

Au terme de ce travail, je tiens à exprimer mes remerciements et mon profond respect à mon directeur de thèse, Jacques-Philippe Saint-Gérard, qui n'a jamais manqué d'attention et n'a jamais cessé de m'encourager à poursuivre ce travail.

Je remercie les élus du Conseil Général de l'Allier qui ont accepté de financer ces recherches. Emmanuel Maugenest s'est toujours montré d'une infinie patience, qu'il en soit ici vivement remercié.

Ma gratitude va aussi à tous les relecteurs (et toutes les relectrices) qui m'ont permis d'ajouter quelques nuances à cette étude, voire de modifier la vision que j'avais de certains points. Mylène Blasco-Dulbecco, Cassian Braconnier, Paul Cappeau, Franck Lebas n'ont pas ménagé leur temps pour discuter les points litigieux, imprécis ou obscurs.

Je remercie les amis qui ne se sont jamais laissés bernier par une période de découragement et qui ont supporté mes exaltations et mes doutes tout en essayant de les modérer.

Je remercie ma famille qui, bien que résistante à toute conversation sur les corpus ou sur les constructions verbales, s'est toujours montrée compréhensive et chaleureuse, comme à son habitude.

Enfin, mes derniers remerciements vont à ceux qui m'ont, durant la première année consacrée à ce travail, accordé leur confiance. A ceux qui m'ont confié leur voix, leurs paroles, un peu de leur temps et parfois un peu plus que tout cela encore. Parmi les personnes dont la voix fait partie de notre corpus, certaines sont aujourd'hui disparues. Ce travail leur est dédié.

TABLE DES MATIERES

Introduction.....	p. 7
1. Corpus et transcriptions.....	p. 14
1.1. Présentation du corpus.....	p. 15
1.2. Enquêtes et transcriptions.....	p. 17
1.2.1. Méthode retenue pour les enquêtes.....	p. 19
1.2.1.1. Choix des critères de sélection des locuteurs.....	p. 20
1. Critère géographique.....	p. 20
1. Présentation du département.....	p. 21
2. Présentation des trois villes.....	p. 22
a. Montluçon.....	p. 23
b. Vichy.....	p. 23
c. Moulins.....	p. 24
2. Critère « niveau d'étude ».....	p. 25
3. Critère âge.....	p. 25
1.2.1.2. Recherche des locuteurs correspondants aux critères prédéfinis.....	p. 26
1. Enquêteur connaissant la localité.....	p. 27
2. Enquêteur ne connaissant pas la localité.....	p. 28
1.2.1.3. Matériel utilisé.....	p. 29
1.2.1.4. Enquête non directive.....	p. 30
1.2.1.5. Conditions d'enregistrement (variables contextuelles).....	p. 31
1.2.1.6. Conventions de transcription.....	p. 33
1.3. Corpus : genre, types et thèmes.....	p. 36

TABLE DES MATIERES

1.3.1.	Historique de la notion de corpus.....	p. 36
1.3.2.	Nécessité d'une réflexion sur le statut du corpus de langue parlée et sur son évolution.....	p. 40
1.3.3.	Définitions (texte, corpus, discours, type, thème).....	p. 42
1.3.4.	Le thème comme base du discours.....	p. 50
1.3.5.	Exemples de requêtes lexicales thématiques Montluçon / Vichy Moulins.....	p. 50
2.	Inventaire et classement des constructions verbales.....	p. 52
2.1.	Les travaux et projets sur les constructions verbales du français.....	p. 53
2.1.1.	Le lexique-grammaire.....	p. 54
2.1.2.	Proton.....	p. 54
2.1.3.	Le Dictionnaire Contrastif des Valences Verbales.....	p. 56
2.1.4.	Les constructions verbales en français parlé. Etude quantitative et descriptive de la syntaxe des 250 verbes les plus fréquents.....	p. 57
2.2.	Pourquoi un inventaire.....	p. 60
2.3.	Définitions.....	p. 66
2.3.1.	Phrase et construction verbale.....	p. 67
2.3.1.1.	Valence et rection.....	p. 69
2.3.1.2.	Extension de la notion de construction verbale.....	p. 75
2.3.2.	Pronoms et pro-formes.....	p. 76
2.3.2.1.	La notion de pronom.....	p. 76
2.3.2.2.	Les pronoms clitiques	p. 86
2.3.2.3.	Les pro-formes.....	p. 91
2.3.3.	Eléments construits.....	p. 94
2.3.3.1.	Un élément construit a une équivalence avec une pro-forme.....	p. 96
2.3.3.2.	Un élément construit peut porter les modalités du verbe.....	p. 97
2.3.3.3.	Un élément construit peut être focalisé.....	p. 98

TABLE DES MATIERES

2.3.4.	La notion de formulation.....	p. 99
2.3.5.	Problèmes de pronominalisation : le cas du verbe <i>aimer</i>	p. 101
2.4.	Cadre méthodologique utilisé.....	p. 103
2.4.1.	L'approche pronominale.....	p. 103
2.4.2.	Ce que l'on retient de l'approche pronominale.....	p. 107
2.5.	Présentation des données.....	p. 108
2.5.1.	Verbes présents dans le corpus.....	p. 108
2.5.2.	Constructions verbales traitées.....	p. 109
2.5.2.1.	Inventaire des constructions verbales : les formes retenues.....	p. 109
2.5.2.2.	La prise en compte de la fréquence.....	p. 111
2.6.	L'inventaire et le classement des constructions verbales	p. 113
2.6.1.	Présentation du concordancier XCOR.....	p. 113
2.6.2.	La base de données.....	p. 116
2.6.2.1.	Les propriétés syntaxiques et distributionnelles.....	p. 117
2.6.2.2.	Présentation détaillée de quelques verbes.....	p. 123
	1. <i>Aller</i>	p. 123
	2. Les verbes <i>être</i> et <i>avoir</i>	p. 129
	a. <i>Etre</i>	p. 130
	b. <i>Avoir</i>	p. 131
	Conclusion et perspectives.....	p. 133
	Bibliographie.....	p. 141
	Index des noms cités.....	p. 154
	Glossaire.....	p. 156

Annexes.....	p. 157
1. Corpus des constructions verbales.....	p. 158
2. Fiches signalétiques des enregistrements.....	p. 503
3. Tables.....	p. 531

Thèse en 2 volumes

Matériel d'accompagnement :

Un CD-Rom comprenant :

- les enregistrements en format mp3,
- les transcriptions,
- un exemplaire de la thèse et des annexes,
- la base de données au format Access2000.

Introduction :

L'objectif de ce travail est de fournir un inventaire et un classement des constructions verbales relevées dans un corpus de français parlé.

Effectuer un inventaire et un classement des constructions verbales dans un corpus de français parlé peut présenter plusieurs intérêts, dans des domaines différents : concernant l'apprentissage du français, la description de données orales attestées permettra sans doute à l'enseignant d'acquérir un point de vue différent de celui que véhiculent souvent les dictionnaires de langue ou les manuels scolaires, et de contribuer à recentrer son activité pédagogique sur l'usage de la langue, plutôt que sur sa norme.

Concernant le Traitement Automatique du Langage (TAL), nous espérons que ce travail pourra permettre aux informaticiens-linguistes de se fonder sur une description claire et cohérente d'un pan du fonctionnement de quelques verbes français, afin d'établir les règles distributionnelles et de parvenir à une formalisation efficace de la langue parlée.

Enfin, nous espérons que ce travail puisse servir à l'intégration de la description du système verbal du français oral dans les dictionnaires de langue, comme l'a fait W. Busse pour ses étudiants allemands et portugais lors de la création du *Französische Verblexicon* (1983) et du *Dicionario sintactico de verbos portugueses* (1994).

1) Corpus

Constituer un corpus de français parlé dans l'Allier présente plusieurs intérêts, le plus général étant l'augmentation du nombre de corpus de français parlé disponibles en France. D'autre part, l'Allier est une localité particulièrement intéressante, puisque le *croissant linguistique*, frontière entre langue d'oïl et langue d'oc, le traverse. Il n'est pas question, bien sûr, d'étudier ici une forme dialectale du français, puisque c'est la langue dite « standard » que nous avons enregistrée. Cependant, l'hypothèse qui sous-tend la constitution de corpus dans ce département est précisément une variation géolinguistique potentielle.

Ainsi, s'il est transcrit en orthographe normée, selon les conventions du GARS, une transcription phonétique est toujours envisageable, un corpus étant, comme nous le verrons au cours de la première partie de ce travail, ouvert à tous les enrichissements et à toutes les annotations possibles.

Nous présentons, dans la première partie de ce travail, le corpus Allier (1.1.), les choix qui sous-tendent la sélection des locuteurs enregistrés et celle des points d'enquêtes, puis la méthode que nous avons utilisée pour les mener à bien (1.2.).

La sélection des locuteurs est importante si l'on souhaite construire un corpus échantillonné. Nous avons retenu trois critères de sélection (1.2.1.1.), qui sont identiques à ceux retenus par l'équipe DELIC lors de la réalisation du projet « Français de référence ». Le critère géographique prend comme points d'enquêtes les trois villes les plus importantes du département de l'Allier : Vichy, Montluçon et Moulins. Le critère « niveau d'études » est partagé entre le niveau collège, le niveau bac et le niveau bac+3. Enfin, nous avons retenu l'âge des locuteurs comme troisième critère de sélection. Ainsi, les locuteurs se répartissent selon trois tranches d'âge : 18-30 ans, 30-65 ans et plus de 65 ans. Les hypothèses associées à chacun de ces critères sont essentiellement fondées sur les phénomènes de variations (géolinguistique – diatopique –, sociolinguistique – diastratique – et diachronique), mais elles n'ont pas été vérifiées dans le cadre de ce travail, le corpus n'étant ici que la base d'une étude syntaxique. Ainsi, la constitution d'un corpus étant relativement longue, nous avons souhaité dès le départ lui apporter un échantillonnage de ce type, en partant du principe qu'il ne servirait pas uniquement au travail présenté. La présentation détaillée des critères de sélection des locuteurs et des intérêts de ce type de corpus pour des recherches dans des disciplines autres que la syntaxe pourra sembler hors de propos, puisque ces critères ne nous ont pas servi ici. Cette incohérence sera explicitée plus longuement dans le courant de la première partie.

Une fois les critères de sélection des locuteurs définis, il s'agit de trouver les individus y correspondant. Nous proposons ici une méthode de recherche, selon que l'enquêteur connaît ou non la localité des enquêtes. La technique des réseaux denses (Labov, 1972 et Milroy, 1980) a déjà été maintes fois éprouvée, nous savons qu'elle est efficace. Elle permet de réduire le phénomène du paradoxe de l'observateur¹, en enquêtant auprès de personnes familières, de connaissances déjà établies, ou de relations indirectes (l'enquêteur connaît A, A le présente à B, et B devient une connaissance de l'enquêteur, qui peut de cette manière effectuer ses recherches en milieu « acquis », et ainsi obtenir un certain naturel

¹ « To obtain the data most important for linguistic theory, we have to observe how people speak when they are not observed » (Labov, 1973 : p. 113)

langagier). Cependant, lorsqu'il n'y a pas de réseau connu, quelques petits points permettent de mener à bien ses enquêtes, que nous exposerons au cours de la partie 1.2.1.2..

La suite de la partie 1.2. présente le matériel utilisé (1.2.1.3.), le type d'enquête effectué (1.2.1.4.), les conditions d'enregistrement (1.2.1.5.) et les conventions de transcription (1.2.1.6.).

La notion même de corpus a fait l'objet de divers colloques et publications lors de ces dix dernières années, les chercheurs en linguistique lui portant un intérêt manifeste, renouvelé d'ailleurs dans le cadre de l'étude des productions orales. La constitution d'un corpus dépend des objectifs que l'on se donne, c'est pourquoi nous avons souhaité réserver une partie de ce travail à une réflexion d'ordre épistémologique sur les diverses questions corollaires de la notion de corpus : les types, les genres, la thématisation. Dans cette partie (1.3.) nous reprenons un historique de la notion de corpus présenté par C. Blanche-Benveniste (*in* Bilger, 2000) (1.3.1.), puis nous proposons une réflexion sur le statut du corpus de langue parlée et sur son évolution (1.3.2.). Dans la partie 1.3.3., nous posons quelques définitions « basiques » (celles de texte, de corpus, de type et de thème). Dans la suivante (1.3.4.), nous montrons l'importance du thème pour l'enregistrement de discours, et signalons une incohérence thématique pour un tiers des enregistrements, qui n'a pas nuit à la cohérence de l'inventaire, comme nous le démontrons au cours de la partie 1.3.5..

2) Inventaire et classement des constructions verbales

Inventorier et classer les constructions verbales dans un corpus de français parlé n'a, à notre connaissance, jamais été fait. De nombreuses études sur les constructions verbales ont été menées, qu'il s'agissent des ouvrages publiés par l'équipe du LADL, de Willems (1981), de Greidanus (1990), ou encore de Proton (2002).

La définition même d'une construction verbale peut s'avérer problématique, tant il est fréquent de douter du statut d'un élément : est-il construit ou non par le verbe, dépend-il du verbe ou de la phrase, est-il ce que Bonami (1999) nomme un argument, ou un ajout ? Appartient-il à la valence, ou à la rection ? Trois tests permettent de distinguer, dans la plupart des cas, un élément construit d'un élément associé. D'une part, un élément construit a une équivalence avec une pro-forme ; d'autre part, il peut porter les modalités du verbe ; enfin, il peut être focalisé (c'est-à-dire introduit dans une tournure en *c'est...que*).

Ce travail permet d'observer la complémentation verbale telle qu'elle apparaît dans l'usage, et non pas dans la norme, comme le font la plupart des dictionnaires, y compris Proton. Ainsi, les constructions verbales relevées dans le corpus sont classées selon un ensemble de propriétés qui nous ont paru pertinentes, que nous détaillerons dans la deuxième partie.

Ce travail s'inspire largement de l'approche pronominale (C. Blanche-Benveniste et al. 1984). Les éléments construits par le verbe se trouvent donc dans la table sous forme pronominale, qu'il s'agisse d'un pronom clitique, ou d'une pro-forme d'une autre nature.

Dans un premier temps, nous présentons quelques travaux et projets importants concernant l'étude des constructions verbales du français, depuis l'entreprise du lexique-grammaire jusqu'à Proton, en passant par les travaux de D. Willems et de T. Greidanus (2.1.).

Après avoir discuté de l'intérêt d'établir un inventaire des constructions verbales attestées en français parlé (2.2.), nous proposons les définitions des principales notions utilisées dans le cadre de ce travail. Depuis quelques années déjà, la notion de phrase proposée par la grammaire dite « traditionnelle » ou « scolaire » est remise en cause par de nombreux linguistes, qui ne la jugent pas pertinente, notamment pour l'étude de la langue parlée, ou encore pour le TAL. Ainsi, nous n'avons pas retenu ici la « phrase simple » comme unité, contrairement au lexique-grammaire, mais la construction verbale (2.3.1.). Cependant, le terme même de *construction verbale* est ambigu : il englobe à la fois les éléments strictement dépendants du verbe, qui appartiennent à la « valence », qui sont dits « obligatoires », et les éléments construits par le verbe appartenant à la « rection », qui sont nommés « facultatifs ». Après avoir défini les notions de valence et de rection (2.3.1.1.), nous montrerons que la distinction entre éléments appartenant à la valence (obligatoires) et éléments appartenant à la rection (facultatifs) n'est pas toujours aisée, mais que le principe même de l'inventaire permet dans bien des cas de lever les difficultés. Toutefois, le terme d'« inventaire » des constructions verbales est problématique ; en effet, dans la proposition suivante :

Jean dit que Marie ment,

il y a plusieurs manières d'envisager ce que l'on nomme *construction verbale* : (1) il peut s'agir de la proposition entière (Jean le dit), et on ne considère alors qu'il n'y a ici qu'une construction ; mais on peut tout aussi bien considérer en plus de (1) que *Marie ment* est une autre construction verbale (2), considérer deux constructions verbales différentes dans cette proposition, et ainsi analyser de façon plus fine et plus précise l'ensemble des verbes. Nous avons retenu ici la deuxième option. Nous expliquerons les modalités de ce choix dans la partie 2.3.1.2..

La notion de *pronom* elle aussi a été très discutée dans la littérature linguistique. Nous observerons dans la partie 2.3.2.1. les problèmes que cette notion pose, puis nous présenterons les pro-formes¹ que nous avons utilisées pour mener à bien notre inventaire et notre classement (2.3.2.2. et 2.3.2.3.).

Au cours de la partie 2.3.3., nous présentons les tests qui permettent de distinguer les éléments construits par le verbe des éléments associés à la construction verbale, tels qu'ils sont présentés dans le cadre de l'approche pronominale : la pronominalisation de cet élément (2.3.3.1.), son aptitude à porter les modalités du verbe (2.3.3.2.), et enfin la possibilité de le focaliser (2.3.3.3.). Nous évoquons brièvement par la suite la notion de formulation telle qu'elle a été envisagée par les auteurs de *Pronom et Syntaxe* (2.3.4.) ; cependant, nous n'avons pas retenu dans ce travail l'étude des formulations et des reformulations possibles d'un verbe. Lorsqu'elles apparaissent dans le corpus (et dans la table), aucune précision particulière n'est apportée.

Le premier test qui permet de reconnaître un élément construit, la pronominalisation, peut s'avérer délicat : nous présenterons à titre d'exemple au cours de la partie 2.3.5. une difficulté que nous avons rencontrée lors de la pronominalisation de certains éléments du verbe *aimer*.

En 2.4., nous présentons le cadre méthodologique que nous avons utilisé, qui s'inspire très largement de l'approche pronominale (2.4.1.). Nous expliciterons les points que nous avons retenu de cette approche dans la partie 2.4.2..

¹ Nous rangeons sous l'étiquette *pro-forme* à la fois les pronoms clitiques et les pro-formes d'autre nature.

La partie 2.5. est consacrée à la présentation des données. Dans un premier temps, nous présentons brièvement le nombre de verbes présents dans le corpus (2.5.1.) ainsi que les constructions verbales que nous avons traitées (2.5.2.1.). Nous avons pris en compte la fréquence d'apparition des constructions verbales (2.5.2.2.). Willems (1981 : pp. 22-23) indique que « [l'influence de la fréquence] dans les études syntaxiques et sémantiques est restée assez faible jusqu'à présent. Elle y joue toutefois un rôle important, particulièrement en ce qui concerne le verbe, qui, dans l'échelle des fréquences, peut atteindre un degré très élevé et se répartir jusqu'aux plus basses fréquences. Intuitivement, nous pouvons en effet vérifier que les verbes les plus fréquents se caractérisent par une syntaxe diversifiée et un sémantisme pauvre, l'inverse étant vrai pour les verbes les moins fréquents. De même, à l'intérieur d'un seul lexème polysémique, la fréquence d'emploi des diverses constructions peut jouer un rôle dans l'analyse sémantique. Fréquence, distribution et sens entretiennent donc des relations étroites. Pour mesurer ce lien de façon objective, il faudrait toutefois disposer, pour l'ensemble des verbes, de leur fréquence par distribution et par sens lexical. » Si la fréquence d'apparition des constructions verbales est relative, puisque nous utilisons un corpus qui est, par définition, limité, elle indiquera cependant quelques grandes lignes qui permettront pour des recherches futures de vérifier les hypothèses de Willems. Nous n'avons pas ici traité ce problème, nous limitant à la tâche indiquée ci-dessus, sans analyser les résultats du traitement fréquentiel.

Dans la partie 2.6., nous présentons de manière détaillée les outils que nous avons utilisés pour l'inventaire et le classement. Dans un premier temps, nous proposons donc une description du concordancier XCOR2000, développé par Jean-Pierre Adam pour les travaux du GARS (2.6.1.), puis nous explicitons au cours de la partie consacrée à la base de données (2.6.2.) les différentes propriétés que nous avons utilisées pour mener à bien ce classement (2.6.2.1.). L'intégration des différentes propriétés utilisées dans la table ne s'est pas faite sans problèmes ; comme le signale Willems (1981 : p. 15), « l'établissement même d'une typologie syntaxique qui permettrait de classer les verbes pose certains problèmes préalables concernant les techniques de découverte des propriétés, leur hiérarchisation, le statut de certaines constructions, telles que les constructions pronominales, impersonnelles, absolues ou passives ». D'autre part, nous présentons quelques verbes de manière détaillée (2.6.2.2.), cette présentation faisant en quelque sorte office de mode d'emploi pour la lecture de la table et du corpus qui lui est associé.

Pour une question d'équilibre, nous avons décidé de faire apparaître la table et le corpus en annexe, ainsi que les fiches signalétiques des enregistrements. Ainsi, la première annexe est consacrée à la table, la deuxième au corpus des constructions verbales, et la troisième aux fiches signalétiques.

1. Corpus et transcriptions

Cette partie est destinée à présenter le travail effectué en préalable à l'analyse syntaxique à proprement parler ; en effet, durant une année, un corpus a été constitué, depuis la « récolte » de français parlé sur le terrain – dans les trois principales villes de l'Allier – jusqu'à la transcription minutieuse de ces relevés. Nous donnerons les références précises des différents sous-corpus lors de sa présentation dans la première partie (1.1.) de ce chapitre.

Le premier enregistrement de « français parlé » date de 1911, date à laquelle Ferdinand Brunot crée les Archives de la parole¹, la « première collection phonographique institutionnelle de France », et qui sera aussi la « première pierre de l'Institut de Phonétique voulu par l'Université de Paris ». Brunot la voulait « un lieu d'enregistrement et de conservation pour les générations futures des manifestations orales de la langue ». Les objectifs de Brunot étaient quadruples, et sont encore d'actualité : le premier était « d'enregistrer de la parole au timbre juste, au rythme impeccable, à l'accent sûr, comme la parole nuancée d'accents faubouriens ou provincial » ; le second, « d'enregistrer les patois et dialectes », aux fins de la constitution d'un atlas linguistique de la France (travail conçu un peu plus tard par Gillieron) ; le troisième était d'« enseigner les langues étrangères », un Français Langue Etrangère avant l'heure, et enfin, le quatrième, qui trouve un puissant écho à l'heure actuelle, le traitement des pathologies du langage. La qualité des enregistrements est à présent assez faible ; se pose ici toute entière la question de la conservation des documents sonores, tant il est communément admis aujourd'hui que notre CD ne sera pas éternel, pas plus que ne l'ont été les disques du début du vingtième siècle.

Depuis lors, des enregistrements de français parlé ont été réalisés, tant en France que dans d'autres pays francophones. Nous aborderons ce sujet lors de la partie 1.3..

A présent, venons-en à la description de la méthode que nous avons utilisée pour construire notre corpus.

On ne part pas muni de son micro au gré des rues, le travail d'enquête doit respecter quelques règles. En effet, la constitution d'un corpus linguistique se fait en trois temps (Delais-Roussarie, E., Durand, J. (éd.) 2003, pp. 91-92) :

¹ <http://gallica.bnf.fr/ArchivesParole/>

- le travail préparatoire qui se fait avant la phase de saisie ou d'enregistrement des données ;
- le travail de saisie ou d'enregistrement des données proprement dit ;
- le travail de mise en forme des données afin de les rendre exploitables (transcription et annotation).

Nous exposerons donc ces trois étapes au cours de la partie 1.2. Enfin, la troisième (1.3.) pose les différents intérêts et types de problèmes relatifs aux corpus de langue parlée, à partir de réflexions d'analystes du discours et de spécialistes des corpus.

1.1. Présentation du corpus

Nous avons constitué durant l'année 2001 un corpus d'environ 100 000 mots, à partir d'enquêtes effectuées à Montluçon, Vichy et Moulins, auprès de 9 locuteurs dans chacune de ces villes (les critères de sélection de ces locuteurs seront précisés dans la partie consacrée aux enquêtes).

La durée de chaque enregistrement est de 30 minutes, dont 15 sont transcrites, selon les conventions données par l'équipe DELIC (*cf.* [1.2.1.6. Conventions de transcription](#)). Ces enregistrements sont donc transcrits et informatisés (son et texte sont disponibles sur CD-ROM), voici leurs références :

	MONTLUCON	MOULINS	VICHY
18-30 ans niveau collège	CHOCM~10	COMMO~14	HOTEV~26
18-30 ans niveau bac	MUMON~44	FANMO~16	JOLIEV~32
18-30 ans niveau bac+3	SANDM~58	FRUIT~22	MARCH~40
30-65 ans niveau collège	MADMO~34	COIFM~12	MONOV~42
30-65 ans niveau bac	PAPMO~50	GILMO~24	QUALV~56
30-65 ans niveau bac+3	MUSIK~46	FRANM~20	JAMES~30
+ 65 ans niveau collège	MAMMO~38	MALLM~36	BIJOU~6

+65 ans niveau bac	FRAMO~18	INSTI~28	NINAV~48
+65ans niveau bac+3	BRUNMO~8	PHYMO~54	PETIV~52

Tous ces enregistrements sont de type « parole privée », c'est-à-dire qu'ils revêtent la forme d'un entretien entre le locuteur interrogé et l'enquêteur (le temps de parole étant occupé à environ 95% par le locuteur interrogé).

Les 9 corpus de Montluçon n'ont pas de thème bien défini : il s'agissait souvent pour le locuteur, par exemple de nous raconter un fait marquant de sa vie ; en revanche, les corpus de Vichy et de Moulins ont été réalisés à partir du thème « petit commerce » (nous expliquerons cette incohérence thématique dans la partie consacrée aux enquêtes).

Nous utiliserons fréquemment les notions de corpus et de sous-corpus. Cependant, le terme *sous-corpus* risquera d'être ambigu, dans la mesure où il peut à la fois référer :

- au premier niveau¹, que nous posons comme le plus « précis », c'est-à-dire celui à l'intérieur duquel on dénombre 27 sous-corpus, soit un par locuteur.

- au deuxième niveau, que nous posons comme le moins « précis », dans la mesure où les variables qui le composent, fondées sur les critères de sélection des locuteurs (3 villes, 3 tranches d'âge, 3 niveaux d'études) multiplient de manière très importante les possibilités combinatoires de recherche. On pourra ainsi sous-catégoriser en fonction de la ville (3 sous-corpus : Montluçon, Vichy, Moulins), en fonction de l'âge (3 sous-corpus : 18-30 ans, 30-65 ans, + 65 ans) ou en fonction du niveau d'étude (3 autres sous-corpus : collège, bac ou bac+3). Mais on pourra aussi sous-catégoriser en associant les différentes variables : les locuteurs de tel âge, de tel niveau d'étude, de telle ville, ce qui donne 3³, c'est-à-dire 27 sous-catégorisations potentielles.

Après cette présentation très générale de notre corpus, nous allons à présent détailler la méthode que nous avons utilisée pour effectuer les enquêtes, tout en précisant lorsque cela sera nécessaire les points qui nous ont paru délicats, ou ceux dont la réalisation s'avère imparfaite.

¹ C'est à ce niveau là que nous ferons référence si rien n'est indiqué ; sinon, nous signalerons le sous-type en note.

1.2. Enquêtes et transcriptions

Plusieurs paramètres doivent être pris en compte lors de la constitution d'un corpus : en ce qui concerne l'enquête de terrain, le choix des locuteurs enregistrés est important si l'on veut rendre compte le plus finement possible de l'état d'une langue à un moment donné de son histoire, et les critères de sélection qui précèdent le choix des personnes interrogées permettent d'introduire diverses variables au cours de l'étude (variable géographique, niveau d'étude, et âge), qui n'ont pas été utilisées au cours de ce travail, mais qui pourraient se révéler intéressantes pour des recherches dans d'autres domaines que celui de la syntaxe, nous pensons ici particulièrement à la sociolinguistique urbaine.

Les enquêtes ont été réalisées en partie suivant les indications données par l'équipe DELIC¹. Lors de l'élaboration du corpus *Français de Référence*, il s'agissait pour elle d'enregistrer des locuteurs selon leur âge, réparti en 3 tranches, leur niveau d'étude réparti également en 3 niveaux, et le type de parole enregistrée (3 types de parole : privée, publique et professionnelle). Nous avons repris les deux premiers critères, l'âge et le niveau d'étude, mais n'avons pas enregistré plusieurs types de parole : nous n'avons que des « paroles privées ». Cependant, nous avons ajouté à cet échantillonnage un critère géographique, qui apparaît dans le *Français de Référence*, mais de manière moins équilibrée (27 corpus pour Paris et sa banlieue, 49 pour le nord de la France et 55 pour le sud). En effet, nous avons enregistré 9 locuteurs dans chacune des trois villes de l'Allier, ce qui fournit à ce corpus un équilibre pouvant permettre, pour des recherches futures, une étude de la variation linguistique dans ce département, qui est situé au cœur de ce que les dialectologues nomment le *croissant*, c'est-à-dire la zone où passe la frontière entre langue d'oïl et langue d'oc, et où l'on trouve, au sud-est, un peu de franco-provençal, comme le montre la carte suivante :

¹ « L'équipe DELIC (Jeune Equipe JE 2328) s'intéresse à la description des structures morphosyntaxiques et lexicales en français, en synchronie comme en diachronie. Sa méthodologie repose sur l'utilisation systématique de grands corpus oraux et écrits, à l'aide d'outils informatiques appropriés (concordanciers, étiqueteurs, outils de gestion et de navigation, etc.). » cf. <http://www.up.univ-mrs.fr/delic/index.html>
DELIC, 2002, Projet « Corpus français parlé de référence », rapport final, non-publié.



— : zone située à l'intérieur du Croissant.

Carte de l'Allier.

Lorsqu'il y a enregistrement de discours, plusieurs méthodes sont possibles, selon l'étude que l'on veut mener à bien, car « les différents types de corpus de langue parlée se définissent généralement par les objectifs qu'ils se donnent : faciliter l'enseignement, faire de la recherche linguistique, collaborer à de plus larges recherches statistiques, participer à des recherches de sociolinguistique, etc. » (C. Blanche-Benveniste, *in* Bilger, 2000 : p. 14). Notre étude portant sur la syntaxe, l'intérêt est de laisser libre cours à la parole des locuteurs, afin de pouvoir observer les unités de discours et plus précisément les constructions verbales que nous observerons dans la deuxième partie de ce travail.

Notre enquête est donc « libre », si ce n'est qu'un thème est défini auparavant, ce qui permet au locuteur de parler relativement facilement. Nous entendons par « libre » le fait que l'entretien n'est pas directif : nous n'avons pas élaboré de questionnaire comme l'avaient fait les enquêteurs travaillant sur le *Corpus d'Orléans* (Biggs et Blanc, 1971). Pendant une demi-heure, le locuteur parle d'un sujet défini avant l'enregistrement. Comme nous le verrons lors de cette partie, « la spontanéité de la prise de parole, qui est un facteur intéressant pour

l'analyse des interactions, n'est pas toujours un bon atout pour l'observation grammaticale. Les conversations à bâtons rompus, qui représentent assez bien les situations de parole spontanées, ne livrent généralement qu'une syntaxe limitée, exploitant des schémas répétitifs et souvent interrompus » (C. Blanche-Benveniste, *in* Bilger, 2000 : p. 20). C'est pourquoi nous sommes intervenue le moins possible dans le discours du locuteur : nous le relançons parfois en posant une question ouverte, ce qui permettait de développer un point précis ou de relancer son discours lorsqu'il y avait essoufflement et que, manifestement, le locuteur était « à court » d'idées.

D'autres paramètres se révèlent importants lors de l'enregistrement : le lieu, la position des locuteurs, l'intervention ou non de l'enquêteur, etc. Nous en présenterons quelques uns au troisième point de cette partie ([Conditions d'enregistrement](#)). Auparavant, observons dans le détail les critères retenus pour la sélection des locuteurs, la méthode de recherche de ces derniers ainsi que le matériel que nous avons utilisé.

1.2.1. Méthode retenue pour les enquêtes

Le recueil des données est une étape non-négligeable de la constitution d'un corpus : d'une part, le travail est très long, qu'il s'agisse des enquêtes de terrain ou de la transcription, et d'autre part, la qualité de l'analyse dépend essentiellement de celle du corpus ; c'est pourquoi il nous semble pertinent de proposer une méthode détaillée du travail d'enquête, qu'il s'agisse des critères de sélection des locuteurs ou des problèmes dus, dans notre corpus, à la cohérence thématique ou ceux inhérents à tout enregistrement de discours.

Cette partie montrera donc, chronologiquement, les différentes étapes de la constitution d'un corpus : dans un premier temps, la définition des critères précédant le choix des locuteurs ; dans un second temps, le travail sur le terrain proprement dit, et les différents problèmes qui s'y présentent, car une fois les critères de sélection des locuteurs définis, il s'agit de trouver les individus correspondants. Comme nous l'indiquions plus haut, connaître la région d'enquête s'avère un avantage, du moins à première vue (nous verrons plus tard que cela peut poser des difficultés particulières). Nous présenterons donc la manière dont nous avons travaillé, en tenant compte du fait que des réseaux de connaissances étaient établis depuis longtemps, puis nous tenterons de proposer une méthode alternative pour les cas où l'enquêteur ne connaît pas la région. Enfin, nous indiquerons le matériel que nous avons utilisé pour mener à bien ces enquêtes.

1.2.1.1. Choix des critères de sélection des locuteurs

Pour être pertinemment exploitable à la fois en syntaxe, mais aussi en sociolinguistique, le choix des locuteurs doit être opéré en fonction de plusieurs critères : c'est ce que l'on nomme l'*échantillonnage* ; dans notre cas, nous avons utilisé les critères d'enquêtes utilisés par l'équipe DELIC lors de l'élaboration du *Corpus français parlé de référence*, au nombre de trois : critère géographique (3 zones urbaines de l'Allier, Vichy, Montluçon et Moulins), critère socioprofessionnel (3 niveaux d'études : collège, bac et bac + 3), critère d'âge (3 tranches d'âge : 18-30 ans, 30-65 ans et + de 65 ans). Ces trois critères peuvent servir de variables dans une étude ultérieure, puisque la variation linguistique peut être analysée d'un point de vue géographique, d'un point de vue sociolinguistique ou encore diachronique (l'hypothèse étant qu'un locuteur de + de 65 ans n'utilisera pas nécessairement les mêmes tournures syntaxiques qu'un locuteur de 18 ans, et que, même si l'on sait que la syntaxe n'évolue pas aussi rapidement que le lexique, des tournures sont délaissées au profit d'autres, etc.). Comme le souligne F. Gadet (1971 : p. 78), « le choix du groupe de sujets tient au but même de la recherche.[...] Il faut chercher à ce que tous les sujets réunissent le plus possibles d'invariants, pour que les différences linguistiques notées puissent s'expliquer par un facteur extra-linguistique unique qui variera selon le but poursuivi. »

1. Critère géographique

Le critère géographique est celui favorisé par les dialectologues. Il permet d'observer les variations linguistiques potentielles d'un lieu à un autre si l'on choisit plusieurs lieux d'enquêtes hétérogènes (Nord/Sud, par exemple), ou, au contraire, il permet d'annuler cette variable si l'on choisit des lieux d'enquêtes situés dans une même région, dans un même département ou dans une même ville. Comme le note F. Gadet (1971 : p. 75) dans un de ses articles consacré aux variations sociales de la langue :

L'étude systématique des différences qu'elles [les variations géographiques] déterminent est du ressort de la dialectologie. Elle ne nous intéresse ici que dans la mesure où nous voyons la nécessité de choisir des sujets de même origine régionale, pour éviter l'interférence des variations linguistiques attribuables à des causes sociologiques avec celles dues à des variétés régionales.

Cependant, il est important de noter que le lieu de notre enquête peut être considéré comme non homogène, du fait de sa situation géolinguistique : le département de l'Allier, situé en partie dans le *croissant*, est à la frontière entre langue d'oïl, langue d'oc et franco-

provençal. Les enquêtes étant menées en zones urbaines, nous pensons avoir limité ainsi les influences dialectales, tant il est vrai que

[un des] principaux facteurs pouvant avoir une répercussion sur la façon de parler [est] la localisation de l'habitat à l'intérieur d'une même région (ville ou campagne). (*ibid.*)

Ainsi, le choix d'effectuer les enquêtes en zones urbaines permet d'unifier ce facteur et par conséquent de limiter l'importance des variations géolinguistiques que l'on aurait eues si les enquêtes avaient été menées à la campagne. Mais leur étude peut se révéler intéressante. En effet, l'emplacement géographique des trois villes à l'intérieur du département peut permettre de comparer le français urbain standard parlé dans la zone d'oïl (Moulins) et dans la zone médiane entre oïl et oc, située dans le croissant (Montluçon au nord et Vichy au sud). Ainsi, ce corpus peut permettre de vérifier l'influence dialectale sur le parler standard, ce que nous ne ferons pas dans le cadre de notre étude, car nous n'utiliserons pas ce critère ; mais, comme nous l'avons mentionné plus haut, plus un corpus est composé de sous-types (qui sont délimités par les différentes variables retenues), plus son exploitation peut se révéler riche pour des études ultérieures.

A présent, voici une description sommaire du département dans lequel nous avons mené notre enquête, suivi de celle des trois villes le composant.

1. Présentation du département

L'Allier est l'un des quatre départements qui composent la région Auvergne, avec la Cantal, la Haute-Loire et le Puy De Dôme. Sa particularité tient au fait qu'il est situé dans le « croissant », zone où passe la frontière entre langue d'oïl (au nord) et langue d'oc (au sud), ce qui rend sa situation particulièrement intéressante pour les linguistes qui voudraient observer des variations géolinguistiques du français dit « standard », comme ont pu le faire les dialectologues pour les dialectes lors de la création de l'Atlas Linguistique de France¹, dans la deuxième partie du XX^{ème} siècle. Pierrette Dubuisson, l'auteur de l'Atlas Linguistique du Centre, indique ceci (1991) :

« On a l'habitude d'appeler Croissant l'aire intermédiaire entre langue d'oc et langue d'oïl à cause de sa forme. Comme on le sait, elle part de la Gironde pour remonter au nord du

Massif Central et rejoindre le franco-provençal. La partie sud du domaine de l'Atlas Linguistique et Ethnographique du Centre (ALCe) recouvre cette aire à la longitude de Montluçon ; c'est là que l'on passe le plus lentement d'une langue à l'autre, puisque le Croissant atteint dans cette région son maximum de hauteur avec une distance, du nord au sud, d'environ 40 km à vol d'oiseau. On peut observer un véritable "dégradé" linguistique dans cette zone d'interférence. Nous sommes aux confins des départements du Cher, de l'Allier, de la Creuse et du Puy-de-Dôme, donc à la jonction du Berry, du Bourbonnais, de la Marche-Limousin et de l'Auvergne. »

En ce qui concerne la syntaxe, ce positionnement ne présente pas un intérêt majeur ; Gaston Tuaillon (1983b : p. 229) indiquait à ce propos que seulement 3% des régionalismes étaient d'ordre syntaxique, contre 97% d'ordre lexical :

« Mais existe-t-il des régionalismes grammaticaux ? Il en existe et d'ailleurs j'en ai déjà cité un, le personnel *y* objet direct neutre. Mais les régionalismes grammaticaux sont beaucoup moins nombreux que les mots régionaux. En relevant les régionalismes du petit village dauphinois où j'habite depuis plus de vingt ans, j'ai pu rédiger 950 articles lexicographiques ou grammaticaux. Le pourcentage des entrées lexicales s'élève à 97%, alors que la grammaire se contente des 3% suivants. »

Nous avons malgré cela retenu le critère géographique, n'émettant pas d'hypothèses *a priori* sur le résultat de l'enquête après dépouillement. En somme, l'enquête doit fournir un corpus aussi riche que possible, afin qu'il puisse être utile à des chercheurs spécialistes d'autres disciplines que la syntaxe, qu'il s'agisse de phonologie, de sociolinguistique, de morphologie, etc.

Voyons à présent quelques caractéristiques des trois villes de ce département.

2. Présentation des trois villes

Le département de l'Allier comporte trois villes majeures d'environ 50 000 habitants (Montluçon, Vichy et Moulins) que nous avons tenu à présenter pour le lecteur qui ne les connaîtrait pas. L'intérêt d'une telle présentation n'est pas tant linguistique qu'historique et sociologique, ces trois villes ayant chacune des caractéristiques particulières, elles apparaissent ainsi complémentaires quant à la composition des zones urbaines du

¹ DUBUISSON P., *Atlas linguistique et ethnographique du Centre*, 3 volumes – 1 La nature, 1971, - 2 L'homme, 1976, - 3 La grammaire, 1971, Paris, éditions du C.N.R.S.

département. Il s'agit donc ici de donner un aperçu socio-géographique des villes entrant dans l'enquête, afin de donner à voir au lecteur la situation urbaine actuelle.

a. Montluçon

Etant originaire de Montluçon, c'est dans cette ville que nous avons commencé l'enquête ; nous pensions alors que les réseaux de connaissances que nous y avons établis nous permettraient de commencer relativement aisément notre travail.

Montluçon a un riche et lourd passé industriel ; comme dans beaucoup de villes de ce type, elle est divisée en deux parties, géographiquement séparées par le Cher : le centre-ville bourgeois, marqué symboliquement par le boulevard de Courtais, qui comprend la vieille ville et plusieurs quartiers périphériques (Saint-Jean, Gare, etc....), et la Ville-Gozet, ancien quartier ouvrier, qui se prolonge en direction de la périphérie urbaine (quartier des Marais). A la frontière sud-est de la ville, en direction de Montmarault, se situe la « Biache », qui comprend entre autre la petite commune de Désertines, et qu'il nous faut mentionner ici dans la mesure où le parler biachais est très marqué, notamment phonologiquement (en ce qui concerne la prononciation des /a/, souvent vélaires, alors qu'ils sont ouverts en centre-ville).

