dans l'exemple (3), ce sont les films et les microfilms qui appartiennent à un centre de documentation; l'interprétation *films* (sans détermination) et *microfilms d'un centre de documentation* est donc à exclure. De même, l'exemple (4) ne peut être interprété comme étant équivalent de *centres de documentation en URSS et bibliothèques*. *Mutatis mutandis*, cette observation est valable pour *automatiques a1* dans (5) et pour *anciens a1 sur s les mathématiques a1* en (6). De même, le syntagme de phrase *textes t0 anglais a1* porte sur le syntagme qui précède — *indexation t1 et condensation t1 automatiques a1*; *textes t0* est le complément verbal de *indexation t1 et condensation t1*.

Grâce à l'effet « embloqueur » du noyau à coordination interne, sa combinatoire interne est réglée par la contrainte suivante : *si le noyau à coordination interne est suivi d'un syntagme de phrase ou d'un bloc de syntagmes de phrase qui porte sur lui, les items T qui le composent doivent tous être des items TV* (items TV : items t1 ou t2 ou, si la notion correspondante admet t1, t3; cf. exemple (5) ci-dessus, où le noyau à coordination interne est suivi d'un syntagme de phrase). Pour le reste, *le comportement du noyau à coordination interne est identique à celui du noyau simple* : la restriction précédente sur les items TV ne joue évidemment pas si le noyau à coordination interne est seulement suivi d'un groupe A (cf. exemple (4) ci-dessus) ou si aucune entité ne porte sur lui; la combinatoire du noyau à coordination interne et les relations auxquelles il participe sont exactement les mêmes que celles du noyau simple; la sous-structure où apparaît chacun des items du noyau à coordination interne s'organise comme toute autre sous-structure (cf. § 1.1.1). Un noyau à coordination interne peut donc constituer à lui seul un syntagme de phrase ou être suivi d'un groupe A. Si les items qui le composent sont des TV, il peut être suivi d'un autre syntagme de phrase. Le syntagme de phrase qui possède un noyau à coordination interne a exactement le même comportement que le syntagme de phrase à noyau simple, et les items A d'un groupe A contractent avec le noyau à coordination interne exactement les mêmes relations qu'avec le noyau simple.

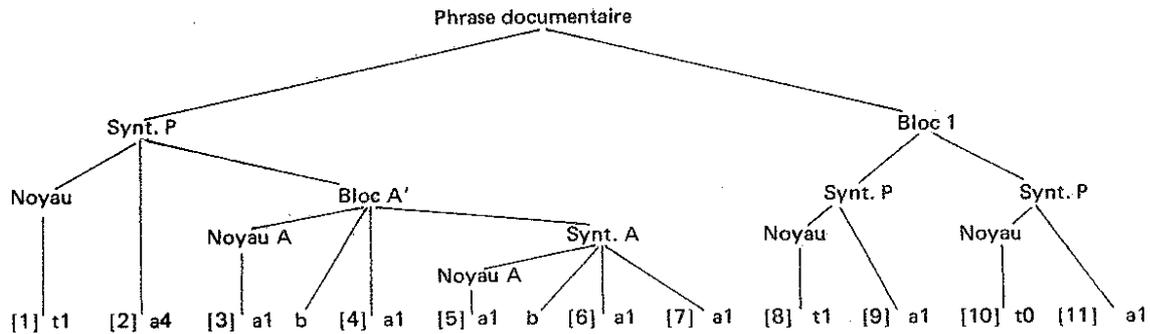
3.5 Récapitulation des relations internes à la phrase documentaire

On sait que la phrase documentaire est composée d'un ou plusieurs syntagmes de phrase, ceux-ci étant constitués d'un noyau et, optionnellement, d'un groupe A. Les syntagmes de phrase se combinent entre eux pour former la phrase documentaire; cette combinatoire détermine la formation de structures emboîtées de plus en plus larges, en commençant par le dernier syntagme, à droite de la phrase documentaire. La *relation de projection*, qui existe entre toute entité X qui porte sur une entité Y, organise les rapports entre ces structures emboîtées : le dernier syntagme à droite porte sur celui qui précède; il constitue avec lui le bloc d'ordre 1, lequel porte sur le syntagme précédent. Le bloc d'ordre 1 constitue avec le syntagme précédent le bloc d'ordre 2, lequel porte sur le syntagme qui précède. Ainsi, successivement, se constituent tous les blocs pour aboutir à la construction des structures emboîtées de la phrase documentaire entière.

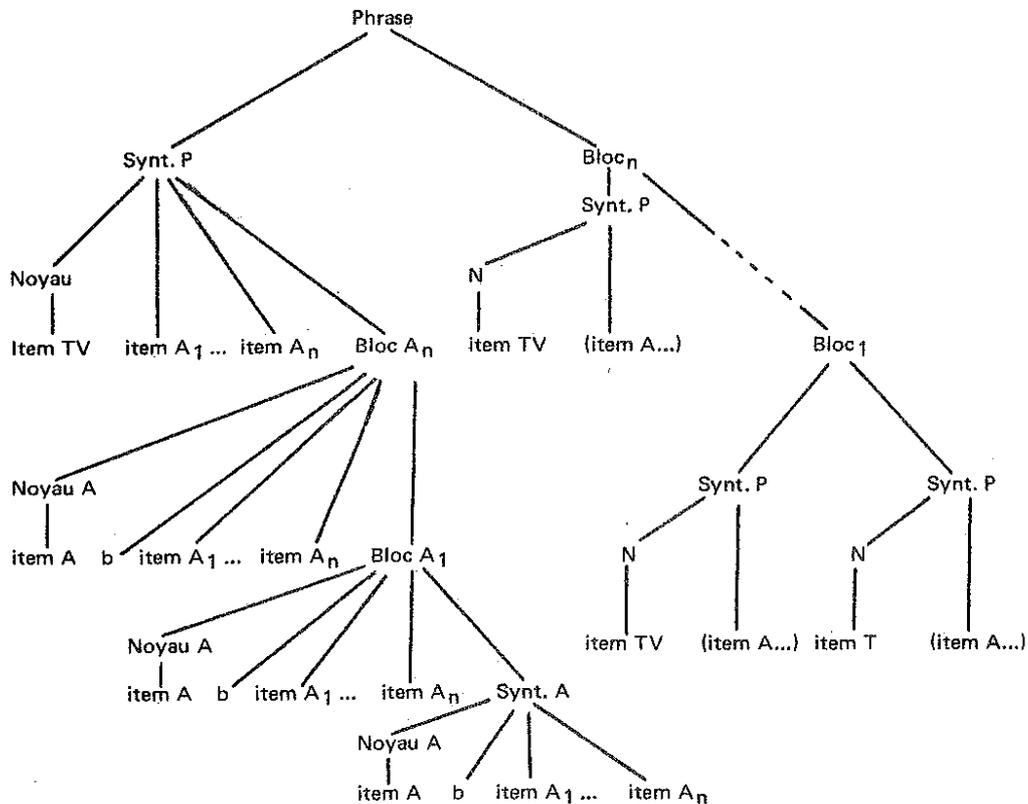
Le groupe A est organisé en fonction du même principe d'emboîtement de structures

de droite à gauche ; c'est également la relation de projection qui organise ces emboîtements. L'organisation du groupe A présente, cependant, deux différences importantes par rapport à celle de la phrase documentaire : (1) un syntagme A ou un bloc A ne porte pas sur un autre syntagme mais sur un noyau, que celui-ci soit noyau A ou noyau de syntagme de phrase ; (2) un item A peut porter, lui aussi, sur un noyau (noyau A ou noyau de syntagme de phrase).

L'organisation de la phrase documentaire, en fonction des syntagmes de phrase qui la composent, et l'organisation du groupe A ont été exposées séparément. Mais il est clair qu'une même phrase documentaire peut présenter des emboîtements aussi bien au niveau des syntagmes de phrase que dans un ou plusieurs groupes A, comme le montre l'exemple suivant, accompagné d'un schéma sur les relations emboîtées dans la phrase documentaire :



Gestion t1 automatisée a4 dans une bibliothèque a1 b centrale a1 d'une université a1 b bilingue a1 en Inde a1
 1 2 3 4 5 6 7
 du prêt t1 interbibliothèque a1 des périodiques t0 étrangers a1
 8 9 10 11



3.6 Les constructions documentaires

Ce qui précède dans ce § 3 a été consacré à la présentation des phrases documentaires et des structures qui les constituent. Mais on sait que V1 incorpore aussi les constructions documentaires.

