

Marie KARAMANOU - GIROUD
UNIVERSITE LUMIERE - LYON 2
INSTITUT DE LA COMMUNICATION

ANNEE 1989 - 1990

MEMOIRE DE DEA

INVESTIGATIONS CONCERNANT LES
METHODES D'INTERROGATION D'UN SERVICE
TELEMATIQUE

Directeur : M. le professeur M. LE-GUERN

SOMMAIRE

Chapitre 1 Introduction	page	1
Chapitre 2 Présentation de VIVA	page	4
Comment se déroule la consultation de VIVA		8
Utiliser un autre mode d'accès à l'information : pourquoi ?		10
Chapitre 3 Les Interrogations	page	13
La première interrogation		13
Deuxième expérience		14
Troisième interrogation		15
Quatrième interrogation		18
Cinquième interrogation		20
Chapitre 4 La grammaire	page	22
Aspect Théorique		22
La morphologie		24
La notion de profil		27
La composante morpho-syntaxique		28
Quelle grammaire ?		29
Chapitre 5 Les propositions	page	34
Les synonymes		38
Les données numériques		41
La suite ...		43
Chapitre 6 Pertinence de l'approche par SN pour la réquisition de l'information textuelle.	page	44
Conclusion	page	47

BIBLIOGRAPHIE

ANNEXES



CHAPITRE 1 INTRODUCTION

Le but de notre travail est le suivant: étant donné un service télématique "grand public", en l'occurrence VIVA 3, comment peut-on optimiser l'interrogation de ce service.

La linguistique a permis de préciser que le vecteur informatif dans une phrase est le syntagme nominal. Il paraît donc approprié d'utiliser le concept descriptif de **syntagme nominal (SN)** pour déterminer le contenu d'une "phrase".

Nous chercherons à montrer comment on peut utiliser le concept de SN pour constituer un outil d'aide à l'interrogation de bases de données textuelles. Nous pensons en outre que certains procédés documentaires utilisés pour des bases de données textuelles, peuvent, sous certaines conditions de réaménagement être également utilisées dans une perspective d'optimisation d'une base de données télématique.

Notre travail comportera les parties suivantes :

- Dans la première partie nous présentons le service VIVA3, tel qu'il existait jusqu' en septembre 90 et plus particulièrement la méthode antérieure d'interrogation. Nous disons "antérieure", parce que VIVA vient de quitter le centre serveur (société G.F.I) qui l'hébergeait jusqu'à alors. Il est probable donc que ce "déménagement" provoque des modifications dans la présentation du journal.
- Dans la deuxième partie, nous présentons le service VIVA3, les rubriques, le type d'informations incluses ainsi que le procédé actuel d'interrogation
- Dans la troisième partie, nous exposons les expériences les plus importantes pour notre travail. Elles sont présentées par ordre de difficulté croissante, permettant ainsi de se rendre compte des types d'interrogation qu'on peut effectuer sur VIVA.
- Dans la quatrième partie, et très brièvement, nous exposons les concepts de base du groupe SYDO et plus particulièrement la grammaire et les procédures élaborées par le groupe dans ces travaux, qui ont comme but l'extraction des SN et leur structuration en base de données utilisable "automatiquement"

- Dans la cinquième partie, nous exposons l'utilisation "pratique" qui peut être faite des SN pour élaborer une méthode d'interrogation à la fois plus efficace et plus maniable pour l'utilisateur. Nous exposons nos remarques par ordre chronologique. En effet et suite au décès du maire de Villeurbanne en décembre dernier, le service a changé un peu. Ces changements nous ont obligés à revoir les propositions que nous avons faites antérieurement sur la construction d'un système d'interrogation. Notre dernière proposition s'aligne à une idée bien antérieure émise par M. LeGuern

Le principe de ce système est une base constituée des SN du texte (d'une édition de VIVA). Cette base peut être en une seule partie, ou en plusieurs parties, joints ou non à des fichiers contenant des données particulières, mais le traitement des données quelles qu'elles soient serait fait à l'aide de la grammaire du groupe SYDO. Il n'y aurait pas de traitement particulier donc pour les données de type "âge" ou "date", comme on l'avait pensé auparavant.

Nous relevons par la même, quelques problèmes pratiques soulevés par l'usage d'un tel système et suggérons quelques précautions ou solutions à prendre en considération. Nous remarquerons enfin, que nos propositions ne peuvent se réaliser qu'à conditions de respecter un certain nombre de consignes superficielles, sans lesquelles aucun système automatique se perfectionné qu'il soit, ne peut fonctionner.

- Dans la sixième partie, nous essayons de démontrer l'intérêt "psychologique" que peut avoir un système d'interrogation par SN. Nous essaierons de démontrer l'affinité qu'il peut y avoir entre un procédé télématique par SN et le procédé humain d'interrogation (télématique ou pas). Il s'agit en effet de montrer qu'outre l'efficacité et rentabilité informatives d'une méthode d'interrogation basée sur les SN, il existe un lien entre celle-ci et l'interrogateur humain qui la rend plus facilement accessible et utilisable par lui. D'où, une raison supplémentaire pour l'adopter...

Nous n'avons pas l'intention d'établir une liste exhaustive des solutions possibles d'optimisation d'un service télématique. Loin de là! Mais nous pensons que notre travail, même s'il reste "sur le papier", permet de se rendre compte que des méthodes différentes de celles utilisées habituellement par les services télématiques, peuvent être

mises en place et rendre l'utilisation du clavier minitel plus attrayante pour l'interrogateur. Il est évident que ce type de travail ne concerne pas les services conviviaux.

Notre travail n'est pas non plus un travail sur une situation de communication. Nous pensons en effet que dans le cas de services tels que VIVA, la communication se passe à un niveau tellement latent, qu'il n'est guère possible d'utiliser ce terme pour désigner la consultation du service. Justement donc, notre travail est un ensemble cohérent de propositions qui permettraient de transformer une situation informative (au sens strict du terme) en une situation de communication, même embryonnaire. Dans cette perspective, notre travail concerne le fameux "destinataire" (ici minitel), à qui nous essaions de donner la possibilité de "répondre" selon les questions qu'on lui pose. La situation d'interrogation est réductrice et n'évitera, en pratique, jamais un certain "bruit" (réponses parasites); mais il est vrai que notre "destinataire" n'est pas le partenaire idéal d'une conversation ordinaire. Il peut malgré tout devenir le partenaire indispensable d'une interrogation "intelligente" et moderne. La société informatique qui sera suffisamment clairvoyante pour investir et adopter un système d'interrogation par SN, bénéficiera d'une technique informative qui se distingue des techniques ordinaires de services télématiques indifferents et indifférentiables. Nous sommes persuadés que le public des utilisateurs ne saura tarder de cautionner une telle démarche.

CHAPITRE 2 PRESENTATION DE VIVA

VIVA est le journal télématique de la mairie de Villeurbanne. Il est attaché au service OVIDE (Office Villeurbannais d'Information et d'Expression) de la mairie. VIVA a commencé à paraître sur les écrans du minitel en octobre 1982. Sa gestion informatique a été assurée jusqu'à présent par la société G.F.I.

Le but de ce journal municipal est de combler le vide créé par l'absence d'un journal, purement villeurbannais; de servir d'intermédiaire entre la mairie, les associations et les citoyens; d'établir un réseau de contacts entre ces derniers (notamment par les sous-parties HELP et STOP), sans pour autant devenir une messagerie, orientation que prennent souvent les service "de contacts".

Le responsable de VIVA, M. Jacques Galliot, nous a par ailleurs affirmé à plusieurs reprises, qu'un service du genre "messagerie-grand-public" ne l'intéressait pas. Le but du journal est, selon M. Galliot, de rendre un "service". Cette notion est primordiale dans la conception de VIVA et dans les options prises par la suite.

La forme originale du journal a subi plusieurs modifications, au cours des dernières années. Ainsi, des rubriques ont "disparu", au profit d'autres jugées plus importantes pour les villeurbannais et autres utilisateurs du journal. Par exemple, la rubrique ASTRO a été remplacée par la rubrique BENNE.

La structure du journal est relativement simple. Les rubriques sont présentées à la page du sommaire (voire Tableau ci-dessous). La structure des rubriques est arborescente et des "passages" sont prévus de certaines rubriques ou parties de rubriques à d'autres. Il est donc facile de connaître la structure du journal.

Tableau 1 SOMMAIRE DE VIVA

Villeurbanne PUISSANCE 3	
	3 URBA
1 ACTUA	
	30 Petites Annonces
10 FLASH	31 MESSages
Revue de Presse	32 Boites A Lettres
11 AGENDA / ville	
12 SORTir LOISir	Questions / réponses
	33 EXPRESS / mairie
13 ETAT civil	34 SANTE / médecin
14 GARDE médicale	
15 BIBLIobus	35 TNP
16 BENNE	36 MAISON du livre
	37 CABLE
2 VILLE	
infos pratiques	38 ABonnez vous
adresses utiles	39 JEUX

n° ou mot- clé :	+ ENVOI

REMARQUE : A l'heure actuelle, suite au changement de centre serveur, nous ne savons pas quelle forme va prendre le contenu du journal. Ainsi, exposerons-nous les données recueillies antérieurement à ce changement.

Par ailleurs, le contenu informatif du journal étant relativement simple et bien structuré, l'utilisateur apprend rapidement à chercher l'information désirée dans la "bonne" rubrique. Il y a peu de chances de se tromper de rubrique.

Le menu du journal est composé de trois parties :

- la partie FLASH, qui contient les... flash d'information, l'état civil, des renseignements sur l'activité culturelle de la ville (cinémas, expositions, concerts,...), les possibilités de stages, etc...
- La deuxième partie, VILLE, contient des informations relatives aux activités et aux décisions prises par la municipalité, ainsi que des indications sur les services et les équipements sociaux (crèches, écoles, services de santé), ...

Un petit problème se pose en ce qui concerne cette deuxième partie. En effet, elle apparaît seulement avec son résumé dans le sommaire et son contenu est complètement obscur pour l'utilisateur (voir tableau 1, ci-dessus). En outre, VIVA a pris la précaution de répéter la partie concernant les services de santé, dans la troisième partie, URBA. Cette petite redondance ne pourrait que faciliter la tâche de personnes en quête de renseignements sur les services médicaux.

- La troisième partie, URBA, comporte des possibilités de contacts, de rédaction de petites annonces, des informations concernant la saison du TNP, la boîte-à-lettre.

En résumé, VIVA est un journal qui est (était?) relativement bien organisé. Le taux de consultation en 1988 est de 2000 à 2500 appels par mois. Fin 1989, le nombre moyen de ces appels était de 2700 ap/mois.

Il y a deux "versions" ("éditions") du journal: une édition en 3615, ouverte à tous (les minitelistes) et une autre version en 3614, réservée aux abonnés du journal.

REMARQUE: Les deux éditions du journal ne sont pas tout à fait identiques : dans l'édition en 3614, on peut trouver des rubriques (questionnaires, informations pour les abonnés), qui ne figurent pas dans la version en 3615. Les informations générales, par contre, sont les mêmes.

Les Rubriques : Les rubriques de VIVA sont, comme nous l'avons déjà signalé, simples. Leur dénomination correspond en général aux informations incluses. Ainsi, nous n'allons pas les présenter. Par contre, il serait utile de donner quelques indications concernant les interrogations relatives à quelques rubriques :

- En août 1988, la rubrique des petites annonces a obtenu le plus grand taux d'interrogations avec 14,5% . La boîte-à-Lettres a eu 13,03% des consultations.
- En juillet 1988, la rubrique SORTIR avait obtenu 37,48% des interrogations. Au sein de cette rubrique, 16,74% des appels concernaient la partie MUSIQUE (les concerts en tout genre).
- En 1989, la moyenne des appels est passée de 2500 appels/mois à 3500 appels/mois.
- Jusqu'à présent, la rubrique FLASH totalisait entre 8 et 9% du nombre total d'appels; les JEUX totalisaient entre 10 et 12% des appels. En tête, arrivaient les rubriques PETITES ANNONCES et BOITE A LETTRES avec 20% environ des appels (nous ne savons pas laquelle des rubriques contribuait le plus à l'obtention de ce taux); la rubrique SORTIR-LOISIRS avait entre 14 -15% des appels par mois.