Il faut ici préciser que le choix des locuteurs ne tient pas compte de leur situation à l'intérieur de la ville (que ce soit à Montluçon, Vichy ou Moulins), qui aurait pu être un critère sociologique il y a encore quelques années, mais qui ne serait plus pertinent aujourd'hui, étant donné l'évolution rapide de la mixité sociale (il est parfois plus « chic » d'habiter un ancien quartier ouvrier réhabilité qu'un quartier bourgeois reconnu comme tel il y a encore vingt ans...).

b. Vichy

Vichy est une ville très marquée par l'histoire. Pendant la seconde guerre mondiale, le gouvernement du maréchal Pétain hésite à s'installer à Bordeaux ou à Vichy. Ville thermale, possédant de nombreux et luxueux hôtels, le réseau téléphonique le plus avancé de France, et une liaison ferroviaire avec Paris, c'est donc à Vichy que s'installe le gouvernement. Les pleins pouvoirs sont votés au maréchal Pétain par les parlementaires dans la grande salle du Casino, et, en 1940, Vichy devient capitale de l'Etat français, jusqu'à la Libération, en 1944.

Ville réputée pour ses cures, elle attire les plus grands, les plus riches, et son économie est prospère, avant et même après-guerre. Son déclin commence au début des années 60. Elle perd alors la majeure partie de sa clientèle, et son commerce décline.

Aujourd'hui encore la ville mise sur le thermalisme et le commerce, sa « force d'attraction » étant le secteur « Santé-Beauté-Forme ».

Cependant, un autre point attractif de la ville est son centre d'étude des langues, le CAVILAM (Centre d'Approches VIVantes des LANGues et des Médias), qui « est un lieu pilote d'enseignement/apprentissage des Langues Etrangères et des Médias. Lieu d'innovation constante, il anticipe, aide à concevoir et accompagne de multiples projets de formation - initiale et continue - des acteurs du Français Langue Etrangère et Seconde à travers le monde ». De ce point de vue, Vichy est un centre international réputé d'apprentissage des langues.

De la même manière que pour Montluçon, le centre-ville est plutôt bourgeois, tandis que les quartiers périphériques sont plus « populaires ».

c. Moulins

Préfecture du département de l'Allier et siège du Conseil Général, Moulins est la ville du département qui résiste le moins bien aux problèmes de dépopulation et de vieillissement de la population.

Le centre-ville, bourgeois et commerçant, souffre d'une mauvaise image de la part des Moulinois : pendant les discussions et les enregistrements (effectués sur le thème du petit commerce), nombreux ont été les locuteurs qui me disaient que Moulins était « ravitaillée par les corbeaux », qu'ils faisaient leurs achats vestimentaires ailleurs (Clermont-Ferrand ou Lyon), faute de mieux... Bien que centre administratif du département, Moulins semble donc, par son commerce du moins, moins dynamique que Vichy et Montluçon.

Après cette explication de ce que nous nommons critère géographique, suivie d'une présentation du département de l'Allier et des trois villes qui le composent, nous allons à présent présenter le deuxième critère de sélection des locuteurs, le niveau d'étude.

2. Critère « niveau d'étude »

Ce critère comprend trois différents niveaux d'études :

- collège
- baccalauréat
- baccalauréat + 3

Il faut cependant préciser que ces trois niveaux sont indicatifs : un individu scolarisé jusqu'à la terminale et qui n'a pas passé l'épreuve du baccalauréat entre dans la catégorie « baccalauréat » ; en revanche, un individu scolarisé jusqu'en seconde entre dans la catégorie « collège » ; un individu ayant fait plus de trois années d'études supérieures entre dans la catégorie « baccalauréat + 3 ».

Selon F. Gadet (1971 : p. 75), le niveau d'étude est un des trois principaux indices d'appartenance d'un individu à une classe sociale, avec le salaire et la profession, qui, selon elle, doivent « être étudiés en bloc, sans perdre de vue que l'un de ces facteurs peut être prépondérant ». Ne retenir que le niveau d'étude pourra donc apparaître réducteur à certains sociolinguistes ; cependant, la profession de chaque locuteur apparaît dans la fiche signalétique qui lui correspond et qui figure en annexe. Ainsi, seul le salaire du locuteur n'apparaît pas.

Ce critère est intéressant pour plusieurs raisons : d'une part, il permet d'éviter de demander au locuteur (qu'on voit souvent pour la première fois) quels sont ses revenus, question qui peut bien sûr être très gênante pour l'« enquêté », mais aussi pour l'enquêteur ; d'autre part, même si la scolarisation n'est pas le seul moyen éducatif ni le seul lieu d'apprentissage de la langue, le niveau d'étude paraît pertinent pour rendre compte de différents niveaux socioprofessionnels.

Passons maintenant à la description de ce que nous appelons « critère âge ».

3. Critère « âge »

Ce critère comprend trois tranches d'âge :

- 18-30 ans

- 30-65 ans

- + de 65 ans

Nous n'utilisons pas ici le terme de génération, qui est fluctuant, et qui se comprenait il y a cinquante ans comme une classe d'âge d'environ 20-25 ans, mais qui aujourd'hui n'est plus interprété de la même manière, les enfants arrivant généralement plus tard (rallongement de la durée des études, progrès de la médecine, etc.).

L'intérêt d'interroger des locuteurs de trois tranches d'âge différentes est de permettre d'avoir une vision assez précise de l'évolution d'une langue au cours du temps. En effet, un locuteur de 70 ans emploiera quelques tournures (et parfois quelques mots issus du patois) que n'emploierait pas un autre locuteur de 18 ou de 40 ans (notamment le passé surcomposé, employé par les locuteurs de plus de 65 ans, mais très rarement par les autres), comme F. Gadet le souligne (*op. cit.*, p. 75) :

Du fait qu'une langue est perpétuellement en évolution, il arrive constamment que certains traits soient réalisés différemment chez les jeunes et chez les vieux. Par exemple en français, il y a une disparition lente, mais progressive, de l'opposition entre /a/ et /ɑ/, entre /ɛ̃/ et /œ̃/ entre le /e/ du futur et le /ɛ/ du conditionnel, chez les jeunes. Au niveau lexicologique, notons un emploi plus systématique de termes argotiques dans la jeunesse.

Nous venons de passer en revue les trois critères de sélection que nous avons retenus pour nos enquêtes. Une fois cette étape achevée, l'enquêteur doit trouver les personnes y correspondant. La tâche se révèle plus ou moins difficile selon qu'il connaît ou non la localité sur laquelle il travaille. Elle peut l'être aussi à cause de tout autre phénomène, que nous n'étudierons pas ici, et qui peuvent être aussi bien liés à la timidité de l'enquêteur, à sa capacité d'adaptation sociale, ou encore à la relation même qu'il entretient avec le locuteur qu'il interroge.

1.2.1.2. Recherche des locuteurs correspondant aux critères prédéfinis

Une fois les critères de sélection établis, commence la recherche de locuteurs y correspondant. Cette étape est sans doute l'une des plus fastidieuse du travail préparatoire à l'enquête, et elle l'est d'autant plus que l'enquêteur ne connaît pas la localité (ce qui n'était pas notre cas).

Nous verrons donc comment trouver les locuteurs, avec deux méthodes sensiblement différentes, selon que l'on connaisse ou non la localité.

1. Enquêteur connaissant la localité

Si l'enquêteur connaît la localité pour y avoir vécu, la tâche lui paraîtra sans doute plus aisée. En effet, des « réseaux de connaissances » y sont souvent établis, qu'il s'agisse d'amis ou même de famille. Se servir de la technique dite des « réseaux denses », utilisée par Labov lors de certains de ses travaux sur le Lower East Side de New-York et par les Milroy à Belfast permet d'éviter ce que le premier nomme le « paradoxe de l'observateur »,

La première étape consiste à trouver dans ce « réseau de connaissances » les personnes répondant au critère socioprofessionnel et au critère d'âge dans chacune des trois villes : voici notre tableau pour Montluçon, à titre d'exemple

18-30 coll	Stéphane M.
18-30 bac	Muriel V.
18-30 bac+3	Sandrine M.
30-65 coll	Madeleine B.
30-65 bac	Patrick G.
30-65 bac+3	Jean C.
+ 65 coll	Marcelle G.
+65 bac	Françoise A.
+65 bac+3	Maurice B.

Sur les neuf personnes apparaissant dans ce tableau, nous en connaissons huit ; la neuvième est un personnage public montluçonnais.

Une fois cette étape achevée, il faut contacter ces personnes, pour leur expliquer le « service » qu'elles pourraient nous rendre en acceptant d'être enregistrées pendant une

demi-heure sur un thème donné. Généralement, ceci est l'étape la plus délicate : beaucoup sont dans un premier temps tentés de refuser (« mais moi, tu sais, je ne sais pas bien parler... »), puis finissent par accepter lorsqu'on leur explique que personne ne parle bien, que nous-même faisons des « fautes », et que l'intérêt pour nous n'est pas tant de traquer les « mauvais usages » que d'observer différentes manières de parler, qui sont recueillies et analysées de manière égale, c'est-à-dire sans aucun jugement de valeur. Nous avons parfois mis beaucoup plus de temps à rassurer les « timides » qu'à les enregistrer.

2. Enquêteur ne connaissant pas la localité

Lorsque l'enquêteur ne connaît pas la localité, la démarche la plus simple et sans doute la plus rapide est sans doute celle qui consiste à se renseigner auprès des organismes publics ou semi-publics, qui sont par nature bien implantés dans la ville et les mieux placés pour l'orienter.

Par organismes publics ou semi-publics, nous entendons :

- les mairies,
- les Conseils Généraux ou Régionaux,
- les associations,
- les établissements scolaires (pour les locuteurs de 18-30 ans),
- les maisons de retraite (pour les locuteurs de plus de 65 ans),
- les Chambres de Commerce et d'Industrie (qui proposent souvent des formations en leur sein, ce qui permet de « toucher » un public assez large (18-40 ans).

Lorsque nous avons commencé l'enquête à Moulins, cette ville nous était complètement étrangère. La première démarche que nous avons effectuée était une simple prise de contact avec l'agent d'accueil de la Mairie, qui nous a volontiers donné les coordonnées de divers établissements scolaires et de maisons de retraite. C'est ainsi que nous avons trouvé l'essentiel de nos locuteurs de plus de 65 ans.

Une fois cette étape achevée, il reste encore à expliquer aux personnels des établissements concernés en quoi consiste notre démarche, ce que nous allons demander aux personnes interrogées, car sans leur accord préalable, il n'est pas question d'aller plus loin. Cette étape s'est très bien passée, les locuteurs moulinois de plus de 65 ans (interrogés dans une maison de retraite) étaient enchantés de donner un peu de leur temps et de partager leurs souvenirs. Qu'ils en soient ici chaleureusement remerciés. A Vichy, c'est avec l'aide de la C.C.I. que nous avons trouvé un locuteur de 18 ans niveau collège, qui faisait alors une formation hôtelière, et qui s'est très gentiment prêté au jeu.

1.2.1.3. Matériel utilisé

Nous avons utilisé, lors de nos enquêtes, le matériel suivant :

- un enregistreur minidisque Sony MZ-R700
- un microphone de table Sony ECM-MS907
- un microphone cravate VIVANCO EM116 (utilisé lors des enquêtes dans les petits commerces, où les personnes interrogées pouvaient continuer à travailler sans être trop gênées par l'enquêteur).

Par la suite, nous avons copié les enregistrements effectués sur des mini-disques sur des CD de données au format mp3 à l'aide d'un logiciel gratuit disponible sur Internet¹ (Electronic Cosmo's MPEG suite), et ajouté à cela les transcriptions au format Word2000². Les fichiers sons et textes étant référencés de manière identique, la consultation se révèle ainsi plus simple, par exemple : les corpus BIJOUV~1.doc et BIJOUV~1.mp3 sont, respectivement, la version transcrite et la version son du même enregistrement.

Cependant, le choix que nous avons fait de réduire la taille du format (le format mp3 est beaucoup plus « malléable » que le format wave, mais il présente l'inconvénient d'être encore plus compressé, donc de ne pas rendre très précisément le signal), qui ne pose pas de problème pour une étude syntaxique, en posera certainement pour un travail en phonétique.

¹ : <http://p2pfr.free.fr/win/convertisseur/MPEGSuite.exe>

² : © Microsoft Corporation, juin 1999.

C'est pourquoi les enregistrements sont conservés sur les supports minidisques ; ils peuvent être extraits vers une machine et traités avec un logiciel adéquat.

Néanmoins, il faut insister sur le fait que les enregistrements sur support minidisque sont moins fiables lors du traitement du signal (dans le cas, par exemple, d'une analyse phonétique) que ceux effectués à l'aide d'un DAT. Dans un ouvrage récent, Elizabeth DELAIS-ROUSSARIE (Delais-Roussarie et Durand (éd.) 2003 : p. 191) présente les avantages et les inconvénients des divers appareils d'enregistrement disponibles sur le marché ; nous ne présenterons ici que ceux qu'elle expose à propos de l'enregistreur minidisque. Les avantages tiennent essentiellement à sa taille réduite (80x20x90 mm) et à la possibilité qu'il offre d'effectuer des enregistrements de 80 minutes avec une qualité maximale en stéréo (dans le cas d'un enregistrement en mono, la durée d'enregistrement peut varier, mais elle est de toute façon au moins deux fois supérieure à la précédente –160 minutes–). Cependant, indique-t-elle (*ibid.*) :

cette miniaturisation n'est pourtant pas sans revers. Tout d'abord la technologie numérique de ces enregistreurs utilise des algorithmes de compression de données. Ceux-ci permettent certes le stockage d'un maximum de ces données sur un minimum de support, mais ils amènent en contrepartie une dégradation du signal sonore par la perte d'information lors de la conversion. Il est vrai que, avec la parole, cette déperdition n'est pas vraiment sensible pour l'oreille humaine. Toutefois, cela peut constituer un handicap lorsque le signal doit subir une analyse et un traitement poussés, par exemple dans le cas d'analyses phonétiques.

Durant les enquêtes, le microphone était placé devant l'enquêté, à une distance d'environ 30 cm, en réglage unidirectionnel, pour éviter les bruits.

1.2.1.4. Enquête non directive

Contrairement à la plupart des enquêtes sociolinguistiques (notamment le *Corpus d'Orléans*), nous n'avons pas utilisé de questionnaire, mais avons laissé libre cours à la parole du locuteur, ce que les sociologues nomment *entretien non-directif*.

Jacqueline FREYSSINET-DOMINJON (1997 : pp. 150-151) définit ainsi l'entretien non directif :

L'entretien non directif, également qualifié d'entretien en profondeur, se place au plus faible degré de structuration de l'échange verbal et au plus haut degré de liberté de parole de la

personne enquêtée. L'enquêteur ayant posé au départ le thème de la recherche, adopte [une] attitude [...] faite d'écoute empathique et de neutralité congruente. [...]

L'usage de ce type d'entretien [...] est recommandé [...] toutes les fois que la recherche porte sur des représentations souvent latentes plutôt que sur des faits précis.

Seul le thème était prédéfini, en l'occurrence le *petit commerce*¹, qui présentait plusieurs avantages : d'une part, l'actualité du sujet (la décrépitude du petit commerce étant nécessairement reliée à la grande distribution, les variations sur ce thème ne manquaient pas) et d'autre part, le fait que tous les citoyens soient des consommateurs plus ou moins conscients nous a permis d'éviter les blancs - périodes pendant lesquelles le locuteur ne sait plus quoi dire -, ou les questions trop fréquentes de l'enquêteur qui empêchent la fluidité du discours, et par là même sa réalisation la plus naturelle possible.

L'intérêt de ce type d'enquête, lorsque l'on travaille en syntaxe, est de favoriser la production de parole la moins « interrompue » possible, permettant ainsi de relever les formes syntaxiques les plus « naturelles », bien que la notion de « naturel » soit discutable lorsqu'un microphone est posé en face du locuteur.

1.2.1.5. Conditions d'enregistrement (variables contextuelles)

Pour qu'un corpus puisse être envisagé et consulté de manière homogène, non-ambiguë et non-équivoque, les conditions d'enregistrement doivent être identiques pour chaque enregistrement, afin de considérer la production orale de chaque locuteur comme égale à n'importe quelle autre. Elles font partie des *variations contextuelles* que mentionne F. Gadet (1971 : p. 76) :

[Les variations contextuelles] comprennent tout ce qui peut déterminer des différences de langage et qui n'est pas directement lié au locuteur dont on analyse le discours : la personne à laquelle il parle (et l'image qu'il se fait de cette personne), ce dont il parle, *l'emplacement dans lequel se déroule la conversation*, les rapports qui unissent les locuteurs [...].

En effet, pour pouvoir comparer deux ou plusieurs objets, ou plus simplement pour décrire un ensemble de textes, il est nécessaire que les propriétés constitutives de chaque production orale soient en partie semblables. Nous disons *en partie* car même si,

théoriquement, les conditions d'enregistrement devraient être identiques pour chaque locuteur, nous avons eu quelques difficultés d'applications théoriques, dans les cas notamment où nous enregistrons des locuteurs sur leur lieu de travail. Ceci est sans doute un des reproches qui pourront nous être adressés, nous en avons bien conscience.

Concernant notre corpus, voici les conditions d'enregistrement :

Lieu d'enregistrement

Lorsque l'enregistrement se déroule chez l'enquêté, ce dernier propose généralement la pièce habituelle de réception la plus calme (cuisine, bureau ou jardin). 24 corpus sont enregistrés ainsi.

Lorsque l'enregistrement ne se déroule pas chez l'enquêté, il est effectué soit sur son lieu de travail à l'aide d'un micro-cravate pour ne pas gêner l'accomplissement de certaines tâches (2 corpus enregistrés ainsi : James~30 et Monov~42), soit à la terrasse d'un café en zone piétonne avec le microphone « traditionnel » (Mumon~44).

Position enquêteur / enquêté

L'enquêteur fait face à l'enquêté, sauf pour James~30 et Monov~42, où la situation d'enregistrement sur le lieu de travail n'a pas permis cette position. Ce choix de face à face a été fortement motivé par des raisons non-scientifiques : ayant un problème auditif bilatéral sérieux et n'étant pas appareillée à cette époque, la lecture labiale était pour moi nécessaire et, dans ce cas, cette situation est la meilleure et la plus confortable.

Ainsi, en essayant d'unifier au maximum le contexte et les conditions d'enregistrement, nous pouvons comparer différents objets ayant ce trait commun, l'idéal étant bien sûr de pouvoir établir une réelle constante à partir de ce contexte et de ces conditions d'enregistrement.

Nous avons vu jusqu'à présent la méthode d'enquête utilisée, depuis le choix des critères de sélection de l'échantillon de locuteurs jusqu'au matériel utilisé et aux conditions d'enregistrement, en passant par une présentation sommaire des lieux de l'enquête. Mais

¹ Nous verrons à la fin de ce chapitre qu'un tiers du corpus n'est pas thématisé ainsi, et nous expliquerons pourquoi nous avons choisi de ne pas le modifier.

après cela commence un long travail d'écoute et de transcription, qui répond lui aussi à certaines règles et conventions, que nous allons présenter dans la partie suivante.

1.2.1.6. Conventions de transcription

« [...]le véritable intérêt d'un corpus réside dans sa transcription. Recueillir de la parole est aujourd'hui d'une déconcertante facilité dès lors qu'on ne se donne pas de réquisits trop extraordinaires. Mais entre l'enregistrement et son exploitation, il faut établir la notation, consultable et comparable à loisir, que la bande magnétique, en cela comparable, dans son maniement, au *volumen* latin, interdit. Si la transcription rebute, elle accompagne aussi la réflexion du chercheur et la lecture la plus avertie du résultat sera le fait de celui qui, pour l'avoir extraite de l'enregistrement, sait la relativité et la valeur de sa version. » (Bergounioux et al., 1992)

Les conventions de transcription que nous avons utilisées sont en partie reprises des indications données par C Blanche-Benveniste et C. Jeanjean (1987) :

ORTHOGRAPHE :

- Transcription en orthographe (avec fidélité à l'orthographe du morphème, même s'il y a infraction à la norme ex. : *les conseils nationaux*).

- Aucun signe de ponctuation sauf :

- Majuscule sur les noms propres
- Guillemets droits pour les titres, avec majuscule sur le premier mot (ex. : "Arnaque, crime et botanique")
- Les signes sont ponctués quand on lit les lettres isolément (S.N.C.F.), non ponctués lorsqu'il s'agit d'un acronyme (CROUS). Le sens des sigles est précisé sur la fiche signalétique.

- Onomatopées transcrites selon l'orthographe du dictionnaire ex ; *euh*, *hum* (cf. l'inventaire).

- Dans les cas où *les noms propres doivent être supprimés* on respectera le codage suivant :

P = patronyme ; *T* = toponyme ; *S* = marque commerciale, nom de société.

S'il y a plusieurs *P* ou plusieurs *T*, chacun reçoit une numérotation.

- ***Amorces, multi-transcriptions et alternances orthographiques :***

un mi- amorce de mot

j- [la notation j'- est interdite]

/d'accord, d'abord/ multi-écoute (hésitation entre plusieurs séquences)

/mais, Ø/ hésitation quant à la présence d'une séquence sonore

il(s) chante(nt) multi-graphies (alternances orthographiques)

(ces, ses)

on (n') a pas

- Sauf indication explicite du contraire, les adjectifs et participes se rapportant au pronom *on* restent au singulier : transcrire *on est parti en Afrique* ; *on est parti tous les deux* ; mais transcrire : *on s'est mises dans de beaux draps*.

- Les mots étrangers sont orthographiés conventionnellement à la norme de la langue d'origine (majuscules aux noms allemands, etc.)

- Orthographe incertaine de certains noms propres (marques, toponymes, etc.) ou de fragments de langue étrangère : transcription en orthographe suivie de la mention {approx.}.

S'il y a ***un seul mot***, noter {approx.} après le mot

S'il s'agit ***d'une séquence plus longue***, elle sera encadrée par {début approx.} et {fin approx.}

PRONONCIATION :

On mettra en introduction les prononciations courantes de types « il y a » ; « tu » devant voyelle ; « puis » prononcé [pi] ; « peut-être » prononcé [ptet].

Il est inutile d'indiquer la prononciation sans schwa pour les clitiques.

–	pause(s)
///	interruption assez longue du discours (justifiée en note)
X	une syllabe incompréhensible
XXX	suite quelconque de syllabe inaudibles ou inorthographiables (dans ce cas, la transcription phonétique est donnée en note)

- *Les chevauchements de paroles* sont soulignés
- *Géminations ou prononciations remarquables* :

Je crois que il ne pas marquer l'élision (par l'apostrophe) si elle n'est pas faite

PROBLEMES D'ECOUTE

- *Les prononciations remarquables* justifiant une transcription phonétique (mots ou suite de mots problématiques sont signalés en note.

LIAISONS :

- On notera une liaison non obligatoire réalisée avec le signe = (ex. : sont = à Paris)
- On notera une liaison fautive en marquant la consonne entre deux = (ex. : quatre =z= amis) ;
- On notera une absence de liaison obligatoire avec le signe • (ex. : je les • accompagne)

NOTATION DES EVENEMENTS NON-LINGUISTIQUES :

Les rires, les bruits sont signalés en note.

1.3. Corpus : genres, types et thèmes

1.3.1. Historique de la notion de corpus

Dans l'ouvrage édité par M. Bilger (2000), C. Blanche-Benveniste propose une présentation historique de la notion de corpus (pp. 12-15), que nous avons jugé utile de rapporter intégralement ici, dans la mesure où elle permet d'observer clairement la situation dans laquelle se trouve la linguistique dite *de corpus* aujourd'hui. Revenant aux origines de l'utilisation de ce terme, l'auteur esquisse son évolution, puis présente les orientations récentes que les linguistes lui ont donné :

Le terme de *corpus* a désigné pendant des siècles des sources documentaires caractérisées par leur exhaustivité : recueils de textes rassemblant exhaustivement tous les documents disponibles pour certains champs d'étude. Le *Corpus Juris* rassemblait tous les documents du droit romain ; le *Corpus des inscriptions grecques et latines* comportait l'ensemble des inscriptions connues ; le *Corpus des catalogues de bibliothèques médiévales* n'avait de valeur que s'il assurait la consultation de tous les catalogues du domaine. L'exhaustivité des données rassemblées faisait partie de la définition. Nul ne pouvait prétendre connaître les données s'il n'avait pas consulté le Corpus.

En matière de langage, le terme pouvait s'appliquer sans encombre à des collections de documents de langues mortes, qu'on pouvait élargir lorsqu'intervenaient de nouvelles découvertes. Sur les langues vivantes, il en va autrement. L'exhaustivité n'est pas possible : qui pourrait prétendre qu'un corpus, aussi vaste fût-il, pourrait contenir toutes les données possibles d'une langue vivante ?

La première utilisation du terme *corpus* appliqué à une collection de données de langue vivante est assez récente. *Le Trésor de la Langue Française* la date de 1961. C'est d'emblée un sens différent du sens classique. Ces corpus linguistiques ne visent pas à l'exhaustivité, mais à quelque chose qui y ressemblerait : une bonne représentativité de la langue.

Dans les années 1960 naissent les premières grandes collections d'enregistrements et transcriptions de langue parlée. Pour l'anglais, *The Survey of English Usage*, et *The Survey of Spoken English*, documentations rassemblées par J. Svartvik et R. Quirk, prendront en 1979 le titre de *A Corpus of English Conversation*. Les auteurs insistent sur un principe important : le matériel qu'ils présentent permettra de faire des études « fondées sur corpus », de nature très différente de celles que menaient leurs devanciers en les fondant « sur des textes ». En effet, l'ambition de leur corpus est de rendre compte de tous les aspects de la langue, à partir d'une collection qui en forme

un échantillonnage raisonné. Les grammaires fondées sur des textes n'avaient pas cette possibilité et pouvaient difficilement se targuer d'avoir travaillé sur des données représentatives de l'ensemble de la langue. Il s'agissait, pour ces nouveaux *corpus*, de mettre au point un répertoire englobant différentes variétés de langue anglaise et les représentant fidèlement. La difficulté de la tâche était de définir les variétés, sans en oublier. L'idéal d'exhaustivité, qui ne pouvait évidemment pas être atteint par la quantité des données, était déplacé et reporté dans l'échantillonnage de ces données. Une centaine de productions, de 5000 mots chacune, prises dans des champs d'activité bien délimités, ont été retenues par J. Svartvik et R. Quirk. C'était le début des « corpus échantillonnés ».

Les données rassemblées dans les années 1950-1960 pour constituer le *Français Fondamental* et le *Portugais Fondamental* ne portaient pas encore le nom de *corpus*. Leurs objectifs étaient différents. Elles voulaient fournir pour l'enseignement des langues vivantes un vocabulaire et une grammaire « de base », calculés en fonction des plus grandes fréquences statistiques. Ce souci de dégager, dans la masse des données linguistiques, des bases utiles pour l'enseignement n'a pas été bien compris en France. Un tel objectif les a à la fois servis et desservis. Comme les auteurs ne cherchaient pas à représenter « toute la langue », on a pu les accuser, surtout en France, de chercher à l'appauvrir.

Quelles que soient les orientations, dès les premières applications aux langues vivantes, le terme de *corpus* désigne non pas simplement des collections de données de langage, mais un choix organisé de ces données. Dans les usages qu'en ont fait les sociolinguistes, à partir des années 1970, le *corpus* désignait un échantillonnage de données choisies en fonction de critères sociologiques et régionaux, comme par exemple le Corpus Sankoff-Cedergren pour le français du Québec, ou le Corpus d'Orléans pour le « français de France ». Les locuteurs sollicités étaient classés d'après leur origine, leur profession, leur sexe, la durée de leur scolarité, leur lieu de résidence et éventuellement d'autres critères analogues, permettant d'assurer que l'ensemble des locuteurs étaient représentés de façon fidèle par un nombre limité de « locuteurs-échantillons ». On espérait que, à partir de ces échantillons, il serait possible de calculer statistiquement la répartition des faits linguistiques comme la prononciation, les choix de lexique ou la grammaire. C'était l'époque où les corpus étaient calculés en fonction des corrélations possibles entre types de locuteurs et types de parlers.

L'étape suivante semble avoir été celle des « corpus de référence ». L'échantillonnage porte alors sur les différents usages attestés, dans différentes activités de langage, quels que soient les locuteurs. Le principe de départ est qu'un seul et même locuteur parle différemment selon qu'il s'adresse à sa famille, à des amis, à un public familier, à un public inconnu, directement, par téléphone ou avec un répondeur, à la radio ou à la télévision, en ayant préparé ou non sa prise de parole. A cette variété des prises de parole se combinent aussi les variétés régionales, qui peuvent être importantes pour les langues de grande diffusion comme l'anglais, l'espagnol, le portugais ou le français. R. Salkie explique ici comment la langue anglaise a été dotée de quantité de corpus

échantillonnés. A. Avila Muñoz expose le problème pour une variante de la langue espagnole. Le français ne s'est pas encore doté, pour la langue parlée, d'un corpus échantillonné.

L'ensemble de ces échantillons vise à représenter toutes les situations de prise de parole, à une époque donnée. En ce sens, le corpus peut « faire référence ». Il comporte aussi bien du langage familier que du langage formel répondant à diverses conventions, qu'aucun locuteur isolé ne peut connaître complètement, par exemple, le langage des médecins, des techniciens, des avocats, etc. Il peut être donné comme modèle à qui veut apprendre divers usages de la langue. L'idéal d'exhaustivité est placé dans les usages de la langue : on espère les avoir tous retenus. C'est ce qu'explique Fernanda Bacelar Do Nascimento pour le *Portugais de Référence*, envisagé dans diverses manifestations sur tous les continents où il est parlé et écrit. Le corpus italien de langue parlée en Toscane vise, dans une région limitée, un inventaire aussi exhaustif que possible des échantillons prosodiques de la langue. On voit que l'exhaustivité peut être située à différents niveaux.

Ces corpus échantillonnés diffèrent de simples collections d'archives parce qu'ils ont été composés et structurés systématiquement. Mais, depuis quelques années, le terme de *corpus* a été utilisé de façon plus large, pour toute collection de textes rassemblée dans des bases de données informatisées. Même si la collecte n'a pas été faite de façon systématique et structurée, l'informatisation peut en faire un usage structuré. L'ensemble des textes d'un auteur peut former un corpus, dans lequel on peut projeter diverses organisations pour la recherche linguistique. La taille des grands corpus permet d'y découper plusieurs sous-corpus structurés. Le terme de corpus est devenu indispensable de celui d'informatisation.

Au début des années 1980, rappelle G. Kennedy, les corpus d'un millions de mots emblaient déjà vastes. La génération des corpus de langue anglaise écrites des années 1980, Cobuild et Longman Lancaster, en comportaient environ 20 millions. Dans les années 1990, on peut envisager des corpus de langue écrite d'un milliard de mots ou plus. Les corpus de langue parlée, qui exigent un long travail d'écoute et de transcription, se tiennent pour l'instant dans des proportions beaucoup plus restreintes, qui ne dépassent pas cinq millions de mots. De nombreuses conventions de transcriptions ont été proposées, selon le type de notation que l'on veut placer sur la transcription.

Les corpus diffèrent entre eux selon les annotations plus ou moins sophistiquées qui les accompagnent. Des plus économes en notation, on dira que ce sont des corpus « nus ». D'autres sont complètement « étiquetés », chaque mot y recevant son étiquette morphologique. D'autres peuvent être accompagnés d'analyses syntaxiques plus ou moins raffinées.

De nouvelles sortes de corpus ont été développées pour l'étude contrastive des langues. Les « corpus parallèles » permettent de placer côte à côte les chapitres, les paragraphes, les phrases, voire les syntagmes correspondants entre deux ou plus de deux langues. De nombreux

corpus portant sur l'acquisition des langues on été composés, surtout pour l'anglais. Avec un corpus assez modeste, conçu pour l'observation des démarches d'apprenants, C. Lagarde montre ici qu'on peut aller assez loin dans l'observation.

Les différents types de corpus de langue parlée se définissent généralement par les objectifs qu'ils se donnent : faciliter l'enseignement de la langue, faire de la recherche linguistique, collaborer à de plus large recherches statistiques, participer à des recherches de sociolinguistique, etc. Nous assistons actuellement à une reconversion complète des études linguistiques, qui ne pourront plus se passer, d'ici peu, de la consultation des ressources de langue informatisées. Dans cette perspective, on peut prévoir une variété de corpus linguistiques encore plus grande que celle que l'on voit se dessiner aujourd'hui. Dans le domaine de la langue parlée, les différentes sortes de corpus se justifient par deux caractéristiques importantes. Comme ils sont informatisés, ils se prêtent facilement à des structurations internes en sous-corpus. Comme ils contiennent des données attestées, dont on peut vérifier les sources, ils engagent à faire un travail d'analyse linguistique qui ne repose pas seulement sur l'intuition mais sur la confrontation avec des données parfois étonnantes, que la simple intuition du linguiste n'aurait pas pu atteindre.

Nous voyons donc, grâce à ce tour d'horizon historique, à la fois l'évolution de l'utilisation du terme *corpus*, et de ce à quoi il réfère. Actuellement, comme le souligne F. Rastier (2001 : p. 86) de la même manière que C. Blanche-Benveniste,

La notion même de corpus doit être affinée, car un corpus n'est pas un ensemble de données : comme toujours dans les sciences de la culture, les données sont faites de ce que l'on se donne [...], et le point de vue qui préside à la constitution d'un corpus conditionne naturellement les recherches ultérieures.

C'est essentiellement la raison pour laquelle nous avons insisté précédemment sur les intérêts que notre corpus pouvait présenter dans divers domaines de la linguistique. Choisir rigoureusement les variables (géographique, âge, niveau d'étude), les respecter lors de l'enregistrement, puis transcrire minutieusement les données orales sont trois des conditions majeures de la qualité de fabrication d'un corpus, et ces conditions sont les garanties apportées à ses futurs usagers.

C'est ce que nous allons montrer dans la partie suivante, consacrée à la réflexion sur le statut du corpus de langue parlée.

1.3.2. Nécessité d'une réflexion sur le statut du corpus de langue parlée et sur son évolution

La linguistique [...] n'a absolument pas défriché la section dont devraient relever les grands ensembles verbaux : longs énoncés de la vie courante, dialogues, discours, traités, romans, etc. car ces énoncés-là peuvent et doivent être définis et étudiés, eux aussi, de façon purement linguistique, comme des phénomènes du langage. [...] La syntaxe des grandes masses verbales [...] attend encore d'être fondée ; jusqu'à présent, la linguistique n'a pas avancé scientifiquement au-delà de la phrase complexe : c'est le phénomène linguistique le plus long qui ait été exploré. On dirait que le langage méthodiquement pur de la linguistique s'arrête ici [...] Et cependant, on peut poursuivre plus loin l'analyse linguistique pure, si difficile que cela paraisse, et si tentant qu'il soit d'introduire ici des points de vue étrangers à la linguistique. (M. Bakhtine, 1978 : p. 59)

C'est ainsi que J.-M. Adam commence l'avant-propos de son ouvrage *Éléments de linguistique textuelle* (1990 : p. 7) : une citation de M. Bakhtine, déplorant le manque d'avancées significative en linguistique du discours. Depuis les années 1980, cependant, avec la création de grands corpus oraux pour le français, ce type d'études semble se mettre en place.

En ce sens, C. Blanche-Benveniste (Bilger, 2000 : pp. 14-15) se montre optimiste et pense que « nous assistons actuellement à une reconversion complète des études linguistiques, qui ne pourront plus se passer, d'ici peu, de la consultation des ressources de langue informatisées » et que « dans cette perspective, on peut prévoir une variété de corpus linguistiques encore plus grandes que celle que l'on voit se dessiner aujourd'hui ». La constitution de corpus échantillonnés et leur large diffusion sont donc les conditions nécessaires du développement de telles recherches. En outre, D. Willems (Bilger, 2000 : pp. 152-153) note : « qu'il s'agisse de corpus multilingues parallèles (de traduction) ou comparables (thématiquement proches), les données élémentaires manquent encore ou sont inutilisables par manque d'harmonisation des supports techniques utilisés », et que « des initiatives – au niveau européen – devraient d'urgence être prises à ce sujet. ».

C'est pourquoi la prise en compte du corpus comme un objet multifaces, composé de plusieurs types de discours (selon le type de parole enregistrée – publique/privée, préparée/non-préparée, narrative/argumentative/descriptive/dialogique... –) de plusieurs locuteurs sélectionnés selon des critères établis clairement doit continuer à se développer, tant au niveau de sa réalisation qu'à celui de la réflexion épistémologique qu'il est nécessaire de lui associer.

D'autre part, comme le signale C. Blanche-Benveniste (Bilger, 2000 : p. 25), les corpus de langue parlée posent une série de défis aux informaticiens spécialistes du Traitement Automatique du Langage (désormais TAL). En effet, s'ils parviennent à des résultats satisfaisants pour l'écrit (littéraire, journalistique), les problèmes de l'oral transcrit demeurent encore : comment traiter, par exemple, dans une chaîne de caractères, les répétitions, les bribes et les amorces, que les transscripteurs soucieux d'adéquation maximale entre l'oral et son support sont tenus de noter ? De la même manière, comment traiter l'absence de ponctuation ? D'autre part, les corpus de langue parlée montrent clairement les limites de la notion de phrase (cf. 2.1.1., [Phrase et construction verbale](#)) : dans ce cas, comment découper l'énoncé en unités analysables ? Sur quelles structures de base se fonder pour le traitement automatique ? Ce sont autant de questions auxquelles les informaticiens-linguistes se doivent de répondre en vue du traitement automatique des corpus (transcrits) de langue parlée. Pour notre part, il nous semble intéressant de poser comme limite la construction verbale telle que nous l'avons définie (2.3. [Définitions](#)) ; mais, si nous ne maîtrisons pas les bases nécessaires en traitement automatique pour nous risquer sur cette question épineuse, il nous semble cependant indéniable que l'observation des régularités distributionnelles de chaque verbe pourrait servir à prévoir et donc à écrire ces règles.

Par ailleurs, nous tenons à signaler une incohérence concernant l'élaboration de notre corpus : l'unité thématique. En effet, le sous-corpus *Montluçon* n'a pas de thème défini : il s'agissait pour les locuteurs de parler d'un point de leur vie, de leur travail pour certains, de leur engagement politique pour tel autre, tandis que les deux autres sous-corpus *Vichy* et *Moulins* ont pour thème le *petit commerce*. Ceci est dû au fait que nous avons pris conscience de l'importance de la cohérence thématique au cours du travail d'enquête, lors d'une entrevue avec Claire Blanche-Benveniste, qui nous a patiemment expliqué cela. En effet, les éléments du lexique utilisés ne sont *a priori* pas les mêmes selon le thème donné : les verbes utilisés pour un discours argumentatif diffèrent par exemple de ceux employés pour une description¹, etc.

¹ Pour la description, nous avons enregistré neuf locuteurs pendant trente minutes à Moulins, sur des descriptions de tableaux, mais ces enregistrements n'ont pas été transcrits. Ils pourraient l'être en vue d'une étude comparée de la parole argumentative et descriptive, ou encore des constructions verbales utilisées pour chacun de ces types de parole.

Compte tenu à la fois du fait que les enregistrements et les transcriptions pour *Montluçon* étaient réalisés lors de la prise de conscience (tardive) de l'intérêt du choix d'un thème identique pour l'ensemble des corpus, et que ce travail est particulièrement long, nous avons décidé de conserver ce sous-corpus. En outre, le travail d'inventaire et de classement des constructions verbales n'étant pas touché, selon nous, par cette incohérence thématique (nous ne recherchons pas d'unité sémantique aux différentes constructions), elle ne nous semble pas rédhibitoire pour l'analyse que nous avons faite. Nous démontrerons cela au cours de la partie 1.3.5. ([Exemples de requêtes lexicales thématiques Montluçon / Vichy Moulins](#)).

Cependant, il convient d'explicitier ici l'intérêt d'une telle unité thématique. Des chercheurs comme Jean-Michel Adam (1990, 1992) et François Rastier (1995, 2001) nous donnent des pistes intéressantes concernant l'analyse du discours, les genres et les types qui permettent de les distinguer. Dominique Maingueneau, (2000) s'intéresse lui aussi aux typologies des genres de discours. Tous les trois proposent des perspectives différentes. Nous n'entrerons pas ici dans leurs détails, cette partie étant plus destinée à la présentation des réflexions actuelles au sujet des « corpus » et des « textes » qu'à la critique théorique sur le sujet.

Nous voulons donc ici présenter le point de vue de ces trois auteurs, ce qui permettra au lecteur de saisir mieux l'erreur que nous avons commise et de la relativiser, en montrant qu'une unité thématique ne garantit aucunement une unité de type discursif, même si elle en favorise quelques uns.

1.3.3. Définitions (texte, corpus, discours, type, thème)

Texte et corpus : un même objet ?