Une construction documentaire est constituée d'une phrase documentaire suivie d'un prédicat et d'une autre phrase documentaire. Le prédicat est constitué par l'association d'un signifié avec un signifiant. La grammaire de V1 détermine à l'avance la liste de prédicats, le signifié de chacun et le ou les signifiants qui peuvent être associés à chacun des signifiés; s'il y a plusieurs signifiants, le documentaliste choisit l'un d'eux (cf. § 5.2).

L'écriture d'une construction documentaire s'effectue selon les principes suivants :

(a) Les phrases documentaires constitutives d'une construction documentaire obéissent aux mêmes conventions de notation que n'importe quelle phrase documentaire à l'exception du détail suivant : la phrase documentaire de gauche ne finit pas par un point, mais par une virgule ;

(b) Le signifiant du prédicat suit la première phrase documentaire et il est lui-même suivi du symbole « r ».

Exemples :

Bibliothèque t0 publique a1 au Mexique a1, en coopération avec r un centre de documentation t3 au Canada a1.

Recrutement t1 de documentalistes t0 spécialisés a4, avec l'objectif d'r automatiser t1 un centre de documentation t3.

Le schéma général auquel obéit une construction documentaire est donc le suivant :

Phrase documentaire, signifiant du prédicat r phrase documentaire.

*

* *

4 RELATIONS ENTRE STRUCTURES GRAMMATICALES

La grammaire de V1 spécifie ou engendre les structures grammaticales. Or, ces structures grammaticales entretiennent entre elles des relations, dont certaines peuvent contribuer aussi à la spécification des contenus significatifs du langage V1. Ces relations entre structures grammaticales sont déterminées par l'anaphorique et par le classement des structures grammaticales en ED et SED.

4.1 La sous-structure anaphorique

Il a été prévu dans le langage V1 un anaphorique. Sa fonction est de renvoyer à une structure grammaticale antérieure et de réintroduire ainsi, dans l'item anaphorique, un contenu significatif qui a été auparavant exprimé dans une autre structure grammaticale. Celle-ci peut soit faire partie de l'ED où apparaît l'anaphorique, soit faire partie d'une autre ED ou constituer à elle seule une autre ED. Exemples (entre guillemets, la paraphrase possible des structures LN qui correspondent aux structures grammaticales comprenant l'anaphorique) :

- (1) 1.1 Échange t1 de documents t0. 2 Coopération t1 internationale a1 par cet échange a1.
« Coopération internationale par échange de documents. »
- (2) 1. Recherche scientifique t0 en Inde a1, en bibliothéconomie a1 et science de l'information a1.
2. Financement t1 de cette recherche scientifique t0.
« Financement de la recherche scientifique en Inde, en bibliothéconomie et science de l'information. »

La composition interne de la sous-structure grammaticale à anaphorique a été présentée au § 1.1.1; cf. aussi fiche n° 4, p. 5. Le comportement de la préposition de complément verbal et du signifiant du spécificateur A est identique dans les sous-structures grammaticales avec ou sans anaphorique et ne sera pas par conséquent présenté ici; sur le développement de l'anaphorique et l'instruction d'écriture correspondante, qui ne concernent que les structures de transition, cf. fiche n° 2, § 2.1.1.3. On s'intéressera donc ici à la *forme anaphorique*, à la *référence*, au *terme repris* et à son *antécédent*, et à l'*indicateur de procédé*.

La *forme anaphorique* a pour rôle de signaler la présence de l'anaphore; elle se présente sous l'une des formes de l'adjectif démonstratif en langue naturelle (*ce, cet, cette, ces*), ce qui permet à l'anaphorique de s'intégrer sans difficulté dans les structures complètes.

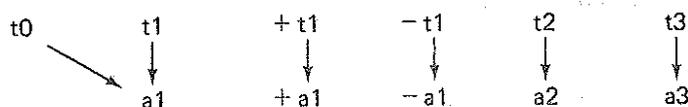
Le langage V1 exclut toute ambiguïté dans l'identification de l'antécédent, à la différence de ce qui peut se produire lorsqu'on utilise l'anaphore en langue naturelle. Lorsque le terme repris peut renvoyer à plus d'un antécédent, la *référence*, qui se note entre parenthèses immédiatement après la forme anaphorique, précise le renvoi et permet à la machine d'identifier correctement l'antécédent (cf. fiche n° 4, p. 6). La machine, dans le cadre de l'aide à l'indexation, doit signaler au documentaliste la nécessité d'introduire une référence en cas d'ambiguïté dans le repérage de l'antécédent. Dans certains cas (cf. modalité III de constitution des contenus significatifs de l'anaphorique, ci-dessous § 4.3.), la référence peut renvoyer à un autre terme repris et non à l'antécédent.

Le *terme repris* est le signifiant — expression ou mot — placé immédiatement après la forme anaphorique ou, éventuellement, après la référence, et avant l'indicateur de procédé. Le terme repris *calque* le signifiant d'un antécédent.

L'*antécédent* d'un terme repris peut apparaître : (1) dans une sous-structure grammaticale, ou (2) dans le prédicat d'une construction documentaire. Si le besoin de distinction se fait sentir, on les appellera respectivement *antécédent non-prédicatif* et *prédicatif*.

L'*antécédent non-prédicatif* d'un terme repris est le mot ou l'expression qui apparaît dans une structure grammaticale qui précède celle où apparaît le terme repris (et, plus particulièrement, dans la position indiquée par la référence, s'il y en a une), 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 (expression qui incorpore un radical discriminant);
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. Le tableau suivant indique cette invariance des traits sémantiques entre items T et items A :



(A noter que l'anaphore de V1, dans sa version actuelle, ne peut pas reprendre les items a4; ceux-ci introduisent des traits sémantiques particuliers, non reproduits par les items T et, par ailleurs, correspondent aux participes passés passifs en LN, formes verbales qui, en LN, ne peuvent pas être reprises en tant que telles par l'anaphore.)

L'*antécédent prédicatif* d'un terme repris apparaît dans le prédicat d'une structure grammaticale qui précède celle où apparaît le terme repris (et, plus particulièrement, dans celle qui est indiquée par la référence, s'il y en a une), prédicat qui est le seul à posséder la caractéristique suivante : le radical discriminant du prédicat est incorporé dans le terme repris.

Le signifié d'un antécédent *non-prédicatif*, antécédent qui apparaît dans une sous-structure grammaticale, est constitué par la notion de l'item qui se situe dans cette sous-structure 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. On appellera *contenu significatif de l'antécédent* le signifié de l'antécédent, prédicatif ou non-prédicatif.

L'*indicateur de procédé*, 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, de même que l'indicateur de procédé placé après l'antécédent est celui qui est exigé par la syntaxe de la structure grammaticale où apparaît l'antécédent.

On trouvera, dans les exemples qui suivent, des illustrations de l'emploi de l'anaphore, et, plus particulièrement, des relations entre antécédent et terme repris; à la suite de chaque exemple, des commentaires permettent de faire le lien entre les principes généraux et les exemples particuliers.

(1) 1. Catalogue t0 d'une bibliothèque a1 b universitaire a1.

2. Cette bibliothèque t0, en relation avec r un centre de documentation t3.

(Assez fréquemment, l'antécédent est suivi d'un des indicateurs T mais le contraire, comme dans cet exemple, est aussi possible : tout item ou, le cas échéant, le spécificateur de notion de tout item, à l'exception des items a4, peut contenir l'antécédent d'un terme repris).

(2) 1. Recherche t1 documentaire a1 automatisée a4 par les bibliothèques a1 b universitaires a1 en France a1.

2. Accès t1 en mode conversationnel a1 de ces bibliothèques a1 aux bases de données t0.

3. Catalogues t0 de ces bibliothèques a1.

(Un même antécédent peut être calqué par plusieurs termes repris, dans la même ED ou, comme c'est le cas de cet exemple, dans des ED différentes. Lorsque l'antécédent et les termes repris appartiennent à des ED différentes, ils peuvent apparaître dans n'importe quelle sous-structure grammaticale de leurs structures grammaticales respectives; en revanche, voir ci-dessous, § 4. 3, la contrainte d'utilisation I, qui porte sur l'antécédent et les termes repris dans la même ED.)