En 1990, les chiffres ont peu changé. Il est possible (à nos yeux, certain) que les statistiques de VIVA soient erronées : le pourcentage ne tenait compte que de la rubrique **consultée en premier**, après la connexion au journal. Les statistiques ne tenaient pas compte des sauts que l'on pouvait faire d'une rubrique à l'autre. D'après M. Galliot, le changement de centre serveur permettra de connaître plus précisément le taux de consultation de chacune des rubriques.

Par ailleurs, M. Galliot nous avait confirmé dans le passé, la centralisation de la rédaction de VIVA. Les Informations parues sont rédigées par l'équipe de VIVA uniquement (sauf la rubrique DELIBERATIONS DU CONSEIL). La saisie se faisait directement sur l'écran minitel (saisie en 40 colonnes) et il n'y avait pas de correction du texte rédigé (ce qui peut expliquer les quelques fautes de syntaxe et d'orthographe).

FLASH

L'info puissance 3
Votre revue de presse locale
en 3 minutes.

1 ROUMANIE: VILLAGE ADOPTE	04/08
2 LA BATAILLE DE STALINGRAD	04/08
3 LE VIRUS DE LA DECOUVERTE	02/08
4 POMPIERS POUR LE MIDI	01/08
5 TOLSTOI: PLANCHER S'EFFONDRE	01/08
6 NOUVELLE ANNEXE DE LA MAIRIE	01/08
7 CARAMBOLAGE POUR LE MAIRE	30/07
8 DEVOIR DE VACANCES	29/07

N° choisi: ENVOI ou tapez Suite
 Sommaire "VIVA" Tapez Sommaire

COMMENT SE DERoule LA CONSULTATION DE VIVA

RAPPEL : Les informations données ici et dans la suite de notre travail, sont antérieures au changement de centre serveur de VIVA. La reprise du journal aura, en principe, lieu le 1 octobre 1990.

L'utilisateur se connecte à VIVA par le 3615 ou 3614, les codes respectifs sont VIVA3 et VIVA2. Après un graphique (très souvent, une image des Gratte-Ciel), on voit s'afficher à l'écran le sommaire du journal (Tableau N° 1). L'utilisateur choisit une rubrique censée l'intéresser. Pour cela, il suffit de taper les deux chiffres relatifs à la rubrique. Celle-ci est sélectionnée et son sommaire s'affiche automatiquement, sous la forme d'une série de titres associés à un chiffre et datés à la date de création de

l'information par VIVA. Les titres dont la page-écran est vide sont marqués NEANT dans le sommaire. Exemple de titres: le sommaire des titres du FLASH (tableau 2, ci-dessus). L'utilisateur tape le numéro qui l'intéresse et la page correspondante, s'affiche. Il lui est possible de consulter d'autres pages d'information à l'aide des fonctions présentées dans la partie inférieure de l'écran. En général, il s'agit des fonctions SUITE, RETOUR ou SOMMAIRE. Certaines rubriques présentent l'avantage d'avoir des zones de saisie où il est possible de taper les caractères spécifiques d'une autre rubrique (rubriques et caractères spécifiques apparaissent en bas de l'écran) et d'y accéder sans passer par le sommaire du journal.

Cette organisation arborescente est, avons-nous dit, simple et efficace à conditions que les titres des articles contenus dans les rubriques soient explicites. Nous faisons cette remarque, car, à plusieurs reprises, nous avons constaté que VIVA formulait les titres des articles de la rubrique FLASH d'une façon très peu explicite, de sorte que les titres n'avaient pas de rapport manifeste avec le contenu de l'article qui leur était joint. Il nous est arrivé souvent de consulter VIVA et d'être obligée, pour diverses raisons, de voir défiler des pages-écrans qui n'avaient rien à voir avec l'information que nous désirions obtenir.

Un autre problème est posé par la mauvaise syntaxe et les fautes de frappe. Il nous semble que la solution de ce problème n'est pourtant pas difficile. Un peu plus d'attention suffirait largement. D'autre part, le responsable de VIVA nous avait précisé qu'il voulait faire de ce service un "vrai service grand-public". Dans cette perspective, il a dû opter pour un style d'expression simple, "oral" pourrions-nous dire en fin de compte. Nous pensons qu'une attention supérieure portée à la syntaxe et à l'usage de la ponctuation, ne nuirait pas à la vocation du journal. Bien au contraire, cette attention serait appréciée par le public et influencerait positivement l'image de marque de VIVA. Nous pensons que ce service aurait un bénéfice bien supérieur à utiliser une syntaxe et une grammaire un peu plus régulières, ce qui par ailleurs, ne l'empêcherait pas d'exercer son style habituel.

UTILISER UN AUTRE MODE D'ACCES A L'INFORMATION : POURQUOI ?

Il se peut qu'à première vue, l'on croit inutile de changer de mode d'accès à l'information chez VIVA : l'utilisateur a à sa disposition un service qui est bien structuré, dont le contenu est clair et simple, relativement facile d'accès. Pourquoi alors proposer des changements?

Il faut tout de suite remarquer, et répéter, que le contenu d'une page de VIVA n'est facile d'accès que lorsque le titre de l'article est explicite, ce qui n'est pas toujours vrai. Dans le cas contraire, comme le montreront les expériences, l'utilisateur peut consacrer beaucoup de temps à la recherche de son information. Qui serait capable de deviner que sous le titre LE MAIRE AU NOM DES PERES se cachait un article concernant des mesures de prévention de la délinquance ? D'ailleurs, il n'appartenait pas au commun des mortels de détecter sous le titre LES NOURRITURES TERRESTRES, un article relatif à un nouveau restaurant, l'énoncé pouvant correspondre au titre d'un roman de Gide. Il nous semble que la clientèle de VIVA, composée dans sa grande majorité de personnes des "classes moyennes", avec un niveau d'instruction BAC, aurait du mal à faire la correspondance entre "restaurant", terme usuel et peu ambigu, et l'expression "LES NOURRITURES TERRESTRES". De même qu'un lien indirect, établi à partir du sémantisme du terme "nourriture", s'il s'avérait exact ne resterait pas moins hasardeux.

Un autre problème qui demanderait l'intervention d'un autre mode d'interrogation est le suivant : il est possible que l'information désirée soit ancienne et, par conséquent, enregistrée à une date antérieure. Pour accéder à cette page d'information, en supposant que le titre n'est pas ambigu, il faut faire défiler plusieurs pages-écran. Il nous semble donc que dans une perspective de modernisation et d'amélioration des performances de VIVA, quelques nouvelles propositions pourraient, et devraient, être étudiées sérieusement par les responsables du service. Remarquons par ailleurs qu'un service qui ne se "renouvelle" pas¹ "vieillira mal".

Il paraît donc utile, et ceci pour deux raisons, d'introduire quelques changements dans le mode d'accès à l'information.

¹ Par renouvellement, nous n'entendons pas les renouvellements inutiles d'aspect extérieur par des graphiques différents de ceux qui existent déjà. L'utilisateur n'est pas éternellement dupe et se détournera d'un service qui n'est pas réellement un.

Au cours de notre réflexion, nous avons trouvé trois solutions qu'il serait relativement aisé à mettre en oeuvre chez VIVA. Nous les exposerons en suivant l'ordre où nous les avons "trouvées". Il est possible que d'autres solutions existent, auxquelles nous n'avons pas pensé. En tout cas, nous proposons ce que notre expérience de VIVA nous suggère de plus facile et aisément acceptable par les utilisateurs; car il est certain qu'un changement radical du mode d'accès peut susciter des réactions, voire des résistances². En effet, les utilisateurs de VIVA sont, nous l'avons déjà dit, des gens "moyens", qui ne vont pas insister plus qu'il ne leur semble nécessaire pour obtenir l'information qu'ils désirent. En cas d'échec répété, ils ne consulteront plus VIVA. Il est nécessaire de tenir compte du facteur "utilisateurs" dans tout système d'information.

Nos propositions restent cependant relativement extérieures à la réalité, dans le sens où les expériences menées par des interrogateurs ne constituent pas un support statistique suffisant. Néanmoins, les expériences sont vraies et constituent des indications, des données à prendre en compte. Le profil des interrogateurs est loin du profil de l'utilisateur moyen de VIVA, tel que nous l'avons établi par des recherches statistiques, en 1988 : nos interrogateurs sont des enseignants en faculté et des étudiants. L'utilisateur moyen de VIVA est :

- un homme de 25 - 30 ans;
- habite à Villeurbanne depuis 1 à 5 ans;
- est célibataire;
- est employé ou cadre moyen;
- n a pas d'enfants;
- niveau d'études BAC.

Ces données sont issues d'une enquête faite par questionnaire, conçu par nous et diffusé par voie télématique en mai-juin 1988. Le questionnaire est passé en 3614 et nous n'avons eu que des abonnés de VIVA, qui n'étaient pas forcément des entreprises. Plusieurs personnes, des particuliers donc, avaient répondu de chez eux. Nous avons eu 104 réponses sur les 1200 abonnés du moment. Nous pensons par ailleurs qu'il y aurait eu peu de variations, en ce qui concerne les résultats de l'enquête, si celle-ci avait eu lieu en 1989.

Nous ne sommes pas en mesure d'apprécier la distorsion créée par la différence entre les interrogateurs et les utilisateurs habituels de VIVA. Il est certain que des enquêtes plus importantes devraient être

2 Résistances internes ou externes au service.

faites afin de déterminer quel type de système choisir³, comment ce système serait-il utilisé et quelles seraient les réactions des utilisateurs face à ce nouveau système. Les expériences que nous avons faites ont néanmoins l'avantage de "mettre le doigt" sur certains problèmes et d'y apporter les solutions les plus proches de la réalité télématique, telle que nous la percevons.

3 Si l'on devrait changer de système d'interrogation...

CHAPITRE 3 LES INTERROGATIONS

Remarque préalable : Le déroulement des expériences était le suivant: nous savions qu'une information était présente dans VIVA, mais que la rédaction de son titre dans le sommaire des titres du journal, ne facilitait pas l'accès à la page d'information correspondante. Ainsi, nous donnions à nos interrogateurs une consigne délibérément ambiguë, présentée une seule fois avant l'expérience et oralement. Ce procédé découlait d'une hypothèse préalable, à savoir que l'interrogateur, en général, a une idée vague de l'information qu'il désire obtenir.

(Si le sommaire des titres du journal ne lui donne pas les indications suffisantes pour accéder à l'information désirée, il est peu probable qu'il y parvienne dès la première tentative. L'interrogateur est donc obligé de parcourir plusieurs pages du journal, ce qui lui occasionne des dépenses. Il paraît évident que si ce type d'incident se produit plusieurs fois, cela éconduit définitivement l'interrogateur de VIVA)

Notre rôle, par la suite, consistait en l'enregistrement du parcours de l'interrogation.

LA PREMIERE INTERROGATION :

La première interrogation avec un utilisateur autre que nous, fut effectuée le 25 juillet 1989. L'interrogateur était un homme, qui sans être un amateur du minitel, connaissait VIVA. Nous lui avons donné la consigne de trouver la date de décès de MME X, que nous savions décédée quelques jours auparavant. Le choix de la consigne n'était pas hasardeux. En effet, nous savions que la structure de VIVA ne permettait pas des interrogations "vagues": le contenu du journal est structuré d'une manière "rigide": l'information couvre un champ restreint, à savoir, la vie municipale. Un utilisateur ne peut pas exiger du journal télématique des réponses de type documentaire (au sens large du terme). Il nous a semblé donc utile de commencer nos expériences en se donnant un but précis et accessible. Il s'agissait de l'objectif le plus simple que nous nous étions fixé.

1. Déroulement de l'expérience - Problèmes pratiques et remarques

En amont du processus d'interrogation, s'est présenté un petit "hiatus" informatique: notre ordinateur n'étant pas équipé d'un modem, il était connecté au minitel à l'aide d'un câble. Il en a résulté un décalage entre l'affichage des pages sur l'écran du minitel et l'affichage complet des pages sur l'écran de notre ordinateur, que nous avons mesuré à 5 secondes. Ce décalage faussait partiellement la mesure du temps que mettait l'utilisateur pour arriver à son but. Pour cette première expérience le décalage total était mesuré à 40 secondes, pour une durée totale de 4 minutes.