Si le texte est associé par tradition à l'écrit et le discours à l'oral, F. Rastier propose de dépasser cette dichotomie en posant comme fondamentale la notion de *support* (2001 : pp. 21-22) :

A l'heure de la dématérialisation numérique, il faut songer à un remaniement, et l'opposition entre écrit et oral doit être dépassée par la notion de *support*. Résumons en une définition : *un texte est une suite linguistique empirique attestée, produite dans un pratique sociale déterminée, et fixée sur un support quelconque*. Ainsi, un texte peut être écrit ou oral, voire présenté par d'autres codes conventionnels, comme le Morse, l'Ascii, etc., et se manifester en

interaction avec d'autres sémiotiques (film, etc.) Ces trois conditions définitoires s'entendent comme suit. (i) Le texte est attesté : il n'est pas une création théorique comme l'exemple linguistique forgé par le linguiste. Cette première condition énonce un principe d'objectivité. (ii) Il est produit dans une pratique sociale déterminée : c'est là un principe d'écologie. Bien que non suffisante, la connaissance (ou la restitution) de cette pratique est nécessaire, ne serait-ce que parce qu'elle assure la délimitation du texte. (iii) Il est fixé sur un support : c'est la condition de son étude critique, supposant débat des conjectures. Cette condition empirique rompt avec le privilège exclusif de l'écrit et rappelle que la substance de l'expression n'est pas définitoire du texte.

Ainsi, nous aurions pu nommer *texte* l'objet sur lequel nous avons travaillé, car, comme le montre J.-M. Adam (1990 : p. 23), un discours équivaut à la somme du texte et de ses conditions de production, tandis qu'un texte équivaut à un discours, mais sans ses conditions de productions. En effet, le support (écrit) que nous avons utilisé pour fixer le discours de nos locuteurs ne prend pas en compte la prosodie, les expressions paraverbales¹. En revanche, en essayant, avec des situations discursives homogènes (du fait du choix de conditions d'enregistrement similaires, dans la mesure du possible, d'un thème (relativement) homogène, ou encore de l'unité du genre narratif), d'unifier les conditions de production, nous pouvons les poser comme « constantes », donc ne pas les considérer comme modificatrices de la production verbale.

F. Rastier soulève le problème que rencontre la linguistique quand elle doit définir la notion de texte, en montrant ses incohérences et les limites de certaines définitions (2001 : p. 28) :

Au mieux, on définit les textes comme des « enchaînements discursifs un peu longs »
(Auroux, 1996, p. 260)

Selon lui, les causes de cette réduction et de cette absence manifeste de précision se trouvent en partie dans le « poids du logicisme sur la construction même de la morphosyntaxe, dont la phrase est l'unité la plus grande » et dans la question platonicienne qui « tourmente encore la sémantique vériconditionnelle », à savoir « Comment le langage peut-il dire le vrai ? ».

¹ Mais des chercheurs, notamment l'équipe belge *Valibel*, proposent des enregistrements d'enquêtes filmées, et peuvent ainsi observer les manifestations extra et paraverbales associées aux productions verbales. D'autre part, il existe des corpus marquant la prosodie : sur ce point, voir C. Blanche-Benveniste et C. Jeanjean (1987).

A la notion de texte nous avons depuis le début de ce travail préféré la notion de corpus, principalement pour la raison suivante : elle est commune à toutes les personnes qui travaillent sur la langue parlée, puisqu'elles fondent leur démarche sur la description d'une transcription de l'oral ; ainsi, le corpus est un concept « aisé » qui permet de dépasser le problème de la définition du texte. Cependant, si chacun d'entre nous a une idée relativement claire de ce qu'est (ou de ce que devrait être) un corpus, il semble ici nécessaire d'en donner une première définition, encore grossière.

Posons une définition sommaire du corpus : *il est un ensemble limité de parole produite par un ou plusieurs locuteurs, cette production pouvant être orale (discours politique, conversation...) ou écrite (lettre, mémoire(s)...)* . F. Rastier (*op. cit.*, pp. 86-87) propose d'affiner la notion de corpus en le caractérisant à l'aide de quatre axes d'évaluation : la *représentativité*, l'*homogénéité*, l'*ouverture* ou la *fermeture* du corpus, et enfin son *entretien*.

La *représentativité* « n'a rien d'objectif et dépend du type d'utilisations prévues », selon F. Rastier. C'est aussi ce que note C. Blanche-Benveniste (*in* Bilger, 2000 : p.15) lorsqu'elle précise que « les différents types de corpus de langue parlée se définissent généralement par les objectifs qu'ils se donnent ». Un corpus destiné à une recherche thématique devra nécessairement être beaucoup plus important quantitativement qu'un autre servant de base de travail en phonétique, et leurs conventions de transcriptions seront, de la même manière, plus ou moins précises. En effet, un corpus de recherche thématique se « contente » d'une transcription orthographique effectuée au niveau du monème tandis qu'un corpus de recherche phonétique doit être transcrit au niveau du phonème (avec un degré de précision supplémentaire). Dans notre cas, une recherche phonétique à partir du corpus Allier est impossible compte-tenu des conventions de transcriptions que nous avons utilisées. Cependant, il convient ici de rappeler que ce corpus, même s'il est pour l'instant fermé (le temps de ce travail), peut être retranscrit de diverses autres manières pour être utilisé dans d'autres domaines de la linguistique, puisque les données orales sont conservées et peuvent être éditées et diffusées.

L'*homogénéité* d'un corpus est liée à trois choses, selon F. Rastier. Tout d'abord, elle est, comme la représentativité, liée aux *types de recherches* auxquelles ce corpus sert ; puis à « l'appartenance à un même discours (juridique, religieux, etc.) », bien que cette

appartenance soit selon lui surestimée¹, car « d'un corpus à l'autre, toutes les classes sont reconfigurées ». Enfin, « l'*homogénéité de genre* doit être privilégiée par défaut », car « en règle générale, les recherches en sémantique des textes doivent porter sur des corpus aussi homogènes que possibles pour ce qui concerne leur genre, ou du moins leur discours : en effet, un texte peut « perdre » du sens, s'il est placé parmi des textes oiseux, car la comparaison avec eux ne permet pas de sélectionner d'oppositions pertinentes. » Pour ce qui nous concerne, nous évoquerons plutôt une homogénéité thématique qu'une homogénéité de genre ou de discours, même si toutes les unités qui forment ce corpus sont homogènes quant à leur genre (récit) et à leurs contraintes discursives de production, car c'est à un niveau encore plus précis (celui du thème) que nous avons souhaité construire ce corpus, même si, comme nous le verrons, la cohérence thématique est imparfaite.

Le troisième axe d'évaluation proposé par F. Rastier est celui de l'*ouverture* ou de la *fermeture* du corpus. Il explique que « les corpus ouverts intéressent notamment des applications de veille technologique ou d'informatique documentaire », et affirme qu'« en revanche, les corpus fermés ont toujours un aspect normatif, car leurs textes sont en quelque sorte institués en canon. » Nous ne sommes pas aussi catégorique que F. Rastier en ce qui concerne la normativité associée aux corpus fermés. En effet, le corpus Allier est pour l'instant fermé, et ce jusqu'à la fin de ce travail, pour des raisons évidentes, qui sont liées à la nature même de notre recherche : un inventaire ne peut s'effectuer que si le « stock » de données est arrêté. C'est le cas de toute recherche descriptive, car la description n'est efficace que si l'objet décrit est posé, arrêté, et envisagé comme tel. Mais l'auteur d'*Arts et sciences du texte* affine cependant son propos, notant qu'« au cours d'une recherche, le corpus de référence et le corpus de travail sont toujours fermés, car ils doivent être prédéfinis », et qu'« en raison de sa méthodologie comparative, la linguistique ne peut travailler utilement que sur des corpus définis. ». Pour autant, nous ne pensons pas que notre corpus puisse avoir un aspect normatif, ou même que les sous-ensembles qui le composent puissent être érigés en canons. Il vaut comme étant un objet dont nous décrivons une partie, et c'est ce à quoi il sert, sans jugement de valeur porté à aucun instant sur la « qualité² » de la production orale que nous avons recueillie. La notion de norme est suffisamment considérée aujourd'hui comme

¹ Il donne alors l'exemple suivant : « sur deux listes de soixante adjectifs issus de deux corpus de dégustation, oral et écrit, seuls une dizaine sont communs ».

² Le terme *qualité* est utilisé ici dans le sens de correspondance à une norme établie, en l'occurrence la norme grammaticale française.

liée à la grammaire, celle d'usage comme liée à la linguistique descriptive, pour que nous n'insistions pas sur ce point. La seule norme à laquelle nous souscrivons est celle de la transcription ; c'est uniquement sur elle que nous pouvons porter un jugement de qualité.

Enfin, le quatrième axe que propose F. Rastier est celui de l'*entretien* du corpus. En effet, même si le linguiste travaille sur un corpus fermé, il est nécessaire, selon lui, de l'« entretenir » régulièrement, car « tout corpus, même fermé, qui ne fait pas l'objet d'une élaboration continue, se périmé, et paradoxalement devient inutilisable s'il n'est pas utilisé ». Un corpus doit être validé, et le processus de validation comprend les étapes suivantes : « présomption unifiante qui préside à la réunion du corpus ; établissement ; enrichissement ; annotations ; commentaires et exploitation ». En guise de conclusion à cette partie, intitulée « *Ressources linguistiques* » ou *corpus* ?, l'auteur « assigne divers objectifs à l'élaboration d'un corpus : tester et améliorer son homogénéité, sa représentativité, son codage ; en produire des sous-ensembles pertinents pour une catégorie de requêtes ; aider à l'analyse sémantique des structures textuelles ». Nous ne reviendrons pas ici sur la pertinence des sous-ensembles que nous avons pu construire grâce aux critères de sélection des locuteurs, puisque nous l'avons expliquée dans la partie concernant cela (1.2.1.1., [Choix des critères de sélection des locuteurs](#)).

Un corpus de langue parlée doit être transcrit pour être étudié, et ce paradoxe fondamental est à l'origine du problème de l'unification des conventions de transcription, qui diffèrent selon l'usage que l'on veut faire d'un tel corpus, comme nous l'avons vu au cours de la partie précédente (1.3.1., [Historique de la notion de corpus](#)). Nous voudrions ici prendre part à ce débat, en signalant certaines difficultés et ce qui nous paraît être un des écueils les plus stériles du problème de la diffusion de ces corpus de langue parlée.

Constituer un corpus est un travail de longue haleine, comme nous l'avons montré au début de cette étude. Par conséquent, les enquêteurs-transcripteurs, ayant passé du temps à transcrire minutieusement leurs enquêtes, ne voient pas toujours d'un bon œil la diffusion de leur travail, et le gardent souvent un peu jalousement. Leur idée est la suivante : ayant passé beaucoup de temps à faire des enquêtes, à les transcrire, à relire, pour se constituer une base de travail, pourquoi devrait-on offrir aux autres le temps que l'on a investi et le résultat de son travail ? A vrai dire, c'est l'idée que nous nous en faisons au début de cette recherche. Mais la lecture de certains articles (notamment ceux que nous citons dans la partie

précédente) nous a conduit à envisager cela d'une toute autre manière : on ne peut pas à la fois regretter le manque de diffusion des corpus de langue parlée et refuser de les diffuser. Ne pas diffuser un corpus, c'est, selon l'expression en usage, scier la branche sur laquelle on est assis. C'est refuser l'élargissement des études sur le français parlé, leur diversité. C'est nier les objectifs de la recherche, et les moyens de les réaliser. C'est pourquoi nous avons décidé de prêter notre corpus aux chercheurs qui nous en ont fait la demande. Ainsi, I. Fujimura¹, de l'Université de Nagoya (Japon), a trouvé quelques exemples intéressants qui lui ont servi lors d'une présentation aux JADT2004. L. Burnard (Oxford University Computing Services) devrait tester notre corpus à l'aide de X/SARA², un logiciel mis au point par son équipe. C. Vaguer (MoDyCo, Paris 10) a trouvé dans ce corpus des expressions utiles pour ses recherches sur les Sprep introduits par *dans*. Ainsi, les échanges réalisés au cours des quelques colloques auxquels nous avons participé se sont avérés fructueux, car les discussions engagées avec les participants nous ont permis de découvrir des approches et des points de vue différents, qui nous ont souvent aidé à sortir de ce que nous croyions être parfois une voie sans issue.

La définition que nous avons proposée plus haut contient en elle-même le principal reproche fait à l'étude de la langue parlée : le fait que le corpus soit nécessairement limité pose le problème de l'exhaustivité de l'étude et par conséquent de sa pertinence. Pour tenter d'apporter une réponse à ce problème, certains linguistes utilisent des corpus très importants (Corpaix comporte aujourd'hui plus de deux millions de mots), en pariant sur le fait que plus il y a d'occurrences d'un même phénomène ou d'une même construction, mieux on observe sa diversité distributionnelle. Mais le problème posé reste toujours celui de la limite : à partir de quand considère-t-on qu'on possède un nombre d'occurrences suffisant pour passer à l'étude ? Peut-on considérer qu'on peut décider d'arrêter un relevé et ainsi présenter une étude non-exhaustive sur un corpus par nature lui-même limité ? Nous avons contourné ce problème en proposant un inventaire, donc un relevé exhaustif de toutes les constructions verbales apparaissant dans notre corpus (telles que nous les avons définies dans la partie [2.3](#).

¹ Fujimura, I., M. Uchida & H. Nakao, 2004, "*De vs des* devant les noms précédés d'épithète en français : le problème de petit ", in *Le Poids des mots* vol 1, (Actes des 7èmes Journées internationales d'Analyse statistique des Données Textuelles), Presses Universitaires de Louvain, p.456-467,

² <http://www.oucs.ox.ac.uk/rts/xaira/Talks/jlc2003.html>

Définitions) ; devant l'ampleur des données à traiter, nous les avons réduites en ne traitant pas les verbes à l'infinitif et les formes en *-ant*. Nous verrons plus précisément au cours de la deuxième partie les formes que nous avons exclues de ce travail.

Après cette réflexion sur le statut du corpus, revenons-en au texte, et envisageons le de manière plus précise : J.-M. Adam (1990 : p. 91) analyse et définit le texte (écrit ou « dérivé » de l'oral) à un niveau encore inférieur à celui qu'utilise F. Rastier :

Un TEXTE est une structure hiérarchique complexe comprenant n séquences – elliptiques ou complètes – de même type ou de types différents.

Nous voyons bien ici que l'unité thématique n'implique pas nécessairement une unité de *type* de discours, puisque, parmi la série que ces types composent, proposée par J.-M. Adam (*op. cit.* : pp. 86-87) (narratif, injonctif-instructionnel, descriptif, argumentatif, explicatif ou conversationnel-dialogal), nous pouvons en trouver plusieurs apparaissant dans un discours produit par un même locuteur :

Généralement, un texte combine des séquences différentes : description dans un récit, récit dans une argumentation ou une conversation, etc.

L'auteur propose alors d'introduire la notion de *séquence* à l'intérieur même de celle de texte, posant qu'un texte se compose de n séquences, différenciées par leur type. Il définit la séquence comme :

- un réseau relationnel hiérarchique : grandeur décomposable en parties reliées entre elles et reliées au tout qu'elles constituent
- une entité relativement autonome, dotée d'une organisation interne qui lui est propre et donc en relation de dépendance/indépendance avec l'ensemble plus vaste dont elle fait partie.

Ainsi, pour chacun des 27 sous-corpus que nous possédons, il est possible d'envisager plusieurs séquences de type différent, que ce soit dans les 9 corpus composant l'ensemble *Montluçon*, ou dans les 18 autres composant *Vichy* et *Moulins*, leur absence d'unité thématique étant atténuée par cette pluralité.

Genres

D. Biber (*in* Bilger, 2000 : p. 225) insiste sur l'importance de la classification des corpus de langue parlée en différents « registers », afin d'observer un usage particulier dans la diversité de ses emplois, ce qui est la condition nécessaire selon lui pour effectuer une analyse complète du fonctionnement de cet usage :

In the present paper, I take a strong position on the importance of register for studies of discourse and grammar, arguing that most functional descriptions of a grammatical feature will *not* be valid for the language as a whole. Rather, characteristics of the textual environment interact with register differences, so that strong patterns of use in one register often represent only weak patterns in other registers. Thus, a complete functional analysis must consider the patterns of use in several registers.

F. Rastier (1989 : p. 37) pose quant à lui :

Il n'existe pas de texte (ni même d'énoncé) qui puisse être produit par le seul système fonctionnel de la langue (au sens restreint de mise en linguistique). En d'autres termes, la langue n'est jamais le seul système sémiotique à l'œuvre dans une suite linguistique, car d'autres codifications sociales, le genre notamment, sont à l'œuvre dans toute communication verbale. [...] Un discours s'articule en divers genres, qui correspondent à autant de pratiques sociales différenciées à l'intérieur d'un même champ. Si bien qu'un *genre* est ce qui rattache un *texte* à un *discours*. Une typologie des genres doit tenir compte de l'incidence des pratiques sociales sur les codifications linguistiques. [...] L'origine des genres se trouve donc dans la différenciation des pratiques sociales.

Citant ce passage, J.-M. Adam pose ensuite l'hypothèse que

Les genres sont probablement des codes « seconds » par rapports aux principes de régularité transphrastique [...]. Les genres relèvent, de toute façon, d'un domaine de recherche extra-linguistique.

Enfin, F. Rastier et al. (1994 : p. 222) proposent une définition arrêtée du genre et du discours, que l'auteur *d'Arts et sciences du texte* reprendra dans ses ouvrages suivants :

[genre] Programme de prescriptions positives ou négatives (et de licences) qui règlent la production et l'interprétation d'un texte. Tout texte relève d'un genre et tout genre, d'un discours. Les genres n'appartiennent pas au système de la langue au sens strict, mais à d'autres normes sociales.

[discours] Ensemble d'usages codifiés attachés à un type de pratique sociale. Ex. : discours juridique, médical, religieux.

Voyons à présent l'intérêt que peut présenter pour notre étude un corpus thématisé.

1.3.4. Le thème comme base du discours

Comme nous l'avons montré lors de la première partie de ce travail, les deux tiers des enquêtes que nous avons menées dans l'Allier ont pour thème le *petit commerce*. L'unité thématique est importante dans la réalisation d'un corpus échantillonné. En effet, à supposer qu'elle conditionne des productions lexicales contenues *a priori* dans un même ensemble, dans un même *champ lexical*¹, elle permet, pour notre inventaire et notre classement des constructions verbales, d'observer les verbes et leur complémentation dans un maximum de contextes¹. En somme, la thématisation d'un corpus permet d'observer une ou plusieurs classes lexicales, liées entre elles par le sujet même du discours. La cohérence de ces classes permet une observation plus fine de leur distribution et de leur fonctionnement, tant il est vrai qu'avec une soixantaine d'occurrences du verbe *acheter*, par exemple, la description de sa complémentation est plus aboutie qu'avec seulement une dizaine d'occurrences.

1.3.5. Exemples de requêtes lexicales thématiques Montluçon / Vichy-Moulins

Pour illustrer notre propos, comparons différentes requêtes appliquées au corpus « Montluçon » en opposition aux corpus « Vichy » et « Moulins » :

La requête [commerce*], qui extrait du corpus à la fois les substantifs au singulier et au pluriel, donne pour Montluçon une occurrence, contre 211 pour Vichy et Moulins.

La requête [#v(acheter)], qui extrait du corpus toutes les occurrences du verbe acheter, à l'infinitif ou conjugué, donne 5 occurrences pour Montluçon, contre 60 pour Vichy et Moulins.

Nous voyons donc bien que le thème retenu pour une enquête servant à la constitution d'un corpus spécifie le lexique employé. Il faudra donc que l'on tienne compte

¹ *acheter*, *vendre*, etc. sont des verbes qui ont une forte productivité dans ce corpus, comme nous le montrerons dans la partie 1.3.5.

de cela si l'on souhaite travailler sur les fréquences des constructions. En effet, la fréquence d'apparition d'un verbe comme *acheter* aurait certainement été plus importante si le thème de Montluçon avait été identique aux autres. Cependant, les 27 corpus étant affectés à des types correspondant aux critères de sélection vus plus haut, il sera aisé de séparer les recherches, et de sélectionner uniquement Vichy et Moulins, pour effectuer un travail d'analyse thématique sur le petit commerce.

L'argument principal auquel nous nous référerons est celui de la non-incidence du thème pour une recherche en syntaxe qui ne vise pas d'autre objectif que celui de la description et du classement. En effet, l'observation du fonctionnement des constructions verbales dans ce corpus n'est pas gênée par cette incohérence thématique, car même si la sélection lexicale opérée par le thème retenu peut s'avérer d'un grand intérêt pour des recherches spécifiquement sémantiques, elle ne modifie en rien l'observation syntaxique à proprement parler. Lorsque nous étudions la complémentation des constructions verbales attestées dans notre corpus de français parlé, nous posons que l'importance du volume du corpus nous permet d'observer les constructions principales. Pour reprendre l'exemple du verbe *acheter*, pour lequel nous trouvons 65 occurrences, dont 60 sont issues des corpus « Vichy » et « Moulins », nous pensons que la trentaine d'occurrences que nous aurions potentiellement trouvées dans « Montluçon » si celui-ci avait eu le petit commerce pour thème n'aurait pas nécessairement apporté de nouveau type de construction.

¹ Ce « maximum » est toutefois limité par la durée de l'enregistrement, de la transcription et par le caractère fini du discours en général

2. Inventaire et classement des constructions verbales

Cette partie a pour but de montrer l'intérêt, les objectifs et les applications possibles de ce travail. L'intérêt d'entreprendre un inventaire des constructions verbales dans un corpus de français parlé est avant tout celui de la description systématique. En effet, comme nous le verrons au cours de cette partie, les travaux entrepris jusqu'à présent sur le verbe se fondent le plus souvent sur le travail des lexicographes avant d'envisager un travail sur corpus (Proton, Melis, 1995 ; Willems, 2002b) et ceux qui se servent de corpus utilisent très souvent des corpus de français écrit associés à du français parlé (Greidanus, 1990). Certes, la difficulté à trouver un corpus de français parlé compte sans doute beaucoup dans ces choix d'orientation. Ainsi, notre travail est l'un des premiers à se servir exclusivement d'un corpus de français parlé pour recenser les constructions verbales qui y apparaissent. La méthode que nous avons choisie pour classer ces constructions, très largement inspirée de l'Approche Pronominale, pourrait permettre à ce travail une complémentarité et un gain de temps pour les projets encore en cours sur le sujet, que leurs objectifs soient identiques ou non.

Les objectifs de cette étude sont de degrés divers. Fondamentalement, ce travail vaut pour ce qu'il est : un inventaire et un classement des constructions verbales dans un corpus de français parlé. Nous avons vu précédemment de quelle manière était constitué le corpus. Nous allons à présent montrer la méthode utilisée pour mener à bien l'inventaire, et le classement. Ce que nous nommons « inventaire » n'en est en fait pas tout à fait un : il s'agit plutôt d'un recensement, dans la mesure où, comme nous le préciserons plus tard, nous n'avons pas retenu ici toutes les formes verbales (par exemple, nous avons exclu les verbes à l'infinitif). Quant au classement, il ne sera utile et utilisable qu'à la condition que les propriétés utilisées soient considérées comme pertinentes. Ceci étant, plusieurs applications sont alors envisageables.

D'une part, il semblerait que les dictionnaires de français (notamment notre référence, le Petit Robert), même s'ils tiennent compte de la langue parlée, ne le font souvent qu'en précisant que la tournure est *familière*. De plus, aucune information n'est mentionnée quant à la fréquence d'emploi de cette tournure, à son importance dans le langage usuel, et pour cause : les corpus de français parlé ne sont pas encore utilisés pour l'élaboration des entrées dictionnaires. Pourtant, cela permettrait d'élaborer un dictionnaire d'usage et de norme, grâce auquel l'utilisateur intéressé par la langue pourrait avoir une idée de son état, de son

système, en comparant la norme (attachée bien souvent à l'écrit) et l'usage (à l'oral). Un tel dictionnaire pourrait aussi servir de base à un apprentissage du Français Langue Etrangère. En effet, les constructions relevées ici apparaissent avec leur fréquence d'emploi dans notre corpus, ce qui permet, entre autres choses, de définir la progression de l'apprentissage : dans un premier temps, les constructions les plus fréquentes, ensuite celles dont la fréquence est moindre.

D'autre part, une collaboration avec des informaticiens-linguistes travaillant sur le Traitement Automatique du Langage (désormais TAL) pourrait s'avérer fructueuse pour la description, et pour la formalisation. En effet, comme le remarque Christophe Benzitoun (2003 : p. 1), les deux sont souvent dissociées, voire senties comme antithétiques, alors que la description pourrait être la base d'une formalisation destinée au traitement automatique, qui se satisfait pour l'instant de la formalisation de la langue écrite, plus aisée à mettre en place, puisque la chaîne est évidente (découpage d'un texte en phrases à partir de l'idée qu'elles commencent par une majuscule et se terminent par un point). Cependant, de nombreux linguistes ont démontré récemment¹ que la notion de phrase telle que nous l'apprenons à l'école n'est pas pertinente ; elle ne vaut pas pour la langue parlée, dans laquelle elle ne commence pas par une majuscule ni ne se termine par un point. D'un autre point de vue, c'est aussi ainsi que la prise en considération de la langue parlée pourrait permettre sans aucun doute un apprentissage plus près de la réalité de la langue, et ainsi encourager les apprenants étrangers à accéder à la langue « quotidienne », plutôt qu'à retenir des formes normatives qui feront sourire les locuteurs natifs.

2.1. Les travaux et projets sur les constructions verbales du français

Depuis la fin des années 1980, de nombreux travaux sur les constructions verbales ont été réalisés ; dans un premier temps, nous en présenterons quelques-uns qui ont retenu notre attention pour l'élaboration de ce travail, en montrant par la suite en quoi le nôtre s'en distingue.

¹ VERBUM, 2002, tome XXIV, n° 1-2, *Y a-t-il une syntaxe au-delà de la phrase ?*, Actes du colloque international, Paris III, septembre 2000, numéro coordonné par M. Charolles, P. Le Goffic et M.-A. Morel.

2.1.1. Le lexique-grammaire

Quelques travaux sur les constructions verbales ont déjà été réalisés. Le plus complet est sans doute celui élaboré au LADL par M. Gross (1975), C. Leclère, A. Guillet et J.-P. Boons (1976). Ce travail contient une classification de tous les verbes simples du français (5 000 verbes divisés en 15 000 emplois et répartis en une soixantaine de tables qui répertorient leurs propriétés syntaxiques et distributionnelles)¹. Mais son utilisation semble restreinte : ces tables, peut-être jugées difficiles à lire, seraient donc sous-employées, notamment dans le domaine de l'enseignement du français. Peu d'ouvrages postérieurs traitant de la description de la complémentation verbale se servent de ces tables : P. Le Goffic et N. Combe McBride ont mené en 1975 une étude sur les constructions verbales fondamentales du français, fondée sur les résultats du LADL. Ceci est l'un des rares travaux effectués à partir des recherches théoriques du LADL, travail qui est par ailleurs critiqué par Tine Greidanus (1990 : pp. 80-81) :

Le Goffic et Combe McBride (1975), livre « essentiellement destiné à ceux qui enseignent ou qui apprennent le français », présente un inventaire des constructions à partir des verbes, adjectifs et substantifs opérateurs du *Français Fondamental* (1^{er} degré). Les auteurs distinguent à peu près 35 constructions verbales différentes [...]. Cela paraît peu. Pour le nom sujet, ils ne font pas la distinction entre nom humain et nom non-humain, contrairement à ce que nous avons fait. On devrait donc peut-être multiplier ces 35 constructions par deux pour rendre plus correcte une comparaison avec nos données. En décrivant toutes les constructions des verbes du FF1, on risque d'inclure des constructions qui ne sont pas fréquentes. Nous avons comparé les 35 constructions de Le goffic et Combe McBRide avec nos 42 premières constructions. On constate que 16 (!) de leurs constructions ne figurent pas parmi nos constructions fréquentes. Il s'agit dans tous les cas de constructions à trois places. Il se peut que l'absence de certaines de ces constructions parmi nos 42 premières constructions soit due au caractère parlé de notre corpus. Il n'est pas improbable qu'en français parlé les constructions verbales soient plus courtes, moins étoffées qu'en français écrit. En français parlé on rencontre de toute façon plus de constructions elliptiques qu'en français écrit.

2.1.2. Proton

Un autre projet s'est donné pour objectif de présenter les valences verbales. Il s'agit de Proton, dictionnaire électronique des valences verbales, élaboré à la K. U. de Louvain sous la direction de K. van den Eynde, qui en est à l'origine. C. Eggermont (analyse des données

du français) et E. Broeders (traitement automatique et la gestion des banques de données) ont collaboré avec lui pour mener à bien cette entreprise, puis L. Melis les a rejoint lors de la dernière phase du projet. Ce dictionnaire, élaboré à partir des entrées verbales du Petit Robert et du dictionnaire de valence de Busse et Dubost (*Französische Verblexicon*), fait apparaître les valences sous leur forme pronominale. Les chercheurs utilisent comme cadre méthodologique l'Approche Pronominale. Ils ont proposé, comme nous le verrons dans la partie consacrée aux pronoms, un inventaire très détaillé des pro-formes pour le français.

Ce dictionnaire automatisé des valences verbales « se présente comme une suite d'entrées consacrées chacune à une formulation verbale² » (Melis : 1995), les formulations verbales n'étant pas étudiées à l'origine à partir de corpus, mais grâce au Petit Robert et au *Französische Verblexicon*. Une étude sur corpus était commencée en 1995, permettant ainsi de confronter leurs données avec celles de corpus, composés de français oral et d'écrits journalistiques.

Les critères qui ont précédé la détermination des formulations verbales, au nombre de quatre, sont les suivants :

- Le premier est le nombre de positions syntaxiques construites par le verbe. C'est un critère que nous avons aussi retenu dans notre inventaire, et qui nous a servi parfois à distinguer, pour un même verbe morphologique, plusieurs fonctionnements différents.
- Le deuxième critère est celui de la composition du paradigme de pro-formes apparaissant dans chaque position. Ce critère sert à prévoir quelles seront les extensions lexicales potentielles ; par exemple, si un paradigme de sujet contient la pro-forme *je*, tous les référents humains pourront s'y substituer. Nous avons aussi retenu ce critère, qui nous a paru fondamental pour constituer une base de données des constructions verbales de notre corpus, dans la mesure où ces paradigmes représentent une sorte de squelette de construction, que le locuteur pourra « charner » avec les éléments lexicaux contenus dans le paradigme de cette pro-forme. Ainsi, à partir de chaque pro-forme, un lien est établi avec les éléments du lexique dont le contenu

¹ cf. <http://infolingua.univ-mlv.fr/LADL/Historique.html>.

² « Par formulation verbale on peut entendre une construction spécifique représentée à l'aide de paradigmes pronominaux et ses reformulations possibles. Un verbe morphologique peut donner lieu à plusieurs formulations.

syntactico-sémantique est identique, pour permettre à l'utilisateur de cette base de données, qu'il soit en apprentissage du français langue étrangère, ou même du français langue maternelle, une utilisation efficace.

- Le troisième critère est celui des reformulations accompagnant la formulation, que nous n'avons pas, quant à nous, retenu.
- Enfin, le quatrième indique les liens de reconstruction entre formulation, non précisé dans notre travail.

2.1.3. Le Dictionnaire Contrastif des Valences Verbales

Un autre projet de dictionnaire de ce type, le Dictionnaire Contrastif des Valences Verbales (CVVD), est en cours à l'université de Gent. Dominique Willems, très attachée aux liens entre syntaxe, lexique et sémantique, le présente ainsi (Willems, 2002b):

« La recherche contrastive à l'Université de Gand remonte à une quinzaine d'années. Elle a débuté à la fin des années 80 par l'établissement d'une grammaire contrastive néerlandais/français/anglais, qui avait pour but une description minutieuse des différences et des similitudes entre les trois langues dans le domaine du SN et de l'organisation de la phrase et l'établissement d'une terminologie grammaticale uniforme et cohérente. Assez rapidement, la nécessité d'un complément lexical s'est fait sentir, en particulier dans le domaine du verbe : les constructions verbales et leur interprétation posent en effet beaucoup de problèmes lors de l'apprentissage des langues étrangères et aucun relevé exhaustif pour ces trois langues n'existe à ce jour. La grammaire lexicalisée est en effet un domaine qui n'est recouvert ni par les grammaires, ni par les dictionnaires.

Dès 1993, l'équipe contrastive gantoise s'est attelée à la tâche d'élaborer un dictionnaire contrastif des valences verbales (dorénavant CVVD) pour les trois langues. En ce qui concerne le nombre de lexèmes verbaux, l'objectif poursuivi est la description

Ceci fait que le dictionnaire français contient 8599 entrées, représentant 3734 verbes morphologiques différents [...]. » (*op. cit.*)

d'environ 500 verbes pour chacune des langues. En novembre 2001, quelque 220 lemmes ont été décrits.

L'originalité du CVVD peut être résumée en trois points essentiels :

(1) Contrairement aux dictionnaires bilingues existants, l'approche est multidirectionnelle et la description exhaustive pour les trois langues (tant sur le plan des structures que des sens verbaux, sur le plan des différences que des similitudes).

(2) L'établissement d'un rapport systématique et explicite entre sens et structure est un des aspects essentiels du CVVD : pour chaque sens distingué, les diverses structures syntaxiques possibles sont énumérées, ce qui permet de mesurer de façon précise pour chaque langue le rapport entre sens et syntaxe.

(3) Le CVVD vise à refléter l'usage réel de la langue contemporaine : les données sont donc établies inductivement à partir d'une analyse de corpus d'exemples réels. La confrontation avec les dictionnaires ne vient que par après et sert essentiellement à vérifier l'exhaustivité de l'analyse et à combler les lacunes éventuelles. La démarche inductive à partir de corpus ajoute au projet une dimension statistique précieuse. »

2.1.4. Les constructions verbales en français parlé. Etude quantitative et descriptive de la syntaxe des 250 verbes les plus fréquents

T. Greidanus (*op. cit.*) a, quant à elle, publié sa thèse en 1990. Son travail est intéressant, car il pose un grand nombre de statistiques et de fréquence d'emploi de telle ou telle des constructions des « 250 verbes les plus fréquents en français parlé ». Sur les 5 corpus que T. Greidanus utilise, seulement 2 peuvent être classés comme appartenant à du

« français parlé » : il s'agit de l'*Enquête statistique sur la langue parlée*¹ et de l'*Etude sociolinguistique sur Orléans*. L'*Enquête statistique sur la langue parlée* est « une liste de fréquence du français parlé, celle qui a servi à l'élaboration du *Français Fondamental* (1^{er} degré) et qui est présentée dans Gougenheim et al. (1967). Des interviews avec 275 témoins ont fourni les matériaux de la liste de fréquence. Les textes dépouillés proviennent pour une petite partie d'enregistrements phonographiques historiques. La plupart des textes ont été recueillis par enregistrements directs au début des années 50, en vue du projet d'élaboration du *Français Fondamental*. » (Greidanus, 1990 : p. 12). L'*Etude sociolinguistique sur Orléans* est « un corpus d'enregistrements du français parlé [qui] a été recueilli à Orléans entre 1968 et 1971. Il y a au total 487 enregistrements qui se divisent en huit catégories : interviews sur questionnaires, reprises de contact informelles, enregistrements spontanés, communications téléphoniques, interviews 'sur mesure', conférences/discussions, enregistrements avec des témoins inconnus, interviews au Centre Médico-Psychopédagogique. » (*op. cit.*, p. 13). Les autres corpus sont de type « écrits » : il s'agit du *Dictionnaire des fréquences*, qui repose sur un corpus de 71 millions d'occurrences issues de textes littéraires du XIX^e et du XX^e siècle, du *Frequency Dictionary of French Words*, fondé lui aussi sur des textes littéraires et journalistiques en prose et enfin du *Vocabulaire du roman français 1962-1968*, fondé sur 25 romans best-sellers publiés entre 1962 et 1968.

Cette étude quantitative et descriptive des constructions verbales du français parlé est réalisée à l'aide du seul *Corpus d'Orléans*. Les chapitres 4 et 5 de cet ouvrage traitent de manière statistique, pour le chapitre 4, les constructions verbales relevées dans la langue parlée, et pour le chapitre 5, la construction impersonnelle. T. Greidanus décrit dans un premier temps les constructions verbales relevées dans le *Corpus d'Orléans*. Elle y relève 241 verbes, qui fournissent « 21 370 occurrences d'une forme verbale conjuguée au mode personnel » (*op. cit.*, p. 53), dont elle retient 18 837 occurrences 'verbe + arguments', mettant de côté les formes figées ou semi-figées (qui sont soit des « opérateurs méta-discursifs, soit d'autres formes verbales lexicalisées »). Elle distingue au cours de son étude quatre types de constructions, selon la voix à laquelle est employé le verbe recteur : active, passive, pronominale à valeur passive et traite la construction impersonnelle séparément, dans le cinquième chapitre.

¹ GOUGENHEIM, G., MICHEA, R., RIVENC, P. et SAUVAGEOT, A., 1967, *L'élaboration du français fondamental* (1^{er} degré), Paris, Didier

La liste des constructions verbales qu'elle a relevées dans le Corpus d'Orléans s'élève à 303 constructions différentes, dont 103 n'ont qu'une occurrence dans le corpus. Nous relevons cette remarque, car nous avons aussi constaté cela dans le corpus Allier. Voici plus en détail la fréquence des constructions qu'elle étudie :

Un tiers des constructions rend donc compte d'une centaine d'occurrences seulement, soit de 0.5% du total. A l'autre extrémité de la liste nous observons le phénomène inverse [...]. La première construction couvre plus de 20% des occurrences, les quatre premières constructions 49%, les dix premières 67%. Les pourcentages de couverture descendent plus lentement pour les fréquences moyennes. Les constructions 11 à 20 couvrent ensemble 12.39% des occurrences, les constructions 21 à 30 6.16%, les constructions 31 à 40 3.62% et les constructions 41 à 50 2.57%. Les 42 premières constructions rendent compte de 90% de toutes les occurrences.

Voici à présent les conclusions qu'elle présente (*op. cit.*, pp. 83-84) :

Une dixième partie environ des occurrences est constituée par des opérateurs métadiscursifs.

Nous avons trouvé 303 constructions différentes. Ces constructions représentent la majorité des constructions possibles. Les dix premières constructions couvrent près de deux tiers de toutes les occurrences. Les 42 premières constructions couvrent 90% des occurrences.

Presque toutes les constructions sont à la voix active. Les constructions dérivées (constructions passive, impersonnelle et pronominales en fonction de passif) occupent une place marginale.

Le sujet 'type' d'une phrase à la voix active est un (pro)nom de personne, le second argument 'type' un (pro)nom de chose. 10% des énoncés seulement ont un troisième argument. L'argument 'type' est alors un syntagme prépositionnel avec un (pro)nom de personne.

La construction passive a une légère préférence pour le sujet animé. Cette préférence n'est pas statistiquement significative quand on se base sur les constructions passives réalisées.

La phrase pronominale à valeur passive a une préférence nette pour le sujet non animé.

L'ellipse est le trait de la langue parlée le plus pertinent quand il s'agit de constructions verbales.

On devrait, dans l'enseignement du français langue étrangère, tenir davantage compte de données fréquentielles concernant les constructions verbales et des constructions caractéristiques du français parlé.

Il n'existe pas deux verbes se construisant de la même manière.

Nous n'avons pas, quant à nous, utilisé les mêmes critères de classement que T. Greidanus. En effet, elle n'utilise pas de cadre théorique particulier : se basant essentiellement sur l'analyse en constituants immédiats, elle décrit, pour les constructions les plus fréquentes, la nature des constituants régis par le verbe (en éliminant les compléments de phrase), leur position par rapport au verbe et enfin, la fréquence d'apparition dans le corpus du type de construction étudié, le nombre de verbes y entrant et un verbe témoin, qui est en fait le verbe possédant le plus grand nombre d'occurrences pour ce type de construction.

2.2. Pourquoi un inventaire ?

Dans l'ouvrage fondateur de la théorie du Lexique-grammaire (1975 : p. 20), Maurice Gross écrivait :

Depuis le début du siècle, les linguistes semblent avoir renoncé à constituer des inventaires. Ceux-ci nous apparaissent pourtant comme fondamentaux.

Les inventaires permettent, entre autres choses, de mesurer relativement précisément l'étendue d'un phénomène dans la langue, même si

Par étendue [d'un phénomène], nous n'entendons pas fréquence d'apparition dans les textes, cette notion n'ayant pratiquement aucun sens statistique pour la syntaxe, mais fréquence d'apparition dans le lexique de la langue. (*ibid.*)

La grammaire est donc envisagée ici « par la fenêtre du lexique », qui sert de base à l'observation systématique des structures de la langue. Par exemple, pour notre inventaire des constructions verbales, le lexique sera le point de départ de l'observation, c'est-à-dire qu'à partir des différents verbes relevés dans le corpus, on observera les structures à l'intérieur desquelles ils apparaissent.