(3) 1. Centre de documentation t3 municipal a1.

2. Historique t0 de ce centre a1.

(L'antécédent du terme repris est le signifiant d'un spécificateur de notion. La sous-structure à anaphorique aurait pu être écrite, au choix du documentaliste : *Historique t0 de ce centre de documentation a1*).

(4) 1. Bibliothèque t0 municipale a1, en coopération avec r un centre de documentation t3.

2. Historique t0 de cette coopération a1.

(Contrairement aux exemples (1) à (3) qui précèdent, l'antécédent est ici prédicatif).

L'anaphorique fonctionne par rapport au noyau à coordination interne selon les principes présentés ci-dessus : chacune des sous-structures grammaticales du noyau à coordination interne peut contenir un antécédent qui aura le même comportement, par rapport au terme repris, ou aux termes repris respectifs, que tout autre antécédent.

- (5) 1. Conservation t1 et microédition t1 de textes t0 anciens a1.
2. Service de microreproduction t3, avec l'objectif de r cette conservation t1 et de cette microédition t1.
- (6) 1. Organismes d'information t3 et organismes de documentation t3 en médecine a1.
2. Ces organismes d'information t3 et ces organismes de documentation t3, en coopération avec r les bibliothèques t0 universitaires a1.
- (7) 1. Organismes d'information t3 et centres de documentation t3 en médecine a1.
2. Ces organismes t3 et ces centres t3, en coopération avec r les bibliothèques t0 universitaires a1.
- (8) 1. Conservation t1 et stockage t1 des documents t0.
2. Salles t0 pour s cette conservation a1 et pour s ce stockage a1.
- (9) 1. Indexation t1 et traduction t1 automatiques a1 des documents t0.
2. Cette indexation t1 par indicateur de rôle a1.

(On observera que, dans les exemples (5) à (8), des termes repris calquent des antécédents non-prédicatifs qui figurent dans chacune des deux sous-structures d'un noyau à coordination interne, tandis que, dans (9), ce n'est qu'un seul des deux antécédents possibles qui est calqué par un terme repris qui apparaît dans une autre ED. Dans (6), à la différence de (7), les termes repris introduisent le spécificateur de notion et le mot ou l'expression qui suivent, pour éviter une ambiguïté qui aurait exigé l'utilisation de la référence, et aussi pour favoriser la lisibilité des structures complètes.)

4.2 Le classement des structures grammaticales

Le *classement de structures* est l'un des types d'informations indexées présentés dans la fiche n°2 (§ 2.3). Le rôle de l'anaphorique ne peut se comprendre pleinement qu'en liaison avec cette notion de classement de structures. On se souvient que les structures grammaticales sont classées par l'indexeur soit comme expressions documentaires (ED), soit comme sous-expressions documentaires (SED), selon que le contenu significatif véhiculé par la structure grammaticale est, par lui-même, pertinent ou non — ou, plus exactement, fortement pertinent ou faiblement pertinent — par rapport à un document donné. L'indexeur effectue ce classement en fonction de la compréhension qu'il a du document. Il crée des ED qui lui paraissent, chacune, correspondre à un bon niveau d'indexation du document, soit en donnant le statut d'ED à une structure grammaticale appropriée, soit en combinant dans une ED, à l'aide de l'anaphorique, plusieurs structures grammaticales non directement pertinentes, qui forment autant de SED (cf. ci-dessous, § 4.3). Dans tous les cas, il procède selon le principe de base suivant : c'est au niveau de l'ED entière que « s'établit le contact » entre le document et l'indexation ; *l'ED — et non la structure grammaticale — est en effet l'unité qui exprime le contenu d'un document*. Prenons l'exemple suivant :

1. Recherche t1 documentaire a1 automatisée a4 par les bibliothèques a1 b universitaires a1 en France a1.
2. Formation t1 professionnelle a1 du personnel t0 de ces bibliothèques a1 affecté à cette recherche a1.

L'indexeur qui a produit cette indexation considère que le document concerné répond pleinement à une question qui porterait sur la *recherche documentaire automatisée par les bibliothèques universitaires en France* ou sur la *formation professionnelle du personnel des bibliothèques universitaires en France, affecté à la recherche documentaire automatisée*. Au contraire, ce document est trop spécifique pour répondre correctement à une question sur la *recherche documentaire automatisée* en général ou sur les *bibliothèques universitaires*. Si l'indexeur avait estimé que le document était suffisamment pertinent pour ces deux dernières questions, il aurait écrit :

1. Recherche t1 documentaire a1 automatisée a4.
2. Bibliothèques t0 universitaires a1 en France a1.
3. Cette recherche t1 par ces bibliothèques a1.

Supposons maintenant que l'indexeur juge qu'un (autre) document répond parfaitement à une question portant sur la *coopération internationale par échange de documents*,

mais ne serait pas vraiment utile à quelqu'un qui chercherait à se documenter sur les *échanges de documents* en général. Il ne peut exprimer cette idée par une structure grammaticale unique qui aurait le statut d'ED, car *échange*, qui serait item A portant sur *coopération*, ne pourrait en tant qu'item A avoir pour complément d'objet direct *documents* (sur cette interdiction, cf. 5.1.2.1.1). Il est donc amené à écrire :

1.1 Echange t1 de documents t0. 2 Coopération t1 internationale a1 par cet échange a1.

L'emploi de l'anaphorique permet de tourner la contrainte grammaticale, nécessaire mais gênante pour l'indexation, qui veut qu'une entité A ne puisse jamais correspondre à un complément d'objet direct en LN. Or, la création de la structure grammaticale *échange t1 de documents t0* ne pose pas de problèmes sur le plan documentaire, puisque son classement comme SED évite de lui attribuer une réalité documentaire qu'elle n'a pas (sur l'utilisation de l'anaphorique pour porter remède à certaines contraintes de la grammaire gênantes pour l'indexation, voir § 5.1.1, 5.1.2.1.1, 5.1.2.1.3).

L'utilisation conjointe du classement de structures et de l'anaphorique permet donc à l'indexeur de jouer sur les différents niveaux d'indexation possibles et de choisir celui qui est le plus pertinent pour le document considéré.

4.3 Le contenu significatif anaphorique

La forme anaphorique et le terme repris, en « reprenant » un antécédent, déterminent le contenu significatif de l'item anaphorique, ou *contenu significatif anaphorique*. Pour abrégé, on utilisera les formules *contenu significatif du terme repris x* au lieu de *contenu significatif anaphorique de l'item où apparaît le terme repris x*, et *terme repris x* au lieu de *item anaphorique où apparaît le terme repris x*. Par la suite, on précisera d'abord les modalités selon lesquelles se constituent les contenus significatifs des termes repris et, ensuite, les contraintes d'utilisation de l'anaphorique, celles-ci devant se formuler en fonction de celles-là. Au préalable, on introduira quelques définitions.

Un antécédent peut être calqué par un ou plusieurs termes repris (par la suite, dans ce § 4 et sauf indication contraire, on parlera toujours des termes repris qui calquent un seul et même antécédent). On appellera *terme repris final* le dernier terme repris qui figure soit dans l'ED où apparaît l'antécédent, soit dans une ED où apparaît un *terme repris initial*. *Terme repris initial* désignera le premier terme repris qui, dans une ED autre que celle où apparaît l'antécédent, calque cet antécédent.

Par *déterminations à droite* d'un item d'un noyau x, on comprend le 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 le contenu significatif du syntagme de phrase ou du bloc de syntagmes de phrase qui, dans la phrase documentaire où apparaît le noyau x, porte sur le syntagme de phrase où apparaît ce noyau (sur les contenus significatifs associés aux entités syntaxiques du langage, cf. § 5).

A partir de ces définitions, il est possible de préciser les *modalités* selon lesquelles se constituent les contenus significatifs des termes repris :

1. - *Le contenu significatif 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.* Exemples (les déterminations à droite incorporées aux contenus significatifs selon la modalité 1 sont, dans les trois exemples qui suivent, en italiques) :

(1) 1.1 Langage t0 pour s l'indexation a1 b automatique a1. 2 Ce langage t0 dans un centre de documentation a3 b en France a1. 3 Relations t0 syntaxiques a1 de ce langage a1.