L'utilisateur s'était connecté rapidement à VIVA (code VIVA3) et avait sélectionné la rubrique ETAT CIVIL assez rapidement. Le terme "assez" concerne la rapidité de la frappe des instructions. Cette rapidité variait notablement d'un interrogateur à l'autre, selon l'habitude que ceux-ci avaient du minitel. Pour obtenir cette même information, nous avons passé moins de 2 minutes.

L'interrogateur a réussi à obtenir sa réponse au bout de 5 pages-écran, en faisant défiler le contenu de la rubrique.

DEUXIEME EXPERIENCE :

La deuxième expérience fut effectuée quelques jours plus tard. Elle était de même type que la première interrogation. Cette fois-ci il s'agissait de savoir quelle pharmacie était de garde le week-end à venir. L'interrogateur était une femme qui ne connaissait VIVA que de nom et n'était pas amateur du minitel. Disons tout de suite, que cette consultation, après une mauvaise manipulation, a été effacée de la disquette la contenant. Nous relevons ici le déroulement de cette expérience d'après des notes prises pendant celle-ci : L'utilisatrice a remarqué en premier lieu que dans le sommaire de VIVA se trouvaient deux rubriques relatives aux services de santé du week-end : dans la partie FLASH, la rubrique GARDE MEDICALE; et dans la partie URBA, la rubrique SANTE URGENCES.

L'utilisatrice a consulté la première rubrique. Sous l'étiquette GARDE MEDICALE, elle trouva au bout de deux minutes (environ), l'information désirée. Signalons tout de suite que le mot PHARMACIE et/ou ses dérivés

étaient absents des titres de la rubrique. Il nous semble que cette absence pourrait conduire certains interrogateurs, qui ne seraient pas habitués au contenu des rubriques de VIVA, à renoncer à la recherche, concluant que la rubrique contiendrait uniquement les noms et adresses des MEDECINS DE GARDE.

TROISIEME INTERROGATION

Cette interrogation a eu lieu le 15 août 1989. Nous avons voulu connaître les propos tenus par l'ex-ministre de la Défense et maire de Villeurbanne, concernant le problème soulevé l'été 89 par la "colère" des gendarmes. Ceux-ci, lassés de la lourdeur hiérarchique, ont choisi de manifester au grand jour leur mécontentement. Nous savions que le maire de Villeurbanne s'était exprimé à ce propos, et nous nous attendions à ce que ses déclarations fassent l'objet d'un flash du journal VIVA3. Ceci d'autant plus que très fréquemment des propos "non officiels" du maire ont déjà fait l'objet de flash d'information. Nous avons donc interrogé VIVA. Le contenu du journal n'avait pas été mis à jour depuis le 30/07/89 et jusqu' au 02/08/89; aucun titre ne laissait supposer que l'information était présente. Sachant que VIVA prend "plaisir" à créer des titres "bizzares", nous avons fait une petite tentative avec le titre LE MAIRE AU NOM DES PERES, un peu au hasard. Le contenu de cette page nous a donné le support d'interrogation de la troisième interrogation. Voici le contenu du flash: LE MAIRE AU NOM DES PERES

LE MAIRE AU NOM DES PERES 11.07.89

Quelle forme doivent prendre aujourd'hui les actions de prévention de la délinquance ? Tel est le thème autour duquel ont planché hier élus et animateurs Villeurbannais. A la rentrée le COVASE devrait être modifié. Les jeunes ne seront plus pris comme uniques interlocuteurs. Une série d'action aura pour objectifs d'aider à la restauration de l'auto-rite parentale. Le maire s'est dit prêt à y jouer de son image.

Remarque : Le texte cité ci-dessus est rapporté tel quel. Nous n'y avons pas ajouté les accents, ni corrigé les erreurs.

La consigne de la troisième expérience était de trouver le(s) article(s) concernant la prévention de la délinquance. L'interrogateur qui a réalisé cette expérience était le même que celui qui a réalisé la première expérience. Cette expérience a eu lieu en deux temps.

Déroulement de l'expérience en août 1989 : Connexion rapide sur VIVA3. L'interrogateur a fait défiler le sommaire des titres du journal. Après 1 minute d'hésitation relativement au choix d'un des titres alors présents, il a déclaré: "Je ne trouve rien" et s'est déconnecté du journal. Déroulement total de l'expérience : 1 min 40".

Nous sommes revenus à cet article quelques mois après, en novembre 1989. Nous avons constaté la présence dans le journal d'autres articles concernant des actes de violence et nous avons voulu savoir ce que faisait l'interrogateur, face à la même question mais avec d'autres éventuelles réponses. Ainsi, avons-nous posé la même question et enregistré le déroulement de l'expérience suivante.

L'interrogateur a fait défiler le sommaire des titres du journal. Il a consulté par la suite et dans l'ordre les articles suivants:

FEMME DE 54 ANS ASSASSINEE 13.11.89

Mme Huguette Garnier Garronnier, agée de 54 ans, a été assassinée en milieu d'après midi, dans l'allée de son immeuble, au 52 crs Emile Zola.

Un crime commis à bout portant, par trois coups de feu, contre cette dame vivant seule qui partait chercher sa petite fille à l'école Lakanal.

Les policiers chargés de l'enquête ne disposaient ce soir que de maigres indices, pas de mobile, pas de signalement précis de l'agresseur.

CHABAGA: ADJOINTS SOLIDAIRES 11.11.89

L'agression dont a été victime l'adjoint à la prévention de Villeurbanne, M.Hocine Chabaga a soulevé l'indignation générale. Quelles que soient les tendances politiques, les responsables villeurbannais s'accordent à trouver revoltant le recours à la violence dans le cadre des débats de société. Marc Fraysse (RPR), Christian Montegu (CDS) et Pierre Vial (FN), l'ont tous déclaré à la presse.

CHABAGA, AGRESSION/PROVOCATION 10.11.89

Veritable agression ou provocation? La question est pour l'instant posée. Adjoint d'origine musulmane, Hocine Chabaga avait le matin même, dans les colonnes d'un quotidien local, pris des positions très modérées sur le voile islamique et la mosquée de Lyon.

Seule l'enquête menée par le commissariat de Villeurbanne, pourra vérifier la thèse d'une éventuelle revanche...

Remarque : Dans cette présentation également, aucun signe de ponctuation, ni accents n'ont été rajoutés par rapport au texte présenté dans le journal.

Après avoir consulté ces trois articles et n'ayant rien trouvé de satisfaisant, l'interrogateur s'est déconnecté du journal. Durée totale de l'interrogation 6 min 40 sec.

Il est à noter que malgré ces deux tentatives, l'interrogateur n'a pas réussi à obtenir l'information désirée, à savoir des informations sur " la

prévention de la délinquance", informations susceptibles de lui être fournies par l'article LE MAIRE AU NOM DES PERES.

Il est certain donc qu'une interrogation faite par des syntagmes nominaux aurait donné accès au texte, chose que le procédé actuel exclue.

Remarque : Bien de mois plus tard, et à l'occasion d'une conversation, nous avons demandé à l'utilisateur de nous dire ce que pourrait lui évoquer un titre de journal comme LE MAIRE AU NOM DES PERES. Après quelques hésitations, l'interrogateur a répondu qu'il s'agirait plutôt d'un cas de pères divorcés...!

QUATRIEME INTERROGATION :

L'interrogation suivante est du même type que la précédente, sauf le fait qu'elle a été réalisée en une seule fois :

QUATRIEME EXPERIENCE :

Elle a eu lieu le 17/10/89. La consigne donnée était : "Prévention des décès d'enfants suite à des accidents domestiques". L'expérience a duré au total minutes et l'interrogateur n'a pas pu accéder à l'information.

Déroulement de l'expérience : Connexion rapide à VIVA 3. L'interrogateur consulte le sommaire des titres du journal. Il a par la suite consulté l'article suivant :

TROIS ENTREPRISES INCENDIEES 18.09.89

Un incendie, d'origine semble-t-il accidentel s'est déclaré cet après-midi dans une entreprise de vernissage.

Le feu s'est rapidement propagé à deux sociétés mitoyennes faisant des dégâts très importants.

La ligne du chemin de fer de l'Est a même dû être coupée durant une heure par mesure de sécurité.

On ne déplore cependant aucun blessé.

se.

Insatisfait, notre interrogateur s'est déconnecté du journal.

L'article visé par cette expérience était le suivant:

ATTENTION MAISON GEANTE 13. 10.89

En France 700 enfants meurent chaque année, chutes, intoxications, brûlures sont souvent causées par des imprudences ou des oublis.

L'information est primordiale, voilà pourquoi Villeurbanne par l'intermédiaire de son service Communal d'Hygiène et de Santé, fait de la prévention un objectif prioritaire du 14 au 20 Octobre, en ayant amené la totalité du gymnase de Cusset en Maison Geante, lieu d'accueil.

Remarque : Il est à noter que dans le texte de VIVA nous trouvons l'expression "700 enfants meurent...", tandis que notre consigne était "décès d'enfants ...". Si le lien sémantique est évident pour l'humain, il l'est certainement bien moins pour la machine. Le syntagme "accidents domestiques" n'est jamais présent dans l'article de VIVA, même si les termes : "chutes, intoxications, brûlures" se rapportent de toute évidence aux accidents domestiques. Le rapport est par ailleurs souligné par le terme "MAISON" présent dans le titre de l'article.

Cette expérience fait ressortir l'importance du traitement des synonymes. En l'occurrence un lien entre "DECES" et "MOURIR" aurait donné accès, entre autres, à cet article également.

Nous avons demandé à l'interrogateur la raison du choix de l'article : TROIS ENTREPRISES INCENDIEES, qui n'avait apparemment pas beaucoup de liens avec la consigne qui lui avait été donnée. La réponse était le lien que l'on pouvait faire entre "ACCIDENT/DECES" et "INCENDIE"; lien très vague certes, mais l'on sait jamais...

En outre, bien de mois plus tard, nous avons demandé à l'interrogateur ce que lui évoque le titre "Attention Maison géante", il

nous a répondu qu'il s'agirait peut-être d'une **publicité pour des nouvelles constructions.**

CINQUIEME INTERROGATION

Cette expérience a été réalisée en deux fois (05 et 06/09/89) par deux interrogateurs différents. Il s'agissait cette fois de trouver l'article concernant un nouveau restaurant qui venait d'ouvrir. La première tentative a été réalisée par notre interrogateur habituel qui a simplement ouvert le sommaire des titres du journal et s'est déconnecté aussitôt. Est-ce par habitude de ne pas trouver ce qu'on lui demandait ?

La deuxième tentative a été réalisée par une future étudiante, qui a trouvé très rapidement le titre qu'on lui demandait (durée totale de cette interrogation : 1 min). Nous lui avons posé la question, à savoir comme a-t-elle deviné que sous la titre "NOURRITURES TERRESTRES" se trouvait un article concernant un nouveau restaurant. Elle nous a dit que le terme "nourriture" l'avais mise dans la voie... Quand nous avons posé la question au premier interrogateur, il nous a répondu qu'il s'agissait du roman de André Gide ("une soirée ou quelque chose autour" de ce roman ou de son auteur)

L'article recherché, était le suivant :

LES NOURRITURES TERRESTRES 02.09. 89

Avec la rentrée des classes, a été inauguré un nouveau restaurant,

"Les nourritures terrestres".

Installé jusqu'à côté de la Maison du Livre, le domaine de Genevieve et Gerard Picot se veut à vocation multiples.

Petit restau sympa, restaurant pour gourmets et gourmands, lieu d'exposition, il entend drainer un public d'actifs à midi et de fêtards le soir, sortant de l'ASVEL ou du TNP.

Remarque : Il est important de signaler que l'enchaînement des écrans se fait soit par la touche minitel SUITE soit en passant par le SOMMAIRE de VIVA, ce qui est très long dans les deux cas, sauf si les articles sont voisins l'un de l'autre. Lors de la troisième interrogation, le consultant a enchaîné les écrans à l'aide de la touche minitel. Il avait remarqué la proximité des ses articles dans VIVA, mais ceci n'est que le cas le plus favorable.