Dans la présentation qu'il fait de l'ouvrage de J.-P. Boons, A. Guillet et Ch. Leclère (1976 : p. 25), M. Gross revient sur ce manque de volonté taxonomique, en insistant sur le fait suivant :

C'est un fait que les linguistes ne construisent plus d'inventaires, et qui plus est, lorsqu'ils produisent un nouveau phénomène ou formulent une règle, ils ne font jamais l'effort d'étudier la

distribution du phénomène ou de la règle par rapport à des lexiques même ne possédant qu'une faible couverture de la langue. Cette activité de vérification élémentaire reviendrait à effectuer une classification, puisqu'en règle générale les observations varient avec les éléments lexicaux. Cette situation curieuse a vraisemblablement des raisons plus idéologiques que scientifiques. Il ne fait aucun doute que l'activité de construction de théories ou de modèles même ultra-simplifiés a, aux yeux d'un public nombreux, bien plus de prestige que celle qui consiste à établir des listes nécessitant efforts et subtilité. C'est certainement dans le caractère prédictif des théories qu'il faut rechercher les différences de prestige : on ne sait pas aussi communément que les classifications prédisent tout autant que les théories.

Cependant, M. Gross (1975 : pp. 25-26) critique l'utilisation de corpus dans le cadre des recherches qu'il mène :

Nous avons voulu réagir contre certaines recommandations qui, déniaient toute valeur à l'utilisation de l'introspection, conduirait à n'utiliser que des corpus lors de la détermination des formes syntaxiques d'une langue. En fait, compte-tenu des objectifs de la grammaire transformationnelle, une telle procédure ne peut être appliquée, l'étude détaillée la plus simple le montre aisément.

Si les linguistes du Laboratoire d'Automatique Documentaire et Linguistique (LADL) utilisent les méthodes transformationnelles afin d'élaborer les différents lexiques-grammaires - ce qui permet, en quelque sorte, une certaine extensivité -, nous avons quant à nous choisi d'élaborer un inventaire à partir d'un corpus de français parlé. Comme nous l'avons mentionné dans la première partie, un des reproches qui pourraient nous être adressé est celui de la limitation de l'observation, étant donné le caractère *a priori* non extensif d'un corpus. Ceci est peut-être en partie ce qui conduit M. Gross à prendre la position que nous venons de noter, outre le fait que dans les années 70, les outils (les concordanciers, par exemple) qui sont actuellement à la disposition des linguistes pour le traitement des corpus n'existaient pas, ce qui supposait alors un traitement manuel systématique. La linguistique de corpus, si elle n'est, en principe, nécessairement pas fondée sur l'extensivité (puisque, comme nous l'avons déjà vu, un corpus est par nature limité), possède tout de même quelques avantages, et non des moindres. D'une part, la description de données attestées permet parfois de découvrir des usages que la seule intuition du linguiste n'aurait pu trouver, comme le signale Claire Blanche-Benveniste (1996 : p. 28) :

Les études faites sur les corpus, et surtout sur les corpus de langue parlée, complètent utilement ce que nous enseigne la simple intuition sur notre propre langue. Comme l'ont remarqué

tous les observateurs, cette intuition est parfois fragile, surtout pour les phénomènes peu étudiés, dont nous n'avons pas une représentation bien nette, et particulièrement pour les faits de variation.

D'autre part, ce type de description permet au chercheur un rapport des plus directs avec la langue qu'il étudie. En effet, la description est toujours dans ce cas description d'un *usage*, au sens saussurien du terme : il s'agit ici d'étudier la langue dans un usage attesté. Nous ne reviendrons pas ici sur ce point, que nous avons déjà traité, si ce n'est pour souligner l'importance de la multiplication de ce type d'investigations, qui, à terme, pourraient conduire à l'élaboration (et à la description) d'une « grammaire » de la langue orale ; en outre, et ceci est un des principaux arguments des linguistes spécialistes de l'étude du français parlé, la description du système de fonctionnement de la langue orale permettra sans doute des avancées significatives dans divers domaines, que ce soit celui de la description grammaticale « pure », des méthodes d'enseignement de la langue (en français langue première ou seconde), ou de la traduction automatique (puisque'aujourd'hui, les principaux axes de recherche en TAL s'orientent à la fois sur la langue écrite et sur la langue orale (les informaticiens linguistes tentent de trouver des procédures automatiques de reconnaissance vocale et de traitement de la parole)).

Ainsi, recenser les différents types de constructions qui apparaissent dans la langue parlée n'est pas vain. De plus, les inventaires sont, depuis l'essor de l'informatique, de plus en plus aisés à effectuer. Avec un corpus informatisé et transcrit de manière adéquate à la recherche envisagée, avec un concordancier et de la patience, tout linguiste peut consigner les structures observables et les classer selon ses objectifs de recherche, qu'ils se situent en morphologie, en dialectologie, en syntaxe, en sociolinguistique, etc., car, comme le soulignait M. Gross (1976 : p. 7) :

L'étude de la syntaxe se pose exactement de la même façon que l'étude de toute famille d'objets importante de par son nombre¹. Les nombres en jeu, ici les nombres de phrases, sont tels qu'il est humainement impossible d'appréhender directement leur organisation, ou même l'existence éventuelle d'une organisation. L'idée naturelle est donc de grouper les objets se ressemblant en classes, et d'étudier les classes constituées. Comme les classes sont moins

¹ A plusieurs reprises, M. Gross compare la taxonomie linguistique à celle de la botanique et de la zoologie, argumentant la nécessité d'inventorier et de classer les objets du lexique de la même manière qu'Adanson a procédé pour les arbres, en les regroupant en familles ou en classes, selon leurs propriétés communes ou au contraire distinctives.

nombreuses que les objets, on peut espérer qu'il sera possible de percevoir l'existence d'une organisation qui les lierait, et d'en déduire une organisation pour les objets de départ.

Si M. Gross (*op. cit.*, p. 13) est très critique quant à la plupart de ses collègues linguistes, qui, selon lui « n'ont pas conscience de ce que sont les procédures scientifiques » et recourent « à des causes extra-linguistiques pour expliquer les variations de jugement d'acceptabilité », sans jamais « faire l'effort matériel important de varier les conditions expérimentales afin de s'assurer de la stabilité du résultat », il est convaincu que :

L'examen systématique du lexique constitue [...] un moyen [...] d'appréhender une langue de façon globale, c'est-à-dire d'en construire une image ayant un certain caractère de généralité. Ce n'est que dans un tel cadre qu'il est possible de détecter des phénomènes massifs, et de les opposer éventuellement à des exemples marginaux ou exceptions. (*ibid.*, p. 9)

Pour en venir au verbe, qui est au centre de ce travail, observons ce que P. Le Goffic présente dans son ouvrage sur la *Grammaire de la phrase française* (1993 : pp. 180-181). Après avoir brièvement caractérisé le verbe, il montre quelques uns des classements que les linguistes ont réalisés :

Un verbe se caractérise, dans le lexique de la langue, par son entourage lexico-syntaxique (quel type de sujet, quelles constructions, avec quels types de compléments), et par son sens, celui-ci étant susceptible de varier selon l'environnement. [...]

Rappelons simplement quelques-uns des principes de classement les plus importants et les plus étudiés :

typologie des aspects lexicaux (types de procès, Vendler)

classements à base logique ou sémantique :

. classement (logique) selon le nombre d'arguments accepté par le verbe (prédicat) : Ø argument : *pleuvoir*, 1 argument : *dormir*, 2 arguments : *battre*, 3 arguments : *donner* ;

. typologie actancielle ou casuelle, à partir des travaux de Fillmore (ainsi *ouvrir* comporte dans sa structure casuelle un agent, un instrument, et un objet)

. du classement des verbes, on passe au **classement sémantique des prédications** : Pottier, par exemple distingue cinq « voix » : l'existential, le situatif, l'équatif, le descriptif, le subjectif, + le causatif, qui s'applique à ces cinq voix ; J. François et d'autres travaillent sur des notions du type état, événement, action ;

INVENTAIRE ET CLASSEMENT

classement au regard de l'énonciation et des actes de discours : verbes constatifs, performatifs, interrogatifs, modaux, ...

classement à base morpho-syntaxique :

. classement selon les possibilités de variation diathétique (présentation respective du procès et des actants) : cf. la classe des verbes actifs (*battre* : *Paul bat Jean / Jean est battu par Paul*), la classe des verbes « neutres » (*casser* : *le vent casse la branche / la branche casse*)

. classement systématique selon les constructions (tables de Gross) :

- 43 tables de constructions nominales (huit mille entrées verbales)

- 18 tables de constructions complétives et infinitives (trois mille entrées verbales), sans oublier

- 20 tables d'expressions verbales figées (*louper le coche*) (vingt mille entrées)¹

Dans ce travail, nous avons donc recensé et classé les constructions verbales des verbes les plus fréquents apparaissant dans notre corpus. L'inventaire a été réalisé à l'aide du concordancier Xcor2000, conçu par Jean-Pierre Adam pour le GARS. Ce logiciel est attaché à un important dictionnaire des verbes français (environ 12 000 verbes recensés et lemmatisés pour l'instant). Ainsi, la première requête que nous avons utilisée pour cet inventaire est celle qui consiste à demander au logiciel d'extraire du corpus tous les verbes y figurant ([#v]). Puis, nous avons effectué la première étape du classement : « ranger » les verbes par ordre alphabétique, à partir de la liste des verbes qui apparaissent sous Xcor à la fin du résultat de la requête dont il s'agit. Pour les 352 verbes « morphologiques » présents dans ce corpus, nous avons donc effectué la requête permettant d'en extraire, pour chacun d'entre eux, toutes les formes (par exemple, la requête #v(*aller*) extrait du corpus toutes les formes (infinitives et conjuguées) du verbe *aller*). Les formes en question apparaissent, bien sûr, dans le résultat de la première requête, mais non-classées (pour reprendre l'exemple du verbe *aller*, ses formes au passé composé – *je suis allée chez le coiffeur* – apparaissent classées à la fois avec la forme *être* au présent, et avec la forme *aller* au participe passé.) Une fois les verbes classés alphabétiquement, selon leur base morphologique, on opère, pour chaque verbe, un classement plus fin, consistant à analyser les différents types de sujets et de

¹cf. Leclère, 1990 : pp. 112-122.

compléments qu'il régit, à partir de propriétés syntaxiques et distributionnelles que nous présenterons dans la partie 2.6.2.1 ([Les propriétés syntaxiques et distributionnelles](#)).

Dans cette partie, nous présentons un inventaire classé des constructions verbales de notre corpus, qui permet d'observer les différents types de complémentations verbales utilisées à l'oral, que ce soit pour un même verbe ou pour un ensemble de verbes.

Pour un même verbe, on observe les différentes constructions que nous avons relevées dans notre corpus. Cette observation permet de constater la complémentation effective dudit verbe. En effet, la valence d'un verbe n'est pas fixée, elle évolue, et des éléments appartenant à sa valence peuvent disparaître, et d'autres y entrer¹.

Pour un ensemble de verbes, on pourra les regrouper s'ils ont le même type de construction, ce qui permettra d'obtenir un classement typologique qui pourrait se révéler d'un usage pratique² pour les enseignants et les apprenants du français langue maternelle ou langue étrangère.

Ainsi, nous essaierons de répondre nous aussi à la question que pose P. Le Goffic (*ibid.*) : « étant donné une construction, pour quels verbes (ou quels types de verbes) la trouve-t-on ? ». Mais nous envisageons aussi bien de pouvoir classer les constructions de telle sorte qu'on puisse aussi bien répondre à la question « étant donné un verbe, quelles sont ses constructions ? », que l'auteur ne veut pas traiter dans son ouvrage, hormis pour le cas exceptionnel du verbe *être*.

D'autre part, ce classement peut servir de point de départ à la constitution d'un dictionnaire des constructions verbales du français parlé, tant nous avons souvent observé que certains verbes avaient, dans notre corpus, des compléments (et des sens) non envisagés par notre dictionnaire de référence³ (Petit Robert). Boons, Guillet et Leclère soulèvent en partie ce problème dans l'introduction de leur ouvrage (1976 : pp. 34-37), regrettant que les dictionnaires d'usage ne mettent en valeur uniquement la parenté sémantique entre deux constructions d'un même verbe, sans noter, lorsqu'elle existe, la parenté structurelle :

¹ Le verbe *précipiter*, par exemple, était intransitif, puis est devenu transitif.

² Pierre Le Goffic et Nicole Combe mac Bride avaient déjà réalisé, à partir du Français Fondamental 1, un « dictionnaire-grammaire » des constructions fondamentales du français. Cf bibliographie.

³ Le verbe *arriver*, par exemple, apparaît avec un complément de type « de + infinitif », avec un sens « locatif » que nous n'avons pas trouvé dans les dictionnaires cités : *j'arrive de travailler*.

Tous les dictionnaires consultés semblent ignorer la relation entre constructions différentes d'un même verbe évoquée ici, et *a fortiori* le fait que les verbes qui admettent les deux constructions ne sont vraisemblablement pas quelconques, mais forment une classe homogène. Ainsi, dans les dictionnaires d'usage, les entrées de chacun des verbes des exemples ci-dessus¹ se contentent de renvoyer les uns aux autres pour ce qui est de la parenté sémantique, sans marquer l'importante parenté structurelle qui les lie.

Les différents problèmes que nous avons rencontrés concernaient essentiellement la méthode de classement, et sa présentation.

Concernant la méthode, nous nous sommes inspirée de celle utilisée pour la description français parlé : l'Approche Pronominale (Karel van den Eynde, Claire Blanche-Benveniste *et al.*), que nous présenterons succinctement au cours de la première partie de ce chapitre.

La présentation en tableau, qui sera peut-être jugée difficilement lisible, permet cependant de synthétiser à la fois les traits des sujets (+/-personnel, asubjetcal, etc.) et des compléments (type P1, P2 etc.), mais aussi les différentes formulations (active, passive, pronominale). De plus, lorsque la réalisation du complément est complexe (par exemple : préposition + SN), elle est précisée en fin de ligne. L'un des avantages de ce tableau est aussi de présenter les éléments construits sur un axe syntagmatique ; en effet, présenter d'abord le type de sujet, puis le verbe constructeur (suivi de sa forme), et enfin le ou les compléments permet de saisir d'emblée la construction décrite, et d'observer directement sa réalisation grâce aux exemples classés du corpus figurant intégralement en annexe. L'observation est d'autant plus aisée que l'exemple est étiqueté de manière identique à la construction à laquelle il est attaché dans la base de données.

2.3. Définitions

Avant de présenter la méthode que nous avons utilisée, il nous semble ici nécessaire de poser quelques définitions. En effet, pour un même objet, les définitions diffèrent fréquemment selon l'angle par lequel il est observé et décrit. Nous expliquerons donc dans

¹ Les exemples illustrent les constructions symétriques qui caractérisent les verbes du type *abonder*, *fourmiller*, etc. :

La religion schismatique grecque abonde d'exigences minutieuses et de prescriptions vexatoires.

cette partie le point de vue que nous avons adopté pour ce travail, en faisant référence le plus souvent aux travaux de l'équipe dirigée jusqu'à peu par C. Blanche-Benveniste (GARS, devenue aujourd'hui DELIC, sous la direction de J. Véronis). Des ouvrages théoriques tels que la *Grammaire de la phrase française* (Le Goffic, 1993), ou la *Grammaire méthodique du français* (Riegel et al. 1994), ou encore la *Grammaire Larousse du français contemporain* (Chevalier et al., 1964) nous serviront aussi de base pour construire ces définitions, ainsi que la présentation du *Dictionnaire de valence Proton*¹.

2.3.1. Phrase et Construction verbale

Nous appelons **construction verbale** une séquence composée d'un verbe et des éléments qu'il construit. Nous verrons dans la partie suivante que les éléments construits (ou régis) par le verbe peuvent être subdivisés en deux catégories, celle de la valence, et celle de la rection.

Nous ne nous servons donc pas dans ce travail de la notion de **phrase**, car sa définition la plus répandue est incompatible avec la description de production orale, comme le soulignent M. Charolles, P. Le Goffic et M.-A. Morel dans la présentation qu'ils font des communications présentées au colloque « **Y a-t-il une syntaxe au-delà de la phrase ?** », tenu à Paris III en septembre 2000 :

La phrase a toujours posé de nombreux problèmes théoriques et pratiques : il n'est que de vouloir délimiter et compter les phrases dans un texte écrit pour que la définition de la phrase s'obscurcisse et que sa réalité devienne problématique, au contact de la coordination, de l'ellipse, de la parataxe, de l'anaphore, ... Et que dire de l'oral, pour lequel la notion de phrase semble inopérante ? (*in Verbum* 2002 : p. 3)

En effet, comme le souligne C. Blanche-Benveniste (*op. cit.* : p. 7)

nous avons intériorisé la notion de phrase apprise à l'école, avec ses manifestations visibles de délimitations graphiques, au point qu'il semble que ce soit une notion naturelle, donnée avec la capacité naturelle du langage. L'assimilation entre le graphique et le linguistique est si forte que la méconnaissance de l'objet phrase, par exemple chez les jeunes enfants en cours d'apprentissage, peut passer auprès des instituteurs pour un défaut dans la capacité même de

Les exigences minutieuses et les prescriptions vexatoires abondent dans la religion schismatique grecque.

langage. (...) La phrase fait partie de ces objets que nous manipulons lorsque nous écrivons. A partir de son existence comme unité graphique, elle fonctionne comme une unité de représentation commune de la langue, et la grammaire scolaire l'a installée dans ce rôle. Mais cela n'implique pas que ce soit une unité efficace pour la description grammaticale. On a certainement intérêt à distinguer, comme l'ont judicieusement proposé Berrendonner et Reichler-Béguelin (1989), l'unité du « savoir pratique » qu'est la *phrase* pour les usagers de la langue, et les unités d'un savoir plus sophistiqué, élaborées pour des tâches d'analyses et de calcul.

La grammaire dite « scolaire » (aussi appelée « traditionnelle ») utilise la phrase comme point de départ de l'étude de la langue. Grammaire « scolaire », donc à vocation pédagogique, et dont le but est la maîtrise de la langue écrite par un public, qu'il soit composé d'enfants apprenant la grammaire de leur langue maternelle (ou de leur langue d'adoption), ou d'adultes apprenant la grammaire d'une langue seconde. La phrase est présentée tout d'abord aux élèves du cycle primaire comme commençant par une majuscule et se terminant par un point. En ce sens, elle est définie effectivement comme étant avant tout un objet graphique. On favorise ainsi la réalisation écrite de la langue au détriment de sa réalisation orale, alors que l'on sait que l'oral est la première face de la langue avec laquelle les enfants sont en contact.

D'autre part, C. Blanche-Benveniste (*ibid.* : pp. 9-10) démontre que les propriétés que la grammaire scolaire² attribue à la phrase sont en fait contenues dans le verbe. Il en est ainsi des diathèses, des modalités et de la rection.

Concernant les diathèses, elle compare deux exemples : le premier, comportant un verbe conjugué à la voix passive et uniquement des compléments régis (*Le facteur a été interrogé par Marie*, exemple cité par le *Bon Usage*), le deuxième comportant lui aussi un verbe à la voix passive, et « un peu plus que le strict domaine verbal », c'est-à-dire des compléments de lieu et de temps non régis par le verbe et une consécutive (*Là-bas, tous les ans, au printemps, les façades des maisons sont blanchies à la chaux, de sorte que cela éloigne quantité de parasites*). Elle montre alors que « le passif ne concerne pas du tout les éléments *là-bas, tous les ans, au printemps* [...] pas plus que *de sorte que cela éloigne quantité de parasites* », et conclut que « le passif est une propriété de la rection verbale,

¹ EYNDE, K. VAN DEN, MELIS, L., MERTENS, P., VANGILBERGEN, L., DEHASPE, L., EGGERMONT, C., BROEDERS, E., *Dictionnaire de valence Proton*, consultable sur Internet à l'adresse <http://bach.arts.kuleuven.ac.be/PA/protonfr.html>

² Les exemples qu'elle utilise sont extraits du *Bon Usage*.

permettant de reformuler les sujets et les compléments selon des règles déterminées, mais [que] ce n'est absolument pas une propriété de l'ensemble de la phrase ». Sa démonstration est en partie identique pour ce qui est des modalités et de la rection verbale.

En étudiant le français parlé, on observe donc immédiatement les limites de la présentation scolaire de la phrase.

Considérer le verbe comme un noyau, autour duquel gravitent des éléments que le verbe construit, envisager sa valence, c'est-à-dire se servir du modèle chimique pour analyser la production verbale est un choix intéressant, d'autant plus qu'il a aujourd'hui fait ses preuves. La théorie élaborée par L. Tesnière, même si elle est souvent critiquée, n'étant pas d'un abord facile, a toutefois permis des avancées significatives en linguistique descriptive.

Après avoir brièvement défini ce que nous nommons construction verbale, nous présenterons donc les notions de rection et de valence, puis nous montrerons que cette notion de construction verbale peut s'envisager à deux niveaux distincts, selon le degré de précision à partir duquel s'effectue l'analyse. Ensuite, nous expliquerons la méthode que nous utilisons pour déterminer quels éléments sont construits (régis) ou non par le verbe. Nous utiliserons pour cela l'Approche Pronominale. Enfin, nous montrerons que l'analyse n'est pas toujours évidente, et que des problèmes de pronominalisation peuvent quelquefois la « bloquer ».

2.3.1.1. Valence et rection

Première approche

De manière générale, une différenciation est faite entre éléments construits par le verbe et entrant dans la valence verbale (éléments qui sont indispensables **à la fois** au fonctionnement syntaxique et sémantique du verbe, nommés par quelques grammaires *éléments essentiels*¹, ou dans la terminologie héritée de la logique *arguments*, ou encore chez Tesnière, *actants*) et éléments construits par le verbe n'y entrant pas (nommés par les mêmes grammaires *éléments accessoires* ou *facultatifs*, ou *ajouts*, et par Tesnière, *circonstants*).

O. Bonami montre dans sa thèse (1999 : pp. 16-79) les limites des notions « syntaxiquement obligatoire » et « sémantiquement obligatoire », et pose ainsi que la

distinction « arguments/ajouts » (valence/rection) n'est pas correcte lorsqu'elle est posée avec les propriétés généralement admises comme étant caractéristiques des arguments. Ces propriétés sont les suivantes :

a. Un argument est syntaxiquement obligatoire ;

b. Un argument est sémantiquement obligatoire ;

c. Un argument est non itérable.

(a) ne tient pas, car, comme le rappelle O. Bonami (*op. cit.* : p. 21), « les verbes admettant la suppression de leur objet direct sont légion », et il donne l'exemple suivant, indiquant toutefois que les effets sémantiques dus à la suppression sont variés :

a. Jean a mangé un gâteau.

b. Jean a mangé.

(b) ne tient pas non plus, car, selon O. Bonami, le caractère sémantiquement obligatoire n'est pas caractéristique des arguments (*op. cit.* : pp. 31-32) : « le fait d'être sémantiquement obligatoire ou facultatif ne semble donner aucune indication quant au statut d'argument ou d'ajout d'un dépendant : si tous les arguments directs sont sémantiquement obligatoires, il existe de nombreux ajouts qui ne sont pas facultatifs, et certains GP [groupes prépositionnels] argumentaux ne sont pas obligatoires ».

(c) ne tient pas, car (*op. cit.* : p. 42) « la propriété d'itérabilité n'est pas non plus opératoire : certains arguments sont itérables, et, pour autant que la notion d'itérabilité puisse être définie de manière précise pour les ajouts, il existe de nombreux ajouts qui ne sont pas itérables ».

Nous ne présenterons pas ici en détail les propositions d'O. Bonami, bien qu'elles soient particulièrement fines et justes. On pourra pour plus de précisions se rapporter à sa thèse, disponible en ligne² (le résumé des propositions est donné à la page 301).

¹ CHERVEL, A., 1977, *Et il fallu apprendre à écrire à tous les petits Français*, Paris, Payot, p. 179, cité par C. Blanche-Benveniste (*ibid.*)

² [\[Fichier PDF \(1,5M\)\]](http://www.llf.cnrs.fr/Bonami/7-Bonami-these99.pdf) <http://www.llf.cnrs.fr/Bonami/7-Bonami-these99.pdf>

Pour la suite, il convient de noter que ce qui est nommé ci-dessus *argument* et *ajout* est nommé respectivement ci-dessous *valence* et *rection* (sens restreint).

Les problèmes de distinction entre valence et rection

La somme des éléments construits (ou régis) valenciels (compléments de valence) et non-valenciels (compléments de rection) est appelée rection verbale. Le terme de rection, comme le fait remarquer Claire Blanche-Benveniste (1981 : pp. 62-63), possède ici une valeur ambiguë qu'il convient cependant de conserver : il signifie « tantôt ce qui englobe valence et rection (au sens large), tantôt ce qui n'est pas la valence, tout en étant régi par le verbe (au sens restreint) : c'est parce que la distinction entre les deux est délicate et qu'il est commode d'avoir un terme couvrant, qui neutralise la distinction. ». Voici une représentation schématisée représentant la valence (en gras), la rection (sens restreint) et la rection (sens large) :

Rection verbale	
Elément(s) de valence	Elément(s) de rection

Figure 1 : schéma de représentation de l'ambiguïté Valence-Rection

L'élément apparaissant en gras dans la deuxième ligne de ce tableau est l'élément dit « indispensable » au bon fonctionnement tant syntaxique que sémantique de la construction verbale. Ce sont les éléments de valence.

Pour illustrer ce schéma, prenons l'énoncé suivant :

Ex1 : Quand je suis malade, je ne parle pas de linguistique.

Pour ce type de construction possible du verbe *parler* ([*X1 parle de X2*], à l'intérieur de laquelle *X1* est un sujet de type [+personnel] et *de X2* un complément de type [+/- personnel]), *Quand je suis malade* sera généralement considéré comme un élément de rection, et *je* et *de linguistique* comme des éléments de valence.

Nous avons retenu dans ce travail le sens large de rection verbale, sans distinguer à l'aide de tests éléments construits appartenant à la valence et éléments construits n'y appartenant pas, pour une raison déjà évoquée par Claire Blanche-Benveniste (*ibid.* : p. 95) :

On ne peut venir à bout, par des procédés systématiques, de la différence entre valence et rection. [...] Le français marque par des procédés morphologiques une partie de la valence des verbes, celle qui se manifeste par des formes *le, lui*, mais pas ailleurs. Cela laisse une grande zone de flou pour tous les autres cas, où l'on peut constamment être victime de "faux-sens", à l'intérieur de sa propre langue.

Plutôt que de trancher, et de prendre des positions théoriques difficilement défendables, nous avons choisi de présenter l'inventaire des constructions verbales de notre corpus en y intégrant tous les éléments construits par le verbe, qu'ils appartiennent ou non à la valence. Ce sont donc tous les éléments de la *rection* au sens large qui seront étudiés ici.

Dans le cadre de notre étude, les constructions verbales sont étudiées en fonction du schéma représentant leur structure pronominalisée, en quelque sorte le « squelette » de la construction verbale lexicale, que nous avons obtenu à partir des exemples de notre corpus.

Par exemple :

CHOCM~10 2,12 *j'ai appris ce que je devais apprendre*

Pronominalisé en : je l'ai appris (cela).

Ici, nous avons deux éléments construits (qui entrent dans la rection verbale, au sens large) : le *je* en P0 (sujet) et le *ce que je devais apprendre* en P1 (Paradigme des clitiques compléments en *le, la, les*). Nous présenterons les tests utilisés pour repérer les éléments construits (ou régis) par le verbe à la partie 2.1.3. (Eléments construits).

L'intérêt de la description pour la distinction valence/rection

Nous avons vu plus haut les limites des tests permettant de distinguer valence et rection (arguments et ajouts), tels que Bonami (1999) les montre.

En étudiant notre corpus, c'est-à-dire en classant les différentes constructions verbales qui en sont relevées, nous ne voulions pas faire la distinction entre valence et rection pour la raison évoquée précédemment. Cependant, il nous est apparu que la distinction s'opérait bien souvent d'elle même par le classement, sans avoir recours aux tests syntaxiques. En effet, pour un verbe donné, s'il est productif (on pose ici qu'un verbe « productif » possède au moins une cinquantaine d'occurrences dans notre corpus), la structure « basique » d'un même type de construction, faisant apparaître les éléments de valence (les arguments), apparaît de

manière constante, alors que les éléments de rection (les ajouts) n'apparaissent pas systématiquement.

Prenons comme exemple le cas du verbe *acheter*. On distingue dans le corpus trois types de constructions verbales différents : le premier est le plus fréquent, avec 44 occurrences. C'est le type *je l'achète* (ou *j'en achète*), qui marque le sujet et l'objet. Le second (17 occurrences) ne marque pas l'objet : *j'achète*. Enfin le troisième (2 occurrences) marque le sujet, l'objet et le bénéficiaire (ou le vendeur) : *je te l'achète* (ou *je t'en achète*). Nous classons les différents types de constructions par rapport à la distribution syntaxique effectivement réalisée du verbe étudié. Par conséquent, l'hypothèse d'un argument sémantiquement obligatoire qui n'apparaîtrait pas dans l'énoncé, mais qui est considéré comme nécessairement présent (dont la place sémantique argumentale est nécessairement saturée, qui fait nécessairement l'objet d'une quantification existentielle), n'est pas ici utilisée dans le marquage des éléments entrant dans la valence. Pourtant, si l'on considère l'exemple suivant :

J'achète

Force est de constater qu'il implique nécessairement un objet :

J'achète => j'achète quelque chose => je l'achète.

C'est ce que Claire Blanche-Benveniste (1981 : pp. 86-89) nomme « réalisation zéro » :

Un élément de valence peut, dans certains cas, avoir une forme de réalisation zéro ; c'est l'exemple classique du verbe *manger*, que l'on peut avoir avec ou sans réalisation de valence

Il ne mange pas - il ne mange pas ça

On a le droit de considérer que dans « il mange » on a une réalisation zéro de la valence et que ce n'est pas la même chose que l'absence de valence ; en effet, à partir de « il mange », on peut enchaîner par coordination :

il mange, mais pas n'importe quoi [...]

alors que ce ne serait pas possible pour un verbe sans valence de ce type :

*elle existe, mais pas n'importe quoi.

Un des problèmes posés par cette analyse est qu'elle permet d'inclure aussi des éléments de rection (au sens restreint). En effet, dans l'exemple suivant, où nous reprenons le verbe *acheter* :

J'achète de la viande, mais pas en grande surface

on peut enchaîner par coordination un locatif (*en grande surface*), qui ne fait pas partie de la valence du verbe. Aussi, ce test de coordination pour repérer éléments de valence à réalisation zéro ne vaut que pour les verbes dont nous savons intuitivement que l'objet (où qu'un autre élément de valence) est latent.

Néanmoins, concernant notre exemple précédent, on peut admettre facilement que le verbe *acheter* possède dans sa valence nécessairement un objet (j'achète Ø, mais pas n'importe quoi). Si l'on pousse la démonstration à l'étape suivante, on peut, de la même manière, considérer que *acheter* implique nécessairement un bénéficiaire (qui peut être co-référent au sujet ou non)¹ :

J'achète ce pull à Marie => je le lui achète

J'achète ce pull (où le bénéficiaire, s'il n'est pas précisé, est compris comme étant indéterminé) => je l' Ø achète, mais pas à n'importe qui (où Ø marque la réalisation zéro)

Aussi, ce verbe ne posséderait qu'une catégorie de construction (dans notre corpus), qui pourrait être représentée comme suit :

Je [te] [l'] achète,

où les éléments entre crochets seraient « sémantiquement obligatoires », mais pourraient ne pas être réalisés sur le plan syntaxique. On voit ici la limite de la définition des éléments de valence considérés comme indispensables au fonctionnement syntaxique du verbe.

D'autre part, on sait que la trivalence est une caractéristique des verbes de don² (classe à laquelle appartient le verbe *acheter*, selon Tesnière (1959 : p. 256)). De la même manière que l'on donne toujours quelque chose à quelqu'un, on achète toujours quelque

¹ on notera ici que dans la structure *je te l'achète*, l'élément *te* peut être interprété comme étant soit le bénéficiaire, soit le vendeur de l'objet, ce qui ne modifie en rien le « test syntaxique » suivant.

² Willems (1981 : p. 106) préfère la notion de « verbes d'attribution ».

chose pour quelqu'un. Cette typologie est lexicale, elle distingue des ensembles lexicaux de verbes. Celle qui nous intéresse dans ce travail est syntaxique, même si nous sommes de toute évidence en accord avec la thèse qui pose que syntaxe, lexique et sémantique sont liés.

Ceci étant, il nous reste à signaler que les éléments de rection (au sens restreint) n'apparaissent pas de manière constante. Ainsi, pour le verbe *acheter*, les pro-formes indiquées entre crochets dans la partie « corpus des constructions verbales » font partie de la rection au sens restreint, ils sont ce que Bonami nomme des *ajouts*. On remarquera que dans tous les cas, l'élément « Ø » apparaît entre crochets, signalant ainsi que la construction apparaît dans le corpus sans élément de rection (sans ajout) – il faut noter que l'élément « Ø » n'a pas ici la même valeur que lorsque nous le présentons plus haut pour expliquer la valence à réalisation zéro. Ainsi, pour la construction du verbe *acheter* la plus fréquemment relevée dans notre corpus (*je l'achète*), on observe qu'elle peut être accompagnée des éléments de rection suivants : [Ø ; ainsi ; quand ; où].

Ainsi, le choix de ne pas distinguer valence et rection au cours de l'étude n'a pas été abandonné, mais, au contraire, s'est avéré pertinent. En somme, lorsque l'on décrit un exemple de constructions verbales attestées, on ne peut, dans un premier temps, venir à bout de cette distinction. On recense alors à la fois les éléments de valence et ceux de rection. A la fin de ce inventaire apparaissent des éléments « détachés », qui ne figurent pas dans toutes les constructions verbales d'un même type, et qui, par conséquent, appartiennent à la rection (au sens restreint d'*ajout*).

2.3.1.2. Extension de la notion de construction verbale

Il y a deux manières de considérer la construction verbale suivante :

Ex : je pense que ça ça arrivera jamais (MADMO~34 12).

1- Soit on l'analyse comme un tout, en pronominalisant *que ça ça arrivera jamais* en *le*, ce qui donne : *je le pense*, tout en précisant la nature (que Phrase) que revêt ce syntagme avant sa pronominalisation. C'est l'analyse la moins fine et la moins précise, qui aurait cependant l'avantage de réduire considérablement le corpus des constructions verbales à traiter dans le cadre de ce travail.

Ainsi, cette analyse ne prend en compte que les constructions des verbes apparaissant

dans les constructions rectrices, et met de côté les constructions verbales des constructions régies.

Nous avons considéré que notre travail perdrait beaucoup de sa richesse si nous retenions cette analyse.

2- En effet, les constructions apparaissant dans les constructions verbales régies (par un autre verbe) peuvent toutes être réduites à des constructions simples (la « que Phrase » de notre exemple peut ainsi être réduite à *ça arrivera jamais*, et étudiée comme telle).

La deuxième analyse possible est donc celle qui consiste à détailler la construction de départ et à la considérer comme un tout composé de deux ensembles, et donc de deux constructions à analyser : la première est la construction complète *je le pense*, la deuxième est le résultat de la transformation de la que-Phrase en construction verbale « indépendante », *ça arrivera jamais*. C'est cette analyse, plus précise et plus fine, que nous avons retenue, et que nous présenterons de manière plus détaillée dans la partie 2.5.2. ([Constructions verbales traitées](#)).

2.3.2. Pronoms et pro-formes

Le terme de « pronom », généralement utilisé par les grammairiens et les linguistes, pose problème dans certains cas. En effet, la définition étymologique qu'on lui attribue traditionnellement (du latin *pro-nomen*, « remplaçant un nom ») se révèle inopérante pour la catégorie des « pronoms personnels sujets ». Nous avons alors décidé de faire un tour d'horizon bibliographique, afin d'observer les différentes propositions émises par les linguistes depuis le 18^{ème} siècle. Cette synthèse sera suivie d'une présentation des différentes pro-formes utilisées pour appauvrir le lexique des constructions verbales relevées dans notre corpus. Un inventaire de ces pro-formes a d'ores et déjà été réalisé par K. van den Eynde et al. pour le lexique de valence Proton. Nous le présenterons donc, tout en précisant les pro-formes que nous avons retenues dans notre classement, car toutes n'ont pas été utilisées.

2.3.2.1. La notion de pronom

Le problème lié à la notion de pronom est très ancien. Il ne date pas de la naissance de la linguistique comme discipline ; on peut le lier à un problème métaphysique lié aux rapports qu'entretient la langue avec le monde. Les mots, classés par les grammairiens dès

l'Antiquité par rapport à leur fonction, puis à leur sens, se répartissent selon les écoles en un nombre limité de classes. Le pronom n'échappe pas à cette classification. Mais pour lui, un problème se pose : celui de sa définition. Etymologiquement, *pro-nomen* remplace un nom (ou un syntagme nominal). Or, tous les mots qui sont étiquetés comme pronoms ne sont pas substituables à un nom. C'est le problème que posent les deux premières personnes des « pronoms personnels sujets », *je* et *tu*. En effet, comment peut-on imaginer remplacer par un nom le *je* dans un énoncé comme : *je te parle de mes soucis* ? On peut à la rigueur y ajouter un pronom disjoint du type *moi* : *moi, je te parle de mes soucis*, encore que, d'un point de vue sémantique, le sens de l'énoncé de départ est modifié. De la même manière, toujours dans la catégorie des « pronoms personnels sujets », que dire de la notion de « personne » pour le pronom de la troisième ? Le *il* dans *il pleut*, ou encore dans [*le livre dont je t'ai parlé*] *il est sorti en librairie* ne réfèrent en aucun cas à une personne. La qualification de « personnel » doit alors être remise en cause pour une partie des pronoms de type *il(s)* ; il faudrait idéalement distinguer à l'intérieur de cette catégorie les pronoms faisant référence effectivement à une personne, et les autres. En outre, notons ici que le terme de *pronom* ne pose pas de problème pour ceux de la troisième personne des sujets. Ainsi, cette sous-catégorie de pronoms, les « personnels sujets », pose deux problèmes, répartis sur deux types de pronoms différents :

1. Le premier réside dans cette notion de *pronom* qui, si l'on s'en tient à la définition étymologique, n'est pas opératoire pour les deux premières personnes du singulier (*je* et *tu*). Les deux premières personnes du pluriel ne posent pas ce problème, dans la mesure où *nous* est toujours composé de *je* + *l(es) autre(s)* et où *vous* est toujours composé de *tu* + *l(es) autre(s)*. En aucun cas, *je* et *tu* n'ont de référence nominale dans le discours.
2. Le deuxième réside dans la notion de *personne*, et concerne une sous-partie des pronoms de la troisième, c'est-à-dire *il*, *ils*, *elle*, *elles*, lorsqu'ils réfèrent à un objet, ou à une situation. En effet, comment parler de pronom « personnel » alors que la notion de personne est absolument absente des référents du discours ?

Un autre problème existe à l'intérieur de cette catégorie des « pronoms personnels sujets », celui de la dénomination du *il* entrant dans une construction dite « impersonnelle ».

En effet, comme l'a remarqué Denis Creissels (1995 : p. 121), la construction dite « impersonnelle » est mal nommée : « La notion concernée par les constructions dites « impersonnelles » est la notion de sujet : une fois cette notion définie de façon rigoureuse, on peut voir que le propre des constructions « impersonnelles » est d'être structurellement dépourvues de constituant nominal identifiable comme sujet. »

Depuis la parution, en 1926, de *La pensée et la langue* par Ferdinand Brunot, le carcan catégorisationnel des parties de discours n'a, semble-t-il, guère évolué. Les mises en garde de Brunot semblaient pourtant pertinentes (1926 : p. XIII) :

Quand on aura fait les révisions nécessaires, que la grammaire dogmatique, purgée de ses fautes, corrigée de ses excès, aura été ramenée à son caractère véritable et réduite à son rôle, qu'elle sera le témoin fidèle de l'usage, la conseillère modeste et judicieuse de ceux qui entendent parler et écrire exactement, toujours éprise de l'ordre et de la règle, mais accueillante aux créations justifiées et aux changements nécessaires, la réforme sera-t-elle faite ? Je ne le crois pas.

L'étude des faits du langage repose, depuis l'Antiquité, sur une classification générale de tous les éléments linguistiques qui peuvent composer une phrase, et qui forment les « parties du discours ». On n'a jamais pu se mettre d'accord sur le nombre de ces « parties », ce qui prouve déjà que le principe adopté n'est pas d'une grande solidité, ni d'une clarté indiscutable. Mais passons. Ces parties du discours, les unes variables, les autres invariables, rangées dans un ordre à peu près immuable, sont étudiées successivement dans leurs formes et dans leur rôle, c'est-à-dire qu'on passe en revue les accidents qui leur surviennent, variations en nombre, en genre, en cas, en personne, etc., et ensuite, qu'on examine leur fonction dans les combinaisons qui forment les phrases. Quelques auteurs ont réuni ces deux parties : morphologie et syntaxe. Le plus souvent, elles sont traitées séparément. Tout le monde a présente à l'esprit la disposition intérieure des chapitres ordinaires : de l'Article, du Nom, etc. Inutile de le rappeler en détail. [...]