« Langage pour l'indexation automatique. Langage pour l'indexation automatique dans un centre de documentation en France. Relations syntaxiques du langage pour l'indexation automatique dans un centre de documentation en France. »

(2) 1.1 Recherche scientifique t0 en Inde a1, en bibliothéconomie a1 et science de l'information a1. 2 Financement t1 de cette recherche scientifique t0.

« Recherche scientifique en Inde, en bibliothéconomie et science de l'information. Financement de la recherche scientifique en Inde, en bibliothéconomie et science de l'information. »

(3) 1.1 Centre de documentation t3 municipal a1, en coopération avec r la bibliothèque t0 universitaire a1. 2 Historique t0 de cette coopération a1.

« Centre de documentation municipal, en coopération avec la bibliothèque universitaire. Historique de la coopération du centre de documentation municipal avec la bibliothèque universitaire. »

II. - *Le contenu significatif 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 de terme repris final, le contenu significatif de l'antécédent, avec, s'il y en a, ses déterminations à droite.*
Exemples :

(1) 1.1 Langage t0 pour s l'indexation a1 b automatique a1. 2 Ce langage t0 dans un centre de documentation a3 b aux USA a1.

« Langage pour l'indexation automatique. Langage pour l'indexation automatique dans un centre de documentation aux USA. »

2. Evaluation t1 de ce langage t0.

« Evaluation du langage pour l'indexation automatique dans un centre de documentation aux USA. »

(2) 1. Langage t0 pour s l'indexation a1 b automatique a1.

2. Evaluation t1 de ce langage t0.

« Evaluation du langage pour l'indexation automatique. »

3. Adaptation t1 de ce langage t0.

« Adaptation du langage pour l'indexation automatique. »

(3) 1. Langage t0 pour s l'indexation a1 b automatique a1.

2.1 Ce langage t0 dans un centre de documentation a3 b aux USA a1. 2 Ce langage t0 adapté a4 b par des chercheurs a1 en France a1.

« Langage pour l'indexation automatique dans un centre de documentation aux USA. Langage pour l'indexation automatique dans un centre de documentation aux USA adapté par des chercheurs en France. »

3. Evaluation t1 de ce langage t0.

« Evaluation du langage pour l'indexation automatique. »

III. - *Le contenu significatif 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.*

(1) 1. Langage t0 pour s l'indexation a1 b automatique a1.

2.1 Ce langage t0 dans un centre de documentation a3 b aux USA a1. 2 Ce langage t0 adapté a4 b par des chercheurs a1 en France a1.

3. Evaluation t1 de ce (2.2) langage t0.

« Evaluation du langage pour l'indexation automatique dans un centre de documentation aux USA adapté par des chercheurs en France. »

(2) 1. Langage t0 pour s l'indexation a1 b automatique a1.

2. Ce langage t0 traduit a4 b par des chercheurs a1 en France a1.

3. Adaptation t0 de ce (2) langage t0.

« Adaptation du langage pour l'indexation automatique traduit par des chercheurs en France. »

Les *contraintes d'utilisation* de l'anaphorique sont les suivantes :

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

Cette contrainte est illustrée par les exemples de la modalité I. Elle implique que les ED suivantes ne sont pas admises dans V1 :

(1) 1.1 Adaptation t1 d'un langage t0 à relations a1 b syntaxiques a1. 2 Ce langage t0 pour s l'indexation a1 b automatique a1.*

(2) 1.1 Bibliothèque t0 au Mexique a1, en coopération avec r un centre de documentation t3 en France a1. 2 Fonds t0 de cette bibliothèque a1.*

En effet, contrairement à ce qu'exige la contrainte I, le contenu significatif qui est associé à *adaptation t1* dans (1) ne réapparaît pas dans *ce langage t0 pour s l'indexation a1 b automatique a1*, où sont reprises seulement les déterminations à droite de *langage t0*. De même, dans (2), le contenu significatif associé à *en coopération avec r un centre de documentation t3 en France a1* ne réapparaît pas dans la deuxième SED, où on ne retrouve que le contenu significatif associé à *au Mexique a1* (les déterminations à droite, conformément à la définition donnée plus haut, p. 35, ne franchissent pas les limites de la PD).

La contrainte d'utilisation I implique que, s'il y a un seul antécédent non-prédicatif dans une ED et que cet antécédent est suivi de plusieurs termes repris dans la même ED, l'antécédent doit apparaître dans la sous-structure grammaticale n° 1 de la première SED et chacun des termes repris, à l'exception du terme repris final, doit apparaître dans la sous-structure grammaticale n°1 des SED suivantes (cf. exemple (1) à la suite de la modalité I, ci-dessus). De même, si, dans une ED, il y a plusieurs antécédents non-prédicatifs et si chacun de ces antécédents, dans la même ED, est suivi de plusieurs termes repris, chacune des SED à l'exception de la dernière doit posséder dans la sous-structure grammaticale n° 1 soit un antécédent, soit un terme repris qui ne soit pas lui-même le dernier terme repris d'un antécédent. Exemple (entre guillemets, la paraphrase possible de la structure LN qui correspond à la SED n° 5) :

1.1 Bibliothèques t0 de médecine a1 b du travail a1. 2 Ces bibliothèques t0 dans les organismes a1 b internationaux a1. 3 Centres de documentation t3 spécialisés en s psychologie a1 b industrielle a1. 4 Ces centres t3 dans les universités a1 b canadiennes a1. 5 Ces bibliothèques t0, en coopération avec r ces centres t3.

« Bibliothèques de médecine du travail des organismes internationaux, en coopération avec les centres de documentation spécialisés en psychologie industrielle des universités canadiennes. »

En résumé, chacune des SED à l'exception de la dernière doit posséder dans la sous-structure grammaticale n° 1 soit un antécédent, soit un terme repris qui ne soit pas lui-même le dernier terme repris d'un antécédent. La machine contrôle le respect de la contrainte I et elle émet un message d'erreur au cas où la dernière SED n'incorpore pas 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 auparavant introduit dans une autre ED que si le contenu significatif anaphorique ainsi créé et le contenu significatif auparavant introduit expriment le même référent.*

Selon cette contrainte, si, par exemple, on a :

1. Bibliothèques t0 publiques a1 en France a1.
2. Financement t0 de ces bibliothèques t0.

ces bibliothèques t0 dans l'ED 2 doit renvoyer aux mêmes entités référentielles que celles qui sont désignées dans l'ED 1. Les ED précédentes ne doivent donc pas être utilisées si le financement, dans l'ED 2, porte sur des bibliothèques publiques en France, lorsque celles-ci ne sont pas référentiellement les mêmes que celles de l'ED 1. La contrainte II s'applique dans tous les cas où un terme repris et sa forme anaphorique introduisent dans une ED un contenu significatif déjà incorporé dans une autre ED ; cf. les exemples qui illustrent les modalités d'application II et III ci-dessus.

4.4 Relations entre structures grammaticales et classement des structures grammaticales : le rôle de l'anaphorique

Compte tenu des indications précédentes, on peut montrer que le rôle de l'anaphorique, dans les relations entre structures grammaticales, est double :

(a) il réintroduit 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 (conformément aux principes exposés au paragraphe précédent) ;

(b) il marque qu'il existe 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 signi-

ficatif précédemment exprimé, et réintroduit par le moyen de l'anaphore. On entend par là que chacun de ces contenus significatifs renvoie à une même entité (ou à un même groupe d'entités) du monde référentiel.

Le rôle de l'anaphorique s'exerce différemment selon qu'il met en relation des SED appartenant à la même ED ou des ED différentes.

Dans le premier cas, la réintroduction du contenu significatif antérieur va permettre de faire entrer dans le contenu significatif de l'item anaphorique des déterminations qui n'auraient pu être exprimées directement au moyen d'une PD unique. On pourra ainsi porter remède à certaines contraintes de la grammaire gênantes pour l'indexation, comme le montre l'exemple déjà examiné au § 4.2 :

- (1) 1.1 Echange t1 de documents t0. 2 Coopération t1 internationale a1 par cet échange a1.