Nous voyons donc qu'en ce qui concerne les quatre dernières expériences, l'interrogateur ne réussit pas toujours à trouver l'information qu'il désire. Pourtant, elle y est toujours, présente dans des SN des textes

CHAPITRE N° 4 : LA GRAMMAIRE

ASPECT THEORIQUE :

Avant toute proposition concrète, il nous a paru nécessaire d'établir le contexte théorique de la démarche. Celle-ci, en effet, s'inspire des travaux du groupe SYDO de Lyon. Les travaux de ce groupe portent sur la recherche de l'information et plus précisément, trouver le moyen le plus efficace d'accéder à l'information incluse dans un document ou ensemble de documents.

Les bases théoriques de la méthode du groupe SYDO se situent dans le courant générativiste chomskien, comme le signale M. Metzger : "... *il me semble utile de rappeler ici les principaux éléments de la grammaire transformationnelle dans le cadre de la théorie générative proposée par N. Chomsky**. En effet, notre démarche et le modèle linguistique sur lequel nous nous appuyons s'inspirent pour une bonne part, du modèle transformationnel" (Thèse d'Etat, Lyon 1988, page 56).

Nous dirons, en bref, que l'objectif de Chomsky est double : Chomsky considère que pour concilier l'explicite et l'explicatif (postulats * pas toujours présents, selon lui, dans les théories linguistiques précédentes) il doit proposer "*une nouvelle définition de ce qu'est une grammaire*".

(O. Ducrot et T. Todorov, *Dictionnaire...* Ed Seuil, Paris 1972, page 58).

Quelle est donc la définition de la grammaire selon ...Chomsky ?

" *C'est un ensemble de règles, d'instructions, dont l'application mécanique produit les énoncés admissibles (=grammaticaux) de cette langue et eux seuls*" (O. Ducrot et T Todorov,....., page 58)

Pour engendrer donc des phrases, nous nous octroyons :

a) un ensemble fini de symboles que l'on appelle "l'alphabet", qui comprend tous les "morphèmes" ** de la langue.

* Nous précisons rapidement que Chomsky est d'accord avec les distributionnalistes dans leur souci d'être explicites. Mais il leur reproche de ne pas essayer d'être explicatifs, en envisageant une théorie globale de la langue.

** au sens que les américains donnent à ce terme

b) Nous définissons à l'intérieur de cet ensemble un symbole de départ que l'on appelle "l'axiome".

c) En dernier, un ensemble de règles qui décrivent des manipulations qu'on peut effectuer sur des ensembles.

Il y a deux types de règles : les règles syntagmatiques (appelés également **règles de réécriture**) et les **règles transformationnelles**, comme par exemple : la règle de réécriture : SV \rightarrow V + SN.

Les règles transformationnelles n'agissent pas sur les symboles mais sur des arbres. Elles permettent des opérations telles que la formation du passif ou la réflexivisation.

Notre travail s'inscrit dans le cadre des règles de réécriture.

Le groupe SYDO a été amené à considérer le concept, ancien mais fort utile, de syntagme nominal, comme le centre théorique de sa démarche. La raison en est simple : le groupe a remarqué que l'information spécifique d'une séquence textuelle se trouve dans les unités (simple ou complexes) que l'on appelle les **syntagmes nominaux**. Il existe donc des unités de discours qui se distinguent par leur caractère référentiel. Ces unités sont principalement "*des occurrences de syntagmes nominaux*" (M. Metzger, Thèse d'Etat, page 26). On peut donc supposer que ces unités constituent une projection (au sens mathématique de terme) de l'univers de référence dans le discours et qu'il est légitime de chercher à les reconnaître dans une perspective de recherche de l'information véhiculée dans le discours.

Nous citons ici une remarque très importante de M. M LEGUERN, et nous reviendrons sur ces réflexions ultérieurement. M LeGuern écrivait en 1984 : "*La finalité du descripteur exclut qu'on puisse l'envisager en faisant abstraction de la valeur référentielle de ses occurrences dans le corpus. Les mots de la langue, en tant qu'ils sont mots de la langue, ne signifient que des propriétés, jamais des entités; ils signifient des attributs, et non des substances, tant qu'ils ne sont pas mis en oeuvre dans le discours. Le descripteur*"^{***} *quant à lui, signifie une entité, une substance au sens de la philosophie d'Aristote. Le descripteur ne peut donc pas être considéré à l'instar des mots de la langue, comme un symbole sans référence*" (M. LeGuern Les descripteurs d'un système

^{***} Descripteur = syntagme nominal extrait de son contexte. Il ne s'agit pas de l'unité lexicale

documentaire : essai de définition, in Bès, Fauchère, Laguenière, *Actes du colloque Traitement automatique des langues naturelles et systèmes documentaires*, cité dans Metzger, 1988, page 26)

Dans la recherche de l'information contenue dans un texte, il s'avère très utile de pouvoir constituer la liste des syntagmes nominaux présents dans celui-ci. Le processus peut être appliqué à des bases de données très structurées (les bases classiques) mais nous allons voir qu'il peut être également utilisé pour optimiser l'interrogation des bases de données "grand-public". Plus spécialement, nous allons voir comment le concept de syntagme nominal peut nous être utile dans la recherche de l'information dans un service minitel et la ... **proximité du processus d'interrogation de la base** (ici VIVA) **avec le processus humain de référentiation**. Nous ne pouvons pas aborder ici l'aspect informatique que suppose l'extraction des syntagmes nominaux d'un texte écrit en français "standard". Nous pouvons néanmoins indiquer que le programme d'extraction des SN a été établi par le groupe SYDO (et plus particulièrement grâce aux travaux des MM Ramburrun et Metzger) en langage informatique PROLOG FOLL ****

Nous exposons ci-dessous, les lignes directrices actuelles du processus d'extraction des SN et nous portons en annexe la grammaire complète établie par M. Metzger dans sa thèse d'Etat, ainsi que la version initiale de celle-ci proposée par M. Ramburrun dans sa thèse de troisième cycle.

LA MORPHOLOGIE:

Le traitement préalable du texte conduisant à la reconnaissance des SN, présuppose la "régulation" de celui-ci.

Dans sa thèse d'Etat (1988), M. Metzger expose longuement les modifications à apporter, dans la perspective de reconnaissance des SN. Nous allons exposer quelques points, pris dans cet ouvrage, qui permettront aux non-avertis de comprendre la méthode générale et surtout de "décoder" la grammaire de reconnaissance des SN.

**** Des travaux ont été effectués à l'aide du langage informatique STARLET, mais nous n'avons encore pris connaissance de leurs résultats.

Dans ses travaux,(à l'initiative des MM Berrendonner et LeGuern), le groupe SYDO reconnaît les catégories suivantes:

" - les principales catégories de SN simple (sans relative ni incise): les catégories F (nominaux), D (prédéterminants), P (prépositions) et une partie de la catégorie W (adverbes).

- les catégories des mots (ou ponctuations ne pouvant pas appartenir à un SN simple. " (J.P Metzger " *SYNTAGMES NOMINAUX ET INFORMATION TEXTUELLE* ", Thèse d'Etat, Lyon 2, octobre 88, page 76).

Ainsi,

A : ...dans la catégorie F, sont inclus les substantifs et les adjectifs.

Afin de pouvoir, malgré tout, établir une petite différence entre eux, le groupe SYDO appelle "**NOM**" les premiers et "**ADJ**" les seconds. Pour les nominaux qu'on arrive pas à intégrer dans l'une ou l'autre de ces deux sous-catégories, une troisième a été établie, celle des **NAN**.

Sont admis également dans cette catégorie (de manière provisoire, comme le souligne M. Metzger) les pronoms toniques : "moi", "toi", "lui", "elle", "elles", "nous", "vous".

En dernier, sont admis dans la catégorie F les **noms propres**. Ceux-ci sont le plus souvent utilisés sans article, sauf dans quelques cas bien précis de noms géographiques.

- Les prédéterminants : D

" *Un prédéterminant correspond, à peu près, à ce qui est habituellement appelé "article". (idem, page 78).*

La catégorie D contient les mots suivants:

" le/la/l'

ce/cet/ cette/ces

mon/ma/mes/ton/ta/tes/son/sa/ses/notre/votre /nos/vos/leur/leurs

différents/différentes

divers/diverses

maint/mainte/maints/maintes

deux/trois ...

un/une

plusieurs

certain/certaine/certains/certaines

chaque

nul/nulle/nuls/nulles

tout/toute/tous/toutes." (idem, page 78-79).

- La catégorie des prépositions P:

Dans cette catégorie il y a :

à/de/

par,

après/avant/avec/chez/contre/dans/dedans/depuis/derrière/dès/

devant/durant/en/entre/envers/hormis/malgré/outré/parmi/pendant

/pour/sans/sauf/selon/sous/dessous/sur/dessus/vers/.

- La catégorie des adverbes **W**, contient les mots qui ne peuvent pas être inclus dans les catégories précédentes

B: Les autres catégories

- Les verbes, V.

On peut définir syntaxiquement la catégorie des verbes en faisant appel aux formes négatives /ne/ et /pas/. Toute forme pouvant apparaître entre /ne/ et /pas/ ou immédiatement après la séquence /ne/-/pas/ est une forme verbale:

/...ne ___ pas.../ (formes fléchies)

/...nepas___.../ (infinitifs) (idem, page 82).

- Les particules préverbales: catégorie Y.

sujets (ppv₁): /je/, /tu/, /il/, /elle/, /ils/, /elles/, /on/, /ce/.

compléments (ppv₂): /me/, /te/, /se/, /le/, /la/, /les/, /l'.

sujets-compléments : /nous/, /vous/.

- Les conjonctions de subordination: Q.

La catégorie se limite à : /que/, /de/, /à/, /si/.

- Les ponctuations, T.

les séparateurs de phrases : /:/, /./, /?/, /!/, /:/.

séparateurs de syntagmes: /,/ , /;/, /-/. (idem, page 84)

- Les sous-catégories DEF. NUM. IND.

Dans la catérogie des prédéterminants, D, on distingue les définis (DEF), les numéraux cardinaux (NUM); la catégorie IND récupère tous les autres:

DEF: le/le/l'/ce/cet/cette/ces/mon/ton/.../leurs/

NUM: deux/trois/ /quelques/ /divers/ /diverses/ /maintes/ /maints/ différents/ différentes.

IND : chaque/ /quelque / /certain/ /tout/ /nul/ /plusieurs/ maint/ /mainte/un/. (idem, page 85).

- Les sous-catégories QUA, AAJ, PRO, TAM.

Des sous-catégories de la classe des adverbes. Elles comportent "les adverbes de quantité et des adverbes modificateurs d'adjectifs (d'intensité)":

QUA : assez/ beaucoup/ guère/ trop/peu /pas/ tellement /tant /autant /moins/ plus/davantage/...

AAJ : assez/bien/for/guère/trop/ peu/ si/ quasiment/ tellement/ tout/. aussi/ moins / plus... " (idem, page 85).

LA NOTION DE PROFIL :

Chaque mot rencontré, doit être classé par rapport à des traits grammaticaux. L'attribution constitue le profil que l'on associe à chaque mot. Cette organisation permet, entre autres, de désambigüiser certains mots a caractère multiple. Ainsi, comme le note M. Metzger, le mot /lui/ peut avoir les traits suivants: être un pronom tonique, ou un pronom preverbal, ou un participe passé (Jean-Paul Metzger, page 93).

Comment se définit la notion de profil ?

Definition : Les traits associés à chaque mot et, par suite, à chaque occurrence étant principalement destinés à être utilisés lors de l'analyse syntaxique (analyse en constituants), ils peuvent être regroupés en fonction du rôle qu'ils jouent vis-à-vis de cette analyse.

Nous avons distingué :

- les traits syntaxiques, qui correspondent à des distributions et qui seront associés aux symboles terminaux de la grammaire de réécriture; il s'agit de la catégorie et des éventuelles sous-catégories NOM-COM, NOM-PRO, NOM-PRP, NAN, ADJ, DEF, NUM, IND.

- Les traits flexionnels de genre, nombre, et personne, qui permettent la prise en compte des contraintes d'accord et qui peuvent être obtenues, pour partie, par une analyse de la forme du mot.

- Les traits "lexicaux", qui correspondent, principalement aux valeurs des variables *DQ* et *RV* décrivant les propriétés rectionnelles, et aux traits "classématiques".

J'appellerai profil chacun de ces trois ensembles de traits associés, une ou plusieurs fois, à chaque mot : profil syntaxique, profil flexionnel et profil "lexical".