Et cependant le nécessaire doit être introduit ; la langue doit entrer dans les grammaires. Par quel moyen ? A mon sens il n'y en a qu'un, mais il suffit. **Entre les formes les plus diverses de l'expression, entre les signes les plus disparates, il y a un lien, c'est l'idée commune que les signes contribuent à exprimer.** Si on la prend pour centre, il ne s'agit plus de choisir entre des rattachements abusifs ou des omissions forcées, tout s'ordonne autour d'elle ; elle groupe les éléments linguistiques venus de toutes parts, et dont d'autres chapitres se trouvent allégés. Tout se complète, s'organise, se classe. De la sorte, **quelques hommes** cesse d'être aux indéfinis, pendant que **des hommes** est à l'article, **une poignée d'hommes** au nom, **vingt hommes** aux noms de nombre ; les expressions de quantités précises ou imprécises se cataloguent dans le langage, comme le font ailleurs les nombres et les mesures.[...]

Il faut se résoudre à dresser des **méthodes de langage, où les faits ne soient plus rangés d'après l'ordre des signes, mais d'après l'ordre des idées**. Ce sont elles qui doivent être classées, non point sans doute en elles-mêmes et pour elles-mêmes, comme elles le seraient pour la psychologie pure, mais en vue de leurs signes et relativement à eux. La scolastique, ici encore, doit mourir.

Mais revenons à présent à cette notion problématique de *pronom*. Un tour d'horizon des points de vue de quelques grammairiens et linguistes du français nous permettra sans doute de mieux comprendre les enjeux et les problèmes que posent la terminologie grammaticale et une improbable réforme terminologique.

Beauzée (1767 : p. 258) indiquait que :

Depuis le temps qu'on parle du Pronom, on n'est point parvenu à le bien connoître ; comme si la nature étoit, dit le P. Buffier, un de ces secrets impénétrables qu'il n'est jamais permis d'approfondir. Pour faire sentir, continue-t-il, que je n'exagère rien, il ne faut que lire le savant Vossius, la lumière de son temps et le héros des grammairiens. Après avoir déclaré (et avec raison) que toutes les définitions, qui avoient été données du Pronom jusqu'alors, n'étoient nullement justes, il prononce que le Pronom est un mot qui en premier lieu se rapporte au nom, et qui en second lieu signifie quelque chose. Pour moi, avec le respect qui est dû au mérite d'un si grand homme, j'avoue que ne comprend rien à la définition du Pronom.

Il semblerait que la définition de cette notion n'ait jamais été mieux éclaircie d'un point de vue syntaxique depuis lors. S'agirait-il alors, comme le dit le P. Buffier, d'un secret impénétrable, ou d'un simple problème d'étiquette terminologique employée à mauvais escient ?

G. Moignet (1965 : p. 9) ne dit rien d'autre :

Le sentiment commun et la grammaire traditionnelle reconnaissent l'existence d'une catégorie grammaticale du pronom, mais, d'une part, il y mettent tant de choses différentes, et, d'autre part, en excluent tant d'autres manifestement liées aux premières (nous pensons notamment à l'article, aux adjectifs démonstratifs et possessifs, etc...) qu'ils ont fait naître dans la pensée de certains que la catégorie *pronom* était artificielle, que le pronom, en tant que tel, ne correspondait à aucune réalité linguistique : quel dénominateur commun peut convenir à *je, celui-ci, leur, quoi, tout, dont, on*, etc... ? **Le pronom serait une catégorie fourre-tout où voisinent des termes très précis et des indéfinis, des mots pleins et des mots vides, des mots autonomes et des mots spécialisés en cas**, etc... F. Brunot (*La pensée et la langue*, p. 173) récuse le mot *pronom* et la définition banale : « mot qui tient la place du nom », parce que, dit-il, souvent des

pronoms ne remplacent pas des noms, mais des adjectifs, des verbes, des phrases (faudra-t-il parler de proadjectifs, de proverbes, de prophrases ?). Il veut substituer à *pronom* le terme *représentant*. Mais il doit bien reconnaître que *je*, *tu* ne peuvent être considérés comme représentant quoi que ce soit, ni *il* dans *il pleut*, *il importe*...

Faut-il donc maintenir la catégorie pronominale, ou la dépecer ?

Il faut y regarder à deux fois avant de récuser un sentiment linguistique bien établi. Dépecer la catégorie peut être une solution de facilité qui dispense d'un effort de définition assez attentif.

Dans un mémoire déjà ancien, Gustave Guillaume définissait le pronom : « la catégorie substituée à celle du nom¹ partout où cette dernière se montre inapte (ou disconvenante), pour quelque raison que ce soit, à satisfaire aux conditions momentanées d'application que le discours lui propose. »

Eluerd (2002 : p. 74) introduit son chapitre sur les pronoms de la même manière :

Dans sa *Grammaire* de 1932, l'Académie Française notait que : « certains pronoms peuvent désigner une personne ou une chose qui n'ont pas été ou ne seront pas exprimées par un nom » (p. 41). Habituellement caustique, le linguiste Ferdinand Brunot note la justesse du propos académique et apprécie la remise en cause de la définition traditionnelle du pronom. Mais il invite à aller plus loin et « à déclarer franchement que dans *je parle*, *je* est simplement une forme de conjugaison destinée à marquer la première personne » (*Observation sur la grammaire de l'Académie française*, Droz, 1932 : p. 43).

L'un des acquis essentiel de la linguistique, l'analyse de l'**énonciation**, prend sa source dans ce genre de remarque. De l'analyse critique de la notion de pronom personnel, E. Benveniste (1966 : p. 232) conclut 1° que *je* et *tu* ne sont pas des pronoms mais des nominaux puisqu'ils désignent directement leur référent dans la situation d'énonciation, 2° que *il* n'est pas vraiment personnel puisqu'il désigne une « non personne », le locuteur absent. *Je* et *tu* sont alors tenus pour caractéristiques de l'énonciation discours, *il* pour caractéristique de l'énonciation récit.

La catégorie des pronoms a donc été largement revue par la linguistique, pour aboutir à des descriptions plus rigoureuses des modes de référence et pour écarter la définition traditionnelle et étymologique (*pro-nomen*) de « mot employé à la place du nom ».

¹ Il faut noter que ceci n'est pas exact : le pronom peut être substitué à autre chose qu'un nom : il peut remplacer par exemple un adjectif comme dans *belle*, *elle l'est*.

Si F. Brunot (1932 : p. 43), comme le souligne Eluerd, soutient l'Académie Française dans sa remise en cause de la notion de pronom, et la pousse jusqu'à dire que *je* n'est qu'une forme de conjugaison pour marquer la première personne, il propose de substituer à la notion de pronom celle de *représentant* (1926 : p. 175). Mais ce terme ne convient pas aux deux premières personnes du singulier, *je* et *tu* : « Au sujet, *je*, *tu*, sujets inséparables du verbe, vraies formes de conjugaison, ne peuvent être considérés comme représentant quoi que ce soit ; *me*, *te*, *moi*, *toi* non plus. Ce sont des nominaux. »

L'avancée permise par la linguistique de l'énonciation est, pour l'analyse et la définition des « pronoms personnels », effectivement importante. Synthétisant les propositions de Brunot et de Benveniste, tout en finalisant la proposition de Moignet¹ (1965 : p. 22), Denis Creissels propose (1995 : pp. 121-122) de « substituer respectivement « élocutif », « allocutif » et « délocutif » aux termes traditionnels de première, deuxième et troisième personne. Pour rendre compte du fait que souvent les formes désignées comme « pronoms de troisième personne » ont un comportement particulier, on peut retenir le terme d'« interlocutif » comme englobant à la fois élocutif et allocutif ».

Cette proposition semble reprendre de manière pertinente et efficace l'ébauche de proposition de Benveniste (1966 : pp. 252-253) : « Quelle est donc la « réalité » à laquelle se réfère *je* ou *tu* ? Uniquement une « réalité de discours », qui est chose très singulière. *Je* ne peut être défini qu'en terme de « locution », non en termes d'objets, comme l'est un signe nominal. [...] la forme *je* n'a d'existence linguistique que dans l'acte de parole qui la profère. [...] La définition peut alors être précisée ainsi : *je* est l'« individu qui énonce la présente instance de discours contenant l'instance linguistique *je* ». Par conséquent, en introduisant la situation d'« allocution », on obtient une définition symétrique pour *tu*, comme l'« individu allocuté dans la présente instance de discours contenant l'instance linguistique *tu* ».

Concernant le *il*, Benveniste (1966 : p. 230) le définit en opposition à *je* et *tu* : « En effet, une caractéristique des personnes « je » et « tu » est leur *unicité* spécifique : le « je » qui énonce, le « tu » auquel « je » s'adresse sont chaque fois uniques. Mais « il » peut être une infinité de sujets – ou aucun. [...] Une seconde caractéristique est que « je » et « tu » sont

¹ « Le système de l'interlocution comporte, en français, la distinction, 1. de trois personnes simples : celle du locuteur, parlant de lui-même : première personne ; celle de l'allocuteur, dont on parle, parlant à lui même :

inversibles : celui que « je » définis par « tu » se pense et peut s'inverser en « je », et « je » (moi) devient un « tu ». Aucune relation pareille n'est possible entre l'une de ces deux personnes et « il », puisque « il » en soi ne désigne spécifiquement rien ni personne. [...] Il n'y a pas aphérèse de la personne, mais exactement la non-personne, possédant comme marque l'absence de ce qui qualifie spécifiquement le « je » et le « tu ». Parce qu'elle n'implique aucune personne, elle peut prendre n'importe quel sujet ou n'en comporter aucun, et ce sujet, exprimé ou non, n'est jamais posé comme « personne ». [...] On voit maintenant en quoi consiste l'opposition entre les deux premières personnes du verbe et la troisième. Elles s'opposent comme les membres d'une corrélation, qui est la *corrélation de personnalité* : « je-tu » possède la marque de personne ; « il » en est privé. La « 3^{ème} personne » a pour caractéristique et pour fonction constante de représenter, sous le rapport de la forme même, un invariant non-personnel, et rien que cela. »

Trente ans plus tard, Creissel (1995 : p. 121) synthétise les propositions de Benveniste et de Brunot, et remet en cause la notion de construction dite « impersonnelle », arguant du fait que la notion de sujet, qui est à la base de cette construction « impersonnelle », en est en réalité absente. En effet, « le propre des constructions « impersonnelles » est d'être structurellement dépourvues de constituant nominal identifiable comme sujet. Il serait donc préférable de les désigner comme « constructions asubjectales » ».

Sur ce point, Beauzée indiquait déjà (1767 : p. 261) : « A l'occasion des *Principes généraux et particuliers de la langue françoise* par M. de Wailly, on trouve dans l'*Année littéraire*, une difficulté contre l'opinion commune, dont l'auteur reconnoît devoir le germe à M. l'abbé de Condillac, et dont il auroit également pu avoir obligation au P. Buffier ou à Sanctius, puisque la nature des Pronoms y est envisagée à peu près sous le même point de vûe. « Il y a, dit le journaliste, trois sortes de Pronoms personnels, *je, me, moi, nous, tu, te, toi, vous*, pour la première et la seconde personne. C'est le cri général de toutes les grammaires... Tous ces mots font les noms de la première et de la deuxième personne, tant au pluriel qu'au singulier, et ne font point des Pronoms. Tout mot quelconque, excepté ceux-ci, appartiennent à la troisième personne ; ce qu'on démontre en ajoutant à un mot quelconque un verbe, qui aura la terminaison de la troisième personne, *Antoine revient, le*

deuxième personne ; celle du délocuteur, ou plutôt du délocuté, dont on parle, sans plus ; 2. de deux personnes doubles [...] »

marbre est dur, le froid se fait sentir, etc. Les mots *je, me, moi*, etc. considérés comme Pronoms, représenteroient donc des noms, et conséquemment des noms de la troisième personne, puisqu'il est certain que la troisième personne s'empare de tout. Or ces mots *je, me, moi*, etc. représentant des noms de la troisième personne, comment feroient-ils des Pronoms de la première personne et de la seconde ? Ces mots sont donc les véritables noms et non les Pronoms de la première et de la seconde personne. »

Dans le chapitre qu'il consacre aux pronoms, Roland Eluerd (2002 : pp. 74-75) en distingue 7 catégories : pronoms personnels, possessifs, démonstratifs, numéraux, interrogatifs (et exclamatifs), indéfinis et relatifs. « Les unités grammaticales qu'on appelle **pronoms** ne sont pas réunies par une structure logique rigoureuse mais par un air de famille.

1. Les pronoms présentent des caractères morphologiques et syntaxiques qui leur permettent :
 - d'être **substituts** d'autres catégories grammaticales ;
 - de remplir les **fonctions du GN** ;
 - et de désigner directement ou indirectement un **réfèrent immédiat**.

Chaque pronom ne possède pas nécessairement tous ces caractères à l'identique. Ils sont répartis diversement selon les types de pronoms et selon les pronoms eux-mêmes.

2. La classe fermée des pronoms présente une morphologie qui est parfois une **quasi-déclinaison**. Selon les types de pronoms, plusieurs marques apparaissent. Ces marques relient le pronom à son antécédent et participent à l'identification de son référent. [...]
3. La syntaxe des pronoms montre qu'ils assument plusieurs rôles.
 - être substitut d'un GN ;
 - Remplir les fonctions du GN ;
 - Introduire une subordonnée relative ;
 - Marquer la phrase interrogative et exclamative ;

- Etre complétés par un adjectif apposé. »

Eluerd ne mentionne pas, hormis brièvement dans l'introduction de ce chapitre, le problème terminologique lié à la notion de pronom dans le cas des deux premières personnes du singulier (*je* et *tu*). Mais son ouvrage n'a pas pour objectif une remise en cause de la terminologie problématique de la linguistique ; au contraire, il s'adresse à un public d'étudiants de 1^{er} et de 2^{ème} cycle de lettres et de linguistique. Il présente l'avantage d'exposer clairement les différentes catégories de pronoms, et, pour chacune d'elle, de noter les problèmes qu'elles peuvent poser au cours de l'analyse linguistique.

Dans *La valence : l'approche pronominale et son application au lexique verbal* (Eynde, Mertens, 2002 : pp. 21), un inventaire des pronoms clitiques, des suspensifs et des paranoms est présenté. Le terme de pronom clitique ne pose pas de problème. En revanche, dans la catégorie des non-clitiques, on trouve des pronoms démonstratifs (*celui-là...*), des pronoms possessifs (*son...*, qui apparaissent définis comme des pro-det¹ de connexion), des pronoms indéfinis (*autant...*). Dans la catégorie des paranoms, on trouve aussi beaucoup de pronoms indéfinis (rien, personne, quelqu'un...). Quant à la catégorie des suspensifs, elle est composée essentiellement de pronoms interrogatifs (*où, d'où, quand, comment, combien, quoi, qui...*). Le tableau représentant l'inventaire de ces pro-formes découpe ces classes « dans un deuxième plan, en fonction du statut syntaxique des éléments, conformément à la classification proposée dans Eynde (1995). »

Ainsi, 5 classes sont dégagées (*op. cit.* : pp. 19-20) :

- « L'élément est du type *pro-syntagme* s'il est proportionnel avec un référent d'une fonction syntaxique particulière. Les pronoms « je, tu, il, on, ils » indiquent ainsi un référent sujet. L'existence de pronoms homonymes tels que « nous, vous, se, ... » figurant dans plusieurs paradigmes (sujet, objet, datif, ...) ne gênent nullement la classification, puisque les éléments sont définis à partir du paradigme entier. Dans le tableau, les homonymes sont accompagnés d'un indice indiquant leur fonction syntaxique.

¹ Pro-déterminant

- L'élément est du type *pro-référent* s'il est proportionnel à un référent sans que sa fonction syntaxique soit marquée. Le pronom tonique « toi » désigne un référent, mais aucunement sa fonction syntaxique. Signalons l'ambiguïté fonctionnelle des prépositions : si le complément est introduit par « à », il peut s'agir d'un objet indirect, d'un circonstant de lieu ou de temps, d'un complément prépositionnel, etc. Si « ça » est un pro-référent, en revanche « contre ça » et « pour ça » sont des pro-syntagmes.
- L'élément est du type « *pro-noyau + déterminant d'identification* » s'il est proportionnel à un référent complexe comportant un noyau déterminé par un démonstratif, un possessif, une construction relative ou un syntagme prépositionnel. Par exemple, si le pronom « lequel » est proportionnel à « ce bureau », le noyau auquel il renvoie est « bureau » et ce dernier est accompagné d'un déterminant pour le déterminer, à savoir le démonstratif « ce ». Dans « lequel », « quel » est déterminant et « le » pro-noyau. Dans « celui-là », le pro-noyau correspond à « lui » et le déterminant d'identification à « ce...là ». (*lequel % ce bureau % celui-là / lequel % mon bureau % celui-là / lequel % le bureau où tu travailles % celui-là / lequel % le bureau près de la fenêtre % celui-là*)
- L'élément est du type « *pro-noyau + déterminant de connexion* » s'il est proportionnel à un référent complexe comportant un noyau déterminé par un possessif. Dans « le mien », « le » est pro-noyau et « mien » un déterminant exprimant la connexion entre le référent indiqué par le pro-noyau et celui impliqué par le possessif. (*le mien % mon bureau*)
- Les « *pro-déterminants d'identification* » (1) et les « *pro-déterminants de connexion* » (2) sont des éléments qui n'interviennent dans la détermination de la valence que pour les constructions à verbe support. Ils interviennent de façon plus régulière dans la détermination de la valence nominale. (1 *prendre la responsabilité de / en prendre la responsabilité / prendre cette responsabilité* 2 *mettre à ma/ta/sa... disposition / mettre à la disposition de...*). »

Après ce tour d'horizon des différentes questions et propositions faites sur la définition et sur le statut du « pronom », voici les différents types de pro-formes que nous avons quant à nous utilisées pour schématiser les constructions verbales du corpus. Tout

d'abord, nous présenterons les paradigmes de clitiques posés par Blanche-Benveniste et al. (1987), puis les pro-formes non-clitiques.

2.3.2.2. Les pronoms clitiques. Les paradigmes de clitiques.

Les clitiques ne posent pas de problème de représentation (hormis les cas particuliers du *je*, du *tu*, et du *il* [asubjectal] dont nous venons d'observer qu'ils ne pouvaient pas revêtir l'étiquette de *pronom*). L'avantage qu'ils présentent est le suivant : venant par définition à gauche du verbe dont ils « symbolisent » la rection, ils y apparaissent « rangés » dans un ordre qui permet dans la plupart des cas de déduire leur rôle syntactico-sémantique. C'est ce qu'ont noté les auteurs de *Pronom et syntaxe*.

L'Approche Pronominale dresse en effet quatre paradigmes de pronoms clitiques¹. Mais avant de les observer, il semble ici nécessaire d'éclaircir cette notion de *paradigme*. C. Blanche-Benveniste et al. (1987 : pp. 80-83) indiquent :

Notre méthode affirme que les clitiques sont de bons indices des constructions verbales. Il est donc indispensable de faire une description générale des constructions qui associent verbes et clitiques. Les paradigmes de clitiques nous livreraient ainsi les paradigmes des constructions verbales.

On entend par « paradigme de clitiques » l'ensemble des clitiques qui commutent auprès d'un verbe dans un même contexte. Ainsi, pour le verbe *voir*, entre *il* et *voit*, on peut trouver :

Me, te, nous, vous, le, la les, en, se.

On dira que cet ensemble constitue un paradigme des clitiques du verbe *voir*.

Si l'on recensait isolément les paradigmes de tous les verbes, on aboutirait à une atomisation de la description qui risquerait de faire apparaître, pour décrire 30 verbes, 20 types de paradigmes différents. Il est donc essentiel de faire apparaître les parentés dans les paradigmes. Ces ensembles paradigmatiques sont caractérisés par des relations de solidarité ; certaines sont générales et valent quel que soit le verbe considéré. Ainsi, *nous* et *vous* sont solidaires ; dans tous les paradigmes où apparaît l'un, l'autre est garanti ; les apparentes exceptions peuvent s'expliquer par des règles indépendantes, à portée très générale ; par exemple, le fait que *nous* est possible dans un contexte où *vous* ne l'est pas :

¹ BLANCHE-BENVENISTE C., DEULOFEU, J., STEFANINI, J. ET VAN DEN EYNDE, K., 1984 (1987 : 2^{ème} édition augmentée), *Pronom et syntaxe, l'approche pronominale et son application au français*, Paris, Selaf, p. 87.

Tu nous regardes, *tu vous regardes

se laissent ramener à une loi générale sur l'incompatibilité entre les personnes simples et complexes. En dehors de l'application de cette règle, *nous* et *vous* apparaissent dans les mêmes contextes, quel que soit le verbe. Il ne sera donc pas nécessaire de mentionner les deux éléments, pour la recherche des caractéristiques de construction. De la même façon, un *lui* est toujours solidaire d'un *me*, quel que soit le verbe ; la forme *lui* peut donc servir à représenter aussi bien un *lui* qu'un *me*.

Les autres solidarités sont particulières et définissent des classes de verbes. La méthode que nous utilisons pour les décrire consiste à dégager de grands ensembles de clitiques, qui sont effectivement représentés par certains verbes, et qui forment une sorte de paradigme maximal, et de considérer les autres comme fragmentaires par rapport à ceux-là. Ainsi, pour le verbe *agiter*, on trouve neuf formes de clitiques compléments possibles : *me*, *te*, *nous*, *vous*, *le*, *la*, *les*, *en*, *se*. Par rapport à celui-ci, apparaissent comme fragmentaires les paradigmes de plusieurs autres verbes, tels que :

<i>Rassembler</i>	nous,	vous,	les,	en,	se
<i>Concerner</i> me,	te,	nous,	vous,	le,	la,	les,	en,	..
<i>Aimer</i>	me,	te,	nous,	vous,	le,	la,	les,	..
se								
<i>Cabrer</i>	me,	te,	nous,	vous,
se								

Le paradigme maximal n'a d'existence que pour autant qu'il est représenté par des verbes réels. Il n'y a donc aucun a priori dans l'établissement de ces paradigmes.

C'est à partir de ces calculs que l'on peut, par induction, poser trois grands ensembles de clitiques compléments (le calcul des paradigmes de sujet devra être calculé à part). Ces trois grands paradigmes sont nommés P1, P2 et P3, et leur tête de liste peut être représentée par *le* pour P1, *lui* pour P2 et *lui* pour P3¹.

Comme les formes de clitiques sont ambiguës, le calcul doit être complété par une étude des doubles marquages, qui est un moyen de vérifier les proportionnalités qu'entretiennent les clitiques, et par là, de les désambigüer : on distinguera ainsi, par exemple, deux *me* et deux *y*, d'après l'emphase :

¹ Nous verrons par la suite que le paradigme P3 est problématique ; en effet, les pronoms associés à leur « emphase », par exemple : *lui_de lui*, sont difficilement réalisables. Ainsi, nous préférons poser ici comme tête de liste le pronom *en(_deça)*, qui pose apparemment beaucoup moins de problèmes.

Il *me* regarde, *moi* - il *me* parle, *à moi*

Il y songe, *à ça* - il y va, *là*

L'analyse doit déboucher sur une analyse en traits, qui permettra de désigner les paradigmes comme des unités à caractéristiques syntactico-sémantiques.

Le terme de pronom clitique désigne les pronoms qui apparaissent à gauche du verbe (sauf à l'impératif où ils apparaissent à droite du verbe). On les appelle aussi pronoms conjoints ou atones (en opposition aux pronoms disjoints, qui apparaissent à droite du verbe, et qui sont aussi nommés toniques). Denis Creissels (1995 : pp. 22-31, pp. 108-125 et p. 158 pour les parties concernant ce sujet) considère les clitiques comme des indices pronominaux de la construction verbale, tout comme les auteurs de *Pronom et syntaxe*. Il indique (*op. cit.*, p. 159) :

L'existence de paradigmes d'affixes verbaux constitués d'indices pronominaux représentant un ou plusieurs arguments du prédicat verbal n'a rien d'universel. Certaines langues ([...] le zarma) n'ont pas d'affixes verbaux du tout, d'autres (on peut citer le japonais) ont un système d'affixes verbaux qui ne manifestent avec aucun constituant nominal la relation d'accord propre aux indices pronominaux.

Le cas de formes verbales intégrant un ou plusieurs indices pronominaux est toutefois beaucoup plus répandu à l'échelle des langues du monde que ne le laisse supposer les descriptions des langues actuellement disponibles.

Une fois montré que les « pronoms personnels conjoints » du français sont en fait des indices pronominaux faisant morphologiquement partie du mot verbal, on doit reconnaître que cette situation est aussi celle du français. Nous avons vu en effet au chapitre I que c'est là un point sur lequel la tradition grammaticale française masque le véritable fonctionnement du français, qui s'est écarté beaucoup plus nettement des autres langues de l'Europe moderne du type d'organisation qu'on trouve dans les langues indo-européennes anciennes pour se rapprocher considérablement d'un type d'organisation que l'on trouve par exemple dans les langues bantoues, avec la possibilité d'antéposer à la base verbale, non seulement un indice de sujet, mais aussi des indices susceptibles de représenter divers types de compléments. C'est le cas de l'ex. (4), où les traits d'union marquent l'appartenance au mot verbal de morphèmes qui sont loin d'avoir l'autonomie que suggère la graphie usuelle du français.

(4) nous-ne-les-lui-donnerions peut-être pas

Voici à présent un inventaire des clitiques utilisés comme indices pronominaux de la construction verbale par l'Approche Pronominale. Ces clitiques sont répartis dans 4 paradigmes :

- P0 est le paradigme comprenant les clitiques sujets [*je_moi, tu_toi, il_lui, il_Ø, elle_elle, on_nous, on_Ø, nous_nous, vous_vous, ils_eux, elles_elles, c'_ça, c'_ceci*], que nous indiquons ici avec leur extension disjointe (*moi, toi, lui, etc.*). Indiquer cette extension présente l'avantage de désambiguïser le clitique en précisant un trait : par exemple, *il_lui* indique un trait [+/-personnel], tandis que *il_Ø* indique un trait [asubjectal]. Les pronoms de 1^{ère} et de 2^{ème} personne sont, par nature, nécessairement personnels.

Denis Creissels (*op. cit.* : p. 122) propose de « substituer respectivement « élocutif », « allocutif » et « délocutif » aux termes traditionnels de première, deuxième et troisième personne », constatant que le terme de *personne* est approprié quant il s'agit des deux premières, mais que tel n'est pas le cas pour la troisième (*il* ne réfère pas nécessairement à une personne). Ainsi, il résume la correspondance entre les termes traditionnels et la terminologie qu'il propose dans le tableau suivant :

interlocutif	élocutif	↔	« première personne »
	allocutif	↔	« deuxième personne »
délocutif		↔	« troisième personne »

Cette proposition avait déjà en partie été proposée par M. Gross (1986 : p. 22) lors de la présentation de ce qu'il a choisi de nommer *particules préverbales (ppv)*¹. Il considérait alors avec Beauzée que la dénomination de *pronom* n'était pas pertinente pour toutes ces formes, dans la mesure où il est parfois impossible de trouver un substantif à leur origine. Par exemple, les « pronoms » des deux premières personnes, *je* et *tu*, n'ont pas de possibilité de substitution par un substantif. La seule forme qui peut leur être accolée est leur forme disjointe, respectivement *moi* ou *toi*. M. Gross proposait alors de nommer ces deux formes *ppv de dialogue* ; ainsi, la proposition de D. Creissels se tient dans cette ligne et la précise.

Dans la base de données, les éléments appartenant au paradigme P0 se situent dans les quatre premières colonnes à partir de la gauche, avant la mention du verbe constructeur. Nous avons distingué quatre catégories de sujets :

¹ *je, tu, il, elle, nous, vous, ils, elles, on, ne, me, te, se, le, la, les, lui, leur, y, en, ce.*

- les sujets de type [+personnel], correspondant à un type *je*,
 - les sujets de type [-personnel], correspondant à un type *il* (qui ne comprend ni les référents humains ni le sujet de type asubjectal),
 - les sujets de type [ça], auxquels nous n'avons pas attribué de valeur syntactico-sémantique particulière, mais que l'on retrouve de manière quasi-systématique dans les constructions pronominales passives, entre autres. Ce type de sujet est souvent couplé à des S_{vinf} ou des Que-phrases. Colette Jeanjean (1982 : p. 118) insistait sur le fait que « *ça* a un statut grammatical qui interdit de le considérer comme une variante d'autres morphèmes, dotée de simples significations superficielles » et qu' « il est révélateur de faits importants pour l'analyse de la construction verbale en français, comme la possibilité d'avoir une *quand*-phrase sujet : *ça* fait mal, *quand il sort le bébé* ».
 - les sujets de type *il* [*asubjectal*], que nous avons déjà évoqué dans la partie précédente. Il s'agit d'un sujet « formel » sans référence à quoi que ce soit, ni dans le discours, ni dans le monde réel.
- P1 est le paradigme comprenant les clitiques compléments [*le, la les, me, te, se, vous, nous*], c'est-à-dire des pronoms ayant une extension disjointe non-introduite par une préposition. Le développement de ces clitiques en *le_lui, le_ça* etc. permet de préciser les traits des compléments (*le_lui, la_elle, les_eux, les_elles* ont un trait [+personnel], c'est-à-dire référant à une personne, alors que *le_ça* a un trait [-personnel], c'est-à-dire qu'il réfère à une action, une situation, un objet¹).

Ex : Max taquine *Paul* > Max *le* taquine, *lui*.

Max regrette *d'avoir fait pleurer Marie*. > Max *le* regrette, *ça*.

- P2 est le paradigme comprenant les clitiques compléments [*y_à ça, y_à celui-ci, y_là, lui_à lui/à elle, me_à moi, te_à toi, nous_à nous, vous_à vous, se_à lui/à elle/à eux/à*

¹ Cette distinction se retrouve sur le plan syntaxique, puisque, comme l'a remarqué C. Blanche-Benveniste, le fait qu'un pronom soit [+personnel] se retrouve dans l'extension disjointe *lui, elle*, qui réfèrent nécessairement, dans cette position post-verbale, à une personne.

elles, leur_à eux/à elles]. Nous avons distingué dans la base de données les P1 [+personnel] des P1 [-personnel]

Ex : Max donne son cours de linguistique *à Marie*. > Max *lui* donne son cours de linguistique. (P1 [+personnel])

Max va à Lyon. > Max *y* va. (Ici, le P1 est noté [-personnel]. Il s'agit plus précisément d'un locatif ; nous n'avons pas intégré cette précision dans la base. Cependant, les locatifs sont toujours accompagnés du cochage de la case *où*, qui permet de désambiguïser leur statut ; de plus, si l'ambiguïté n'est pas levée immédiatement, on pourra se tourner vers le corpus, qui montrera alors qu'il s'agit d'un locatif ou d'un autre type de pronom.)

- P3 est le paradigme comprenant les clitiques compléments [*en_de ça/de lui d'elle/d'eux, lui_de lui, leur_d'eux/d'elles, me_de moi, nous_de nous, vous_de vous, se*].

Ex : Max parle *de linguistique*. > Max *en* parle.

2.3.2.3. Les pro-formes

Nous rangeons sous l'étiquette « pro-forme » tous les éléments qui, à l'intérieur d'une construction verbale, peuvent « appauvrir le lexique » des éléments construits (les clitiques, que nous venons d'énumérer, en font partie). Claire Blanche-Benveniste (2002b : p. 61) propose de préciser cela avec l'exemple suivant :

Ex. : Dans un premier temps, les cations adsorbés neutralisent la charge négative à la surface du colloïde.

Comparant la démarche de l'Approche Pronominale à celle d'un lecteur non averti de cet exemple, elle précise que ce dernier peut « toujours en retenir une version lexicalement pauvre, mais qui 'tient debout' : Dans un premier temps, ils neutralisent la charge négative là». Ainsi, « l'Approche Pronominale propose, en utilisant une démarche semblable, d'étudier la complémentation verbale en conservant le lexique du verbe mais en appauvrissant systématiquement le lexique des éléments dépendants du verbe. »

Les pro-formes comprennent les pronoms *clitiques*, les pronoms *toniques* (aussi appelés *disjoints*, qui permettent de désambiguïser le clitique, comme dans *je l'aime*, {*lui ; elle*} ou encore *j'y pense* {à *lui* ; à *elle* ; à *ça*}), les *suspensifs* (les formes interrogatives, souvent utilisées par la grammaire scolaire pour la reconnaissance des fonctions des constituants, comme *qui*, *quoi*, *comment*, à *qui*, à *quoi*) et les *paranoms* (*partout*, *toujours*, *nulle part*, *rien*, *personne*...).

Inventaire des pro-formes

Un inventaire de ces pro-formes a été établi par K. van den Eynde *et al*, pour l'élaboration du dictionnaire Proton¹. Il regroupe pronoms clitiques, non-clitiques, des suspensifs et des paranoms. Voici le tableau qu'ils présentent (Eynde, Mertens, 2002 : p. 24) :

P0	P1	P2	P3	PL	PT	PM	PQ	PP (...)
je nous on elle il il en il en...Q ils que qui que_il ça ceci celui-ci ceux-ci	0 le la les te en Q vous ceci combien que qui ça se rec. se refl.	0 y lui leur se réc. se réfl. lui_ton eux ceci celui-ci ceux-ci l'un l'autre qui quoi ça	0 en lui leur se réc. se réfl. lui_ton eux ceci celui-ci ceux-ci combien l'un l'autre qui quoi ça	0 y en où là ici là-bas prep ici prep là prep là-bas prep où	0 quand prep maintenant prep quand prep à c m-l. à ce mom-là	0 comment comme ça	0 combien	0 qui quoi ça ceci lui_ton eux celui-ci ceux-ci combien elle-même soi-même l'un l'autre là contre là-dedans là-dessous là-dessus prep []
il_deinf il_qpind il_qpsubj ça(de_inf) ça(Inf) ça(qpsubj)	ça(de_inf) ça(indq) ça(qpind) ça(qpsubj) le(de_inf) le(indq) le(qpind) le(qpsubj) le(sipind) le(à_inf)	ça(qpind) ça(qpsubj) ça(à ce qpi) ça(à ce qps) ça(de_inf) ça(qpind) y(de_inf) y(qpind) y(qpsubj)	ça(de ce qpi) ça(de ce qps) ça(de_inf) ça(qpind) ça(qpsubj) ça(sipind) en(de ce					ça(Inf) ça(p. ce qps) ça(qpsubj)

¹ EYNDE, K. VAN DEN, MELIS, L., MERTENS, P., VANGILBERGEN, L., DEHASPE, L., EGGERMONT, C., BROEDERS, E., *Dictionnaire de valence Proton*, consultable sur Internet à l'adresse <http://bach.arts.kuleuven.ac.be/PA/protonfr.html>

		y(à ce qpi) y(à ce qps) y(à _inf)	qpi) en(de ce qps) en(de _inf) en(qpind) en(qpsubj) en(sipind)					
--	--	---	--	--	--	--	--	--

Liste des étiquettes employées pour représenter des pronoms.

Dans ce tableau, nous observons que des ajouts ont été effectués par rapport aux paradigmes définis dans *Pronom et syntaxe*. En effet, PL, PD, PT, PM, PQ et PP sont des paradigmes qui ont été créés pour l'élaboration de Proton. K. Van den Eynde et al. (*op. cit.*, pp. 14-15) définissent ainsi ces cinq paradigmes :

PL est le locatif (le complément actanciel de lieu).[...]

Viennent enfin les compléments de temps, de manière et de quantité, désignés par les étiquettes PT, PM et PQ, respectivement.[...]

Certains prédicateurs comportent également un complément prépositionnel (PP) ou plusieurs. Ceux-ci sont caractérisés par la préposition ou par l'ensemble de prépositions qu'ils sélectionnent.[...]

Les actants PL, PT, PM et PP se distinguent des circonstants correspondants (de lieu, de temps, de manière) par le fait qu'ils sont spécifiques au prédicateur.

Un syntagme prépositionnel introduit par « à » peut être P2 (« à moi »), PL (« à Paris »), PT (« à six heures »), PM (« à sa manière »), etc. Le même type d'ambiguïté se présente pour les SN introduits par « de ». D'une manière générale, en cas d'ambiguïté fonctionnelle la fonction sera déterminée à partir du paradigme proportionnel complet. Le groupe prépositionnel sera analysé PL dans « il va à Paris » en raison de sa relation avec « y, là, où, *lui » ; il sera analysé P2 dans « il offre ça à Jean » en raison de sa relation avec « *y, *là, *où, lui ».

Quant à nous, nous n'avons pas utilisé les actants PL, PT, PM et PP, car nous avons préféré indiquer l'« extension lexicale » de la pro-forme lorsqu'elle paraissait ambiguë¹. Nous avons ainsi, pour reprendre l'exemple de K. van den Eynde et al. :

Ex : Il va à *Paris*. > Il y va. (où y = à + SN[-pers])

¹ Cependant, si ambiguïté il y a dans la base de données, elle est nécessairement éclaircie par le corpus.

Qui se distingue de :

Ex : il va *faire ses courses*. > Il y va. (où $y = SVinf\Omega$)

Si nous n'avons pas repris la dénomination proposée par K. van den Eynde et al., nous avons cependant mentionné ces divers paradigmes dans la base, qui apparaissent sous les étiquettes *ainsi*, *où*, *combien*, *pourquoi*, *quand*. En revanche, il n'y a pas dans notre travail de distinction faite entre *actants* et *circonstants*, autrement dit entre éléments entrant dans la valence d'un verbe et éléments ne faisant partie que de la rection au sens restreint, pour des raisons déjà évoquées.

Les éléments qui entrent dans la construction d'un verbe peuvent donc être remplacés par une pro-forme, qui peut être soit un pronom clitique soit une pro-forme d'autre nature, ce que nous allons observer dans la partie suivante.

2.3.3. Éléments construits

On appelle éléments construits (par le verbe), ou éléments régis (par le verbe) les éléments qui possèdent les trois propriétés suivantes, que nous expliquerons dans cette partie :

- 1- Un élément construit a une équivalence avec une pro-forme ;
- 2- Un élément construit peut porter les modalités du verbe ;
- 3- Un élément construit peut être focalisé.

Tous les éléments qui ne répondent pas à ces trois propriétés ne sont pas des éléments construits par le verbe ; C. Blanche-Benveniste les appelle « éléments associés », en opposition à ce que l'analyse ancienne appelait complément de phrase. La critique de cette terminologie tient essentiellement au fait que « ce qu'on appelle généralement complément de phrase¹ se définit surtout par des réponses négatives aux propriétés que peut avoir un complément de verbe. En tout premier lieu, il faut rappeler qu'un verbe sélectionne ses compléments, alors qu'une phrase est hors d'état de le faire » (2002a : p. 13).

¹ Nous avons nommé *associés*, en vrac, tous ces compléments dont les relations s'exercent hors du domaine de la rection verbale (Blanche-Benveniste et al., 1990).

Les éléments associés n'entrent pas dans la base de données que nous avons constituée. Voici un exemple d'élément associé :

Ex : *Evidemment*, Max aime la linguistique.

Dans cet exemple, nous observerons l'élément *évidemment*.

- 1- Il a certes bien une équivalence avec une pro-forme : on peut le remplacer par *ainsi*, comme dans l'exemple suivant :

Ex. *Ainsi*, Max aime la linguistique.

- 2- En revanche, il ne peut pas porter les modalités du verbe :

Ex : *Max n'aime pas la linguistique *évidemment*, mais naturellement.

- 3- Et il ne peut pas être focalisé :

Ex : *C'est *évidemment* que Max aime la linguistique.

Evidemment sera donc considéré comme un élément associé, et ne sera pas intégré à la base de données, dans la mesure où il n'est pas représentatif d'un paradigme d'argument du verbe (contrairement à *la linguistique*, dans l'exemple proposé), mais porte sur l'énonciation.

D. Creissels (1995 : p. 150), dans un chapitre consacré aux adverbes, propose

« un test très simple pour faire apparaître de manière sûre des différences quant au point d'incidence structurel d'« adverbes » qui à première vue semblent participer de façon semblable à la construction des énoncés : un « adverbe » intégré à la structure prédicat-arguments peut être mis sous la portée de la négation en étant précédé (pas forcément de façon immédiate) du morphème négatif *pas* comme dans l'ex. (80) ; par contre, un adverbe incident à l'unité phrastique en tant qu'unité d'énonciation, s'il n'est pas incompatible avec la négation, se place avant le morphème *pas*, comme dans l'ex. (81).

(80) il n'est PAS parti RAPIDEMENT

(81) il n'est certainement pas parti

Le terme de « modalisateur » est proposé ici pour étiqueter les unités ayant les propriétés syntaxiques illustrées ci-dessus par l'« adverbe » *certainement*. On dit parfois « adverbe de

phrase », mais ce terme me semble à proscrire, car il implique l'existence d'une catégorie « adverbe » dont les modalisateurs constitueraient une sous-catégorie, ce qui est contesté ici. »

Mais voyons à présent la réalisation des trois propriétés auxquelles doivent répondre les éléments construits.