Mais on ne peut pas dire dans ce cas que l'anaphorique marque l'identité référentielle : la relation avec le contenu du document et, par conséquent, avec le monde référentiel ne se réalise vraiment qu'au niveau de l'ED entière et, plus précisément, au niveau du contenu significatif de la dernière SED, qui intègre le contenu significatif des SED précédentes. On voit donc que dans les relations entre SED d'une même ED, seul fonctionne le premier rôle de l'anaphorique.

Il en va autrement lorsque l'anaphorique établit une relation entre deux ED. Soit l'exemple (2) :

- (2) 1. Recherche t1 documentaire a1 automatisée a4 par les bibliothèques a1 b universitaires a1 en France a1.

2. Formation t1 professionnelle a1 du personnel t0 de ces bibliothèques a1.

L'emploi de l'anaphorique dans la deuxième ED marque nettement que les bibliothèques visées dans cette ED sont les mêmes que celles dont on a pu dire, dans la première ED, qu'elles pratiquaient (ou que certaines d'entre elles pratiquaient) la recherche documentaire automatisée. Un lien assez souple se crée entre les deux ED. Dans l'exemple considéré, on peut dire que la formation professionnelle du personnel concerne un ensemble de bibliothèques *où se pose le problème de la recherche documentaire automatisée*, précision qui n'est pas indifférente sur le plan documentaire. Ces liaisons entre ED représentant le contenu d'un même document pourront être utilisées par la machine, dans le fonctionnement du réseau sémantique, pour fournir des réponses par *association d'idées*.

Dans le cas où l'anaphorique établit une relation entre deux ED, il sert donc à la fois à marquer une identité référentielle et à réintroduire un contenu significatif antérieur (ici : « bibliothèques universitaires en France »). La réintroduction du contenu significatif antérieur peut avoir des effets pratiques plus ou moins développés :

(a) si l'anaphorique renvoie à un antécédent qui n'a pas de déterminations à droite, seul est réintroduit le contenu significatif de l'antécédent. Exemple :

- (3) 1. Construction t1 en France a1 de bibliothèques t0.
2. Financement t1 de ces bibliothèques t0.

L'effet pratique de la réintroduction est donc très limité et la fonction réelle de l'anaphorique est dans ce cas essentiellement de marquer l'identité référentielle entre les bibliothèques de l'ED 1 et celles de l'ED 2 (en permettant, par exemple, à la machine de rapprocher par association d'idées l'item *en France a1* de la première ED et l'item *bibliothèques t0* de la seconde).

(b) plus fréquemment, l'anaphorique va permettre, en application des principes vus au § 4.3, la réintroduction d'un contenu significatif qui ne se limite pas à celui de l'antécédent. C'est le cas de l'exemple (2), où le contenu significatif de l'ED 2 est « formation professionnelle du personnel des bibliothèques universitaires en France ». L'identité référentielle avec les bibliothèques auxquelles se réfère l'ED 1 est marquée, mais en plus l'indexeur se voit offrir une commodité d'écriture, assez comparable à celles que procurent les instructions d'écriture : il n'a pas besoin d'écrire effectivement les items *universitaires a1 en France a1*.

(c) parfois l'anaphorique permet de réintroduire un contenu significatif qui ne pourrait s'exprimer directement à partir d'une PD unique et qui a été construit dans une ED antérieure à partir de plusieurs SED; ou encore, il permet de construire un tel contenu significatif en s'appuyant sur une ED précédente. Les effets pratiques du rôle de réintroduction jouent alors à plein, comme à l'intérieur d'une ED, ainsi qu'on le voit dans l'exemple suivant :

- (4) 1. Echange t1 de documents t0 scientifiques a1.
2. Coopération t1 internationale a1 par cet échange a1.

L'identité référentielle entre entités apparaissant dans des ED différentes doit être notée par l'indexeur à l'aide de l'anaphorique chaque fois que le document lui-même permet de la constater sans ambiguïté et que l'indexeur la juge pertinente sur le plan documentaire. Lorsque l'identité référentielle n'existe pas ou lorsque le document ne permet pas à l'indexeur de la constater utilement et sans ambiguïté, on ne trouvera donc entre les entités comparables de chaque ED qu'une simple *identité linguistique*. Cette identité linguistique signifie que, dans chaque ED, on renvoie à des entités du monde référentiel qui peuvent être désignées par les mêmes notions ou combinaisons de notions, éventuellement associées aux mêmes traits sémantiques et aux mêmes relations apportés par les indicateurs de procédé. Une recherche par association d'idées pourra reposer, dans le fonctionnement du réseau sémantique, sur la possession en commun de traits sémantiques associés aux notions et/ou aux indicateurs de procédé.

*
* *

5 LES CONTENUS SIGNIFICATIFS DANS LE LANGAGE V1

Aux § 1 et 1.1.1, il a été établi une distinction entre les entités du langage documentaire qui sont *complètes* et celles qui sont *schématiques*; les premières (construction documentaire, phrase documentaire, syntagme de phrase, etc.) ont des articulations particularisées, introduites par le documentaliste, tandis que les secondes ne présentent que les emplacements réservés à ces articulations (emplacements des articulations libres). On sait aussi (cf. § 1) que l'expression *contenu significatif* est réservée aux entités schématiques, tandis que *contenu* est employé pour faire référence au signifié des entités complètes.

Dans le § 3, consacré aux contraintes combinatoires et aux relations internes de la phrase documentaire et de la construction documentaire, on a constamment présenté des entités complètes, alors qu'on décrivait les relations dans des entités schématiques. Cette démarche est justifiée par des raisons pédagogiques et de lisibilité; il a été possible d'en faire usage grâce à la valeur ambiguë des expressions telles que *phrase documentaire*, *syntagme*, etc. (cf. § 1.1.1) et elle se justifie par la naturalité des structures grammaticales de V1. La démarche adoptée a, certes, permis de présenter les relations internes des entités schématiques; mais, maintenant, il se pose le problème de savoir quel est le contenu significatif de ces entités schématiques. Autrement dit, si on est confronté aux phrases documentaires complètes suivantes :

- (1) Indexation t1 automatique a1 des périodiques t0 anglais a1 d'une bibliothèque a1 b universitaire a1.
(2) Indexation t1 automatique a1 dans une bibliothèque a1 b universitaire a1 des périodiques t0 anglais a1.

il est très facile, en effaçant les indicateurs de procédé, d'être tenté de les comprendre comme, respectivement :

(1') Indexation automatique des périodiques anglais d'une bibliothèque universitaire.

(2') Indexation automatique dans une bibliothèque universitaire des périodiques anglais.

Mais on sait que les articulations, telles que les suffixes, les indicateurs de genre et de nombre, les prépositions, les articles, etc., ne relèvent pas des entités schématiques et ne contribuent pas au contenu significatif de la phrase. De ce fait, des phrases documentaires complètes qui sont différentes par leurs articulations sont identiques en tant que phrases documentaires schématiques; autrement dit, deux ou plus de deux contenus différents peuvent correspondre à un même contenu significatif. On doit donc se poser la question de savoir quel est le contenu significatif des structures grammaticales schématiques, car tout le processus documentaire se fera en fonction de ce contenu significatif et non de celui qui résulte simplement du fait d'effacer les indicateurs de procédé d'une phrase documentaire complète. Autrement dit, il faut spécifier les contenus significatifs des phrases documentaires schématiques suivantes, qui sont sous-jacentes aux phrases documentaires complètes précédentes (les emplacements des articulations libres ne sont pas indiqués) :

(1'') INDEXA t1 AUTOMATIQ a1 PERI t0 ANG a1 BIBLIOTHEQ a1 b UNIVERSIT a1.

(2'') INDEXA t1 AUTOMATIQ a1 BIBLIOTHEQ a1 b UNIVERSIT a1 PERI t0 ANG a1.

Il est clair qu'une simple inspection des structures schématiques ne suffit pas pour que le documentaliste détermine leurs contenus significatifs. Mais ceux-ci peuvent être spécifiés par quelques règles simples, qu'on présentera dans ce paragraphe.