Ainsi le mot /*définition*/ a pour profil syntaxique : (F, NOM-COM),

profil flexionnel : (FEM, SIN),

Profil lexical : (DVB, ACC). "

(Jean-Paul Metzger, page 95).

LA COMPOSANTE MORPHO-SYNTAXIQUE

Les catégories ci-dessus ne sont opérationnelles qu'après un certain "prétraitement". Cette opération permet d'obtenir de certaines séquences, des séquences équivalentes qui peuvent être intégrées dans la catégories. Nous allons citer quelques cas, les plus courants. Les problèmes que génère la régulation morpho-syntaxique sont brillamment exposés dans la thèse de M. Metzger.

- les articles contractés : /au/ --> /à/ /le/
/ du/ --> /de//de/ ...
- les relatifs composés : / auquel/ --> /au//quel/
/ desquelles/ --> /des// quelles/...
- les mots négatifs : /aucun/ --> /pas/ /un/
/ aucune/ --> /pas/ une/...

L'analyse des mots en /que/ pose beaucoup de problèmes. Nous allons relater les solutions qu'apporte M. Metzger. Ainsi, (page 125) les séquences qui se trouvent à gauche, subissent un traitement qui permet de faire apparaître /que/ suivie de ...

- "/ le chien qui aboie/ --> le chien que lui aboie/
/ le chien dont je parle/ --> /le chien que de lui je parle/
/ la ville où je vais/ --> / la ville que là je vais/ ...".

Nous répétons qu'il est, à nos yeux, inutile d'exposer la totalité des problèmes que posent les catégories, la régulation morpho-syntaxique. Et nous n'avons rien dit des problèmes posés par l'analyse flexionnelle. Le but de ce bref exposé était de présenter quelques aspects de l'analyse à laquelle est soumis le texte par l'analyseur du groupe SYDO. L'étape suivante sera la détermination de la grammaire qui, à partir des résultats obtenus par l'analyse préalable, dressera la liste des syntagmes nominaux du texte. On voit que le prétraitement est très important pour le succès de l'étape suivante, à savoir celle de la reconnaissance des SN.

QUELLE GRAMMAIRE ?

Nous avons voulu établir des règles d'une généralité suffisante, qui permettraient de traiter les deux types de données, à savoir, les titres (ou parties fonctionnant comme des titres), et les textes "pleins". M. LeGuern nous avait fait remarquer que la grammaire des titres devrait faire partie de celle des textes "pleins". Ainsi, par exemple : les conditions suivantes s'appliqueraient qu'aux textes pleins :

N" --> NOM-PRO

N" --> NOM-PRN

Nous avons remarqué toutefois que dans les parties de VIVA, autres que les titres proprement dits, nous avons rencontré assez souvent des phrases qui auraient pu très bien constituer des titres. Nous pensons donc qu'il serait utile de le traiter de la même manière que les titres. Reste à savoir comment pourrait être réalisé un tel projet d'un point de vue informatique.

La grammaire établie par M. M Ramburrun dans sa Thèse de troisième cycle, permet de traiter les séquences sans prédéterminant. Allions-nous pour autant adopter cette grammaire dans sa totalité ? Notre réponse est négative et ceci pour une raison déjà relevée par M. Metzger dans sa Thèse d'Etat : la grammaire de M. Ramburrun ne permet pas de faire la distinction, pourtant capitale, entre SP et le EP. Exemple de syntagme prépositionnel (SP) : "de la CEE" (thèse de M. Ramburrun, page 62). Le fonctionnement logico-sémantique de "de la CEE" n'est pas le même que "à quartz" (expansion prépositionnelle, EP).

pris dans l'exemple : "montre à quartz" (M. Metzger, pages 167-168). Ainsi, sommes-nous obligés de recourir à des clauses permettant de faire la distinction entre les syntagmes prépositionnels (SP) et les expansions prépositionnelles (EP). Sur ce point, nous avons adopté les clauses établies par M. Metzger :

$N \rightarrow N + EP$ (règle n° 17)

$EP \rightarrow P + N'$ (règle n° 31) (Jean-Paul Metzger, page 168)

Nous ne pouvons pas cependant accepter la totalité de cette grammaire : cette grammaire est "prédéterminative" dans sa conception même.

Nous avons adopté donc les clauses primaires de réécriture des SN établies par M. Ramburrin. A ceci près que nous imposons, à la suite de M. Metzger, les contraintes suivantes :

$N'' \rightarrow N'$ si et seulement si les deux clauses suivantes sont remplies : a savoir : $N'' \rightarrow NOM-PRO$ et $N'' \rightarrow NOM-PRP$.

Donc, si et seulement si nous avons à faire avec un nom propre ou un pronom.

Cette condition n'est pas valable dans le cas du titre.

La clause $N'' \rightarrow N'$ s'applique à notre corpus par le fait que les prédéterminants sont maintes fois absents du corpus. A l'aide de la règle $N'' \rightarrow N'$ nous pouvons saisir des séquences du genre : "POMPIERS (POUR LE MIDI)", que nous considérons comme des N'' avec un élément 0 en amont.

Comment pouvons-nous traiter la séquence : "...POUR LE MIDI" ? Nous constatons qu'il s'agit d'un syntagme prépositionnel ou se trouve inclus un SN "le midi". Pouvons-nous appliquer à cette séquence la règle $N'' \rightarrow N'$? Afin de préserver la distinction entre les syntagmes prépositionnels et les expansions prépositionnelles, nous ne pouvons pas appliquer la règle de réécriture ci-dessus dans ce cas, à savoir quand un N'' est inclus dans un syntagme prépositionnel.

"...pour éviter une ambiguïté importante dans tout syntagme prépositionnel (catégorie SP) la préposition doit être suivie d'un

prédéterminant (ou assimilé = expansion quantitative), ce qui implique la suppression de la règle $N'' \rightarrow N'$." (Jean-Paul Metzger, page 168).

On écarte donc la possibilité de trouver un élément 0 après une préposition et l'on évite ainsi la confusion due aux règles suivantes :

$N'' \rightarrow N'$

$SP \rightarrow P + N''$

$EP \rightarrow P + N'$

Or, nous avons besoin de la règle $N'' \rightarrow N'$ pour les raisons que nous avons évoqué antérieurement. Afin de "joindre" les deux positions, dont nous avons également besoin, et qui semblent à priori s'exclure mutuellement, nous proposons la **non-application de la règle $N'' \rightarrow N'$ quand le N'' est régi par un SP.**

Remarques :

D'après des remarques faites par M. Jean-Paul Metzger, nous nous proposons de créer une nouvelle catégorie, provisoirement appelée B'' (ou mieux encore N''_{bis}) qui permettrait d'extraire les parties de titre ou de texte qui n'ont pas de prédéterminant. Cette catégorie serait posée au même niveau de généralité et aurait les mêmes constituants qu'un N'' comme le conçoit dans sa grammaire M. Ramburrun, c'est-à-dire, sans prédéterminant. Ainsi, nous réservons la notation N'' explicitement pour la catégorie des syntagmes nominaux avec prédéterminant. Un titre donc comme "POMPIERS POUR LE MIDI" pourrait être formalisé à l'aide de l'écriture suivante :

$S \rightarrow N''_{bis} + SP.$

Cette écriture permet, pensons-nous, de lever l'ambiguïté générée par l'écriture de N'' en N' et $N.$

La réécriture du syntagme prépositionnel serait la suivante :

$SP \rightarrow P + N''$

La réécriture d'une expansion prépositionnelle serait :

$EP \rightarrow P + N''_{bis}$

La cohérence donc, de la grammaire ne peut être assurée sans les clauses de réécriture initiale de SN et à l'aide de cette nouvelle catégorie :

$N'' \rightarrow D' + N'$

$N''_{bis} \rightarrow N'$

$N' \rightarrow N$

Il suffirait de vérifier si cette nouvelle catégorie est compatible avec les autres catégories proposées par M. Metzger dans sa thèse d'Etat.

Remarque : Les données suivantes sont très fréquentes dans le corpus de VIVA :

- les expressions comme "adolescents de 13 à 17 ans";
- les adresses.

Nous pensons que ces données doivent être traitées par la grammaire. Elles posent en effet des problèmes qu'il serait intéressant et constructif de résoudre.

Considérons d'abord les expressions du genre "adolescents de 13 à 17 ans". Notre problème n'est pas tant l'absence de prédéterminant devant "adolescents" que la compréhension que l'on pourrait avoir de la séquence du type : "de ___ à ___" devant le mot "ans". Nous faisons appel à la catégorie D-QUA définie par M. Ramburrun dans sa grammaire.

Comment pourrait être analysée cette séquence ?

de	13	à	17	ans
P	D-QUA	P	D-QUA	F

Il nous semble que deux analyses sont possibles :

- [[de 13 à 17] ans] où la séquence "de 13 à 17" a un rôle d'épithète, impliquant un continuum, bien défini (ce n'est pas incompatible), dans ses limites bien précises. Cette séquence est à rapprocher avec ce que B Pottier appelle "morphèmes antéposés" (Bernard Pottier *Linguistique Générale*, Klincksieck, Paris 1974, page 132). En effet, nous pouvons considérer que "de" et "à" fonctionnent comme des morphèmes antéposés en dépendance locative notionnelle.

- [de [13 [à 17 ans]]] où l'on constate un emboîtement de SP. Cette analyse ne nous parait pas valable. Nous ne voyons pas en effet, à quoi peut correspondre, du point de vue de sens, la suite : prep + D-QUA+ SP. D'autre part, cette analyse ne nous permet pas de rendre compte du rôle de la séquence " de 13 à 17" sur le substantif.

Nous optons pour l'instant pour la première solution.

Remarque : Il serait utile également, dans une perspective d'interrogation, de pouvoir insérer dans les possibilités de traitement de la grammaire, les données du type "adress". Nous avons une brève suggestion à faire ici : l'analyse des adresses (après élimination de la virgule entre la partie numérique et le reste), ressemble étrangement à cette fameuse catégorie **K**, jusement évincée de la grammaire d'extraction des syntagmes nominaux. Serait-il possible de la réintroduire avec des clauses bien spécifiques d'utilisation ? Cela mérite l'attention.

Remarque : M. Metzger me signale que le pluriel, dans la séquence "POMPIERS POUR LE MIDI" n'est pas tout à fait "innocent". L'absence d'article dans ce cas, ne peut pas être considérée de la même manière que dans la séquence écrite au singulier, comme par exemple : PREVENTION DE LA DELINQUANCE (article enregistré en décembre 1988). Entre ces deux exemples, nous sentons une différence importante. Dans le premier exemple, l'article est relatif à l'envoi dans le Midi de la France d'une équipe de pompiers. Il est légitime de supposer que la forme complète du titre pourrait être "Des pompiers pour le Midi". Dans ce cas, nous avons à faire avec ce que R. Martin appelle "l'extraction aléatoire non-universelle", où le choix se porte sur une équipe de pompiers, et non pas sur une autre formation. D'autre part, ce n'est pas tous les pompiers qui seront convoqués pour aller dans le Midi.

En ce qui concerne le deuxième exemple, il y a lieu de supposer que sa forme complète serait "La prévention de la délinquance". Suivant toujours R. Martin, nous pensons que l'article a un usage "intentionnel" dans ce cas. Le substantif est donc pris "dans les propriétés qu'il signifie" (R. Martin, *Pour une logique du sens*, ed PUF, Paris 1983, page 160).

Notre grammaire traiterait ces deux cas, "pompiers" et "prévention" de la même manière, à savoir elles les aurait catalogués dans la catégorie N^{bis}. Or, leur différence logique est importante. Serait-il intéressant de réserver un petit traitement particulier, pour des cas comme ceux-ci ?

Nous remarquons également que l'attribution d'un article à une séquence qui n'en a pas, ne se fait pas toujours de manière aussi aisée. Des ambiguïtés existent et il est fréquent d'avoir deux solutions possibles simultanément.

CHAPITRE 5 LES PROPOSITIONS

Nous allons revenir sur les expériences que nous avons faites et nous exposerons les résultats auxquels nous avons abouti.

Etablir un système de traitement d'un corpus comme VIVA ne fut pas d'une facilité évidente, vu que VIVA elle-même a subi des modifications pendant la rédaction de ce mémoire. Ainsi, avons-nous décidé de présenter la suite chronologique de notre démarche et les étapes que nous avons franchies.