2.3.3.1. Un élément construit a une équivalence avec une *pro-forme*

- Les pronoms clitiques : cf. [2.3.2.2.](#)
- Les *pro-formes* d'autre nature :

Tous les éléments construits pas le verbe n'ont pas nécessairement d'équivalence avec un pronom clitique. Dans ce cas, il est possible de les remplacer par un autre type de *pro-forme*. Comme le signalent les auteurs du *Français parlé, études grammaticales* (C. Blanche-Benveniste et al., 1990 : p. 42) :

Pour utiliser systématiquement les « pronoms », nous avons dû élargir la liste des éléments qu'on classe généralement sous ce vocable ; il vaudrait mieux du reste parler de « *pro-formes* », car ils ne sont pas seulement en relation avec des « noms ». Nous admettons comme pronoms, aussi bien les pronoms clitiques (dits « pronoms personnels atones »), comme *je, tu, il, elle, on, nous, vous, ils, elles, me, te, se, le, la, les, lui, leur, y, en*, que les formes comme *là, ainsi, alors, ceci, cela, celui-ci, celui-là, qui, que, quoi, où, comment*, ou les séries *quelqu'un, quelque chose, quelques part, personne, rien, nulle part*.

Nous avons donc utilisé diverses *pro-formes* lors de la pronominalisation des éléments construits de notre corpus : pronoms clitiques, mais aussi pronoms démonstratifs, adverbes... Par exemple :

- *ça*

Ex : Max aime *étudier la linguistique*. > Max aime *ça*.

- *Ainsi*

Ex : Max lit le livre de linguistique *en pensant à Marie*. > Max lit le livre de linguistique *ainsi*.

- *Où*

Ex : Max lit le livre de linguistique *à la bibliothèque*. > Max lit le livre de linguistique *où*.

- *Combien*

Ex : Max lit *trois* livres de linguistique. > Max *en* lit *combien*. (dans ce cas-là, *combien* marque le statut de quantifieur du clitique *en*)

- *Pourquoi*

Ex : Max lit le livre de linguistique *parce qu'il veut comprendre le fonctionnement du langage*. > Max lit le livre de linguistique *pourquoi*.

- *Quand*

Ex : Max lit le livre de linguistique *à partir de 10 heures*. > Max lit le livre de linguistique *quand*.

2.3.3.2. Un élément construit peut porter les modalités du verbe

Un élément construit peut être dans la portée de la modalité verbale. Nous allons l'observer à l'aide de l'exemple suivant. Nous utiliserons la modalité négative.

Ex : Max ne donne pas son cours de linguistique à Marie.

Nous avons vu au cours de la partie précédente que les éléments *son cours de linguistique* et *à Marie* avaient une équivalence avec une pro-forme, plus précisément avec un pronom clitique.

Voyons à présent s'ils peuvent porter la modalité négative du verbe :

En ce qui concerne son cours de linguistique :

Ex : Max ne donne pas *son cours de linguistique* mais *son cours de physique* à Marie.

Son cours de linguistique est manifestement dans la portée de la modalité négative du verbe *donner*, il possède donc cette propriété.

En ce qui concerne *à Marie* :

Ex : Max ne donne pas son cours de linguistique à *Marie* mais à *Anne*.

A *Marie* est aussi dans la portée de la modalité négative du verbe *donner*, il possède aussi cette propriété.

Observons à présent la dernière propriété que doivent posséder les éléments construits, la focalisation.

2.3.3.3. Un élément construit peut être focalisé

Un élément construit peut être focalisé, c'est-à-dire qu'il peut entrer, par exemple, dans une tournure en *c'est...que*.

Reprenons le même exemple pour illustrer cette propriété :

Ex : Max donne son cours de linguistique à Marie.

Et observons encore une fois les éléments *son cours de linguistique* et à *Marie*.

Son cours de linguistique :

Ex : c'est *son cours de linguistique* que Max donne à Marie.

A Marie :

Ex : c'est à *Marie* que Max donne son cours de linguistique.

Les deux éléments *son cours de linguistique* et à *Marie* sont donc bel et bien construits par le verbe *donner*. Ils possèdent tous deux les trois propriétés des éléments construits : ils ont une équivalence avec une pro-forme, ils peuvent porter la modalité du verbe, et enfin, ils peuvent être focalisés.

Comme nous l'avons présenté plus haut, les éléments ne possédant pas une de ces propriétés ou n'en possédant aucune ne sont pas construits par le verbe. Ils sont nommés *associés*.

2.3.4. La notion de formulation

Utilisées par l'équipe du GARS et par K. van den Eynde dans sa présentation du dictionnaire de valence Proton, les notions de formulation et de groupes de formulation servent essentiellement à définir l'ensemble des différents types de constructions possibles pour un même verbe morphologique. Ainsi, un verbe peut avoir plusieurs ensembles de valences : le verbe *écrire* peut se construire avec trois éléments de valence, un sujet, un objet, et un bénéficiaire, comme dans l'exemple *je le lui ai écrit (à cet ami, qu'il devait venir)*, ou avec seulement deux éléments, un sujet et un complément, comme dans l'exemple *ça écrit mal (ce stylo, sujet de type ceci, et mal, complément de type ainsi)*. Un groupe de formulations est donc le résumé des différentes formulations possibles d'un même verbe. Comme le soulignent les auteurs du *Français parlé, étude grammaticales* (C. Blanche-Benveniste et al., 1990, pp. 44-45) :

Un verbe peut se manifester avec différents ensembles de valence ; ainsi, le verbe *écrire* en a, entre autres, deux :

tu le lui as écrit (à cet ami, qu'il devait venir) > avec un sujet personnel de type *je, tu, il*, et deux compléments, *le* et *lui*

ça écrit mal, ce stylo écrit mal > avec un sujet non personnel, de type *ceci, cela*, et un complément de type *ainsi*.

Les deux valences de *écrire* ne diffèrent pas seulement par leurs sujets, mais par les compléments : dans le deuxième cas, il serait impossible d'avoir *le* et *lui* :

*ce stylo lui écrit mal qu'il doit venir

*ce stylo le lui écrit mal

Nous appelons « formulations » ces ensembles de valence. Il est bien connu que les verbes les plus fréquents de la langue ont des formulations très diversifiées. Le problème se pose de savoir comment traiter ces diverses formulations ; pour *écrire*, l'intuition conduirait à dire qu'il s'agit d'un même verbe, doté de deux formulations apparentées ; tandis que pour *dépendre*, il s'agirait de deux verbes entièrement distincts, dans : *il dépend le jambon du plafond (il le dépend de là) / cela dépend entièrement de vous (cela en dépend)*. [...]

Nous appelons « groupes de formulations » les ensembles de valence qui ont des parentés manifestes et qui semblent se rapporter à un même verbe. Un groupe de formulation typique serait l'exemple de Maurice Gross (1976) : *le jardin grouille d'insectes / Des insectes grouillent dans le*

jardin. Le terme *en/ils*, qui apparaît alternativement comme complément et comme sujet est reconnaissable à son trait de complexité : on doit avoir *grouille d'insectes*, au pluriel, et non *d'un insecte*, et de même *les insectes y grouillent*, au pluriel et non *l'insecte y grouille*. Les deux formulations semblent exprimer la même relation, en la prenant soit par le bout du *jardin* soit par le bout des *insectes*.

Ce que nous nommons « groupes de formulations » a été identifié, sous d'autres vocables, par de nombreux auteurs. Nous traitons de cette façon, par exemple, les « verbes causatifs » :

Il brûle la forêt, il la brûle

La forêt brûle, elle brûle

La formulation longue (avec sujet et complément) est considérée comme « causative » par rapport à la plus courte (seulement un sujet).

L'intérêt de l'utilisation de ces groupes de formulations est essentiellement classificatoire. En effet, repérer une formulation puis sa ou ses reformulations possibles permet de simplifier la description du verbe. Comme le disait Maurice Gross, il faut parvenir à classer les éléments ayant des propriétés communes, et c'est ce que permettent les groupes de formulations, en regroupant dans une même série un même verbe employé soit avec X en sujet et Y en complément, par exemple, soit avec Y en sujet et X en complément (et où pour les deux formulations, X=X et Y=Y).

Dans l'article de K. van den Eynde et P. Mertens (2002 : p. 10), la méthode de détermination des groupes de formulations est clairement exposée. Nous nous contenterons ici de rapporter la définition qu'ils donnent de la notion de formulation :

La *formulation* n'est autre que le tableau distributionnel des pronoms attestés pour un prédicateur donné. C'est l'ensemble des paradigmes d'éléments spécifiques au prédicateur, où chaque paradigme comporte l'ensemble des pronoms (clitiques ou non) pouvant apparaître dans la position donnée, après élimination des restrictions sur l'ordre linéaire et après résolution des paradigmes complémentaires. Elle offre déjà une caractérisation partielle de la valence du prédicateur : elle explicite les restrictions¹ sur le contenu du paradigme (*il pleut* / **{elle, on} pleut*), sur le nombre et la nature des dépendants (leur fonction syntaxique) (*je lui offre celui-là* / **je lui dors celui-là*), ainsi que sur leur caractère obligatoire ou facultatif (*je les lui offre*, *je les offre* / **j'offre*).

¹ Ces restrictions étaient appelées *micro-grammaire des pronoms* dans Blanche-Benveniste *et al.* (1984).

Dans le cadre de ce travail, les différentes formulations et les groupes de formulations seront classés uniquement s'ils apparaissent dans le corpus. Nous n'avons pas jugé nécessaire d'utiliser les transformations pour rendre notre corpus plus important qu'il n'est déjà, d'autant plus que, comme nous l'avons déjà indiqué, les deux transformations¹ auxquelles nous avons eu recours ne sont pas génératrices de structures nouvelles, elles permettent simplement d'« éclaircir » la structure de départ. Cependant, l'extension du corpus existant par transformations pourra éventuellement faire l'objet d'un travail de recherche fondé sur l'inventaire et le classement proposé ici, ce qui permettrait d'avoir, pour les quelques centaines de verbes morphologiques de ce corpus, un inventaire exhaustif de leurs complémentations possibles, en tenant compte, cette fois-ci, de l'intuition du linguiste et non plus de la réalisation effective de la production verbale. Pour valider cette extension par transformations, et éviter les critiques consistant à dénier au linguiste la possibilité de faire confiance à son intuition, nous pourrions envisager d'établir un questionnaire recensant toutes les constructions « nouvelles » (ne figurant pas dans l'inventaire de départ), et les soumettre au jugement des locuteurs enregistrés lors de l'enquête.

2.3.5. Problèmes de pronominalisation : le cas du verbe *aimer*

Il est utile ici de préciser que les exemples choisis ci-dessus sont aisément pronominalisables. Cela n'est pas toujours le cas. Dans un corpus de français parlé, rien n'est moins évident par exemple que de pronominaliser certains des éléments : on bute sur le choix du type de pro-forme à utiliser.

Pour le verbe *aimer* par exemple, lorsqu'on trouve une phrase du type *j'aime le lapin*² dans laquelle le sens de *le lapin* est clairement celui de « viande », de « produit de consommation », on peut difficilement pronominaliser *le lapin* avec *le* : *le lapin*, je l'aime, car cela reviendrait non pas à considérer *le lapin* comme une viande qu'on apprécie, avec un trait [-personnel], mais plutôt comme un animal de compagnie, avec un trait [+personnel]. On utilisera ici la pro-forme *ça* : *le lapin*, j'aime ça.

Cependant, contrairement à ce que l'exemple du lapin pourrait suggérer, il n'y a pas pour le verbe *aimer* de répartition nette en ce qui concerne la pronominalisation entre les

¹ Il s'agit de la pronominalisation et de la réduction des subordonnées en indépendantes

² Cf. Marina Yaguello, *Petits faits de langue*, Paris, Le Seuil, pp. 35-39.

traits [+personnel] et [-personnel]. De prime abord, on pourrait croire qu'un élément possédant un trait [+personnel] se pronominalise en *le la les*, et qu'un élément possédant un trait [-personnel] se pronominalise en *ça*. Il n'en est rien : lorsqu'un élément [-personnel] est individualisé, ou rapporté à un élément [+personnel], il est pronominalisé en *le* :

Ex : *ton père aime son métier* L1 ben tout du moins euh ses parents *le* faisaient du côté de ma mère (FRUITM~1 12,7)

> ton père l'aime, son métier

De la même manière, un élément [+personnel] caractérisant une catégorie peut être pronominalisé en *ça* :

Ex : Les enfants grognons, j'aime pas *ça*

Dans notre corpus, nous avons trouvé des exemples problématiques comme :

Ex : j'aime pas les vendeuses généralement c'est des connes (JOLIEV~1 8,15)

Comment pronominaliser *les vendeuses* ?

Doit-on utiliser le clitique *les* : *les vendeuses, je les aime pas*, ou la pro-forme *ça* : *les vendeuses, j'aime pas ça* ? Est-ce que *les vendeuses* correspond à une catégorie de personnes, ou est-ce que ce sont certaines vendeuses, particulièrement ? Dans le premier cas, on pourra pronominaliser en *ça*, alors que dans le second, on utilisera *les* :

Ex : en général, les vendeuses, j'aime pas *ça*

Ex : les vendeuses de cette boulangerie, je les aime pas

Pour trancher, une observation du contexte est souvent indispensable et, dans bien des cas, elle s'avère efficace. Nous ne présenterons pas ici tous les exemples pour lesquels la pronominalisation est délicate ; cependant, il semblerait que l'utilisation de telle ou telle pro-forme est, dans de pareils cas, conditionnée par la sémantique du verbe étudié. Il s'agit donc de déduire du contexte le type de pro-forme adéquate pour montrer la différence sémantique entre deux constructions similaires d'un même verbe.

2.4. Cadre méthodologique utilisé

La méthode dont nous nous sommes inspirée est celle de l'Approche Pronominale, qui se base sur la description de corpus de français parlé et en tire des observations sur l'usage de la langue orale.

2.4.1. L'Approche Pronominale

L'Approche Pronominale apparaît dans les années 1970. Elaborée par Karel van den Eynde et Claire Blanche-Benveniste, elle a pour but de permettre l'élaboration d'une grammaire descriptive complète du français.

Dans l'article de K. van den Eynde et de P. Mertens (2002 : p. 3-4), on trouve une description sommaire mais efficace de l'A.P. :

L'approche pronominale [...] est une méthode de description syntaxique qui préconise l'étude systématique de la valence à partir de la relation de *proportionnalité* qui existent entre les dépendants lexicalisés d'une part, et d'autre part, les éléments de référence minimale que sont les *pronoms* (ceux de la grammaire scolaire ainsi que certains éléments de statut comparable, d'habitude considérés comme des adverbes : 'là, alors, où, quand, comment, pourquoi...'). Tout en étant des dépendants de référence minimale, les pronoms permettent de former avec le prédicateur des énoncés complets. Contrairement aux entités syntagmatiques (GN, GV,...), aux fonctions syntaxiques (SUJET, OBJET,...), ou aux rôles (AGENT, PATIENT,...), les pronoms sont des éléments appartenant à la langue même et dès lors directement contrôlables : on peut porter un jugement de grammaticalité sur les énoncés qui en comportent. Mais surtout, l'inventaire des pronoms est limité et fermé. Ces observations sont à l'origine du principe de base de l'AP : dans l'établissement de la valence d'un prédicateur, on exploite cette proportionnalité pour ramener le nombre énorme de combinaisons entre éléments lexicaux à un nombre limité de combinaisons entre pronoms. Autrement dit, le nombre restreint de pronoms permet de vérifier de façon systématique et exhaustive leurs combinaisons avec les prédicateurs, sans recourir à l'interprétation de certaines propriétés (les traits sémantiques utilisés en syntaxe) posées par le linguiste. La possibilité d'employer tel ou tel pronom est en effet significative : les pronoms révèlent des propriétés fondamentales qu'un prédicateur impose à ses dépendants.

S'inspirant de l'analyse des langues bantoues, dans lesquelles le verbe répète ses actants sous formes d'affixes, C. Blanche-Benveniste et K. van den Eynde posent la structure [clitiques + verbe] comme structure de base, à laquelle le lexique correspondant aux traits des

paradigmes pronominaux est proportionnel, démontrant ainsi ce que Raymond Queneau indiquait dans « Connaissez-vous le chinook ? » (1965 : pp. 55-57) :

Les linguistes ont, depuis longtemps, découvert que le français parlé contemporain présentait quelque ressemblance avec certaines langues très éloignées par leur structure des langues indo-européennes. Ainsi, en chinook, la phrase est construite de telle sorte qu'une première partie contient toutes les indications grammaticales (c'est-à-dire les « morphèmes ») et la seconde toutes les données concrètes (les « sémantèmes »). Cette construction de la phrase ne ressemble en rien à celle du français écrit mais, ainsi que l'a fait remarquer Vendryes, elle est fréquente en français parlé. Pour reprendre ses exemples, on ne dit pas par exemple : « *Ta cousine n'a pas encore voyagé en Afrique* », mais : « *Elle n'y a pas encore voyagé, ta cousine, en Afrique* » ; on commence par énoncer les signes grammaticaux abstraits, le « résumé algébrique de la pensée », puis on emplit cette forme vide avec des désignations de choses et de faits précis.

Pour exposer de manière plus détaillée les principes de cette approche, nous rapportons en partie l'article¹ que Claire Blanche-Benveniste et Karel van den Eynde ont proposé dans les *Cahiers de lexicologie* en 1978, qui préfigurait l'ouvrage fondateur de cette approche, *Pronom et syntaxe*², et qui s'appuyait sur la thèse de Claire Blanche-Benveniste³ :

ESSAI DE SYNTHESE

1. L'unité première

L'unité syntaxique première est constituée par une relation établie entre un élément constructeur, le verbe, et les termes qu'il construit. Le constructeur de relations, la relation et les termes ne sont pas nécessairement représentés dans l'énoncé par des items distincts observables directement.

2. Mécanisme de détermination de l'unité première

Les termes de la relation sont analysables en *traits*, unités premières de marques relationnelles et de sémantisme. La composition des termes peut être modifiée par changement d'un trait, ajout d'un trait, suppression d'un trait. Ceci est valable pour tous les termes, y compris le verbe. Les éléments concrets de l'énoncé qui fournissent la meilleure approximation des termes sont les pronoms.

¹ EYNDE, Karel Van den & BLANCHE-BENVENISTE, C., 1978, « Syntaxe et mécanismes descriptifs: présentation de l'approche pronominale », in *Cahiers de Lexicologie*, 32, 12-14.

² BLANCHE-BENVENISTE C., DEULOFEU, J., STEFANINI, J. et VAN DEN EYNDE, K., 1984 (1987 : 2^{ème} édition augmentée), *Pronom et syntaxe, l'approche pronominale et son application au français*, Paris, Sela.

³ BLANCHE-BENVENISTE C., 1975, *Recherche en vue d'une théorie de la grammaire française. Essai d'application à la syntaxe des pronoms*, Paris, Champion

Plusieurs formulations sont possibles pour une même relation entre verbe et termes ; chacune de ces formulations se caractérise par rapport aux autres par une combinatoire de différence dans la composition des termes.

L'ensemble des formulations de la relation entre verbe et termes forme un *groupe de formulations*. Un verbe est défini par son groupe de formulations, et la composition de ses termes dans toutes ses formulations.

On appellera *bloc verbal nucléaire* les différentes réalisations discursives de ces formulations. Le noyau verbal entre en tant que tel en relation avec des éléments discursifs. Dans la mesure où ces éléments de relation ne sont pas caractéristiques des groupes de formulation du verbe, on ne les considérera pas comme termes mais comme éléments non-nucléaires, ce qui semble être le cas pour des éléments comme « quand » en français.

3. Les niveaux de représentation de l'unité première

La représentation des termes se fait à différents niveaux, dans différentes catégories grammaticales de la langue. Entre les diverses représentations, à plusieurs niveaux, d'un même terme de la relation existe une relation stable dite de proportionnalité.

Sont considérés comme termes de la relation verbale uniquement les éléments qui entrent dans cette proportionnalité, c'est-à-dire ceux qui sont représentés dans au moins deux catégories. Les éléments qui ne sont représentés que par une catégorie ne sont pas considérés comme termes. On peut citer comme exemples d'éléments « hors proportion » pour le français :

C'est lui – il y en a – je lui en veux – qui vive – la nuit tous les chats sont gris –

On ne peut préjuger des catégories de proportionnalité dans les différentes langues. Pour le français on pourrait distinguer les catégories suivantes : les clitiques ; les pronoms terminaux (moi, toi, quelqu'un, rien...) ; les pronoms lexicalisables (lui, elle, ainsi, là, tel...) ; les semi-lexicaux (celui-, celle-...) ; les formes en qu- ; les constituants lexicaux ; le noyau verbal à sujet, traité comme terme ; le noyau verbal sans sujet représenté, traité comme terme.

Chaque catégorie a ses règles de réalisation (en français il faudrait des règles distinctes pour la place des clitiques, des constituants lexicaux, des formes en qu-, etc.).

La représentation d'un terme peut être limitée à certaines catégories (par exemple un terme pourrait ne pas être représentable sous forme de clitique), ou peut exiger une représentation par deux catégories simultanément.

4. La proportionnalité, lien entre les différents niveaux de représentation

Comme les représentations des termes en catégories sont reliées par la proportionnalité, les relations posées pour une catégorie sont, de ce fait, posées pour toutes les autres catégories de représentation, et pour tous les types de formulation appartenant au groupe de formulations. La grammaire spécifiera par des marques et des règles de passage entre catégories les différentes réalisations.

Pratiquement, on a intérêt à porter les marques sur la catégorie qui représente le corpus le plus restreint (en français, les clitiques), et à calculer les règles de passage à partir de cette catégorie. Les relations et les groupes de formulations dégagés à partir des clitiques seront garantis pour toutes les catégories de représentation, en particulier pour les constituants lexicaux.

5. Les opérations, reformulations à caractère généralisable

Dans l'ensemble des formulations, on constate qu'il y a certaines possibilités de constructions qui ne sont pas le propre d'une catégorie de verbes et qui ne sont pas affectées par les reformulations ; ce serait le cas par exemple de la négation en français. On a intérêt à ne pas les considérer comme faisant partie du groupe de formulations, à ne pas en faire un terme de la relation, vu leur caractère de très haute généralité et le fait que leur présence ou leur absence n'affecte en rien la composition des autres termes. Nous les appellerons des *opérations*. Pour le français on pourrait poser des opérations de négation, d'interrogation sur un terme, de restriction en *ne...que...*, de binarisation (lui, je connais), de redondance (lui, je le vois), d'interrogation en « est-ce que », d'extraction (c'est lui que je vois).

On doit faire état de contraintes portant sur ces opérations :

- Certaines ne peuvent porter que sur un verbe nucléaire, telle l'interrogation en *est-ce que* et la binarisation (*je veux est-ce qu'il vienne, *je souhaite lui connaître).
- Certaines ont des limites de catégories : elles ne peuvent porter que sur des termes réalisés selon des catégories données ; par exemple la restriction en *ne...que* exclut la catégorie des clitiques, des formes en *qu-* et des noyaux verbaux à sujet.
- D'autres exigent une représentation en deux catégories au moins, par exemple la redondance « *lui, je le vois* ».
- Certaines opérations sont compatibles entre elles, d'autres non, par exemple la redondance entre clitique et élément d'une autre catégorie, et l'opération de restriction en *ne...que*, « *je ne le vois que lui »

6. L'unité première et le discours

Les unités syntaxiques premières, accompagnées de leurs éléments non-nucléaires éventuels, sont liées entre elles, dans les énoncés, de diverses façons : successions, coordination, subordination, etc. Ces diverses liaisons sont soumises à des règles de linéarisation particulières. L'unité de phrase ne sera décernée qu'à ce niveau.

C'est au niveau des liaisons que se manifeste le phénomène de « possibilité de co-référence ».

Les liaisons sont organisées en « phases » d'énoncés, qui sont les unités d'analyse du discours. La phase peut avoir des démarcatifs d'ordre divers, tels que différence de pause, d'intonation, insertion de phatiques... L'analyse de la co-référence entre français montrera que la rupture dans les marques de co-référence va de pair avec une démarcation en phases.

2.4.2. Ce que l'on retient de l'Approche Pronominale

- Le choix du verbe

Comme le soulignent les auteurs de *Pronom et syntaxe* (1987 : p. 24), on a le choix, lorsqu'on élabore une syntaxe, entre deux notions de départ : le verbe ou la phrase. Nous avons choisi de retenir le verbe pour plusieurs raisons.

La première est liée à la nature même de cette recherche. Etudier les constructions verbales suppose que l'on s'appuie sur l'élément constructeur, et que l'on observe les éléments qu'il construit. C'est donc au niveau du verbe, et non de la phrase, que s'articule nécessairement la description.

Les autres raisons qui ont déterminé ce choix sont celles données dans l'ouvrage sus-cité (pp. 24-25). Choisir le verbe permet :

- de partir pratiquement de la réalité syntaxique la plus simple, celle qui est constituée par le verbe doté de ce qu'il lui faut comme éléments adjacents pour faire un énoncé [...].

- de situer notre analyse dans le prolongement de l'analyse classique distributionnelle (Z. Harris, 1951) : après l'unité « morphème », on dégage l'unité de construction qui est le verbe construit, et non la phrase.

- d'atteindre le discours plus facilement, en installant une relation qui va du verbe au discours, sans passer par la phrase. [...] Nous nous éloignons donc du type d'analyse qui se fonde sur l'unité phrastique, elle-même justifiée souvent par le rapport prédicatif entre deux termes

En outre, comme nous l'avons vu dans la partie consacrée à l'extension de la notion de construction verbale, nous étudierons chaque verbe constructeur, même si la construction verbale dont il est le noyau est elle-même régie par un autre verbe.

- Pronominalisation et relation de proportionnalité

On pose qu'un élément construit possède une équivalence avec une pro-forme, et entretient avec elle une relation de proportionnalité. Par exemple, la pro-forme clitique *le* est considérée potentiellement proportionnelle, pour un verbe donné, à un SN +/- personnel, ou à un S_{vinf} introduit par une préposition, ou encore à une Que-phrased, comme le montrent les exemples suivants :

Je le vois, cet homme (où *le* = SN +personnel)

Je le vois, ce livre (où *le* = SN -personnel)

Je l'envisage sérieusement, de partir avec lui au bout du monde (où *l'* = de + S_{vinf}Ω)

Je le sais, que tu ne viendras pas ce soir (où *le* = Que-Pind)

Ainsi, plusieurs « réalisations lexicales » sont envisageables selon les pro-formes utilisées pour représenter les différents types de constructions relevés dans le corpus. Dans le cas où, pour un même verbe, nous avons relevé dans le corpus des constructions dont la représentation pronominale est identique, mais dont la « réalisation lexicale » était différente, les réalisations trouvées dans le corpus apparaissent en fin de ligne pour la construction étudiée. Par exemple, *j'y vais, à Lyon* et *j'y vais, faire les courses* se distingueront dans la base par les colonnes [à + SN-p] pour le premier exemple, et [S_{vinf}Ω] pour le deuxième.

2.5. Présentation des données

2.5.1. Verbes présents dans le corpus

Dans ce corpus, on trouve 352 verbes morphologiques, possédant chacun une ou plusieurs constructions. Nous distinguons différents types de constructions en fonction des propriétés syntaxiques et distributionnelles des divers éléments construits par le verbe. Par exemple, le verbe *aller* possède cinq catégories principales de constructions, selon qu'il régit

un syntagme verbal à l'infinitif (*je vais faire les courses*), un syntagme prépositionnel ou adjectival (*je vais au marché / ce pull me va bien*), ou selon les traits de ses sujets (*je vais bien / ça va bien*). Certaines d'entre elles sont sous-catégorisées de manière à préciser un ou plusieurs de leurs traits (la construction notée 1-[abc], comprend en 1a le trait +personnel du sujet, en 1b, le trait –personnel du sujet et un syntagme verbal à l'infinitif avec ses propres compléments (ce que Maurice Gross notait SvinfΩ) et en 1c le trait –personnel du sujet et un syntagme verbal à l'infinitif sans compléments (Svinf)).

2.5.2. Constructions verbales traitées

2.5.2.1. Inventaire des constructions verbales : les formes retenues

La requête [#v] donne 15 749 résultats, dont 1 900 verbes à l'infinitif que nous n'étudions pas dans ce travail. Il reste donc 13 849 constructions, parmi lesquelles on trouve des « déchets », c'est-à-dire des lexèmes reconnus comme des formes verbales par Xcor2000, mais qui n'en sont pas : *bouffe*, *communions*, *plus*, etc. C'est un des problèmes que rencontrent les informaticiens-linguistes lorsqu'il s'agit de différencier deux formes homographes qui ne sont pas de même nature dans un corpus qui n'est pas étiqueté morpho-syntaxiquement (« taggé »). On peut compter entre 10 et 15 % de déchets. Nous ne les avons pas comptés systématiquement, ce qui explique cette approximation.

Le nombre total de constructions restantes est donc d'environ 12 000, duquel il faut encore déduire le nombre des formes au participe présent, que nous n'avons pas pris en compte, et qui s'élève à 236. Il reste alors un peu moins de 12 000 constructions verbales à traiter.

Parmi ces constructions, nous n'avons pas retenu les verbes qui apparaissaient dans des « chaînes d'hésitations » du type *il achète il achète le journal*. Dans ce type de constructions, nous n'avons retenu que la deuxième partie de la chaîne, à savoir *il achète le journal*, sans prendre en compte la première partie. Nous n'avons pas compté systématiquement le nombre de verbes apparaissant en premier dans une chaîne d'hésitations, cependant, les verbes *être* et *avoir* à eux seuls apparaissent 600 fois, ce qui nous laisse penser que le nombre total de verbes apparaissant dans un tel contexte est assez important.

D'autre part, le nombre total de constructions relevées par XCOR donne comme verbe tous les éléments d'une chaîne verbale. Par exemple, dans l'exemple suivant : *on a été*

nommé par le Ministère (PETI 15, 11), XCOR relève trois verbes (*avoir*, *être* et *nommer*) alors que nous n'en comptons qu'un seul (*nommer*). En somme, la requête #v sous XCOR propose un comptage morphologique, compte et relève toutes les formes apparentées morphologiquement à un verbe, alors que nous ne retenons dans cet inventaire et ce classement uniquement les verbes constructeurs, c'est-à-dire que nous opérons un comptage syntaxique. Nous n'avons pas compté les chaînes verbales à deux, trois ou quatre éléments ; aussi, il nous est difficile d'évaluer le nombre de verbes apparaissant comme non-constructeurs.

Nous n'avons donc pas retenu les auxiliaires et les modaux, qui ne sont pas, par définition, constructeurs. Le verbe *aller* modal est le seul à apparaître dans la base de données et dans le corpus, afin de montrer les ambiguïtés qui existent parfois entre *aller* constructeur et *aller* modal. Ces ambiguïtés se retrouvent souvent avec les autres verbes du même type, par exemple *devoir* dans : *Jean doit prendre sa douche*, qui peut signifier aussi bien l'obligation (*devoir* est alors constructeur, on peut pronominaliser *prendre sa douche* en *le*, il peut porter les modalités du verbe et être focalisé) que la probabilité (*devoir* est alors modal, aucun des trois tests ne fonctionne) ; généralement, ces ambiguïtés sont levées par le contexte.

Les verbes *être* et *avoir*, très productifs, puisqu'il représentent 2 795 occurrences pour le premier, et 2 134 pour le second, n'ont pas été triés manuellement, mais automatiquement. Il doit donc subsister quelques erreurs et quelques manques dans le corpus. A cette approximation près, les auxiliaires *être* et *avoir* représentent environ 2 300 occurrences, et les verbes constructeurs *être* et *avoir* 2 080. Il faut noter aussi que l'auxiliaire *être* a parfois une construction très proche du verbe *être* construit avec un adjectif (cf. Gaatone, 1998 : pp. 52-54).

Ces précisions apportées, voici, pour le corpus des constructions verbales traitées, les pourcentages des représentations actif/passif : parmi les constructions verbales retenues (une fois éliminées celles à l'infinitif et les formes en *-ant*), les CV actives représentent 96,8 % de la totalité des CV, et les CV passives 2 %. Nous avons calculé le pourcentage des CV pronominales passives à part : elles occupent 1,2 % du total des CV. A l'intérieur des formes actives, nous avons intégré les verbes pronominaux réfléchis et réciproques, qui forment respectivement quant à eux 2,8 % et 0,4 % du total des CV étudiées.

Ainsi, pour un total d'un peu moins de 7 000 constructions verbales analysées, nous avons 6 692 CV à la voix active « pure », 197 pronominales réfléchies, 27 pronominales réciproques, 78 pronominales passives et 145 passives.

Nous avons choisi, dans le cadre de cette étude, de comptabiliser le nombre de constructions verbales pour chaque verbe et pour chaque type de construction si le verbe en possède plusieurs. Cette étape présente l'avantage de donner une indication sur la fréquence d'emploi de tel ou tel type de construction par rapport à un autre, dans le cadre de notre corpus. Il faut cependant préciser que la thématization du corpus, et le fait que ce dernier soit limité ne donnent pas d'indication absolue sur la fréquence d'emploi des constructions. Il faut donc observer les chiffres avec prudence, en tenant compte du fait qu'un autre corpus, thématized de manière différente, ou plus important en taille, donnerait sans doute des résultats différents. Cette précaution d'emploi indiquée, observons à présent ce que nous signifions en employant le terme de « fréquence ».

2.5.2.2. La prise en compte de la fréquence

Comme nous l'avons déjà mentionné, plus un verbe est fréquent, plus il a de chances de posséder des types de constructions différents. Cette caractéristique nous a paru fondamentale pour retenir l'étude de la fréquence. Ainsi, nous avons comptabilisé le nombre d'occurrences de chaque type de construction pour chaque verbe, et l'avons indiqué dans le corpus et dans la base de données.

Tous les verbes apparaissant dans le corpus et dans la base de données sont entrés avec un ou deux nombres. Si le verbe n'a qu'un nombre qui lui est accolé, cela signifie que nous n'avons relevé qu'un type de construction dans notre corpus (ce qui ne veut pas dire qu'il n'en existe pas d'autres, *a priori*) et le nombre qui apparaît alors est celui des occurrences de ce verbe trouvées dans le corpus. Si le verbe est suivi de deux nombres, comme par exemple *appeler 1-23*, cela signifie que *appeler* possède au moins deux types de constructions différents, et que le premier de ces types apparaît 23 fois dans notre corpus.

Prendre en compte la fréquence présente quelques avantages :

- pour l'enseignement des langues, cela permet de trier les verbes de manière décroissante, pour apprendre en premier lieu le fonctionnement des verbes les plus

fréquents, puis, au fur et à mesure des progrès de l'apprenant, celui des verbes moins fréquents ; la notion de fréquence reste bien entendu limitée et valable uniquement pour le corpus utilisé (dans un autre corpus, thématisé de manière différente, ou encore faisant intervenir des « genres de discours » différents, la fréquence d'un verbe donné pourrait se trouver radicalement modifiée) ;

- pour le TAL, cela permet de prévoir, pour un verbe donné, quels sont les types de schémas fréquemment utilisés et quels sont ceux qui apparaissent de manière exceptionnelle, afin d'établir une hiérarchie des règles opératoires.

2.6. L'inventaire et le classement des constructions verbales

2.6.1. Présentation du concordancier Xcor¹

Xcor est un concordancier développé « pour exploiter par informatique le corpus de langue parlée du GARS » (Adam, J.-P., 2000 : 121). Jean-Pierre Adam nous a très gentiment laissé utiliser son programme. Nous allons ici en donner une présentation sommaire, sans proposer de comparaison avec d'autres concordanciers existants. En effet, notre objectif ici n'était pas d'évaluer l'efficacité d'un programme, mais de l'utiliser au mieux afin d'automatiser le plus possible le classement des constructions verbales.

Les principes de fonctionnement de Xcor ont été exposés par J.-P. Adam (*op. cit.*) :

On cherche à trouver tous les segments de texte s'appariant avec un modèle vu comme une suite de filtres élémentaires s'appliquant sur les mots du segment. Chaque filtre élémentaire définit une condition qui doit être vérifiée par les mots du segment. Conceptuellement, ce modèle est appliqué sur l'ensemble des groupes de mots du texte, et les segments qui « passent » au travers du filtre sont notés, avec leur contexte. Pratiquement, le mécanisme d'indexation accélère grandement cette recherche. Le système de base opère au niveau des signes graphiques (mots, amorces, codes de transcription) tels qu'ils résultent du découpage initial. Le corpus est totalement indexé sur ces signes, permettant un accès très rapide. A ce niveau, nous n'avons introduit ni découpage morphologique, et encore moins de « marquage morpho-syntaxique » (tagging) de tout le corpus. Nous avons adopté une démarche pragmatique en plusieurs niveaux successifs :

1. Une requête de base est une séquence de filtres sur les formes graphiques découpées du corpus

2. Au niveau suivant, un dictionnaire (facultatif) attribue des « catégories » aux « mots », selon le schéma :

Forme ; Catégorie ; Lemme ; Marqueur

Et on peut interroger sur les formes ayant certaines des caractéristiques définies par « Marqueur ». Exemple : #v_futur (verbes au futur).

¹ Jean-Pierre Adam, concepteur de Xcor, s'est montré d'une gentillesse et d'une patience infinie. Qu'il en soit ici vivement remercié

INVENTAIRE ET CLASSEMENT

3. Il n'y a pas d'analyse arborescente complexe qui porterait sur des segments « fermés » du corpus. Il n'y a d'ailleurs pas de « phrases », pas de ponctuation dans le corpus [...]. Une grammaire d'analyse robuste sur ce type de corpus reste à écrire, si elle est jugée nécessaire.

4. Par contre, on peut isoler des portions de texte par des filtres encadrant des espaces non analysés qu'on peut récupérer en vue d'une analyse locale ultérieure. On peut envisager d'écrire des grammaires partielles permettant l'analyse de ces segments, plus complexe que ce qui est permis par le système de requêtes de Xcor. [...]

Une **requête** est une combinaison de **filtres élémentaires** que l'on applique sur les mots du corpus, ne laissant passer que les séquences de mots vérifiant une condition. Par exemple la requête de deux éléments

#v(faire) #v

sélectionne toutes les séquences du corpus constituées d'une forme conjuguée ou infinitive du verbe "faire", suivi d'une forme conjuguée d'un verbe quelconque. Il ne s'agit pas d'une recherche séquentielle sur les milliers ou millions de mots du corpus. Le corpus est complètement indexé. La recherche utilise cet index pour pointer directement sur les mots, et tient compte des fréquences d'occurrence de ceux ci. Tout se passe en mémoire ou sur disque, en temps réel, et n'y a pas de tri.

Filtres :

Type de filtre	Syntaxe	Exemple
mot complet	<mot>	Famille
mot tronqué	<*chaîne> ou <chaîne*>	*issant qu* (que qui qu)
mot quelconque	<*>	*
mot commençant par une majuscule	<.maj>	
espace de 0 à n mots	<@n>	@5
espace de 1 à n mots	< <=n >	<= 4
espace d'exactly n mots	<=n>	=5
mots appartenant à une classe	<#classe> ou <#classe(lemme)>	#v #v(finir) #v_inf #v_futur #v_ppass
liste de filtres élémentaires	<filtre> <filtre>	machine * *er #v(faire) #v
disjonction d'une liste de filtres	.ou(<filtre> <filtre>)	.ou(moi toi il elle lui nous ...)

Exclusion	~<filtre>	machine à ~laver
-----------	-----------	------------------

Les réponses s'accompagnent d'un contexte gauche et droit de longueur arbitraire, et s'accumulent dans un fichier qui peut être consulté et imprimé. Chaque ligne de la réponse indique le fichier-corpus et l'emplacement dans ce fichier. L'interface graphique permet d'examiner un contexte plus large de la réponse, (qui n'est limité que par la taille du fichier).[...]

On peut définir des **dictionnaires** selon un modèle très simple. Une entrée est un quadruplet du type

Forme ; Catégorie ; Lemme ; Marqueur

Pour le lexique verbal par exemple "Forme" est une forme conjuguée de l'infinitif "Lemme" ; "Marqueur" est un code qui représente les mode, temps, personne

(par exemple idpr2 = indicatif présent 2 personne singulier). Le filtre #v(lemme) s'apparie avec toute forme conjuguée de l'infinitif "lemme", (Catégorie = v)

Ce système est très général et "lemme" est vu comme le représentant modifié par "marqueur" des occurrences de "Forme" appartenant à "Catégorie".

Les différentes requêtes que nous avons utilisées pour l'inventaire ont déjà été évoquées. La première, #v, permet d'extraire du corpus tous les verbes. Xcor n'étant pas taggé, certains mots sortants comme verbe avec cette requête n'en sont pas. Nous avons trouvé, par exemple, de très nombreuses occurrences de *plus*, reconnu par le logiciel comme participe passé de *plaire*, mais qui en réalité est dans le corpus un quantifieur.

Nous avons effectué la seconde requête pour récupérer toutes les formes d'un même verbe, la première requête n'étant pas satisfaisante pour cela. Ainsi, à partir de la liste des verbes données par Xcor à la fin de la première requête, nous avons pour chacun d'entre eux demandé toutes leurs formes, à l'aide de la requête #(infinitif), ou *infinitif* représente le verbe à l'infinitif dont nous demandons au logiciel d'extraire les occurrences.

Une fois cette longue série de requêtes lancée, nous avons, pour chacun des verbes trouvésn effectués un classement plus fin en fonction des différents types de compléments que nous trouvions. Ce classement s'est fait manuellement, dans la plupart des cas.