On sait qu'une construction documentaire est constituée de deux phrases documentaires mises en rapport par un prédicat et qu'une phrase documentaire est composée de syntagmes qui sont eux mêmes composés d'items; on sait aussi que syntagmes et items s'organisent dans la phrase en structures emboîtées. L'item possède un contenu significatif (cf. ci-dessous). Il y aura un contenu significatif au niveau de l'ensemble de la phrase documentaire et de la construction documentaire, mais il y aura aussi des contenus significatifs intermédiaires, au niveau des structures qui se situent entre la phrase documentaire globale et les items individuels. Par ailleurs, on sait aussi (cf. § 4) que V1 spécifie certaines relations entre structures grammaticales, notamment par le biais de l'anaphorique et du classement en ED et SED des structures grammaticales. Or, ces relations entre structures grammaticales contribuent également à spécifier les contenus significatifs du langage V1, ce qui sera exposé dans ce § 5 en liaison avec les contenus significatifs des structures grammaticales. C'est pour cette raison qu'on abordera dans ce § 5 les contenus significatifs du langage V1 : on présentera aussi bien les contenus significatifs spécifiés par la grammaire de V1 qui relèvent des items individuels, d'une structure grammaticale globale — phrase ou construction documentaire — et des structures intermédiaires que les contenus significatifs qui relèvent des relations entre structures grammaticales.

5.1 Entités de la phrase documentaire avec contenu significatif

On sait que l'item, qui n'est ni une entité schématique ni une entité complète, est constitué soit d'une notion associée à un radical discriminant et d'un indicateur de procédé, soit d'un terme repris, à la suite d'une forme anaphorique, suivi d'un indicateur de procédé (item anaphorique). Les articulations autour des deux types d'item s'organisent dans le cadre de la sous-structure grammaticale, comme il a été indiqué au § 1.1.1.

Au § 4.3, on a présenté les modalités de constitution du contenu significatif de l'item anaphorique. Le contenu significatif d'un item sans anaphorique est déterminé par le spécificateur de notion (éventuellement), la notion associée au radical discriminant et l'indicateur de procédé. *Le contenu significatif d'un item sans anaphorique résulte donc de trois composantes : le signifié du spécificateur de notion (qui peut ou non exister), la notion et les traits sémantiques introduits par l'indicateur de procédé.*

Il faut remarquer que tous les éléments d'une sous-structure grammaticale ne déterminent pas le contenu significatif d'un item, avec ou sans anaphorique. On sait que les articulations libres ne le font pas, mais le développement de l'anaphorique, le spécificateur A et

la préposition éventuelle du complément verbal ne le font pas non plus. Les deux derniers participent cependant au contenu significatif d'une entité schématique en introduisant des précisions soit dans la relation d'attribut (cf. § 5.1.2.1.2) soit dans la relation verbale (cf. § 5.1.2.2), mais cette participation à la détermination du contenu significatif ne se fait pas au niveau de l'item individuel.

On considérera le contenu significatif de l'item, avec ou sans anaphorique, comme l'unité de contenu significatif de la phrase documentaire : tous les autres contenus significatifs de la phrase documentaire sont le résultat d'une mise en rapport des contenus significatifs des items.

La mise en rapport des contenus significatifs des items individuels est déterminée, le cas échéant, à un premier stade, par la particule *et*, placée entre les items d'un noyau à coordination interne. Ensuite, et c'est le cas le plus courant dans toute structure grammaticale avec deux ou plus de deux items, la mise en rapport — qu'on appellera *composition* — des contenus significatifs des items individuels est déterminée par la relation de projection entre une entité X et une entité Y qui porte sur elle.

On aura par conséquent des contenus significatifs résultant d'une composition des contenus significatifs des items individuels, déterminée par la relation de projection, dans les entités suivantes :

- (1) le dernier syntagme A, à droite du groupe A;
- (2) le bloc A;
- (3) le syntagme de phrase;
- (4) le bloc de syntagmes de phrase;
- (5) la phrase documentaire de deux ou plus de deux syntagmes de phrase.

La composition de (1), (2) et (3) s'effectue à l'intérieur du syntagme de phrase ; en revanche, la composition en (4) et en (5) s'effectue à l'extérieur du syntagme de phrase, soit entre syntagmes de phrase, soit entre syntagme de phrase et bloc de syntagmes de phrase. Il est à remarquer que la suite A et le groupe A, en tant que tels, ne possèdent pas un contenu significatif : ce sont des entités internes à la phrase documentaire, mais elles ne portent pas sur une autre entité.

5.1.1 Le contenu significatif du noyau à coordination interne

Le contenu significatif du noyau à coordination interne résulte de la mise en rapport des contenus significatifs des items qui le composent, reliés par la particule *et*. Cette mise en rapport s'effectue comme opération préalable à celle de la composition ; pour effectuer celle-ci, le contenu significatif d'un noyau à coordination interne est traité comme le contenu significatif d'un noyau simple. Le contenu significatif du noyau à coordination interne exprime une entité complexe, formée par les entités exprimées par les items individuels, entités qui sont considérées ensemble, comme formant un tout, ou dans leur interdépendance, et non comme une succession ou juxtaposition d'individualités indépendantes. Le rapport qui, au niveau du noyau à coordination interne, réunit en une entité composite le contenu significatif des items individuels correspond en général, en langue naturelle, au signifié « avec ». C'est, en effet, le signifié de la conjonction *et* qui apparaît dans *Donnez-moi un whisky et soda*, ou *Jean et Marie se sont mariés*, ou *Pierre et Jacques sont venus* (à condition qu'ils soient venus ensemble et/ou que la venue de l'un soit étroitement liée à la venue de l'autre) ; ce n'est pas, en revanche, le *et* qui apparaît dans *Molière a écrit L'Avare, Tartuffe et Les Femmes savantes*, ou dans *Louis XIV a gouverné en France et Cromwell en Angleterre* (sur la façon de rendre cette différence entre les deux sens de *et* dans V1, voir le § 5.3).

Le rôle de la particule *et* est de permettre la création, au moment même de l'indexation et pour répondre aux besoins de l'indexeur, de cette entité nouvelle de caractère complexe qui n'a pas été déterminée a priori dans le lexique de V1 et que représente le noyau à coordination interne. Alors que les notions du lexique de V1 et les items qu'elles forment avec les indicateurs de procédé ont un contenu significatif déterminé, pour lequel tous les uti-

lisateurs du système sont censés avoir la même définition, les entités nouvelles créées à l'aide de la particule *et* n'ont pas un contenu significatif contrôlé et défini à priori, mais sont interprétées avec une certaine marge d'appréciation par chaque utilisateur, à partir du contenu significatif des termes joints. Cette *mise en rapport* du contenu significatif de chacun des constituants pour construire le contenu significatif du noyau à coordination interne diffère par ses effets de la *composition* déterminée par la relation de projection. Dans ce dernier cas, les déterminations à droite d'une entité donnée apportent une restriction au contenu significatif de cette entité ; cette restriction est exprimée de manière explicite : *bibliothèque t0 municipale a1* désigne une bibliothèque qui possède tout le contenu significatif apporté par l'item *A municipale a1*. Au contraire, un noyau à coordination interne ne se définit pas complètement par la relation que le langage établit entre le contenu significatif de chacun des constituants de ce noyau, mais ces contenus significatifs interfèrent d'une manière non prévisible par règle générale, qui dépend des relations qui peuvent concrètement s'établir dans le monde référentiel entre les entités représentées. On comparera avec ce qui se passe en langue naturelle lorsque l'on crée un nom composé tel que *location-vente* : le mot ne désigne pas la simple addition d'une location et d'une vente, mais un contrat original qui possède *certaines* caractéristiques de la location et *certaines* caractéristiques de la vente. Le noyau à coordination interne permet donc d'introduire dans V1, à côté du lexique défini à priori, un facteur de souplesse et de créativité facilitant l'adaptation à des situations nouvelles.

Le contenu significatif du noyau avec coordination interne explique les règles assez strictes de sa combinatoire ; on se souviendra, en effet (cf. § 3.4), que le noyau avec coordination interne possède exactement le même comportement que le noyau simple. Or, ceci n'est possible que par le caractère « embloqueur » de la particule *et* : les contenus significatifs individuels forment un bloc compact au niveau du noyau à coordination interne, ce qui permet de le traiter comme un noyau simple.

L'utilisation de l'anaphorique permet cependant de créer de manière indirecte des noyaux à coordination interne dans lesquels un item reçoit une détermination (une entité A ou un complément verbal) qui lui est propre et qui ne porte pas sur le noyau dans son ensemble. Exemple :

(1) 1.1 Traduction t1 automatique a1. 2 Cette traduction t1 et cette indexation t1 de documents t0 techniques a1.