Dans un premier temps, nous avions un corpus constitué des différentes interrogations et expériences faites par les interrogateurs et nous-même, citées dans le chapitre 3. Cet état des choses (de VIVA) a duré jusqu'en décembre 89, où le maire de Villeurbanne Charles Hernu est mort. Son décès a été la cause de plusieurs modifications dans la vie publique et par voie de conséquence, dans l'organisation du service télématique. Des projets furent, pour quelque temps abandonnés, le contenu et la présentation du journal ont été changés : pendant quatre mois, VIVA ne présentait plus son habituelle page de sommaire des titres. L'interrogateur arrivait directement dans la page-écran flash la plus récente. S'il voulait obtenir une information antérieure, il devait faire "SUITE" ou "RETOUR" et voir défiler toutes les pages intermédiaires. Gênant! Nous avons posé la question à M. Galliot et il nous a répondu que "tout le monde" (était) "en attente dans la municipalité...on ne pouvait plus faire comme avant. (Il était donc utile) que les gens tombent dans le dernier flash directement".

Pendant près de trois mois, VIVA est restée en attente. Les décisions ne pouvant plus être prises comme auparavant, le journal a perdu son style habituel de rédaction et présentation. (à notre avis, d'ailleurs, il ne l'a pas entièrement retrouvé depuis) : les articles étaient centrés sur les décisions du conseil municipal; certains articles étaient repris textuellement des journaux lyonnais. On ne sentait plus l'aspect convivial dont VIVA nous avait habitué jusqu' alors.

Nous allons exposer les propositions auxquelles nous avons abouti dans leur ordre chronologique.

Tout d'abord, qu'avons-nous remarqué lors des deux premières expériences, qui sont, rappelons-le, très simples ? Elles sont effectivement très simples, c'est-à-dire, l'information désirée a été

obtenue très facilement. Il est à noter que les services du type de VIVA sont à priori destinés à ce type d'utilisation. Même si, une partie de l'information est "fatale", (pour reprendre la formulation de F. Jakobiak dans *Maîtriser l'information critique*, Paris 1988, page 29) et donc inutile à l'utilisateur, celui-ci parvient rapidement à l'information "utile" (idem, page 31).

La question qu'on s'est posé a trait à la diminution de cette information "fatale", que la structure hiérarchique de VIVA impose à priori. Pour y parvenir, il fallait modifier le système existant et, justement écarter cette structure hiérarchique. Le système de traitement de l'information par le SN répond à ce besoin de changement. Nous pensons en effet, que la recherche par l'intermédiaire de SN permettrait de cibler la recherche sur la (ou les) page(s) qui contiendrait l'information. Faire défiler le sommaire d'une rubrique, sans comprendre les titres des parties, c'est là d'un obstacle qui peut être contourné par un système d'interrogation inspiré des travaux du groupe SYDO. Pour ce système, il s'agirait d'établir, à l'aide d'une application en PROLOG, un fichier contenant tous les SN des textes de VIVA. L'utilisateur ne se trouverait plus passif face à un système qui aurait tout "décidé" pour lui. Il prendrait part à la recherche de l'information en formulant lui-même, en tapant lui-même sa question.

Comment serait-il construit ce fichier des SN de VIVA ? Notre première proposition consiste donc à établir ce fichier unique où les SN du corpus figureraient par ordre alphabétique ou par ordre de page-écran¹. C'est là le traitement le plus simple. En cas de réponses multiples, le système poserait la question à l'utilisateur si ce dernier veut visualiser (afficher à l'écran) une partie ou la totalité des réponses données par le système. Ce système présuppose que l'interrogateur va utiliser (pour poser sa question) la même formulation que celle déjà existante dans VIVA. Dans le cas contraire, le système serait incapable de donner une réponse. Pour pallier à cet handicap, il serait possible d'utiliser le procédé des tronçatures, très fréquent dans les systèmes documentaires. Nous pensons qu'il serait possible de demander à

¹ Il peut être plus commode, même s'il est plus encombrant, d'intégrer dans le système toutes les occurrences des SN, avec un numéro de page-écran correspondant à chaque occurrence...

l'utilisateur de taper à chaque fin de mot un caractère spécifique. Cette opération pourrait faciliter la tâche du système en ce qui concerne les formes les plus simples de féminin ou du pluriel et contourner les problèmes créés par une simple erreur de frappe. Mais il est bien évident que ceci est loin de résoudre tous les problèmes qui se posent au moment de l'interrogation.

Ce système avait alors à nos yeux , un seul inconvénient : il nous paraissait assez lourd et n'exploitait pas les possibilités qu'offrait la structure de VIVA². Nous pensions en effet qu'il aurait été possible et souhaitable d'exploiter le type de données propres à VIVA, afin de faciliter la tâche. VIVA a trois type de données :

- les informations bien structurées dont les données se présentent sous une forme précise, par exemple : GARDE MEDIALE;
- les textes pleins, par exemple , les articles du FLASH;
- les données numériques par exemple, dates, âge,...

Il nous paraissait judicieux de ne pas traiter ces trois types de la même manière....³

Suite aux troisième et quatrième interrogations plusieurs autres remarques et réflexions sont venues s'ajouter à notre première idée.

Nous avons voulu connaître les propos du maire à propos de la colère des gendarmes. Au lieu de la colère des gendarmes, nous avons rencontré... la colère du maire dans LE MAIRE AU NOM DES PERES!

Nous avons alors pensé qu'il serait intéressant de découper notre fichier initial en deux : un fichier pour les SN des titres et un autre pour les SN des textes. Cette initiative a une raison bien simple : si, dans sa question, l'utilisateur introduit une séquence équivalente à celle déjà existante dans un titre, la tâche du système sera simplifiée et les risques de reponses parasites diminués par la même. Si le traitement du fichier des SN-titres ne donne aucun résultat, on peut continuer la recherche par le fichier SN-texte. Si une réponse, ou supposée telle, est reperée par le système, celui-ci doit demander à l'utilisateur s'il veut

² Nous pensions alors qu'il serait possible d'utiliser l'organisation de VIVA afin d'élaborer un système d'accès à l'information plus élaboré.

³ La notion même de SYNTAGME NOMINAL ne trouvait pas son plein emploi dans les données de ce genre, d'après M. Jean-Paul Metzger. M. Metzger pensait en effet il pourrait y avoir "des raccourcis".

visualiser cette "réponse". Il est bien évident que les délais de réponse du système doivent être les plus courts possibles (on est sur minitel).

Nous avons également remarqué que plusieurs titres de VIVA présentaient la structure suivante : "séquence 1" + ";" / ":" + "séquence2". Cette organisation du titre permettrait de distinguer une partie "thème", la première partie, et une partie "rhème", la deuxième, dans le sens où la thème est ce dont on parle et le rhème, ce qu'on en dit. Cette propriété pourrait être exploitée afin d'établir notamment l'ordre dans lequel pourraient être considérés par le système les SN relevés du corpus. En ce qui concerne le fichier SN-texte, nous souhaitons y trouver les SN dans l'ordre alphabétique.

Le problème qui se (re)pose est celui de la formulation des questions par les interrogateurs. L'usage d'un caractère de troncature ne nous apparaissait pas suffusant et cela même si l'utilisateur pose sa question simplement.

Pour résumer nos remarques ultérieures à propos des deux dernières interrogations, nous dirons que :

- les SN devraient être classés par ordre d'apparition (gauche à droite pour les SN isolés, par ordre croissant pour les SN emboîtés).
- Le système procède par comparaisons entre les SN de la question et les SN du fichier SN-titres. Il range la réponse (s'il en a une) dans un fichier temporaire "réponse" et continue les recherches avec les SN suivants de la question. L'intersection entre les réponses déjà obtenues, pourrait donner un résultat qui serait proposé à l'utilisateur. S'il n'y a pas de résultat un message serait affiché.
- En cas de plusieurs réponses, il faudrait affiner la recherche en interrogeant le fichier SN-texte et plus précisément les parties du fichiers SN-texte correspondant aux parties déjà sélectionnées par le système. Cette correspondance peut être réalisée, par exemple, par reconnaissance de numéro de page.

En suivant la réflexion de M. Metzger (notice 3 page 36) il serait également possible de procéder au découpage suivant : un premier fichier qui serait composé des noms, prénoms adresses et dates. Nous pensions qu'il serait plus commode qu'un programme "nettoie" d'abord les données des mots vides (prépositions, etc). Un deuxième fichier comporterait tous les SN des textes et titres. Ainsi, pour une question

comme "prévention de la délinquance", le système pourrait saisir, à l'aide du SN "la délinquance", tous les contextes où ce syntagme se trouve. Il en dresserait la liste et la présenterait à l'utilisateur. Celui-ci pourrait donc choisir le contexte qui lui semble être le plus proche de ce qu'il désire obtenir.

Mais que faire devant une question qui contient "la criminalité" ? Le problème se pose indifféremment, quelque soit le découpage des données et la petite astuce de la troncature serait par définition inutilisable pour les synonymes. La seule solution que nous avons envisagé à ce moment était la construction d'un fichier "synonymes". Si le système ne pouvait plus trouver une réponse parmi les SN déjà relevés dans sa base, il pourrait interroger le lexique des synonymes. L'utilisateur pourrait choisir dans une liste de termes et relancer la recherche avec l'un d'eux. Là aussi se posent des problèmes d'accord en genre et en nombre.

LES SYNONYMES :

La synonymie est une relation qui dépend beaucoup du contexte plus que les autres rapports de sens (hyponymie, antonymie, ...)

"La synonymie peut avoir deux acceptions différentes : ou bien deux termes sont dits synonymes quand ils ont la possibilité de se substituer l'un à l'autre dans un seul énoncé isolé (pour un mot donné, la liste des synonymes (synonymie absolue) ou bien deux termes sont dits synonymes quand ils sont interchangeable dans tous les contextes, et alors il n'y a pratiquement plus de véritables synonymes sinon entre deux langues fonctionnelles (par exemple en français, en zoologie, la nomenclature scientifique et la nomenclature populaire offrent de nombreux exemples de synonymie absolue) "

(Dictionnaire de Linguistique, ed Larousse, 1973, page 476).

Les synonymes susceptibles d'être inclus dans notre fichier ne sont pas a priori en relation de synonymie absolue. Des relations particulières les articulent les uns aux autres. Une relation fondamentale est celle qui lie un terme "générique" à un autre terme "spécifique".

"On dit d'un mot qu'il est générique quand il sert à dénommer une classe naturelle d'objets dont chacun, pris séparément, reçoit une dénomination particulière. Ainsi, le mot "poisson" est générique d'une

classe dont les membres sont le maquereau, la sole, la truite etc. " (idem, page 229).

Dans notre corpus, le concept de "criminalité" est en rapport de générique à spécifique avec le concept "délinquance". Le lien qui lie "nourriture" à "restaurant" est d'un autre type.

B. Pottier, dans *Linguistique Générale*, propose le concept du "sémiogramme" : "*Le sémiogramme sera la visualisation du réseau sémantique d'un morphème, en compétence.*

(a) Soit le lexème autobus

(i) Relation d'opposition

S_n S_m

autobus /v/ autocar

(si S_n = /interurbain/ : S_m = /intraurbain/).

(ii) Relation d'inclusion :

S_n

autobus => véhicule

Ces deux relations paradigmatiques signifient :

l'autobus n'est pas un autocar

l'autobus est un véhicule

(iii) Relation de participation

S_n

autobus

S_n

/intraurbain/

synthèse de "véhicule qui circule en ville"

(iv) Relation d'association

S_n

autobus

S_x

/pour aller à l'école/

(Bernard Pottier, *Linguistique Générale*, ed Klincksieck, Paris 1974, pages 103-104).