2.6.2. La base de données

Nous avons utilisé, pour présenter le classement des constructions verbales de notre corpus, le logiciel de traitement de base de données Access. Les colonnes représentent :

- les traits des sujets,
- le verbe constructeur,
- la voix du verbe,
- son « degré » de pronominalité s'il en a un,
- les paradigmes de compléments régis,
- leur extension lexicale, lorsque la pro-forme représente un Sprep (on lit alors quelle préposition est utilisée, suivie de quel type de syntagme (nominal ou verbal à l'infinitif), ou une Que-phrased (au subjonctif ou à l'indicatif), ou encore un syntagme quelconque.

Nous développerons et expliciterons ces différentes colonnes dans la partie [propriétés syntaxiques et distributionnelles](#).

Dans cette base de données, un peu plus de 600 enregistrements apparaissent, qui synthétisent les différentes constructions des verbes de notre corpus. Une présentation en tableau fixe nous est apparue réductrice. En effet, l'un des intérêts de ce logiciel est de pouvoir questionner la base, pour chercher tel ou tel type de verbes ayant telle ou telle propriété, ou telle ou telle distribution. Le classement pourra apparaître relativement sommaire, le nombre de propriétés étant assez faible ; cependant, le corpus des constructions verbales étant ici-même présenté de manière intégrale, un utilisateur soucieux de plus de précision pourra, s'il le souhaite, modifier la table en fonction des propriétés qu'il souhaite voir apparaître. C'est donc une base de données dynamique qu'il faut considérer ici, modifiable selon les besoins de ses utilisateurs, de la même manière que le corpus que nous avons présenté dans la première partie de ce travail.

Access a une réputation de logiciel « compliqué » auprès des personnes que l'informatique rebute un peu. Il est vrai que ses possibilités sont immenses, tout comme le

logiciel de traitement de texte du même éditeur. Cependant, nous n'avons utilisé que des objets sommaires, facilement manipulables, et nous avons regroupé toutes les constructions inventoriées en une seule table, à partir de laquelle peuvent être effectuées des requêtes par formulaires, permettant de trouver tous les verbes répondant aux propriétés sélectionnées.

2.6.2.1. Les propriétés syntaxiques et distributionnelles

La notion de propriété est fondamentale pour ce travail. La plupart des linguistes l'utilise, en y attachant une importance plus ou moins grande, selon le modèle théorique qu'ils suivent. Par exemple, Boons, Guillet et Leclère (1976 : pp. 169-183) présentent de manière très détaillée les deux types de propriétés qu'ils utilisent pour mener à bien leur Lexique-grammaire des phrases simples en français :

La méthode est d'inclure le verbe à examiner dans un nombre aussi grand que possible de phrases différentes, et de déterminer par jugement d'acceptabilité quelles sont les combinaisons acceptées ou refusées. Dans tous les cas, cela revient à étudier les compatibilités d'un verbe avec un maximum d'éléments différents. Dans le cadre de la phrase simple – c'est-à-dire à un seul verbe – les éléments susceptibles de varier sont de deux natures : les *éléments lexicaux* et les *structures*.

On aura alors au moins deux types de propriétés, associés chacun à un type de variable. On appellera les premières *propriétés de distribution* et les secondes *propriétés de structure*. La combinaison de ces deux niveaux fournit des propriétés complexes dont le but est de vérifier si un verbe accepte une distribution donnée dans une construction donnée, et vice-versa. [...]

[Les propriétés de distribution] ont pour but d'étudier les distributions acceptées par un verbe, et s'appliquent aux éléments qui lui sont attachés. [...] On éprouvera les propriétés dans une seule structure à la fois : la distribution acceptée par un verbe dans une structure donnée peut être *a priori* différente de celle acceptée par le même verbe dans toute autre de ses structures. [...] La procédure consistera alors à déterminer dans un premier temps quels types de structures admet un verbe donné, puis à établir pour chacune d'elle la distribution acceptée. [...].

Dans ce travail, nous avons utilisé aussi ces deux types de propriétés, qui ne se trouvent cependant pas sous la même dénomination.

Les propriétés distributionnelles sont représentées dans la table par les colonnes représentant les éléments régis sous la forme pronominale (les quatre colonnes des sujets (+/- pers, ça, il asubjectal) et les douze colonnes des compléments (P1 +/- pers, P2 +/- pers, P3 +/- pers, ainsi, cela, où, combien, pourquoi, quand). Elles permettent de la même manière

d'observer les éléments qui entrent dans la distribution d'un verbe dans une construction donnée. On observe aussi que la distribution acceptée par un verbe dans une construction donnée peut être différente de celle acceptée par un autre type de construction du même verbe. C'est précisément ce qui permet de différencier plusieurs *formulations* d'un même verbe. Pour une description plus détaillée de ces éléments, on se reportera aux parties [2.3.2.2.](#) et [2.3.2.3.](#)

Les propriétés de « structures » sont les dix colonnes qui apparaissent dans la base de données à la droite de la colonne du verbe constructeur. Il s'agit :

- du fait que certains verbes ont des constructions de type **causatif** ou **factitif** (Riegel et al., 1984 : pp. 229-232) :

« [Des] structures de phrases [...] peuvent être enchâssées dans la structure [N₀ + *faire* [____]] où le référent du sujet N₀ représente la cause ou l'agent du procès décrit par la phrase originale. En entrant dans cette construction dite causative ou factitive, la phrase de départ subit une double modification. Son sujet nominal est postposé à son verbe qui lui-même se met à l'infinitif : *Ce mélodrame fait* + [*Margot pleure*] → *Ce mélodrame fait [pleurer Margot]*. Ce type de constructions appartient à une classe plus vaste de constructions infinitives qui comprend les verbes *faire, laisser, sentir, voir, entendre*. »

Nous avons dû ajouter à cette liste le verbe *regarder* qui entrait dans le même type de construction :

Ex. 1 : je regardais les gens passer (MUMON~44 8,3)

- du fait que certains verbes, à côté de leur sens « plein », se « combinent avec un nom, un adjectif ou un groupe prépositionnel pour construire une forme complexe fonctionnellement équivalente à un verbe » (Riegel et al., 1984 : p. 232). Ces verbes sont nommés « verbes supports » :

Ex. 2 : moi j'ai peur que ça se dégrade (MADMON 7,4)

La pronominalisation montre immédiatement que l'on a pas affaire à un verbe « normal ». En effet, si l'on pronominalise l'exemple ci-dessus, on ne peut pas réduire l'élément *peur* : j'ai peur que ça se dégrade >* je l'ai. En revanche, on peut pronominaliser la

que-phrase : j'ai peur *que ça se dégrade* > j'*en* ai peur, ce qui montre ainsi qu'*avoir peur* fonctionne comme une forme verbale complexe.

Remarque : dans la locution verbale *avoir l'air*, utilisée assez fréquemment à l'oral, beaucoup de locuteurs attachés à l'idée de norme reprennent leurs interlocuteurs si ces derniers ont le malheur d'accorder l'adjectif suivant *air* (*malin, heureux, etc.*) avec le sujet et non pas avec *air*. Ainsi, pour eux, l'adjectif est perçu comme étant épithète du substantif *air*, et devrait donc s'accorder au masculin singulier. Or, on observe à l'aide de la pronominalisation que nous avons bien ici une locution verbale complexe : j'ai l'air *malin* > *je l'ai > j'ai l'air *ainsi*, où *ainsi* pronominalise l'adjectif.

- de la mention d'un **régionalisme** :

Le corpus étant « récolté » dans l'Allier, quelques régionalismes syntaxiques apparaissent, notamment le pronom neutre *y* employé à la place du pronom *le* ou *ça* (Tuaillon, 1983b), comme dans les exemples suivants :

Ex. 3 : la bohème euh moi j'*y* trouve bien (MADMO~34 13,3)

Ex. 4 : un sirop après tu mélanges pas la pulpe hop tu *y* mets dans ton machin et puis hop c'est bon (CHOCM~10 11,6)

Cependant, ces cas restent très rares, et sont souvent issus d'un même sous-corpus.

- de la propriété qu'ont certains verbes d'être **recteurs faibles**, c'est-à-dire « moins constructeurs », comme c'est le cas pour les verbes en incise, par exemple (Blanche-Benveniste, 1989) :

Ex. 5 : enfin quand **je dis** capacités intellectuelles c'est pour é- éviter d'utiliser le mot intelligence (PAPMON~1 14,5)

Ex. 6 : on est assez euh **je pense** aimable (MONOV~42 12,12)

Ces verbes sont caractérisés par l'impossibilité de remplacer le pseudo-complément par un clitique (?*quand je le dis* / **je le pense on est assez aimable*). De plus, lorsque le verbe est en incise, il est bloqué dans la modalité assertive (**quand je ne dis pas capacités intellectuelles* / **on est assez je ne pense pas aimables*). Ces verbes ne sont donc pas

constructeurs au même titre qu'un verbe qui répond aux tests mentionnés ci-dessus. Nous les avons toutefois inclus dans la base, pour permettre, en s'appuyant sur les exemples du corpus, de montrer les différences qui existent entre recteurs forts et recteurs faibles, dans l'éventualité d'une application future, qu'il s'agisse d'enseignement ou de lexicographie.

- de la propriété qu'ont quelques verbes de régir des **éléments de valence solidaires** (Blanche-Benveniste *et al.*, 1984 : pp. 64-67 et Blanche-Benveniste *et al.*, 1990 : pp. 46-47),

- que l'un de ces éléments soit **pivot**, comme dans l'exemple suivant, où l'on a deux termes de valence pour le verbe recteur *donner* (un P2 de type *nous* et un P1 de type *le*), dont le P2 du verbe recteur est aussi le P0 du verbe régi *manger*

Ex. 7 : c'est ici **ce** qu'on **nous** donne à manger (MALLM~36 18,16)

> on **nous le** donne à **manger**

- ou que l'on trouve une **solidarité entre deux termes**, comme c'est le cas dans l'exemple suivant, entre *le* et *mal*, où l'on ne pourrait pas trouver *le* sans *mal* :

Ex. 8 : ça **le** fout **mal** qu'elle soit pas là (MAMMO~38 9,7)

> *ça le fout

- de la **voix (active / passive)** :

Ex. 9 : je l'achèterais pas inconsciemment – parce que même en grande surface et je suis pas la seule je prends je regarde (COIFMO~1 13,12)

Dans cet exemple, le verbe est à la voix active.

Dans l'exemple suivant, le verbe est à la voix passive :

Ex. 10 : le fils de mon fils qui a été abandonné euh aussi bien par sa mère quand il était petit que par son père quand il était adolescent (FRAMO~18 15,8)

On se fonde ici sur la définition du passif que donne Gaatone (1998 : p. 27) : « est dit passif tout participe passé dont le support n'est pas le premier argument de son lexème verbal, et est raccordable à ce support par *être*, indépendamment du temps-aspect ».

En 1993, Gaston Gross, dans la présentation des articles de la revue *Langages* n°109, consacrée au passif, indique que « le passif a fait l'objet ces dernières années d'une certaine désaffection après son succès des débuts de la grammaire générative où son caractère de généralité l'a fait considérer comme la transformation par excellence. La prédominance de considérations strictement syntaxiques élaborées à partir d'un nombre d'observations limité a donné l'illusion de cette généralité. Ce n'est pas que l'on ait passé sous silence les exceptions, on n'était simplement pas en mesure de rendre compte de leur diversité. Les travaux qui ont été effectués depuis, qu'ils considèrent le passif comme une transformation ou qu'ils postulent qu'une réduction actantielle par effacement du sujet, ont comme point commun d'accorder une importance croissante au lexique et de mettre l'accent sur le rôle discursif du passif. Les règles syntaxiques générales, comme celle qui dit que le passif nécessite la présence d'un verbe événementiel, butent sur un grand nombre de contre-exemples. La tentation a été quelquefois de jeter le bébé avec l'eau du bain et de ne pas relier du tout actif et passif. Ce serait oublier que les cas de blocage sont beaucoup moins nombreux que ceux où la règle s'applique. Entre ces deux extrêmes, la solution consiste à établir des classes de verbes sémantiquement homogènes, à observer leur comportement au regard du passif et à trouver des régularités explicatives. »

Dans le cadre de ce travail, nous n'avons pas cherché à pousser plus loin les analyses sur le passif existantes. Le problème des diathèses verbales est très complexe ; les grammaires et autres manuels de linguistique ont encore des difficultés à le traiter. En effet, si la voix passive est traditionnellement considérée comme une catégorie du verbe, la diathèse verbale due à des formes passives affecte en revanche l'ensemble de la phrase. Pour une synthèse sur le passif, on se reportera à Riegel et al. (1994 : pp. 433-444).

Cependant, des études récentes permettent de considérer le passif sous un regard nouveau (Blanche-Benveniste, 1986 ; Gaatone, 1998 ; Le passif, 2000).

- du niveau de **pronominalité** d'un verbe. S'il est accidentellement pronominal, c'est-à-dire si la construction pronominale n'est que l'une de ses constructions possibles, il peut être :

- **pronominal réfléchi**, auquel cas le pronom de type *se* est nécessairement coréférent au sujet :

Ex. 11 : tu t'aperçois que tu manges pas tout (COIFM~12 17,6)

> * tu m'aperçois que tu manges pas tout

- **pronominal réciproque**, et dans ce cas, le pronom de type *se* « exprime une multiplicité de relations « croisées » entre les éléments d'un ensemble. [...] Aussi le sujet prend-t-il obligatoirement la forme d'un pluriel, d'un singulier collectif ou d'une conjonction de sujets » (Riegel et al., 1994 : p. 257). Dans l'exemple suivant, c'est un sujet de forme « singulier collectif » que nous observons :

Ex. 12 : on **se** causait de d'une euh d'une place à l'autre ben maintenant personne se connaît (MALLM~36 11,5)

- **pronominal passif**, et dans ce cas, on ne peut pas faire « une lecture réfléchie ou réciproque où le sujet serait l'agent du procès [...] » (Riegel et al., 1994 : p. 258). Le sujet est toujours, dans notre corpus, de type – personnel ou de type ça, mais, comme le font remarquer Riegel et al., on pourrait aussi trouver un sujet dont le référent est un être humain (*une coiffeuse expérimentée se recrute à prix d'or*), qui sera désindividualisé.

Ex. 13 : j'y avais jamais vu moi que **ça** s'était allumé comme ça de partout (MADMO~34 9,3)

- les verbes **essentiellement pronominaux** sont précédés du pronom *se*. Ils n'ont pas d'autre emploi possible. Nous n'avons trouvé qu'un seul verbe de ce type dans notre corpus. Il s'agit de *se souvenir*. C'est la raison pour laquelle nous n'avons pas créé de colonne spécialement pour ce verbe, mais avons préféré mentionner son statut directement dans le corpus, en le faisant précéder de *se* à l'entrée qui lui correspond et dans la colonne « verbe constructeur » de la base de données :

Ex. 14 : j'allais chercher mon lait pour mes enfants et **je me souviens** j'ai un souvenir j'ai traversé la route le jour où les Allemands sont = arrivés (INSTI~28 12,5)

Les éléments qui apparaissent en colonne après les compléments (c'est-à-dire après *quand*) sont les réalisations lexicales des compléments pronominalisés : il s'agit souvent de Sprep, de Svinf Ω/\emptyset , ou de Qu-Phrases. Il ne s'agit donc pas de propriétés du verbe étudié, mais des structures internes des éléments pronominalisés. Ces extensions lexicales permettent de préciser des constructions particulières (par exemple, qu'un verbe accepte dans sa distribution une Que-phrase au subjonctif, ou un Svinf Ω/\emptyset , que ces deux éléments soient ou non pronominalisables sur le même paradigme)

2.6.2.2. Présentation détaillée de quelques verbes

Certains verbes sont très productifs dans notre corpus. Ils présentent des possibilités de constructions plus importantes que ceux qui n'apparaissent que plus rarement. Au cours de cette partie, nous voulons présenter de manière détaillée le verbe *aller*, dont certaines constructions semblent relever d'une spécificité de l'oral, puis les verbes *être* et *avoir*, les plus fréquentes dans notre corpus..

1. Le verbe *aller*

1- La première catégorie comprend les constructions dans lesquelles *aller* est suivi d'un syntagme verbal à l'infinitif accompagné ou non de complément(s) et qu'il est dans ce cas :

- a. soit **recteur**, auquel cas *aller* accepte un pronom de type *y* parfois ambigu (il est locatif, tout en reprenant le Svinf Ω , dans l'exemple suivant¹), son sujet est toujours de type +personnel (dans notre corpus) (cat. « 1a » dans le tableau) :

Ex. 1 : tu vas ***faire des stages à Yssingeaux*** (CHOCM~10 9,7)

➤ j'y vais (là, pour ça)

¹ Tous les exemples attestés de cette catégorie ne comprennent pas nécessairement de locatif régi par *aller*, c'est pourquoi le complément de type *où* apparaît entre parenthèses dans le tableau p. 47, comme substitution possible de *y*. Sans locatif, le Svinf Ω/\emptyset se pronominalise en *y* : *je vais faire mon stage* > *j'y vais*, d'où l'ambiguïté relative à ce pronom lorsqu'un élément locatif est présent.

- b. soit il est utilisé avec un emploi **modal**, auquel cas la pronominalisation en *y* est impossible, comme le notait déjà Xiao-Quan Chu (1987 : p. 76). Le sujet de la CV peut être soit +personnel (type *je*, ex. 2), soit –personnel (type *ça*, ex. 3), et le Svinf peut être accompagné ou non de compléments (cat. 1a', 1b et 1c dans le tableau). Dans ce cas, *aller* n'a pas de valence propre : c'est alors le verbe à l'infinitif qui est recteur de ces éléments. Deux tests permettent de repérer ce type de constructions (C. Blanche-Benveniste et al., 1990 : pp. 199-202) : si *aller* est en emploi modal (*il va pleuvoir*), il n'est pas conjugable au passé composé (**il est allé pleuvoir*), et il n'y a pas de substitution possible sur son paradigme (**il court pleuvoir*) ; s'il ne l'est pas, il accepte ces propriétés (*il va faire ses courses* > *il est allé faire ses courses* > *il court faire ses courses*).

Ex. 2 : le vendeur il va **pouvoir aussi vous conseiller** (HOTEV~26 12,16)

- **j'y vais* (pouvoir vous conseiller) (sujet +personnel)
- **je suis allé pouvoir vous conseiller*
- **je cours pouvoir vous conseiller*

Ex. 3 : j'espère que *ça va durer* enfin je sais pas (MADMO~34 4,6)

- **ça y va* (sujet –personnel)
- **ça est allé durer*
- **ça court durer*

En dépouillant notre corpus, nous avons remarqué une régularité intéressante dans la distribution du verbe *aller* : suivi d'un infinitif, il a toujours un emploi modal lorsque cet infinitif est lui-même un modal (*pouvoir, vouloir...*), un auxiliaire d'aspect (*commencer, venir...*), un verbe construisant un adjectif¹ (*être, devenir...*), le verbe *avoir*, ou enfin un verbe de déplacement (*monter, partir, venir...*). Notons ici que pour cette catégorie 1, lorsque

¹ Verbes que la grammaire traditionnelle nomme « verbes d'état »

le trait du sujet est [-personnel], *aller* est toujours dans un emploi modal, alors que ses emplois sont répartis lorsqu'il possède un sujet de type [+personnel].

D'autre part, il semblerait que dans le cas où *aller* est suivi d'un verbe à l'infinitif (exceptés ceux cités au paragraphe précédent) et d'un locatif, il soit toujours employé dans son sens « plein », comme dans l'exemple déjà cité :

tu vas faire des stages **à Yssingeaux** > tu **y** vas faire des stages

La présence contextuelle presque systématique d'un locatif dans ce type de constructions nous conduit à penser que le français parlé a tendance à désambiguïser ces relations de dépendance en « redonnant » au verbe *aller* son sens locatif. Cependant, pour ce type de constructions, deux interprétations sont toujours en concurrence : le locatif peut être dépendant soit du verbe *aller*, soit du verbe à l'infinitif :

tu vas faire des stages **à Yssingeaux**

- tu **y** vas faire des stages (c'est dans ce cas le verbe *aller* qui régit le locatif).
- tu vas **y** faire des stages (c'est ici le verbe *faire* qui régit le locatif).

Nous avons décidé d'inclure ce type de constructions (*aller* emploi modal + Vinf) sous l'entrée du verbe *aller* car, même s'il n'est pas constructeur dans ce cas, sa distribution est très proche de celle de la catégorie 1a. Ainsi, l'entrée 1b permet de distinguer les emplois modaux des emplois « pleins » du verbe *aller*.

2- La deuxième catégorie comprend les constructions à l'intérieur desquelles *aller* régit un syntagme prépositionnel, et pour lesquelles il prend son sens « plein » de verbe locatif. Les 7 sous-catégories sont effectuées essentiellement en fonction de la présence ou non d'éléments régis se rajoutant à la structure de « base », qui est la suivante :

Ex. 4 : j'allais **chez mon boucher** (INSTI~28 3,13)

- j'**y** allais

3- La troisième catégorie est composée :

- a. des constructions dans lesquelles *aller* régit soit seulement un sujet de type *ça* (cat. 3a) :

Ex 5 : les sorbets **ça va** c'est quelque chose de relativement simple CHOCM~10 12,12

- b. des constructions dans lesquelles le verbe *aller* régit un élément de type *ça* et un élément de type *ainsi* (cat. 3b) :

Ex. 6 : le soixante-dix pour euh enrober **ça va très bien** parce que c'est très liquide CHOCM~10 16,13

➤ **ça va ainsi**

4- La quatrième catégorie comprend les constructions dans lesquelles *aller* régit un sujet de type *ça*, un élément de type *te*, et un élément de type *ainsi* (cat. 4) ;

Ex. 7 : qu'on leur disent ben vous êtes belles **ça vous va bien** vous êtes euh ça elles aiment les clientes FANMO~16 20,8

➤ **ça vous va ainsi**

5- La cinquième catégorie est composée de constructions de type *s'en aller*, qui semblent figées (cat. 5). Le sujet est de type +/- personnel :

Ex. 8 : j'y vais je prends et **je m'en vais** (COMMO~14 4,3)

➤ je m'en vais

Ex. 9 : la rue **qui s'en va au théâtre** la rue Bréchimbault (MALLM~36 9,2)

➤ Elle (-pers) s'en va

6- Enfin, la sixième catégorie comprend les constructions dans lesquelles *aller* ne régit que son sujet, et dans lesquelles il est toujours à l'impératif (cat. 5) :

Ex. 10 : comment ça s'appelle les Bootlegers les mmh **allons** bon Claude Luter Maxime Sauri La Ferrière (MUSIK~46 7,5)

Présentation du tableau :

Toutes les propriétés utilisées ne figurent pas dans ce tableau, pour une question de place. Notre base de données est constituée à l'aide du logiciel Access, qui ne permet pas de partage avec Word ; cela ne facilite évidemment pas la consultation de la base hors environnement informatique.

Nous ne faisons figurer ici que les propriétés requises pour le verbe *aller*.

Les colonnes grisées sont liées entre elles de la manière suivante : les quatre colonnes solidaires sont celles de la réalisation du syntagme prépositionnel pronominalisé en *où*. On observe par exemple que pour le deuxième type de constructions (2 [a b c d e f g]), les quatre prépositions [à, en, dans, chez] peuvent apparaître dans le corpus en tête de ce Sprep.

+p	-p	ca	Verbe ALLER	active	modal	p1 +/- pers	p2 +pers	p2 -pers	p3 -pers	ainsi	où	comment	pourquoi	quand	à	en	dans	chez	+SN +p	+SN -p	+Vinf ac	+Vinf ss
+			1a-165	+				+			(+)				(+)	(+)	(+)	(+)			+	
+			1a'-48	+	+																+	
	+	+	1b-42	+	+																+	
	+	+	1c-3	+	+																	+
+			2a-92	+				+			+				+	+	+	+	+	+		
		+	2b-2	+				+			+				+	+	+	+	+	+		
+			2c-62	+				+			+			+	+	+	+	+	+	+		
+			2d-18	+				+		+	+				+	+	+	+	+	+		
+			2e-5	+				+		+	+			+	+	+	+	+	+	+		
+			2f-19	+				+		+	+		+		+	+	+	+	+	+		
+			2g-4	+				+			+		+	+	+	+	+	+	+	+		+
		+	3a-11	+																		
		+	3b-13	+						+												
		+	4-2	+			+			+												
+	+		5--9	+		+			+													
+			6-3	+																		

Table des différentes constructions du verbe *aller*.

Pour une aide visuelle à la lecture de ce tableau, voici pour chaque construction un exemple tiré du corpus :

1-a Je vais vinfΩ (+/- élément locatif) (verbe « plein »)

CHOCM~10 9,7 mais bon sinon tu vas faire des stages à Yssingeaux

1-a' Je vais vinfΩ (emploi modal)

HOTEV~26 12,16 le vendeur **il va pouvoir aussi vous conseiller**

1-b ça va vinfΩ (emploi modal)

MUMON~44 11,12 enfin quand je vais rentrer **ça va être nickel** quoi donc voilà

1c-ça va vinf (emploi modal)

MADMO~34 4,6 j'espère que **ça va durer** enfin je sais pas

2-a j'y vais, là

INSTI~28 3,13 **j'allais chez mon boucher** il savait ce que je prenais

2-b ça y va, là

MUSIK~46 17,5 à l'occasion mais **ça allait pas bien plus loin**

2-c j'y vais quand

MADMO~34 9,12 l'exposition des guitares **j'y allais presque tous les** enfin pour dire **tous les jours**

2-d j'y vais ainsi

BRUNMO~8 13,6 **on y est = allé tous les quatre** et euh on a gagné les élections

2-e j'y vais quand ainsi

MAMMO~38 8,13 elle a gagné le voyage mais **quand tu vas en voyage c'est avec d'autres directeurs**

2-f j'y vais pourquoi

COIFM~12 9,14 **ils vont pas chez les autres pour se présenter**

2-g j'y vais quand pourquoi

COMMO~14 3, j'y vais **tous les lundis parce que ben la semaine XXX je prends je travaille de neuf heures à dix-neuf heures**

3-a Ça va

CHOCM~10 12,12 les sorbets **ça va** c'est quelque chose de relativement simple

3b Ça (ou P0-pers) va ainsi

CHOCM~10 16,13 le soixante-dix pour euh enrober **ça va très bien** parce que c'est très liquide

4- Ça te va ainsi

FANMO~16 20,8 qu'on leur disent ben vous êtes belles **ça vous va bien** vous êtes euh ça elles aiment les clientes

5- je m'en vais / Po –pers s'en va

- Je m'en vais :

COMMO~14 4,3 euh j'y vais je prends et **je m'en vais** et en fait je vais toujours dans la même grande surface

MALLM~36 9,2 la rue **qui s'en va au théâtre** la rue Bréchimbault

6- allez

MUSIK~46 7,5 comment ça s'appelle les Bootlegers les mmh **allons** bon Claude Luter Maxime Sauri La Ferrière

2. Les verbes *être* et *avoir*

Comme nous avons déjà eu l'occasion de le mentionner, les « verbes » *être* et *avoir* se répartissent leurs emplois entre auxiliaires et verbes constructeurs. Nous n'avons pas, pour ces deux cas, trié de manière très fine leurs occurrences, et n'avons présenté qu'un classement assez grossier, qui mériterait pas la suite d'être approfondi et affiné. Par exemple, concernant le verbe *avoir*, la tournure dite présentative en *il y a SN* possède plusieurs sous-catégories de constructions, selon que le SN est ou non suivi d'un SVinf Ω/\emptyset précédé de la préposition *à* ou qu'il est suivi d'une que-Phrase. De la même manière, la première catégorie du verbe *avoir* (*je l'ai*) pourrait être affinée pour les mêmes raisons. Or, dans notre corpus et notre base de données, ces catégories sont indistinctes. En effet, vu l'importance de la

fréquence d'apparition de ces formes, nous avons renoncé, pour l'instant, à les compter et à les analyser de manière très précise. Cela fera sans doute l'objet d'une recherche future.

a. Être

La requête **#(être)** donne 2 795 réponses. Après avoir retiré les occurrences dues à des hésitations, dans lesquelles le verbe *être* est répété (par exemple comme dans *c'est c'est vrai qu'il faut faire attention*) et qui sont approximativement au nombre de 300, il faut distinguer les occurrences de *être* auxiliaire et de *être* verbe constructeur parmi les 2 500 constructions restantes. L'auxiliaire, à première vue, représente un peu plus de la moitié des emplois du verbe *être*, c'est-à-dire environ 1 350 occurrences sur 2 500 (54%). Nous n'avons pas effectué ce tri à la main, mais en nous aidant de XCOR, auquel nous avons lancé les requêtes suivantes : **#v(être) #v_ppass / #v(être) * #v_ppass / #v(être) * * #v_ppass / #v(être) * * * #v_ppass**, qui permettent de trouver dans le corpus les occurrences de *être* suivies d'un participe passé, soit directement, soit après un élément, soit après deux éléments, soit après trois éléments. Aussi, pour ce verbe, les calculs ne sont pas aussi précis que pour les autres, et les résultats ne valent pas preuve. Toutefois, après une première observation, nous avons relevé (parmi les 1 150 constructions dans lesquelles *être* est constructeur) quatre catégories, qui pourraient sans doute être encore affinées :

- la première, du type *j'y suis / il (-personnel) y est* ou *ça y est*, possède 91 occurrences relevées. Les requêtes utilisées pour trier automatiquement ces occurrences sont les suivantes : **#v(être) #s / #v(être) * #s**, qui demandent toutes les formes de *être* suivies d'une préposition, soit directement (**#v(être) #s**), soit après un mot quelconque (**#v(être) * #s**). Nous n'avons pas lancé la requête demandant plus de mots entre le verbe *être* et la préposition, car, en regardant les résultats du total de **#v(être)**, nous n'avons pas trouvé d'occurrence significative.
- La seconde est du type *je le_cela suis / c'est ça*. On trouve 388 occurrences de ce type. Les requêtes utilisées pour trouver ces formes sont les suivantes : **#v(être) #d et #v(être) * #d**, qui recherchent dans le corpus respectivement toutes les formes du verbe *être* suivies par un déterminant et toutes les formes du verbe *être* suivies par un mot quelconque puis par un déterminant. Il est inutile de demander un nom après le déterminant, puisque la présence même de ce dernier implique celle du second. Nous avons aussi lancé la requête **#v(être) #n**, qui recherche dans le corpus toutes les

formes du verbe être suivies directement d'un nom (*je suis marchand de jouet*). Pour cette requête, le nombre de réponses données par XCOR est très largement supérieur aux formes correspondant réellement à cette requête dans la mesure où beaucoup de formes sont homographes (la négation avec le substantif *pas*, des adjectifs avec des substantifs, par exemple *malheureux*). Dans ce cas-là, un second tri est alors effectué à la main. Dans cette catégorie, un affinage devrait être réalisé, permettant d'analyser les constructions dans lesquelles le SN apparaît seul, de celles dans lesquelles il est lié à un SVinfΩ/Ø ou à une que-phrased (*il y a ça à faire / il y a mon père qui est malade*).

- La troisième est du type *je suis ainsi / c'est ainsi*, dans laquelle *ainsi* représente un adjectif. C'est la catégorie la plus importante pour le verbe *être*, puisqu'elle possède 636 occurrences. Les requêtes que nous avons utilisées pour mener à bien ce tri sont les suivantes : **#v(être) #a** et **#v(être) * #a**, qui recherchent dans le corpus respectivement toutes les occurrences du verbe être suivies d'un adjectif, et toutes les occurrences du verbe être suivies d'un mot quelconque et d'un adjectif. Nous n'avons pas retenu la requête **#v(être) * * * #a**, qui nous renvoyait beaucoup plus de constructions dans lesquelles l'adjectif était lié à un nom, et non pas au verbe *être*.
- La quatrième, la moins fréquente (28 occurrences), est du type : *c'est pour quoi / c'est pour qui*. Nous avons lancé la requête **#v(être) #s** et **#v(être) pas #s** qui cherchent respectivement dans le corpus toutes les occurrences du verbe *être* suivies d'une préposition, et toutes les occurrences du verbe *être* suivies de *pas* (négation) et d'une préposition. Cette requête s'est avérée trop imparfaite. Après avoir regardé le résultat, il nous est apparu que la seule préposition devant être retenue dans ce cas était *pour*. Nous avons donc ensuite précisé notre demande, sous la forme des deux requêtes suivantes : **#v(être) pour** et **#v(être) pas pour**, qui recherchent dans le corpus toutes les occurrences du verbe être suivies de *pour* et toutes celles suivies de *pas* (négation) et de *pour*.

b. Avoir

La requête **#v(avoir)** donne 2 134 réponses. Nous avons procédé pour le classement de ce verbe de la même manière que pour *être* : nous avons tout d'abord retiré les occurrences de *avoir* dans lequel il est répété, suite à des amorces ou des répétitions. Nous en avons trouvé à peu près 250. L'auxiliaire *avoir* représente à peine plus de la moitié des

emplois du verbe, soit environ 950 occurrences (50,55%). Pour trouver les occurrences de *avoir* auxiliaire, nous avons lancé les requêtes suivantes sous XCOR: **#v(avoir) #v_ppass / #v(avoir) * #v_ppass / #v(avoir) * * #v_ppass / #v(avoir) * * * #v_ppass**, qui permettent de trouver dans le corpus les occurrences de *avoir* suivies d'un participe passé, soit directement, soit après un élément, soit après deux éléments, soit après trois éléments. Parmi les 930 occurrences restantes, dans lesquelles *avoir* est constructeur, on trouve différentes catégories de constructions (qui pourraient elles-aussi être affinées) :

- La première est le type *je l'ai / j'en ai*. Nous en avons relevé 295 occurrences, parmi lesquelles il est possible qu'il figure encore des verbes supports (*avoir besoin, avoir envie...*). La requête utilisée sous XCOR pour trouver ce type de construction est la suivante : **#v(avoir) #d**. Elle permet de trouver dans le corpus toutes les occurrences de *avoir* suivi d'un déterminant. De la même manière que pour l'auxiliaire, nous avons lancé trois autres requêtes en demandant entre le verbe et le déterminant un, deux puis trois éléments. Dans cette catégorie, il se peut que l'on trouve des verbes supports ; en effet, nous en avons isolés quelques uns, mais peut-être pas tous. Une observation manuelle serait indispensable pour une étude des verbes supports dans ce cas-là.
- La deuxième catégorie est celle de la tournure présentative en *il y a / il y en a*. Elle figure 550 fois dans le corpus. La requête permettant de l'isoler est la suivante : **il y #v(avoir)**.
- Enfin, la troisième est celle des verbes supports : *avoir peur, avoir envie, avoir besoin, avoir l'impression*. Les requêtes utilisées sont les suivantes : **#v(avoir) N, #v(avoir) * N, #v(avoir) * * N, #v(avoir) * * * N**. N est en fait entré comme *peur, besoin, envie, impression*.

Conclusion et perspectives :

L'inventaire et le classement des constructions verbales dans un corpus de français parlé et la prise en compte de la fréquence d'apparition des occurrences permet de mettre à jour une autre vision de la complémentation verbale en français parlé. En effet, Gadet (2003 : p. 12-13) revient sur l'intérêt et sur l'apport de l'étude des corpus de langue parlée pour le développement des connaissances des réalités linguistiques : « les progrès techniques, du matériel d'enregistrement comme du traitement documentaire des données, ont introduit de nouvelles perspectives de travail sur la langue parlée et sa variabilité. Dans le même temps, la demande sociale apportait de nouvelles exigences de connaissances sur la langue telle qu'elle se parle vraiment. La conjonction de ces deux facteurs a favorisé le développement des études de la langue parlée à partir de grands corpus. [...] C'est aussi vers une demande d'échantillons authentiques qu'ira, à partir des années 60, l'expansion de l'enseignement de « Français Langue Etrangère » (FLE) : détacher les apprenants des seules connaissances littéraires et livresques, jusqu'à valoriser une oralité quasi-native, informée de la variation. [...] Les grands corpus de langue parlée de mise à disposition générale susceptibles d'exploitations diversifiées demeurent ainsi assez peu nombreux, surtout compte-tenu du rôle qu'ils jouent pour modifier les perspectives d'analyse de la langue, en particulier quant à la fréquence, quant aux distributions et aux restrictions, pour un nouvel éclairage des rapports entre lexique et grammaire, ou pour les spécificités linguistiques selon les genres, discursifs, contextuels ou sociaux. »

Actuellement, l'essor des recherches linguistiques sur corpus, qu'il s'agisse de langue écrite ou parlée, est considérable. Cependant, c'est un phénomène relativement nouveau dans l'histoire des sciences du langage : il date d'une vingtaine d'années environ. Cet essor est intrinsèquement lié au développement des nouvelles technologies, et notamment de l'informatique. En effet, la création d'outils permettant le traitement automatique de données a offert l'opportunité aux chercheurs de travailler sur une masse d'informations difficilement traitables jusqu'alors, les moyens engagés, nécessairement humains, coûteux, étant insuffisants pour parvenir à de tels objectifs. Mais si les progrès technologiques ont, de fait, accéléré le processus de recherche, une réflexion de fond sur le statut même de ces données, sur leur organisation et sur la manière dont il fallait les traiter était nécessaire. Elle a été engagée, et d'importantes contributions ont été fournies par les linguistes et les informaticiens depuis lors. Lorsque l'on étudie un système linguistique, qu'on le décrit,

qu'on l'analyse, plus le nombre de données est important, et plus l'analyse est délicate, puisque la multiplication même de ces données engendre la potentialité de leurs différences. Ainsi, le traitement des grands corpus de langue orale, puisque c'est ce type qui nous intéresse ici, a déjà montré ses intérêts : la description du système de l'oral, la démonstration que la description de l'usage remet en cause la notion de « faute », et celle que l'étude de la variation permet de mieux comprendre notre langue, etc. Nous ne reviendrons pas sur ces avancées, déjà constatées et relatées dans des ouvrages récents (Bilger, 2000 ; Gadet, 2003).

Un corpus est défini par les objectifs que le linguiste se donne. Cette idée, énoncée maintes fois, que ce soit par F. Rastier ou par C. Blanche-Benveniste, présuppose celle d'une construction, qui elle-même présuppose une réflexion. La première partie de ce travail montre donc quelles ont été les questions que nous nous sommes posées avant d'effectuer les enquêtes de terrain, les choix, les problèmes qu'elles ont soulevé et les solutions que nous leur avons trouvées. Ainsi, après avoir brièvement présenté le corpus, qui apparaît annexé sous forme de fichiers audio et de fichiers textes et que chacun pourra consulter à sa guise, nous avons présenté la méthode utilisée pour mener à bien ces enquêtes : sélection des locuteurs (l'échantillon sélectionné n'étant pas un reflet de la population, mais bien une sélection), recherche des locuteurs correspondants aux critères retenus, matériel utilisé, type d'enquête, conditions d'enregistrement et enfin conventions de transcriptions. Les problèmes que cette méthode peut poser sont soulevés au cours de la fin de cette partie, qui permet de saisir l'importance et la complexité de notions comme celle de texte, de type de discours, ou encore de thème choisi pour l'élaboration du corpus.

Un corpus, nous l'avons observé, n'est, par définition, fermé que le temps du travail pour lequel il a été constitué. Ainsi, bien des améliorations pourraient être apportées à présent, tant du point de vue de la saisie (quelques fautes de frappe subsistent encore) que du rétablissement de la cohérence thématique. De nouvelles enquêtes pourraient être menées à Montluçon, ayant pour thème le petit commerce, afin d'unifier le thème choisi. D'autre part, nous avons enregistré 18 locuteurs pendant une demie-heure sur un autre thème. Il s'agissait d'une description de tableaux. Ces locuteurs sont les mêmes que ceux enregistrés à Vichy et Moulins sur le petit commerce. Une autre perspective de développement de ce corpus serait une transcription phonétique, qui nécessiterait cependant des moyens très importants, étant donnée la finesse et la minutie que ce travail présuppose. Ce travail pourrait peut-être s'intégrer dans un projet du type de celui initié par J. Durand, B. Laks et C. Lyche (Toulouse-

Le Mirail, Paris 10, Université d'Oslo : Phonologie du Français Contemporain (PFC)). Cela permettrait sans doute d'utiliser de manière pertinente les richesses que le croisement des critères de sélection des locuteurs peut offrir, puisque, comme nous l'avons déjà signalé, ce travail ne prend pas en compte la variation, il n'est qu'un point de départ descriptif, qui nécessiterait bien des approfondissements.

Nous avons souhaité, avant même que ce travail soit présenté publiquement, « diffuser » notre corpus. Peut-être était-ce prématuré ; cependant, l'honnêteté de l'utilisation qu'en ont fait les personnes concernées nous a rassurée, et les liens créés à cette occasion, qu'ils soient d'ordre strictement professionnels ou qu'ils débouchent sur des échanges à la fois amicaux et scientifiques, nous ont permis de prendre conscience de la force de la communauté scientifique unie autour de projets et d'intérêts communs.

La constitution d'un corpus de langue parlée laisse rarement l'enquêteur indifférent, puisqu'il doit faire preuve d'une écoute empathique afin d'assurer à l'enquêté des conditions de discours aussi faciles que possible. Les hommes et les femmes interrogées se livrent souvent bien au-delà des espérances du scientifique. Ainsi, sur le thème du petit commerce et des grandes surfaces, la parole des personnes âgées interrogées, qui sont pour certaines aujourd'hui disparues, révèle à présent un discours précieux, qui dépasse largement le cadre de la linguistique, et qui pourrait sans doute intéresser sociologues ou philosophes, par exemple.