(2) 1.1 Traduction t1 automatique a1 de résumés t0. 2 Cette traduction t1 et cette indexation t1 dans les centres de documentation a3 b en France a1.

Cette possibilité peut jouer parallèlement pour plusieurs items dans un même noyau à coordination interne :

1.1 Traduction t1 automatique a1. 2 Indexation t1 assistée a4. 3 Cette traduction t1 et cette indexation t1 dans les centres de documentation a3 b en France a1.

Il faut noter toutefois que l'obligation de ne joindre que des items TV, lorsqu'un syntagme de phrase ou un bloc de syntagmes de phrase porte sur le noyau à coordination interne, ne peut être tournée par le recours à l'anaphorique.

5.1.2 La relation de projection

Lorsqu'une entité A porte sur une entité B, la *relation de projection* s'établit entre l'entité A et l'entité B. La relation de projection remplit une fonction générale, la même partout où elle est utilisée : c'est celle de former un contenu composite à partir des deux entités qu'elle met en rapport. En effet, la formule *A porte sur B* indique que le contenu significatif de A doit former avec le contenu significatif de B un autre contenu significatif, composite, qui résulte de : (1) le contenu significatif de A, (2) celui de B, (3) le type d'entité (syntagme de phrase, syntagme A, item A, etc.) que sont A et B. Il est clair que la relation de projection ne s'établit qu'entre entités à contenu significatif (cf. § 5.1).

On sait que la relation de projection remplit son rôle de composition soit à l'intérieur du syntagme de phrase, soit à l'extérieur. A l'intérieur du syntagme de phrase, la relation de projection s'établit *entre* item A, syntagme A ou bloc A *et* noyau A ou noyau de

- (3') [Diagramme] annexe à un document.
 noyau complément
 adjectival' déterminatif
- (4') [Documents] traduits par les documentalistes d'un centre de documentation.
 noyau complément déterminatif

(II) *La relation d'attribut, lorsqu'elle s'établit entre un item A, ou un syntagme A, ou un bloc A, et un noyau de syntagme de phrase, correspond à celle qui existe en langue naturelle entre un complément déterminatif et un noyau nominal dans un syntagme nominal.* Exemples :

Les structures grammaticales de V1 suivantes :

- (1) Traduction t1 automatique a1.
 noyau item A
- (2) Périodiques t0 traduits a1 b par la bibliothèque a1.
 noyau syntagme A
- (3) Bibliothèque t0 en Allemagne a1.
 noyau item A

correspondent respectivement aux structures suivantes en langue naturelle :

- (1') Traduction automatique.
 noyau complément
 nominal déterminatif
- (2') Périodiques traduits par la bibliothèque.
 noyau complément déterminatif
 nominal
- (3') Bibliothèque en Allemagne.
 noyau complément
 nominal déterminatif

Les principes (I) et (II) ci-dessus sont soumis à la restriction suivante : *la relation d'attribut ne peut jamais correspondre à une structure LN où le complément déterminatif est paraphrasable par un complément d'objet direct ou par un complément qui correspond en V1 à un complément verbal.*

On sait, en effet, d'après les exemples de complément déterminatif d'un syntagme nominal présentés au § 1.1.2, que la relation du complément déterminatif avec le noyau nominal est susceptible, en langue naturelle, d'être interprétée de plusieurs manières : complément de lieu (*bibliothèque en Allemagne*), de matière (*couteau en acier*), d'appartenance (*livre de Pierre*), de qualification (*fleurs agréables*), etc. Parmi ces possibilités d'interprétation, existe celle d'attribuer au complément déterminatif la fonction de complément d'objet direct ou d'objet indirect ou de régime verbal, lorsque le noyau nominal est un nom verbal. Exemples :

Les structures LN suivantes :

- (1) La traduction des documents
 (2) L'indexation des textes
 (3) L'accès à la bibliothèque

sont susceptibles d'être respectivement paraphrasées par les structures LN :

- (1') Traduire des documents.
 On a traduit *ou* on traduit des documents.
- (2') Indexer des textes.
 On a indexé *ou* on indexe des textes.
- (3') Accéder à la bibliothèque.
 On a accédé *ou* on accède à la bibliothèque.

où *des documents* et *des textes* sont des compléments déterminatifs paraphrasables par un complément d'objet direct et *à la bibliothèque* est un complément déterminatif paraphrasable par un complément de régime verbal qui doit s'exprimer en V1, étant donné l'entrée lexicale *accéder*, par un complément verbal. Or, d'après la restriction ci-dessus, capitale

pour le fonctionnement de V1, ces compléments déterminatifs en langue naturelle ne peuvent pas s'exprimer en V1 par la relation d'attribut et, par conséquent, les structures grammaticales suivantes ne sont nullement admises par la grammaire de V1 :

- (1'') Traduction t1 des documents a1 *
- (2'') Indexation t1 des textes a1 *
- (3'') Accès t1 à la bibliothèque a1 *

Il faut écrire :

- (1''') Traduction t1 des documents t0.
- (2''') Indexation t1 des textes t0.
- (3''') Accès t1 à la bibliothèque t0.

La conséquence de cette interdiction apparaît plus particulièrement quand on veut employer un noyau A qui correspond en langue naturelle à un nom verbal. Si ce nom verbal a lui-même un complément d'objet direct ou un complément d'objet indirect ou de régime verbal exprimable en V1 par un complément verbal, ce dernier ne peut s'exprimer sous la forme d'un item A, placé à la suite de *b* et portant sur le noyau A. Pour tourner cette difficulté, il faut avoir recours à l'anaphorique. En effet, si l'on veut exprimer en V1 les expressions LN suivantes :

Historique de la traduction des documents
Gratuité de l'accès aux banques de données

il faudra respectivement écrire :

- 1.1 Traduction t1 des documents t0. 2 Historique t0 de cette traduction a1.
- 1.1 Accès t1 aux banques de données t0. 2 Gratuité t0 de cet accès a1.

5.1.2.1.2 La relation d'attribut avec spécificateur A et avec une seule entité portant sur le noyau

On sait que dans l'articulation d'un item A on peut introduire un spécificateur A (cf. § 3.3.2), la fonction de celui-ci étant de préciser la relation d'attribut. Le spécificateur A possède un signifié associé à un signifiant. On sait aussi (cf. § 3.3.1) que tous les items A d'une suite A sont des compléments A du noyau, noyau A ou de syntagme de phrase, qui précède la suite et que, par conséquent, le premier item A, à gauche, d'un syntagme A ou d'un bloc A est complément A du noyau sur lequel porte le syntagme A ou le bloc A auquel il appartient.

Le contenu significatif déterminé par la relation d'attribut avec un spécificateur A est presque entièrement analogue à celui qui est déterminé par la relation d'attribut sans spécificateur A. Les principes I et II ainsi que la restriction qui les a suivis, présentés au § 5.1.2.1.1, restent en effet valables avec l'adjonction de la précision suivante : « lorsque cette relation en langue naturelle est paraphrasable par le signifié du spécificateur A ». Ces principes ainsi complétés donnent les principes III et IV ci-dessous :

(III) *La relation d'attribut, lorsqu'elle s'établit entre un item A précédé d'un spécificateur A, ou un syntagme A ou un bloc A dont le complément A est précédé d'un spécificateur A, et un noyau A, correspond à celle qui existe en langue naturelle soit entre un complément déterminatif et un noyau nominal, soit entre un complément déterminatif et un noyau adjectival, lorsque cette relation en langue naturelle est paraphrasable par le signifié du spécificateur A.*

(IV) *La relation d'attribut, lorsqu'elle s'établit entre un item A précédé d'un spécificateur A, ou un syntagme A ou un bloc A dont le complément A est précédé d'un spécificateur A, et un noyau de syntagme de phrase, correspond à celle qui existe en langue naturelle entre un complément déterminatif et un noyau nominal dans un syntagme nominal.*

V1 incorpore, dans sa version actuelle, deux spécificateurs A. Ce sont :

- | | Signifié : | Signifiants : |
|-----|--------------|--|
| (1) | « domaine » | spécialisé(e)(s) en, spécialisé(e)(s) dans, portant sur, sur |
| (2) | « fonction » | pour, à |