Par exemple : dans notre corpus nous avons rencontré la lexie "troisième âge". Son réseau, selon le schéma de B. Pottier serait la suivant :

PARADIGMATIQUE	jeunesse (opposition)	
vieillesse (inclusion)	troisième âge	"relève la tête" (ass)
	inactifs (participation)	SYNTAGMATIQUE

Il nous paraît souhaitable de structurer le fichier des "synonymes", selon les relations que nous venons de citer. Ces relations en effet, permettent de delimitier le nombre de termes susceptibles d'être inclus dans le fichier en raison de leur insertion dans un des cadres définis par elles. Il ne s'agit pas pour nous ici d'élaborer un modèle de la synonymie. Par contre, nous proposons d'établir un lexique des synonymes de la manière suivante :

- relever statistiquement les termes les plus fréquents dans VIVA;
- établir les synonymes pour les termes qui peuvent en recevoir, par la constitution des relation d'inclusion, d'opposition, de participation ("extraction selective; R. Martin) et d'association pour chacun d'eux.

L'utilisation des synonymes pourrait être la suivante :

- l'utilisateur formule sa question.
- Si le système ne peut donner une réponse, il demande au lexique des synonymes le mot correspondant "générique".
- A partir de ce mot, on peut relancer la recherche, soit en cherchant les SN incluant le terme "générique", soit en recherchant les SN des autres termes spécifiques.

Remarque : En pratique l'utilisation de ce lexique poserait quelques problèmes dont il faut faire état.

Reprenons l'exemple : /la prévention de la délinquance/

/ la prévention de la criminalité/

Nous constatons en effet qu'entre les termes "criminalité" et "délinquance" il existe une concordance en genre (féminin) et en nombre (singulier). Le premier syntagme existe dans le corpus. Le second serait la question que poserait l'utilisateur au système. Or, rien ne nous assure que le terme générique ou spécifique de la séquence établie par l'utilisateur sera en accord de genre et de nombre avec le terme existant déjà dans le texte. Autrement dit, si dans le texte on parle de "véhicules" (masculin, pluriel) et que l'interrogateur utilise sa "voiture" (féminin, singulier), le système aura quelque peine à faire la correspondance, si un traitement spécifique n'est pas établi.

Dans notre exemple, le système devra :

- "reformuler" la question de l'utilisateur en remplaçant /la voiture/ par /le véhicule/. Cette opération ne pose pas de problèmes car le terme/véhicule/ ne doit pas subir d'autres modifications pour être inséré dans le syntagme. Il est possible par contre que la séquence totale, ait besoin d'autres ajustements.

- Détecter si le nouveau syntagme obtenu, se trouve dans le texte. Si ce n'est pas le cas, il y a deux possibilités :

(a) utiliser les troncatures afin d'obtenir les formes les plus proches;

(b) détecter dans le texte les formes se rapprochant de celles de la question et en dresser la liste.

LES DONNEES NUMERIQUES :

A ce niveau de notre réflexion, nous considérons comme données numériques les données suivantes : date, âge, taille, adresse. Il est évident que ces données ne sont pas strictement numériques ⁴ Nous avons déjà signalé que ce type de données constitue une grande part de l'information véhiculée par VIVA et que l'on pouvait pas donc les négliger.

- Les dates : nous avons déjà dit que les dates n'ont pas une écriture canonique dans VIVA. On trouve par exemple, des formulations aussi différentes que "le 04.02.89", "29.03", "19/12", "le mercredi 25 janvier 89", "le 15 decembre". L'utilisateur humain ne peut pas se tromper sur la signification de chacune de ces dates et peut établir une

⁴ elles sont alphanumériques, si l'on utilise le jargon informatique.

correspondance dans la forme. La machine a besoin de plus d'indications afin d'établir une relation entre ces différentes formulations. Nous proposons l'écriture suivante pour les données de type "date" : DET+ Nom de jour + données numériques (4).

Le système doit reconnaître cette forme. En cas d'absence d'une des données, des éléments 0 devraient être considérés à sa place. Serait-il possible de considérer la séquence "date" ci-dessus comme ayant la forme d'un SN. Nous avons vu là, comme nous l'avions déjà indiqué la possibilité de réutiliser la catégorie K, établie par M. Ramburrun et qui avait posé quelques problèmes à l'analyseur du groupe SYDO. Nous avons pensé éviter ces problèmes et analyses parasites en la réservant uniquement aux données de type "date" et à elles seules, afin qu'il n'y ait aucune ambigüité.

Remarque : Les données de type "âge" se présentent souvent sous la forme "13 a" ou "13/17 a".

Il nous semble nécessaire d'exiger une canonisation de l'écriture des données de ce type, qui pourrait être, par exemple, la suivante : "13 ans". De même "13 à 17 ans. Le problème qui serait posé à l'analyseur serait le suivant : supposons qu'un interrogateur demande, par exemple, les stages de ski pour des adolescents de 16 ans. Qu'aurait fait le système ? :

Le système est constitué de trois fichiers, comme indiqué précédemment. Dans ce cas, les données "numériques" seraient séparées des autres données. Pour les données de type "âge" et plus spécialement pour des données créant un intervalle [13, 17], un traitement informatique spécifique serait nécessaire. Rappelons ici que nous avons proposé de considérer la séquence [13 à 17] comme ayant une fonction d'épithète sur le terme "ans". Cette analyse nous a permis d'isoler cette séquence et de lui imposer un traitement informatique.⁵

En effet, sans un tel traitement, l'utilisateur aurait eu comme réponses toutes les pages concernant les stages de ski et, dans ce cas, il n'y aurait eu aucun gain de temps et d'efficacité par rapport au système actuel. Si, par contre nous demandions au système de tenir compte des données

⁵ Dans le cas des trois fichiers, nous pouvons éventuellement ne pas traiter l'aspect "lexématique" de la séquence "âge" et traiter directement les données numériques. Le défaut d'une telle structure est que différents fichiers subissent des traitements différents par des programmes différents, ce qui multiplie les temps de réponse.

numériques, le nombre de réponses serait sensiblement inférieur et l'utilisateur pourrait trouver bien plus facilement l'information désirée.

LA SUITE ...

Effectivement, suite aux modifications apportées à VIVA après le décès du maire, plusieurs points de nos propositions ont été revus, corrigés ou évincés. Il est ici important de dire que la solution que nous tenons actuellement, est à l'origine de M. LeGuern. En effet, et depuis bien longtemps, il nous avait suggéré de ne garder qu'un seul fichier, le seul et unique fichier des SN du texte, titres et textes pleins inclus. Nous pensons actuellement que c'est là la meilleure solution. Le système des fichiers, avec des traitement en PROLOG nous paraît trop lourd pour un petit (quand-même) service dont le contenu et /ou la présentation peuvent subir des modifications importantes. A quoi aurait servi un tel système pendant la période de quatre(et plus) mois, pendant lesquels VIVA n'avait pas de sommaire ? Le système serait bien plus simple en utilisant un seul fichier SN, et comportant des clauses spéciales pour le traitement des données de type "âge" ou "date". Ces clauses fonctionneraient comme des clauses de reconnaissance de SN et ceci dans un but purement pratique. Par exemple : une clause concernant l'âge pourrait avoir grossièrement, la forme suivante : AGE --> J+ ANS, où J serait sous une forme numérique ou sous une forme plus complexe avec préposition (J).

Il est bien évident qu'il ne faudrait pas augmenter occasionnellement les clauses de la grammaire soustendant ce système. Les clauses concernant l'âge et la date sont secondaires mais nécessaires vu la présence de ces données dans le corpus. En dernier, nous maintenons l'organisation générale proposée pour les synonymes et leur traitement.

CHAPITRE 6 PERTINENCE DE L'APPROCHE PAR SN POUR LA REQUISITION DE L'INFORMATION TEXTUELLE

Dans une optique interprétative (et non générative) des textes, quel outil morpho-syntaxique choisir pour décrire non pas exhaustivement un texte mais pour le saisir à son niveau proprement informatif?

Si l'on considère le sémantisme d'un texte comme la construction d'un modèle partiel du monde, son niveau informatif consiste en ce qu'il le détermine référentiellement, après en avoir construit une représentation globale, telle qu'elle est prescrite par le *système fonctionnel*, les normes *sociolectales* et *idiolectales* (François Rastier *Sémantique interprétative* PUF, Paris 1987, page 40) Or un texte décrit des procès autant que des entités "substantielles"

"Ma fille, ne vois-tu rien venir?" Outre l'identité du destinataire, cet énoncé ne renvoie précisément à ... "rien". Il définit seulement le type d'un événement possible dont la réalisation correspondra à l'instanciation d'une valeur de vérité. Mais la nature de l'événement est ici implicite et requiert une connaissance du contexte (à la fois comme co-texte et comme entour, au sens de RASTIER). Pour qui ignore ce contexte ce n'est pas le SV qui sera d'abord pertinent, mais bien les **SN précisant les types d'actants et leurs attaches déterminées au monde.**

Puisqu'il s'agit de représentation du monde et donc de connaissances, pourquoi ne pas s'adresser aux recherches sur les réseaux sémantiques? Ces recherches semblent donner une nouvelle vigueur aux recherches en sémantique linguistique, en les confirmant et en leur apportant de nouveaux outils (CF biblio/le numéro de la revue *Langages*). Toutefois, ces recherches consistent plus en une notation des connaissances, qu'en une étude de l'usage qui en est fait. Par exemple, les recherches sur les mécanismes de réquisition des connaissances ne distinguent pas entre la recherche de connaissances et la recherche d'informations. Convenons de dénommer "interrogation" celle-ci, et "réflexion" celle-là. La réflexion, (considérée ici comme recherche et non comme construction de connaissances), est censée se dérouler en suivant lesdits "réseaux sémantiques". Les connaissances ainsi mobilisées

appartiennent au savoir d'un sujet plus ou moins abstrait. Elles ont une valeur gnoséologique (plutôt qu'informative) dont l'impact concret n'est pas évalué. En revanche, dans le cas de l'interrogation d'une base de données, ce qui est recherché est moins une connaissance qu'une information, c'est à dire un nouveau savoir sur l'état du monde extra-linguistique, et extra-cognitif. Ou encore (pour reprendre une distinction de M. Desclés dans la revue citée), la "réflexion" se meut parmi des valeurs référentielles (entités posées à l'intérieur d'énonciations); tandis que **l'interrogation cherche l'accès à des référents proprement dits** (entités extra-linguistiques). Certes une base de données demeure une donnée sémiotique dont seule la mise à jour exige le recours à des référents extérieurs. Il reste que l'interrogation d'une donnée sémiotique demeure une démarche qui articule des procédures linguistiques et des procédures cognitives d'un autre type. Une donnée sémiotique, en tant qu'informative, doit être articulable à la démarche basale d'interrogation et par conséquent elle doit refléter la structure de l'accès aux référents (notamment, un texte fictionnel doit simuler cette structure référentielle de base) Précisons : l'interrogation d'une donnée textuelle est une sous-classe de l'interrogation cognitive; la base textuelle doit donc refléter la structure référentielle à laquelle l'interrogation cognitive est ordonnée. Or l'accès à des référents comporte, outre une étape de formulation verbale, des étapes instrumentales ordonnées finalement à la perception. Si je veux savoir l'heure, je peux me contenter de la réponse qu'on me donne. Je devrai alors accorder une présomption de confiance à mon interlocuteur qui aura dû vérifier l'heure par des procédés extra-verbals avant de me la transmettre. Bref, afin d'acquérir une information authentique sur le monde, je dois articuler mes activités logico-verbales à des activités d'un autre type. Question psycholinguistique que l'on peut formuler schématiquement ainsi: **quelle est l'opération linguistique de base qui permet au langage de se connecter à la réalité extra-linguistique?** Comme il est impossible d'étudier séparément le fonctionnement langagier d'une part, et la structuration du réel extra-linguistique d'autre part, c'est de l'intérieur du langage qu'il faut le voir se dépasser vers son extérieur. On pourra donc considérer la démarche onomasiologique interrogative comme un bon indicateur de la démarche semasiologique à visée référentielle. Comme nous le notions plus haut, au cours d'une interrogation préliminaire et cursive d'un texte, les SV

devraient être négligés, si ce texte est lu dans une visée informative (bien entendu, seules des enquêtes expérimentales pourraient spécifier cette hypothèse, avant même de la confirmer ou de l'infirmier) Cela parce que les SV sont présupposants., de même que sur le plan des "schèmes d'entendement", chers à B.Pottier, les comportements présupposent des entités. En effet, même si "l'événement" est l'unité informative élémentaire, celle-ci se constitue en deux étapes. La première étape, pré-informative, consiste à localiser l'information possible. Et cela se fait en déterminant des entités ou supports susceptibles de comportements, ou apports. Si les entités (et donc les SN sur le plan verbal) ont cette priorité dans l'accès à l'information, c'est qu'elles correspondent aux référents de type nominal (ou "substances" au sens d'Aristote). Celles-ci ont une stabilité extra-linguistique qui fonde leur saillance perceptive et leur prégnance biologique (au sens de R.Thom repris par Pottier), lesquelles les constituent en entités-repères et supports de l'action. Précisons encore une fois: la lecture à visée informative (spécialisation de la lecture ordinaire), n'est pas une lecture globale cursive, mais bien une lecture sélective qui négligera toutes les informations de décodage n'ayant pas une valeur informative référentiative. Pour ce faire, elle devrait se centrer sur les SN.