La deuxième partie est consacrée à la description de la méthode et des outils utilisés pour effectuer l'inventaire et le classement des constructions verbales dans notre corpus.

Dans la première partie (2.1.), nous présentons quelques travaux effectués plus ou moins récemment sur les constructions verbales : le lexique-grammaire (2.1.1.), Proton (2.1.2.), le Dictionnaire Contrastif des Valences Verbales (2.1.3.) et enfin la thèse de T. Greidanus (2.1.4.).

Dans la seconde partie (2.2.), nous montrons l'importance du système de l'inventaire pour la description des constructions verbales, et mettons en perspective les propos de M.

Gross (1975) à ce sujet, en les actualisant par la prise en compte des progrès de la linguistique dite « de corpus », liés aux avancées technologiques récentes.

La troisième partie (2.3.) est consacrée au travail définitoire. Ainsi, nous tentons de montrer les limites de la notion de phrase et les avantages que présente le choix de la notion de construction verbale comme unité d'analyse (2.3.1.). Nous revenons dans cette partie sur les notions de valence et de rection, tout en précisant la difficulté que nous avons parfois à délimiter la valence d'un verbe, comme l'a démontré O. Bonami (1999) en réfutant la théorie selon laquelle un argument est à la fois syntaxiquement et sémantiquement obligatoire (2.3.1.1.). Nous précisons ensuite à quel niveau nous avons envisagé l'analyse des constructions verbales, dans la mesure où il existe des constructions verbales rectrices et des constructions verbales régies par les premières (2.3.1.2.).

Par ailleurs, la littérature linguistique montre les limites de la notion de pronom (2.3.2.) ; comme ce travail se fonde sur l'approche pronominale, nous avons ici jugé utile de les décrire ici, en revenant sur ce qu'en ont écrit des théoriciens comme Beauzée, Gross ou Creissels (2.3.2.1.). Ensuite, nous précisons quels types de « pronoms » nous ont servi pour effectuer la réduction pronominale des éléments qui entrent dans la construction verbale (2.3.2.2. et 2.3.2.1.).

La partie 2.3.3. est consacrée à délimiter le type d'éléments qui sont considérés comme étant construits par le verbe. Ainsi, nous montrons les trois tests qui permettent de révéler un élément construit : la pronominalisation (2.3.3.1.), la portée des modalités verbales (2.3.3.2.) et enfin la focalisation (2.3.3.3.).

Au cours de la partie suivante (2.3.4.), nous présentons brièvement la notion de formulation, que nous n'avons pas utilisée dans ce travail, mais qui est au cœur des travaux de l'approche pronominale, puis nous montrons que le premier test permettant de reconnaître un élément construit par le verbe, la substitution par une pro-forme, n'est pas toujours aisé à réaliser, et que nous pouvons parfois buter sur le choix de la pro-forme adéquate.

Après avoir, au cours de la partie 2.3., posé les jalons définitoires de cette étude, nous présentons le cadre théorique de l'approche pronominale (2.4.1.), puis nous expliquons ce que nous en avons retenu et ce que nous avons exclu (2.4.2.).

Dans la partie 2.5., nous présentons nos données : les verbes présents dans le corpus (2.5.1.), les types de constructions verbales que nous avons traités (2.5.2.1.), puis nous expliquons les différents intérêts que présentent la prise en considération de la fréquence dans ce type de travail.

Dans la dernière partie (2.6.), nous présentons les outils qui nous ont servi à élaborer cet inventaire et ce classement. Ainsi, dans un premier temps, nous présentons le concordancier XCOR2000, créé par Jean-Pierre Adam pour le GARS (2.6.1.), puis nous expliquons les principes et le fonctionnement de la base de données (2.6.2.), en détaillant les propriétés syntaxiques et distributionnelles que nous avons retenues (2.6.2.1.), puis en prenant pour exemple trois verbes : *aller*, *être* et *avoir*.

L'inventaire et le classement des constructions verbales dans ce corpus a révélé quelques problèmes inhérents à toute description de ce genre. D'une part, comme nous l'avons signalé, ce que nous nommons *inventaire* n'en est pas exactement un : il s'agit ici d'un recensement de toutes les formes verbales conjuguées, exclusion faite des infinitifs et des formes en *-ant*.

Nous avons toutefois relevé à la fois les constructions verbales rectrices et régies, considérant que dans un exemple construit comme :

Max aime que Marie fasse des gâteaux

deux constructions verbales apparaissent, la première (*Max aime cela*) régissant la seconde (*Marie en fait*). Nous avons choisi de considérer le deuxième type de construction comme étant potentiellement indépendante, et nous l'avons transformée afin qu'elle le devienne (dans cet exemple, nous avons ainsi remplacé le subjonctif par l'indicatif). Cependant, la trace de la construction régie apparaît dans le relevé de la rectrice, puisque la table précise, dans un cas comme celui-ci, l'extension lexicale de la pro-forme utilisée pour réduire *que Marie fasse des gâteaux*. La pro-forme utilisée dans ce cas (*cela*) est accompagnée en fin de ligne par l'extension QuePsubj, qui indique la présence d'une Que-Phrase au subjonctif.

Une des difficultés majeures que nous avons rencontrées lors du classement est celle de la délimitation des propriétés prises en compte. En effet, qu'il s'agisse des propriétés de distribution (qui permettent de savoir quels types de compléments un verbe admet) ou des

propriétés de structure (qui indiquent dans quel type de structure syntaxique un verbe apparaît), le problème se pose de savoir où arrêter la finesse de la description. Les recherches sur les propriétés verbales sont aujourd'hui très avancées, notamment grâce aux travaux lancés par M. Gross pour l'élaboration du lexique-grammaire des verbes français. Mais une description de la complémentation verbale à partir de données orales attestées diffère du formalisme du lexique-grammaire, que ce soit du point de vue de la méthode utilisée que des objectifs visés. En effet, si nous avons, dans le cadre de ce travail, fournit une description à partir de données « réelles », le lexique-grammaire permet une description plus systématique dans la mesure où les exemples, construits par les chercheurs, sont transformés selon des principes définis et jugés grammaticaux ou non. Ainsi, notre description de la complémentation verbale est limitée de fait par le corpus et par les principes descriptifs que nous nous sommes fixés. Si les objectifs du lexique-grammaire sont de fournir un inventaire détaillé du fonctionnement de la grammaire par la fenêtre du lexique, et de proposer, pour l'ensemble du lexique, sa distribution et son fonctionnement syntaxique, les nôtres, bien que plus humbles, sont de même nature.

Nous ne pouvons pas ici présenter de résultats issus d'une analyse très fine, puisque, nous l'avons vu, ce classement n'avait pas pour objectif d'étudier très précisément tel ou tel type de construction, mais plutôt de dresser un premier tableau de la complémentation verbale dans notre corpus. Cependant, quelques résultats assez grossiers se dégagent : pour 352 verbes, nous avons environ 7 000 constructions verbales étudiées, ce qui donne une moyenne d'une vingtaine d'occurrences par verbe. Nous avons 632 enregistrements différents dans notre table, ce qui ne signifie pas que nous ayons 632 types de constructions différents, puisque les enregistrements sont fondés sur le verbe constructeur, et non sur le type de construction. Ainsi, le nombre moyen de constructions différentes par verbe est de 1,8.

Sur les 7 000 constructions analysées, nous en avons relevé 6 692 à la voix active « pure », 145 à la voix « passive », 197 sont pronominales réfléchies, 27 pronominales réciproques et 78 pronominales passives. Les constructions verbales actives représentent donc 96,8 % de la totalité, les passives, 2 %. Les pronominales passives occupent 1,2 % du total des CV. Les verbes pronominaux réfléchis et réciproques représentent quant à eux respectivement 2,8 % et 0,4 % des CV étudiées.

D'autre part, les emplois en tant qu'auxiliaires des verbes *être* et *avoir* sont en nombre supérieur à ceux de verbe constructeur : les auxiliaires représentent, *être* et *avoir* confondus, 2 300 occurrences, tandis que les verbes constructeurs sont approximativement 2 100. Ainsi, les auxiliaires forment 52,5 % du total des occurrences des verbes *être* et *avoir*, et les verbes constructeurs, 47,5 %, ce qui correspond bien à la répartition observée par divers linguistes auparavant.

Si nous souhaitons que le corpus que nous avons constitué soit ouvert, il en est de même pour la base de données. En effet, la table des constructions verbales doit être ici considérée comme une base de travail relativement sommaire, pour plusieurs raisons : d'une part, le verbe étant au cœur du système du français, les multiples problèmes de traitement qu'il pose n'ont pas tous été abordés, et les propriétés intégrées sont sans doute en nombre insuffisant pour permettre d'effectuer des analyses très fines. D'autre part, cette table étant liée au corpus par un système de classement alpha-numérique, les chercheurs intéressés par d'autres propriétés verbales que celles qui apparaissent dans cette étude ont la possibilité d'accéder aux données pour en ajouter ou en supprimer.

Willems (1981 : p. 12) indique que « la recherche syntaxique vise avant tout à la découverte des propriétés distributionnelles et transformationnelles les plus cachées et leur intégration dans un système cohérent de description », et que « c'est dans ces études, rigoureuses et détaillées, qu'on retrouve l'embryon d'une sémantique syntaxique, la description minutieuse des données syntaxiques laissant entrevoir des prolongements sémantiques possibles ». Nous n'avons pas, dans le cadre de ce travail, intégré d'analyse sémantique à partir des types de constructions verbales que nous avons classés. Cependant, s'il s'agit là d'un manque évident, les liens entre lexique, syntaxe et sémantique seront l'objet de recherches futures à partir du classement proposé ici.

Nous espérons que cette description permettra aux informaticiens-linguistes travaillant en TAL d'avoir une base de données reflétant le français parlé et ses différents types de constructions verbales, et ainsi de pouvoir, par exemple, mesurer la distance entre production écrite et production orale. D'autre part, ce classement peut être une base intéressante pour la création d'un dictionnaire des constructions verbales en français parlé ; en effet, bien que le corpus soit thématisé et par nature limité, la prise en compte de la fréquence et le lien entre la

table des constructions et le corpus classé des CV sont une base de travail solide pour une étude lexicologique.

Bibliographie

I]Dialectologie :

- BLANCHE-BENVENISTE C., 1991, « La difficulté à cerner les régionalismes en syntaxe », in Salmon éd., *Variétés et variantes du français des villes*, Paris-Genève, Champion, Slatkine.
- BRUN-TRIGAUD, G., 1990, *Le croissant, le concept et le mot, Contribution à l'histoire de la dialectologie au XIX^e siècle*, Centre d'Etudes Linguistiques Jacques Goudet, Série Dialectologie 1, Université Jean Moulin, Lyon III.
- DUBUISSON P., *Atlas linguistique et ethnographique du Centre*, 3 volumes – 1 La nature, 1971, - 2 L'homme, 1976, - 3 La grammaire, 1971, Paris, éditions du C.N.R.S.
- DUBUISSON P., 1991, « Zones de contact oïl/oc, sommet du Croissant », Actes du colloque *Les Français et leurs langues*, publication de l'Université de Provence Aix-Marseille I.
- ESCOFFIER S., 1958, *La Rencontre de la langue d'oïl, de la langue d'oc et du francoprovençal entre Loire et Allier*, Paris, Les Belles Lettres.
- GALLARD P., 1995, « Diversité des usages et idiolectes, ou la nécessité de deux niveaux d'analyse : l'exemple du poitevin », *La Linguistique*, vol. 31, fasc.1.
- GALLARD P., 1996, « La conscience linguistique chez les jeunes patoisants des Deux-Sèvres : réflexions sur les normes », *Travaux de Linguistique*, n° 7, Université d'Angers.
- GARDETTE P., 1983, *Études de géographie linguistique*, publiées par la Société de Linguistique Romane avec le concours de la Fondation Georges Guichard, par les soins de B. Horiot, M.-R Simoni et G. Straka, Strasbourg.
- LEROND A. (Dir.), 1973, « Les parlers régionaux », *Langue Française*, n° 18, Paris, Larousse.
- POTTE J.-C., 1987, *Atlas linguistique et ethnographique de l'Auvergne et du Limousin*, 2 volumes, Paris, éditions du C.N.R.S..

- REZEAU P., 1998, « Albert Dauzat et la variation géographique en français d'Auvergne et d'ailleurs », Actes du colloque *Albert Dauzat et le patrimoine linguistique auvergnat*, publiés avec l'aide du Conseil Régional d'Auvergne, Thiers.

- STRAKA G. et TAVERNET G., 1977, *Les Français régionaux*, Paris, Klincksieck.

- TUAILLON G., 1967, « Principes pour distinguer français et francoprovençal », *Revue de Linguistique Romane* n° 37, pp. 292-296.

II] Sur le français parlé :

- BLANCHE-BENVENISTE C., 1997b, « La notion de variation syntaxique dans la langue parlée », *Langue Française*, n° 115, *La variation en syntaxe*, Paris, Larousse.

- BLANCHE-BENVENISTE C., 1997a, *Approches de la langue parlée en français*, Paris, Ophrys, coll. « L'essentiel français ».

- BLANCHE-BENVENISTE C., 1993, « Une description linguistique du français parlé », *Le Gré des Langues* n° 5, Paris, L'Harmattan, pp. 8-29.

- BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K., 1990, *Le Français parlé, études grammaticales*, Paris, éd. du C.N.R.S..

- BLANCHE-BENVENISTE C., 1987, « Syntaxe, choix de lexique et lieux de bafouillage », *DRLAV*, pp. 36-37.

- BLANCHE-BENVENISTE C., JEANJEAN C., 1987, *Le Français parlé, transcription et édition*, Paris, INALF, Didier Érudition.

- BLANCHE-BENVENISTE C., DEULOFEU, J., STEFANINI, J. et VAN DEN EYNDE, K., 1984 (1987 : 2^{ème} édition augmentée), *Pronom et syntaxe, l'approche pronominale et son application au français*, Paris, Sela.

- BLANCHE-BENVENISTE C., coll. et al., 1982, « Des grilles pour le français parlé », *Recherches sur le français parlé* n° 2, Aix-en-Provence, Publications de l'Université de Provence.

- BLANCHE-BENVENISTE C., 1975, *Recherche en vue d'une théorie de la grammaire française. Essai d'application à la syntaxe des pronoms*, Paris, Champion.
- BLASCO-DULBECCO, M., 1999, *Les dislocations en français contemporain, Etude syntaxique*, Paris, Champion, coll. « Les français parlés, textes et études ».
- EYNDE, Karel Van den & BLANCHE-BENVENISTE, C., 1978, « Syntaxe et mécanismes descriptifs : présentation de l'approche pronominale », in *Cahiers de Lexicologie*, n° 32, 3-27.
- GADET F., 2003, *La variation sociale en français*, Paris, Ophrys, coll. « L'essentiel français ».
- GADET F., 1997b, (2^e éd.), *Le Français ordinaire*, Paris, A. Colin, coll. U, série « Linguistique ».
- GADET F., 1997a, « La variation, plus qu'une écume », *Langue Française*, n° 115, *La variation en syntaxe*, Paris, Larousse.
- GADET F., 1971, « Recherches récentes sur les variations sociales de la langue », *Langue Française*, n° 9, *Linguistique et société*, Paris, Larousse, pp. 74-81.
- GOUGENHEIM, G., MICHEA, R., RIVENC, P. et SAUVAGEOT, A., 1967, *L'élaboration du français fondamental (1^{er} degré)*, Paris, Didier.
- GUIRAUD P., 1965, *Le français populaire*, Paris, P.U.F., coll. Que sais-je ?.
- LANGUE FRANÇAISE, n° 60, 1983, *La Phonologie des usages du français*, Paris, Larousse.
- LANGUE FRANÇAISE, n° 93, février 1992, *Enquêtes, corpus et témoins*, Paris, Larousse.
- QUENEAU, R., 1965, *Bâtons, chiffres et lettres*, Paris, Gallimard.
- RECHERCHES SUR LE FRANÇAIS PARLE n°18, 2004, *Autour du corpus de référence du français parlé*, Publications de l'Université de Provence.
- ROUBAUD, M.-N., 2000, *Les constructions pseudo-clivées en français contemporain*, Paris, Champion.

- TUAILLON G., 1983b, « Régionalismes grammaticaux », *Recherches sur le français parlé* n°5, Aix-en-Provence, Publications de l'Université de Provence, pp. 227-239.
- TUAILLON G., 1983a, *Matériaux pour l'étude des régionalismes du français ; les régionalismes du français parlé à Vourey, village dauphinois*, Paris, Klincksieck.
- WALTER H., 1982, *Enquête phonologique et variétés régionales du français*, Paris, P.U.F..

III] Méthodologie :

1- Enquêtes :

- BERTHIER, N., 1998, *Les techniques d'enquêtes en sciences sociales, Méthodes et exercices corrigés*, Paris, A. Colin, Coll. « Coursus ».
- BLANCHET, P., 2000, *La linguistique de terrain, Méthode et théorie, Une approche ethno-sociolinguistique*, Rennes, P.U.R., Coll. « Didact Linguistique ».
- BULOT, T., 2004, *Lieux de ville et identité. Perspectives en sociolinguistique urbaine*, Paris, L'Harmattan.
- DURAND, J., DELAIS-ROUSSARIE, E., 2003, *Corpus et variations en phonologie du français, méthodes et analyses*, Toulouse, Presses Universitaires du Mirail, coll. « Interlangues linguistique et didactique ».
- FREYSSINET-DOMINJON, J., 1997, *Méthodes de recherche en sciences sociales*, Paris, Montchrestien, Coll. « AES ».
- GADET, F., 2003, « Derrière les problèmes méthodologiques du recueil des données », revue électronique *Texte*, <http://www.revue-texte.net/Inedits/Inedits.html>.
- GUIBERT, J. et JUMEL, G., 1997, *Méthodologie des pratiques de terrain en sciences humaines et sociales*, Paris, A. Colin, Coll. « Coursus ».
- LABOV, W. 1972. *Sociolinguistic Patterns*, Philadelphia : University of Pennsylvania Press. Traduction française, *Sociolinguistique*, 1976, Paris, Editions de Minuit.

- MILROY, J., 1980, *Language and social networks*, Oxford, Blackwell.
- MUCCHIELLI, A. (dir.), 1996, *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*, Paris, A. Colin.

2- Syntaxe :

- BENZITOUN, C., *Un langage pivot pour articuler description et formalisation : l'exemple des verbes introducteurs de « que-phrases »*, Colloque RECITAL 2003, Batz-sur-Mer, 11-14 juin 2003.
- CREISSELS, D., 1995, *Eléments de syntaxe générale*, Paris, PUF, coll. « Linguistique nouvelle ».
- GROSS, M., 1975, *Méthodes en syntaxe, Régime des constructions complétives*, Paris, Hermann.
- LE QUERLER, N., 1994, *Précis de syntaxe française*, Caen, Presses de l'Université de Caen.

IV] Actes de colloques :

- LE FRANÇAIS PARLE, CORPUS ET RESULTATS, 2000, Actes du colloque international, Université de Copenhague, 29-30 octobre 1998, publiés par H.-L. Andersen et A.-B. Andersen, *Etudes Romanes* n° 47, Museum Tusculanum Press, University of Copenhagen.
- LE PASSIF, 2000, Actes du colloque international, Institut d'Etudes Romanes, Université de Copenhague, 5-7 mars 1998, *Etudes Romanes* n° 47, Museum Tusculanum Press, University of Copenhagen.
- VERBUM, 2002, *Y a-t-il une syntaxe au-delà de la phrase ?*, Actes du colloque international, Paris III, septembre 2000, tome XXIV, n° 1-2, numéro coordonné par M. Charolles, P. Le Goffic et M.-A. Morel.

V] Subordination :

- ALLAIRE, S., 1973, *La subordination dans le français parlé devant les micros de la Radiodiffusion, Etude d'un corpus*, Paris, Klincksieck.
- *Subordination, Subordinations*, Travaux Linguistiques du CERLICO, Rennes, P.U.R.
- MULLER, C., 1996, *La subordination en français, Le schème corrélatif*, Paris, Armand Colin, coll. U, série « Linguistique ».
- SANDFELD, Kr., 1936, *Syntaxe du français contemporain, Les propositions subordonnées*, Genève, Droz.

VI] Constructions verbales, verbes et opérateurs :

- BLANCHE-BENVENISTE, C., 2002b, « La complémentation verbale : petite introduction aux valences verbales », *Travaux neuchâtelois de linguistique* n°37, pp. 47-73.
- BLANCHE-BENVENISTE, C., 2002a, « Phrase et construction verbale », *Verbum* tome XXIV, n° 1-2, Nancy, Presses Universitaires de Nancy, pp. 7-22.
- BLANCHE-BENVENISTE, C., 1989, « Constructions verbales « en incise » et rection faible des verbes », *Recherches sur le français parlé* n° 9, Aix-en-Provence, Publications de l'Université de Provence, pp. 53-74.
- BLANCHE-BENVENISTE, C., 1986, « La notion de contexte dans l'analyse syntaxique des productions orales : exemples des verbes actifs et passifs », *Recherches sur le français parlé* n° 8, Aix-en-Provence, Publications de l'Université de Provence, pp. 39-57.
- BLANCHE-BENVENISTE, C., 1981, « La complémentation verbale : valence, rection et associés », *Recherches sur le français parlé* n° 3, pp. 59-98.
- BLANCHE-BENVENISTE, C., BOREL, B., DEULOFEU, J., DURAND, J., GIACOMI, A., LOUFRANI, C., MEZZIANE, B. et PAZERY, N., 1979, « Des grilles pour le français parlé », *Recherches sur le français parlé* n° 2, Aix-en-Provence, Publications de l'Université de Provence, pp. 163-208.

- BONAMI, O., 2001, « Complémentation et structure du lexique », in J. Chuquet et M. Paillard (éds), *Travaux linguistiques du CERLICO* 15, pp. 11-29, Rennes, Presses Universitaires de Rennes. [\[Fichier PDF \(820K\)\]](#)
- BONAMI, O., 1999, *Les constructions du verbe : le cas des groupes prépositionnels argumentaux*, Thèse de doctorat, Université Paris 7. [\[Fichier PDF \(1,5M\)\]](#)
- CAPUT J., CAPUT, J.-P., 1969, *Dictionnaire des verbes français*, Paris, Larousse.
- CHU, X.-Q., 1989, « Examen de la structure des enchaînements des verbes modaux », *Recherches sur le français parlé* n°9, Aix-en-Provence, Publications de l'Université de Provence, pp. 75-86.
- DEULOFEU, J., 1977, « La syntaxe et les constructions binaires », *Recherches sur le français parlé* n°1, Aix-en-Provence, Publications de l'Université de Provence, pp.30-61.
- DUGAS, A., 2000, *Le verbe et ses prépositions*, Outremont, Editions Logiques.
- DUGAS, A., MANSEAU, H., 1996, *Les verbes logiques*, Montréal, Editions Logiques.
- EYNDE, K. van den, MERTENS, P., 2002, *La syntaxe du verbe, l'approche pronominale et le lexique de valence PROTON*, Louvain, document preprint n° 174, téléchargé à l'adresse suivante : <http://bach.arts.kuleuven.ac.be/pmertens/papers/proton2002.pdf>
- FRANÇOIS, D., 1983, « Normes orale et norme écrite : l'exemple de *c'est* et *il y a* », *Etudes de grammaire française descriptive*, Heidelberg, J. Groos, pp. 64-68.
- GAATONE, D., 1998, *Le passif en français*, Paris, Bruxelles, Duculot.
- GIACOMI, A., VERONIQUE, D., 1982, "A propos de *il y a/il y en a*", *Le français moderne*, 50-3, pp.237-242.
- GIRY-SCHNEIDER, J. 1986. "Les noms construits avec *faire* : compléments ou prédicats ?", *Langue Française* n° 69, Paris, Larousse, pp. 49-63.
- GIRY-SCHNEIDER, J., 1978, *Les nominalisations en français. L'opérateur ... « faire »... dans le lexique*, Genève, Droz.

- GREIDANUS, T., 1990, *Les constructions verbales en français parlé. Etude quantitative et descriptive de la syntaxe des 250 verbes les plus fréquents*, Tübingen, Niemeyer, coll. « Linguistische Arbeiten ».
- JEANJEAN, C., 1982, « Qu'est-ce que c'est que "ça" ? Etude syntaxique de "ça"-sujet en français parlé : la construction "quand P + ça" », *Recherches sur le français parlé* n° 4, Aix-en-Provence, Publications de l'Université de Provence, pp. 117-151.
- JEANJEAN, C., 1979, « Soit y'avait le poisson soit y'avait ce rôti farci, étude de la construction *il y a* dans la syntaxe du français », *Recherches sur le français parlé* n°2, Aix-en-Provence, Publications de l'Université de Provence, pp.121-160.
- LE GOFFIC, P., 1993, *Grammaire de la phrase française*, Paris, Hachette Université, Coll. « Langue française ».
- LE GOFFIC, P., COMBE MCBRIDE, N., 1975, *Les constructions fondamentales du français*, Paris, Hachette/Larousse, Coll. « Le français dans le monde-B.E.L.C. ».
- MELIS, L., 1990, *La voie pronominale, le systématique des tours pronominaux en français moderne*, Paris, Louvain-la-Neuve, Duculot, coll. « Champs linguistiques ».
- MELIS, L., 1995, « Les dictionnaires automatisés des valences verbales du français et du néerlandais développés à la K. U. Leuven. Présentation », in *Quarterly Newsletter of the Constrative Grammar Research Group of the University of Gent*, n° 4.
- POURADIER-DUTEIL, F., 1997, *Le verbe français en conjugaison orale*, Frankfurt, Peter Lang
- REQUEDAT, F., 1980, *Les constructions verbales avec l'infinitif*, Paris, Hachette, « F », coll. Recherches/Applications.
- TESNIERE, L., 1959, *Eléments de syntaxe structurale*, Paris, Klincksieck.
- VALLI, A., 1977, « Rection verbale complexe et effets de 'transparence' », *Recherches sur le français parlé* n°1, Aix-en-Provence, Publications de l'Université de Provence, pp.62-86.
- WILLEMS, D., 1981(1^{ère} éd. en 1975), *Syntaxe, lexique et sémantique. Les constructions verbales*, Gand, Publications de la Faculté des Lettres de Gand.

- WILLEMS, D., 2002a, « Classes et relations verbales. A la recherche d'une typologie intégrée. », in J. Chuquet & M. Paillard (eds.) *Morphosyntaxe du lexique* 1. (Travaux linguistiques du Cerlico, 15), pp. 285-301.

VII] Problèmes liés à la notion de pronom

- BEAUZEE, N., 1767, *Grammaire générale ou exposition raisonnée des éléments nécessaires du langage, pour servir de fondement à l'étude de toutes les langues*, Paris, Barbou (éd. consultée : facsimilé de l'éd. de 1767, 1974, Stuttgart-Bad Cannstatt, Friedrich Frommann Verlag Günther Holzboog KG.), pp. 258-287.

- BENVENISTE, E., 1966, *Problèmes de linguistique générale*, Paris, Gallimard, pp. 251-257.

- BLANCHE-BENVENISTE C., DEULOFEU, J., STEFANINI, J. et VAN DEN EYNDE, K., 1984 (1987 : 2^{ème} édition augmentée), *Pronom et syntaxe, l'approche pronominale et son application au français*, Paris, Selaf, pp. 206-238.

- BRUNOT, F., 1926, *La pensée et la langue*, Paris, Masson, pp. 173-177.

- CREISSELS, D., 1995, *Eléments de syntaxe générale*, Paris, PUF, coll. « Linguistique nouvelle ».

- DUBOIS, J., 1965, *Grammaire structurale du français, nom et pronom*, Paris, Larousse, coll. « Langue et langage ».

- ELUERD, R., 2002, *Grammaire descriptive de la langue française*, Paris, Nathan Université, pp. 74-95.

- GROSS, M., 1986, *Grammaire transformationnelle du français*, vol. 1, *Syntaxe du verbe*, Malakoff, Cantilène, Coll. « Systématique de la langue française » (1^{ère} éd., 1968, Paris, Larousse).

- MOIGNET, G., 1965, *Le pronom personnel français, essai de psycho-systématique historique*, Paris, Klincksieck.

- PINCHON, J., 1986, *Morphosyntaxe du français, étude de cas*, Paris, Hachette Université, coll. « Langue, Linguistique, Communication », pp. 46-133.

- PINCHON, J., 1972, *Les pronoms adverbiaux en et y, leurs emplois et leurs valeurs en français classique et moderne*, Genève, Droz.

- QUENEAU, R., 1965, *Bâtons, chiffres et lettres*, Paris, Gallimard.

VIII] Linguistique discursive :

- BERTHOUD, A.-C. et MONDADA, L., 2000, *Modèles du discours en confrontation*, Bern, Peter Lang, Coll. Sciences pour la communication.

- ROULET, E. (Dir.), FILLIETAZ, L., GROUBET, A., avec la collaboration de BURGER M., 2001, *Un modèle et un instrument d'analyse de l'organisation du discours*, Bern, Peter Lang, coll. Sciences pour la communication.

IX] Linguistique textuelle, analyse du discours, genres et types :

- ADAM, J.-M., 1992, *Les textes : types et prototypes, récit, description, argumentation, explication et dialogue*, Paris, Nathan Université, coll. Fac linguistique.

- ADAM, J.-M., 1990, *Éléments de linguistique textuelle, théorie et pratique de l'analyse textuelle*, Liège, Mardaga.

- MAINGUENEAU, D., 2000, *Analyser les textes de communication*, Paris, Nathan, coll. « Lettres sup. ».

- MAINGUENEAU, D., 1987, *Nouvelles tendances en analyse du discours*, Paris, Hachette, coll. « Langue, Linguistique, Communication ».

- RASTIER, F., 2001, *Arts et sciences du texte*, Paris, P.U.F., coll. « Formes sémiotiques ».

- RASTIER, F., dir., 1995, *L'analyse thématique des données textuelles, l'exemple des sentiments*, Paris, Didier Erudition, coll. « Etudes de sémantique lexicale ».

- RASTIER, F., CAVAZZA, M., ABEILLE, A., 1994, *Sémantique pour l'analyse. De la linguistique à l'informatique*, Paris, Masson, Coll. « Sciences cognitives ».

- RASTIER, F., 1989, *Sens et textualité*, Paris, Hachette.

BIBLIOGRAPHIE

- RASTIER, F., 1988, « Macrosémantique et syntaxe », *L'information grammaticale* n° 37, Paris.

X] Corpus :

- BERGOUNIOUX (G.), BARADUC (J.) et DUMONT (C.), 1992, « L'Etude sociolinguistique sur Orléans (1966-1991) ; 25 ans d'histoire d'un corpus », in *Langue Française* n°93, Paris, Larousse.
- BIGGS, P. et BLANC, M., 1971, « L'enquête sociolinguistique sur le français parlé à Orléans », in *Le français dans le monde*, n° 85, décembre 1971, pp. 16-25.
- BILGER, M., (ed.), 2000, *Corpus. Méthodologie et applications linguistiques*, Paris, Honoré Champion et Presses Universitaires de Perpignan.
- BLANCHE-BENVENISTE, 1996, « De l'utilité du corpus linguistique », in *Revue française de linguistique appliquée*, vol. I-2, Amsterdam, De Werelt.
- DELIC, 2004, « Présentation du *Corpus de Référence du Français parlé* », *Recherches sur le français parlé*, n° 18, Aix, PUP, pp. 11-42.

XI] Lexique-grammaire :

- BOONS, J.-P., GUILLET, A., LECLERE, Ch., 1976, *La structure des phrases simples en français. Constructions intransitives*, Genève, Droz.
- GROSS, G., VIVES, R., 1986, « Les constructions nominales et l'élaboration d'un lexique-grammaire », in *Langue Française* n° 69, pp. 5-27.
- GROSS, M., 1986 (1^{ère} édition 1968), *Grammaire transformationnelle du français*, vol. 1, *Syntaxe du verbe*, Malakoff, Cantilène, Coll. « Systématique de la langue française ».
- GROSS, M., 1975, *Méthodes en syntaxe, Régime des constructions complétives*, Paris, Hermann.
- LAPORTE, E. et al., 1999, *Langages* n° 133, *Lexique-grammaire des adjectifs*, Paris, Larousse.
- LECLERE, Ch., 1990, « Organisation du lexique-grammaire des verbes français », *Langue Française* n° 87, Paris, Larousse, pp. 112-122.

- MEYDAN, Métiyé, 1999, *Transformation des constructions verbales et adjectivales, élaboration du lexique-grammaire des adjectifs déverbaux*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.

XII] Dictionnaires et grammaires :

- BUSSE, W. (dir.), 1994, *Dicionario sintactico de verbos portugueses*, Coimbra, Almedina.
- BUSSE, W., DUBOST, J.-P., 1983, *Französische Verblexicon*, Stuttgart, Klett.
- CELERIER, P., MAILLARD, J.-P., 1979, *Dictionnaire des Structures Fondamentales du français* (D.S.F.), Paris, CLE International.
- CHEVALIER, J.-C., BLANCHE-BENVENISTE, C., ARRIVE, M. et PEYTARD, J., 1964, *Grammaire Larousse du français contemporain*, Paris, Larousse.
- TRESOR DE LA LANGUE FRANÇAISE (T.L.F.), dictionnaire consulté sur Internet à l'adresse : <http://atilf.inalfr.fr/dendien/scripts/tlfrv4/showps.exe?p=combi.htm;java=no>;
- EYNDE, K. VAN DEN, MELIS, L., MERTENS, P., VANGILBERGEN, L., DEHASPE, L., EGGERMONT, C., BROEDERS, E., *Dictionnaire de valence PROTON*, consultable sur Internet à l'adresse <http://bach.arts.kuleuven.ac.be/PA/protonfr.html>.
- EYNDE, K. van den, MERTENS, P., 2002, *La syntaxe du verbe, l'approche pronominale et le lexique de valence PROTON*, Louvain, document preprint n° 174, téléchargeable à l'adresse suivante : <http://bach.arts.kuleuven.ac.be/pmertens/papers/proton2002.pdf>
- RIEGEL, M., PELLAT, J.-C., RIOUL, R., 1994, *Grammaire méthodique du français*, Paris, P.U.F., coll. « Linguistique nouvelle ».
- ROBERT, P., (texte remanié et amplifié sous la direction de REY-DEBOVE, J. et de REY, A., 1996, *Version électronique du Nouveau Petit Robert, Dictionnaire analogique et alphabétique de la langue française*, Paris, Dictionnaires Le Robert.
- WILLEMS, D., 2002b, « Sémantique et syntaxe contrastives. Le dictionnaire contrastif des valences verbales : présentation et illustration » in E. Castagné (ed.) *Pour une modélisation*

de l'apprentissage simultané de plusieurs langues apparentées ou voisines, Publications de la Faculté de Lettres de Nice, CID Diffusion, Paris, pp. 45-66.

XIII] Logiciels (références et articles) :

1- Logiciels :

- Xcor 2000, concordancier créé par Jean-Pierre ADAM pour le traitement informatique des corpus de langue parlée du GARS.
- Access 2000, juin 1999, logiciel de base de données, © Microsoft Corporation.
- Word 2000, juin 1999, traitement de texte, © Microsoft Corporation.

2- Articles :

- ADAM, J.-P., 2000, « Exploitation informatique des corpus de langue parlée au GARS », *in Corpus. Méthodologie et applications linguistiques*, éd. Par M. Bilger, Paris, Honoré Champion et Presses Universitaires de Perpignan, pp. 121-131.

Index des noms cités

- Adam, J.-M. : 40, 42, 43, 48, 49
- Adam, J.-P. : 12, 64, 113, 137
- Bakhtine, M. : 40
- Beauzée, N. : 79, 82, 89
- Benveniste, E. : 81, 82
- Benzitoun, C. : 53
- Bergougnioux, M. : 33
- Biggs, P. : 18
- Bilger, M. : 9, 18, 19, 36, 40, 41, 44, 49, 134
- Blanc, M. : 18
- Blanche-Benveniste, C. : 9, 10, 18, 19, 33, 36, 39-41, 44, 61, 66-68, 71, 73, 86, 91, 94, 96, 99, 103, 104, 119-121, 124, 134
- Bonami, O. : 9, 69, 70, 72, 75, 136
- Boons, J.-P. : 54, 60, 65, 117
- Broeders, E. : 55
- Brunot, F. : 14, 78-82
- Burnard, L. : 47
- Busse, W. : 7, 55
- Charolles, M. : 67
- Chevalier, J.-C. : 67
- Chu, X.-Q. : 124
- Combe McBride, N. : 65
- Creissels, D. : 78, 81, 88, 89, 95, 136
- Delais-Roussarie, E. : 14, 30
- Dubost, J.-P. : 55
- Dubuisson, P. : 21
- Durand, J. : 14, 30, 134
- Eggermont, C. : 54
- Eluerd, R. : 80, 81, 83, 84
- Eynde, K. van den : 54, 66, 76, 84, 92-94, 99, 100, 103, 104
- Freyssinet-Dominjon, J. : 30
- Fujimura, I. : 47
- Gaatone, D. : 110
- Gadet, F. : 20, 25, 26, 31, 133, 134
- Gillieron, J. : 14
- Gougenheim, G. : 58
- Greidanus, T. : 9, 10, 52, 54, 57, 58, 60, 135
- Gross, G. : 121
- Gross, M. : 54, 60-64, 89, 99, 100, 109, 121, 129, 136, 138
- Jeanjean, C. : 33, 90
- Labov, W. : 8, 27
- Laks, B. : 134
- Le Goffic, P. : 54, 63, 65, 67
- Lyche, C. : 134
- Maingueneau, D. : 42
- Melis, L. : 52, 55

- Mertens, P. : 84, 92, 100, 103
- Milroy, J. : 8, 27
- Moignet, G. : 79, 81
- Morel, M.-A. : 67
- Queneau, R. : 104
- Riegel, M. : 67, 118, 121, 122
- Tesnière, L. : 69, 74
- Tuailon, G. : 22, 113
- Vaguer, C. : 47

Glossaire¹

Approche Pronominale : cadre théorique d'analyse syntaxique qui utilise les pronoms pour donner une typologie des différents types de construction (ou de rection) verbale.

Corpus : ensemble limité de parole produite par un ou plusieurs locuteurs, cette production pouvant être orale (discours politique, conversation...) ou écrite (lettre, mémoire(s)...).

Élément construit (ou régi) : terme utilisé par l'Approche Pronominale pour désigner un élément construit par un verbe recteur, caractérisé par une proportionnalité à une pro-forme, une organisation paradigmatique, la possibilité d'entrer dans plusieurs dispositifs et de prendre les modalités du verbe.

Modalités : relations marquées par des morphèmes d'affirmation, de négation, d'interrogation, de restriction.

Pro-forme : terme générique qui désigne tout élément qui entre en relation systématique avec du lexique et qui correspond à des noms, à des adjectifs, à des adverbes, à des verbes, à des Que-Phrases.

Proportionnalité : dans l'Approche Pronominale, équivalence syntaxique entre un élément construit et son homologue pronominal.

Rection : désigne les éléments construits par le verbe, qu'ils entrent ou non dans la valence de ce dernier.

Valence : sous-partie de la rection verbale qui caractérise le sens et la construction minimale du verbe. Désigne le ou les éléments de la construction verbale qui ont une relation de proportionnalité avec une pro-forme.

¹ Ce glossaire est largement inspiré de celui de Blasco-Dulbecco (1999 : pp. 209-212)

Verbe recteur : lexème verbal qui sélectionne les éléments spécifiques d'une construction syntaxique.

Inventaire et classement des constructions verbales dans un corpus de français parlé dans l'Allier

Résumé :

Nous proposons de nous intéresser à une méthode de classement des constructions verbales relevées dans un corpus de français parlé (environ 15 000 constructions inventoriées).

Cette méthode présente quelques avantages : une représentation en tableau permet d'avoir une vision synthétique des types de constructions relevés dans notre corpus et de les classer selon la forme du verbe et les traits syntactico-sémantiques des sujets et compléments qu'il régit. En outre, le résultat de ce travail pourrait servir de base à un essai de typologie verbale, effectué en fonction de la complémentation des verbes étudiés.

Nous présentons en détail la méthode utilisée pour ce travail depuis la constitution du corpus (enquêtes de terrain et transcriptions), jusqu'à l'inventaire et au classement des constructions verbales.

Inventory and classification of verbal constructions in a French spoken corpus based in Allier

Abstract:

We put forward a method of classification of verbal constructions coming from a French spoken corpus (about 15,000 listed constructions).

The advantages of this method are the following : a representation in a table permits to have a synthetic vision of the types of constructions found in our corpus and to classify them according to the form of the verb and the syntactico-semantic features of the subjects and the complements which it governs. Moreover, the result of this work could be used as a basis for a verbal typology, carried out according to the complementation of the studied verbs.

We will present in detail the method used for this work from the constitution of the corpus (investigations and transcriptions), to the inventory and the classification of verbal constructions.

Sciences du langage

Mots-clefs : verbe, constructions verbales, syntaxe, français parlé, corpus.

Laboratoire de Recherche sur le Langage, 4 rue Ledru, 63 057 Clermont-Ferrand Cedex 1