La contrainte d'utilisation du *spécificateur « domaine »* est la suivante : *l'item A qui est précédé par ce spécificateur est complément A d'un noyau qui porte le trait sémantique [+ document] et/ou le trait sémantique [+ domaine]*. (Des notions telles que *bibliothèque, centre de documentation, recherche scientifique*, possèdent le trait sémantique [+ domaine]). L'item A précédé du spécificateur « domaine » exprime le domaine ou champ qui est traité par l'objet exprimé par le noyau ou sur lequel l'objet exprimé par le noyau conserve de l'information. Exemples :

Bibliothèque t0 spécialisée en s médecine a1
Thèse t0 sur s les microfilms a1

La contrainte d'utilisation du *spécificateur « fonction »* est la suivante : *l'item A qui est précédé par ce spécificateur est un item a1 ou + a1 dont la notion correspondante admet un t1*. Cet item a1 précédé du spécificateur « fonction » indique la fonction à laquelle est destiné l'objet exprimé par le noyau dont il est complément A; en d'autres termes, l'item a1 ou + a1 précise à quoi sert l'objet exprimé par le noyau. Exemples :

Langage t0 pour s l'indexation a1
Machine t0 à s traduire à1
Étagères t0 pour s le stockage a1

On peut comprendre, à la suite de ces exemples, dans quel sens l'utilisation d'un spécificateur A doit correspondre à la relation établie en langue naturelle entre un complément déterminatif et son sujet ou son noyau nominal ou adjectival, lorsque cette relation est paraphrasable par le contenu du spécificateur A. En effet, dans les structures LN suivantes :

Thèse sur les microfilms
Langage pour l'indexation

sur les microfilms et *pour l'indexation* sont bien des compléments déterminatifs de, respectivement, *thèse* et *langage*, mais, restant toujours dans le cadre de cette relation, *sur les microfilms* indique plus précisément le champ sur lequel porte *thèse* et *pour l'indexation*, à quoi sert le *langage*. C'est dans ce sens que la relation de *sur les microfilms* avec *thèse* est paraphrasable par le signifié « domaine » et la relation de *pour l'indexation* avec *langage* est paraphrasable par le signifié « fonction ».

5.1.2.1.3 La relation d'attribut avec ou sans spécificateur A et avec plusieurs entités portant sur le noyau

On sait (cf. § 3.3.1) que plusieurs entités A peuvent porter sur un même noyau. En effet, plusieurs items A ou plusieurs items A et un syntagme A ou plusieurs items A et un bloc A peuvent porter sur un même noyau, noyau A ou noyau de syntagme de phrase. Chacun des items A qui portent sur le noyau, et l'item A qui, dans le syntagme A ou le bloc A, est complément A du noyau, peuvent ou non être précédés d'un spécificateur A. Le principe général qui règle ces cas est le suivant :

(V) *Lorsque plusieurs entités A portent sur un même noyau, noyau A ou noyau de syntagme de phrase, chacune parmi elles établit :*

(a) *avec ce noyau, la même relation qui se serait établie si cette entité était la seule à porter sur le noyau ;*

(b) *avec les autres entités A qui portent sur le même noyau, une relation qui correspond, en langue naturelle, à celle de la juxtaposition.*

Exemples : les structures grammaticales suivantes :

- (1) Traduction t1 automatique a1 en anglais a1.
- (2) Bibliothèque t0 publique a1 en Inde a1.
- (3) Document t0 sur s l'agriculture a1 d'une bibliothèque a1 b publique a1 en Inde a1.

correspondent respectivement aux structures LN suivantes :

- (1') Traduction, automatique, en anglais
- (2') Bibliothèque, publique, en Inde
- (3') Document, sur l'agriculture, d'une bibliothèque, publique, en Inde

Grâce à la relation de juxtaposition qui existe entre les entités qui portent sur un même noyau, une modification de l'ordre dans lequel apparaissent les items A portant

5.1.2.2 La relation verbale

On sait (cf. § 5.1.2) que la relation verbale s'établit entre syntagme de phrase et syntagme de phrase ou entre syntagme de phrase et bloc de syntagmes de phrase; elle est par conséquent une relation extérieure au syntagme de phrase. Le principe VI, qui règle la caractérisation du contenu significatif déterminé par la relation verbale, reçoit deux formulations (VI a) et (VI b), la première correspondant au cas le plus général.

(VI a) *La relation verbale correspond, dans le cas le plus général, à celle qui existe en langue naturelle entre un complément d'objet direct et un infinitif, ou entre un complément déterminatif, paraphrasable par un complément d'objet direct, et son noyau nominal.*

Exemples : les structures grammaticales suivantes :

- (1) Indexation t1 automatique a1 des textes t0 techniques a1.
- (2) Reproduction t1 par microfiche a1 des documents t0 anciens a1.
- (3) Gestion t1 de l'indexation t1 automatique a1 des textes t0 techniques a1.

correspondent, respectivement, aux structures LN qui suivent :

- (1') Indexer automatiquement des textes techniques
Indexation automatique des textes techniques
- (2') Reproduire par microfiche des documents anciens
Reproduction par microfiche des documents anciens
- (3') Gérer l'indexation automatique des textes techniques
Gestion de l'indexation automatique des textes techniques

(VI b) *Dans certains cas, relativement peu fréquents, la relation verbale correspond à celle qui existe en langue naturelle entre le complément d'objet indirect ou le complément de régime verbal et un infinitif, ou entre un complément déterminatif, paraphrasable par un complément d'objet indirect ou par un complément de régime verbal, et son noyau nominal.*

Exemples : les structures grammaticales suivantes :

- (1) L'abonnement t1 d'une bibliothèque a1 à une revue t0.
- (2) L'accès t1 à une salle t0 de lecture a1.

correspondent respectivement aux structures LN qui suivent :

- (1') L'abonnement d'une bibliothèque à une revue
- (2') Accéder à une salle de lecture
L'accès à une salle de lecture

Le choix de la relation verbale exprimant soit un objet direct (cas général), soit un complément d'objet indirect ou de régime verbal (cas peu fréquent) n'est pas laissé au documentaliste : c'est le lexique de V1 (cf. § 2.2) qui le fixe pour toutes les entrées lexicales qui admettent un t1. Si l'entrée appartient aux cas particuliers (la relation verbale exprimant un objet indirect ou de régime verbal), le lexique fixe également la préposition qui doit être obligatoirement écrite par le documentaliste ou les prépositions parmi lesquelles il peut choisir.

5.2 Le contenu significatif de la construction documentaire

La relation de projection agit aussi dans le cadre de la construction documentaire, et de manière fort simple : *le contenu significatif de la phrase documentaire de droite porte sur le contenu significatif de la phrase documentaire de gauche*. Il est à remarquer que c'est le contenu significatif de la phrase documentaire tout entière qui porte sur le contenu significatif de l'autre phrase documentaire tout entière. Ceci implique qu'une construction documentaire comme :

Adaptation t1 du recrutement t1 de documentalistes t0 spécialisés a1 , avec l'objectif d'automatiser t1 un centre de documentation t3.

ne peut jamais correspondre à une structure LN telle que :

Recrutement de documentalistes spécialisés avec l'objectif d'automatiser un centre de documentation ; adaptation de ce recrutement.

mais doit correspondre à une structure LN telle que :

Adaptation du recrutement de documentalistes spécialisés ; cette adaptation se fait avec l'objectif d'automatiser un centre de documentation.

Chacun des contenus significatifs des phrases documentaires constitutives d'une construction documentaire est identique au contenu significatif de cette phrase documentaire considérée isolément : il est en effet construit en fonction des mêmes principes que ceux qui règlent la composition des contenus significatifs des phrases documentaires isolées.

Le contenu significatif de la construction documentaire dans son ensemble correspond à une structure LN où le contenu d'une sous-structure LN correspondant au contenu d'une phrase documentaire est en relation avec le contenu de l'autre sous-structure LN correspondant au contenu de l'autre phrase documentaire, et cette relation est paraphrasable par le signifié du prédicat de la construction documentaire.

La version actuelle de V1 incorpore les prédicats suivants :

signifié	signifiant
« objectif »	avec l'objectif d(e)
« moyen »	au moyen d(e)
« relation »	en relation avec
« coopération »	en coopération avec
« influence »	influençant, provoquant, entraînant