CONCLUSION

Notre travail est un travail **modulaire**. Son but était de rendre compte de la possibilité nouvelle d'interrogation d'une base de données grand-public, telle que VIVA, et de l'intérêt que cette nouvelle méthode peut présenter pour l'utilisateur humain. En pratique, cette méthode ne peut aboutir sans un minimum de conditions pratique en amont...

EN AMONT...

Nous devons souligner qu'avant tout changement de système d'interrogation, un certain nombre de "défauts" doivent **disparaître** de VIVA.

Le premier défaut est celui des fautes de frappe. On me rétorquera qu'aucun document n'est à l'abri des erreurs de frappe. C'est exact. Mais lors de nombreuses consultations de ce journal télématique, nous avons rencontré beaucoup de fautes de frappe. Il nous semble qu'une correction double doit survenir avant le passage définitif de l'enregistrement dans le menu. Le problème des fautes de frappe risque, par ailleurs, de rendre inopérantes les solutions que nous proposons. La petite astuce des troncatures ne peut résoudre qu'un nombre limité de problèmes. Ce système est d'ailleurs destiné à contourner les erreurs, inévitables, de frappe que pourrait faire l'utilisateur. Celui-ci, rédigeant rapidement sa question, est plus susceptible de commettre des fautes (de français ou de frappe). Si à ces fautes, très probables de l'utilisateur, s'ajoutent les fautes existant dans les pages-écrans de VIVA, l'ensemble compromet sérieusement la réussite de l'interrogation. Il est donc indispensable que le service aie des textes sans faute (autant que possible).

Un deuxième point qu'il faut soulever, c'est la pauvreté de la typographie. Aucun accent n'y apparaît, ce qui introduit un nombre d'ambiguïtés considérable. Le service gagnerait en clarté, en efficacité et donc en nombre de clients, à établir ses textes dans une écriture plus "standart" du français.

48

Notre troisième remarque va également dans ce sens: utiliser de manière plus "correcte" les signes de ponctuation.

Voici un exemple de texte, pris dans notre enregistrement de décembre 88:

SERGE LAZAREVITCH TRIO LE 26/01
Lazarevitch diplome de le celebre
"Beklee School" de Boston, une
etonnante musicalite.

a 21H00,
au "61", ...

Le nombre des ambiguïtés générés par le manque d'accents et la mauvaise ponctuation, est disproportionné par rapport à la taille de cette petite information.

Un dernier point qui mérite d'être soulevé: les erreurs dans la segmentation de mots, compromettent l'efficacité du système. Si le texte de VIVA est directement saisi sur un clavier du type minitel, quelle peut être la raison d'une saisie comme "ajour - hui" ? (voir notre expérience du 25 août 88). Il est évident qu'un effort est à faire.

Ces différents points soulevés se posent en amont de toute modification du système, afin de lui garantir un minimum de succès.

LE SYSTEME

On pourrait nous reprocher le fait que nous n'avons pas cherché à établir une base de données, fondée sur le modèle relationnel, dont les lignes maîtresses ont été exposées par M. Metzger dans la dernière partie de sa thèse d'Etat.

Il y a plusieurs raisons à cela:

- nous l'avons déjà dit plusieurs fois et au risque de nous répéter encore, nous soulignons le fait que VIVA est un journal municipal (donc un journal à faible envergure). Ceci veut dire que VIVA n'a pas d'intérêt à garder des données si ce n'est pour un usage interne dans le service . Il est donc **inutile** de construire pour Viva un système de base de données géré par un système de genre INFORMIX.
- Un autre défaut, lié à la nature des systèmes documentaires, est que ceux-ci demandent l'usage d'un langage d'accès spécial. Les questions sont formulées dans un langage qui est cryptique pour le grand public. Les réponses le sont tout autant. Il est alors **hors de question**, en toute logique, de proposer un tel système. Il n'aurait aucune chance d'aboutir dans l'état actuel des choses.
- Le responsable du service ne peut concevoir qu'un système auxiliaire, qui serait annexe au système déjà existant, et qui permettrait aux personnes qui connaissent bien VIVA d'avoir une réponse précise. Donc, un système secondaire avec un but bien précis. Il serait gratuit, d'un point de vue opérationnel, de proposer à VIVA une base de données, qui nécessiterait par ailleurs des investissements en personnes et matériel, qui eux ne seraient nullement gratuits.

Les propositions que nous avons faites, tiennent compte de la réalité de VIVA et surtout des possibilités du service. Les solutions que nous proposons peuvent être facilement réalisées par des informaticiens, pour qui elles n'ont rien de nouveau, sauf peut-être l'aspect linguistique du traitement de l'information et les infidélités de celui-ci sur la conception informatique.

Les solutions que nous proposons ont été conçues pour VIVA et valent pour VIVA et éventuellement pour d'autres services du même genre. . Nous n'avons nullement l'intention de tenter de les appliquer à n'importe quel service télématique de genre "journal".

EN AVAL...

L'impact effectif de nos propositions ne peut être évalué, faute d'application. Sur le plan de la recherche, il est certain qu'une partie importante de travail reste à faire; à savoir quelques décisions à prendre sur des questions théoriques, notamment linguistiques et sur la

mise en oeuvre pratique où certains points de notre système seraient amenés à être modifiés afin de mieux répondre aux attentes.

Ceci est inévitable et n'enlève rien à l'intérêt du système comme guide de la mise en oeuvre pratique.

Le facteur le plus difficile à cerner est l'utilisateur de VIVA, à qui on demande de changer d'habitudes d'interrogation, d'apprendre à formuler une question d'une certaine manière, à utiliser des signes, etc... L'utilisateur aura besoin d'un certain temps d'adaptation et une certaine campagne d'information auprès du public s'avère indispensable. Cette campagne devrait mettre l'accent sur la "responsabilité" que le système octroie à l'utilisateur, celui-ci n'étant plus obligé de subir (et de payer) des pages d'information qui ne l'intéressent guère. Il pourrait désormais imposer un choix et demander une réponse. Notre système n'étant pas omnipotent, la présence de la version actuelle de VIVA peut résoudre éventuellement quelques cas problématiques.

Nous pensons que les services télématiques qui adopteront des systèmes de traitement de l'information nouveaux, (à savoir des systèmes qui associent l'informatique à la linguistique ou éventuellement d'autres sciences) seront certainement favorisés. La méthode d'interrogation à l'aide des SN a l'avantage supplémentaire d'être proche de l'utilisateur humain psychologiquement. Il est très difficile de démontrer le mécanisme de génération de l'interrogation et plus généralement du passage du cognitif à linguistique. On ne pouvait donc y parvenir dans le cadre de ce mémoire. Cette question nous préoccupe beaucoup. En fin, nous voulons signaler que dans leur article, intitulé "Activités mnémoniques et nature linguistique des associations", publié dans le *Bulletin de Psychologie*, G. Oléron, J. Ségui et A. Lieury ont fait état de leurs expériences. Une de leurs hypothèses de départ, qui a été confirmée par les résultats de l'expérience était :

" Les paires de mots qui constituent une séquence syntaxique minimale de la langue (paires «syntagmatiques») doivent être plus facilement retenues¹ que celles qui ne constituent pas une telle séquence", (page 1946). Notons que par "paire" syntagmatique", ils considèrent

¹ C'est nous qui soulignons ce terme. Les séquences nominales (nom adjectif) sont mieux mémorisées que les autres types de séquence (par exemple nom - nom).

51

" l'association «château féodal» qui "constitue une séquence syntaxique minimale de la langue;" (idem page 1946).

L'intérêt d'un mode d'interrogation par SN est évident. Reste à savoir quels services auront l'audace d'investir dans des projets de ce genre.

BIBLIOGRAPHIE

- Beaugrande (de) R. Les contraintes générales qui affectent les processus de compréhension
Bulletin de Psychologie, tome 35, N°356, pages 683 - 694.
- Berrendonner A. *Cours Critique de grammaire générative*. P.U.L. Lyon 1983.
- Caverni Jean-Paul
Claude Bastien,
Guy Tiberghien *Psychologie Cognitive. Modèles et Méthodes*. P.U.G. Grenoble. 1988.
- Chevalier Jean-Claude... *Grammaire Larousse du français contemporain* Ed Larousse, Paris, 1964.
- Descles Jean-Pierre Réseaux Sémantiques : La nature logique et linguistique des relateurs, in *LANGAGES*, septembre 1987, pages 55-77
- Dijk (van) Teun A. Attitudes et compréhension de textes. *Bulletin de Psychologie*, tome 35, N° 356, pages 557 - 569.
- Dubois J. L. Guespin, ... *Dictionnaire de Linguistique*, Ed Larousse, Paris, 1793.
- Ducrot O. et T. Todorov *Dictionnaire encyclopédique des sciences du langage* Ed Seuil, Paris, 1972.

- Ehrlich Stéphane Construction d'une représentation de texte et fonctionnement de la mémoire sémantique. *Bulletin de Psychologie*, tome 35, N° 356, pages 659 - 671.
- Hormann Hans Les déterminants latents de la compréhension. *Bulletin de Psychologie*, tome 35, N° 356, pages 741 - 747.
- Jakobiac François *Maîtriser l'information critique*, Paris 1988
- Kekenbosch Christiane Identification de relations sémantiques et mémorisation. *Bulletin de Psychologie*, tome 35, N° 341, pages 887 - 890.
- Klix F, J. Hoffmann et E. van der Meer. Le stockage de concepts et leur utilisation cognitive. *Bulletin de Psychologie*, tome 35, N° 356, pages 533 - 543.
- Mandel Jean M. Recherches récentes sur les grammaires de récit. *Bulletin de Psychologie*, tome 35, N° 356, pages 705 - 715.
- Metzger Jean-Paul *Syntagmes nominaux et information textuelle. Reconnaissance automatique et représentation.* Thèse d'Etat, Université Lyon 1, octobre 1988.
- Morton John Le lexique interne. *La Recherche*, N° 143, avril 1983, pages 643 - 649.
- Oleron G. J. Ségui et A. Lieury Activité mnémiques et nature linguistique des associations *Bulletin de Psychologie*, tome 23, pages 1945 - 1959.

- Pottier Bernard *Linguistique Générale*, éd Klincksieck, Paris, 1974.
- Pottier Bernard Linguistique et Intelligence Artificielle LANGAGES, septembre, 1987 , pages 21-31.
- Rastier François *Sémantique Interprétative* P.U.F Paris, 1987.
- Rastier François Représentation du contenu lexical et formalismes de l'intelligence artificielle. in *LANGAGES* septembre 1987, pages 79 - 101.
- Rey Brigitte La mémorisation des mots catégorisés et non catégorisés. Etude génétique. *Bulletin de Psychologie*, tome 26, pages 458 - 463.
- Sabah G. Un système de compréhension du langage naturel. *Bulletin de Psychologie*, tome 35, N° 356, pages 789 -797.
- Sabah G. *L'IA et le langage*. Tome 1 Ed Hermès, Paris, 19
- Sanford A.J et S. Garrod Vers la construction d'un modèle psychologique de la compréhension du langage écrit. *Bulletin de Psychologie*, tome 35, N° 356, pages 643 - 647.
- Vezin Jean-Francois et Liliane *Compréhension de textes et intégration cognitive.* *Bulletin de Psychologie*, tome 35, N° 356, pages 649 - 657.

Voutsinas Dimitri

Interrelation des structures affectives et cognitives. *Bulletin de Psychologie*, tome 33, N° 343, pages 51 - 60.

Thèse
JP Metzger



BIBLIOTHEQUE DE L'ESSIB



801289B