Linguistische Arbeiten 243

Herausgegeben von Hans Altmann, Peter Blumenthal, Herbert E. Brekle, Hans Jürgen Heringer, Heinz Vater und Otmar Werner

Tine Greidanus

Les constructions verbales en français parlé

Etude quantitative et descriptive de la syntaxe des 250 verbes les plus fréquents

Max Niemeyer Verlag
Tübingen 1990

CIP-Titelaufnahme der Deutschen Bibliothek

Greidanus, Tine : Les constructions verbales en français parlé : Etude quantitative et descriptive de la syntaxe des 250 verbes les plus fréquents / Tine Greidanus. – Tübingen : Niemeyer, 1990.
(Linguistische Arbeiten ; 243)
NE: GT

ISBN 3-484-30243-7 ISSN 0344-6727

© Max Niemeyer Verlag GmbH & Co. KG, Tübingen 1990
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Printed in Germany.
Satz: Lichtsatz Walter, Tübingen-Lustnau.
Druck: Allgäuer Zeitungs-Verlag, Kempten.
Einband: Heinrich Koch, Tübingen.

Préface

Nous voudrions exprimer ici notre reconnaissance à toutes les personnes sans qui nous n'aurions pu réaliser les recherches rapportées dans cet ouvrage.

Le professeur Bernard Al a bien voulu diriger notre recherche. C'est lui qui, connaissant notre goût pour un travail ayant une dimension appliquée, nous en a suggéré le sujet. Les discussions que nous avons eues avec lui nous ont souvent ouvert de nouvelles perspectives. Il a lu avec soin les chapitres que nous lui avons soumis et a contribué, par ses réflexions constructives, à en améliorer le contenu et la rédaction. Qu'il soit assuré de notre gratitude.

Le professeur Claire Blanche-Benveniste a accepté dans un stade déjà plus avancé des recherches d'être co-directeur de thèse. Ses suggestions, ses remarques et ses encouragements nous ont été très précieux.

Ce travail doit beaucoup à l'apport de plusieurs informaticiens. Sans Onno Huber cet ouvrage ne serait pas ce qu'il est devenu, tant sur le plan du contenu que sur celui de la forme. Il a créé les nombreux programmes nécessaires pour informatiser les données. C'est lui qui nous a initiée aux secrets, parfois impénétrables, du programme de traitement de texte LaTeX et a fait les adaptations nécessaires à nos besoins particuliers. Qu'il en soit remercié de tout coeur! Nos remerciements vont également à Arie Dirk Schenkeveld, qui nous a assistée au début de l'entreprise, et à Paul Janssen, qui a aidé à éditer le lexique.

Deux chapitres de cet ouvrage ont une composante statistique. Nous remercions Sophie de Schaepdrijver et Kees Mandemakers, qui nous ont conseillée dans le choix des tests statistiques utilisés dans le premier chapitre et qui ont exécuté les analyses. Harald Baayen a bien voulu appliquer à nos données un modèle statistique développé par lui. Nous lui devons la figure 4.2.

Jacqueline van Baardewijk-Rességuier nous a beaucoup obligée en acceptant de relire le manuscrit complet. Elle a ainsi contribué à améliorer le texte. Nous l'en remercions vivement.

Que soit enfin remerciée la Faculté des Lettres de l'Université Libre d'Amsterdam, qui, en nous accordant un congé d'études, nous a permis de mener à bien cette recherche.

Nous dédions ce livre à la mémoire de nos parents.

Table des matières

Introduction 1

1 Matériaux analysés 7
- 1.1 Verbes étudiés . . . 7
  - 1.1.1 Introduction . . . 7
  - 1.1.2 Nombre de verbes à étudier . . . 7
  - 1.1.3 Quels verbes? . . . 10
  - 1.1.4 Listes de fréquence utilisées. Procédures de sélection appliquées . . . 11
    - 1.1.4.1 Le Dictionnaire des fréquences . . . 11
    - 1.1.4.2 Le Frequency Dictionary of French Words . . . 12
    - 1.1.4.3 Le Vocabulaire du roman français 1962–1968 . . . 12
    - 1.1.4.4 L'Enquête statistique sur la langue parlée . . . 12
    - 1.1.4.5 L'Etude sociolinguistique sur Orléans . . . 13
    - 1.1.4.6 Principales caractéristiques des cinq corpus . . . 13
  - 1.1.5 Remarques sur les cinq listes de verbes. Sélection des verbes retenus pour notre recherche . . . 13
- 1.2 Le corpus d'Orléans . . . 20

2 Le domaine du verbe 23
- 2.1 Introduction . . . 23
- 2.2 Les trois types de compléments . . . 24
  - 2.2.1 Les arguments du verbe . . . 24
  - 2.2.2 Les compléments du syntagme verbal . . . 24
  - 2.2.3 Les compléments de phrase . . . 25
- 2.3 Comment identifier les trois types de compléments? . . . 26
  - 2.3.1 Eléments régis par le verbe et éléments non régis . . . 26
  - 2.3.2 Arguments et compléments du syntagme verbal . . . 27
- 2.4 Compléments retenus dans notre recherche . . . 29

3 Principes d'analyse 31
- 3.1 Principes généraux . . . 31
- 3.2 Les constituants arguments du verbe . . . 33
  - 3.2.1 Le syntagme nominal . . . 33

3.2.2 Le syntagme prépositionnel 35
3.2.3 Le pronom 36
3.2.3.1 Le pronom personnel 36
3.2.3.2 Le pronom interrogatif 36
3.2.3.3 Le pronom clitique 'se' 37
3.2.4 La subordonnée 38
3.2.5 La phrase 39
3.2.6 L'infinitif 39
3.2.7 L'adjectif 40
3.2.8 L'adverbe 42
3.3 Les opérateurs métadiscursifs 42
3.4 Les compléments du syntagme verbal 43
3.5 Autres décisions concernant le dépouillement du corpus 43
3.5.1 Les phrases à construction passive, à construction impersonnelle et à construction pronominale à valeur passive 43
3.5.2 Place occupée par deux éléments solidaires 44
3.5.3 Coordination et juxtaposition 45
3.5.4 La phrase négative et la phrase interrogative 46
3.5.5 L'ordre des constituants 47
3.5.6 Temps et mode 47
3.5.7 Constructions caractéristiques de la langue parlée 47
3.6 Quelques remarques concernant la réalisation pratique du dépouillement . . 49
3.7 Le lexique 50

4 Les constructions verbales dans la langue parlée 53
4.1 Introduction 53
4.2 Constructions verbales relevées dans le corpus d'Orléans 54
4.3 Construction active et constructions dérivées 67
4.3.1 Introduction 67
4.3.2 La construction active 67
4.3.2.1 Remarques générales 67
4.3.2.2 La place à gauche du verbe 69
4.3.2.3 La première place à droite du verbe 71
4.3.2.4 La seconde place à droite du verbe 73
4.3.3 La construction passive 75
4.3.4 La construction pronominale à valeur passive 77
4.4 Constructions caractéristiques de la langue parlée 79
4.5 Les constructions fondamentales 80
4.6 Caractéristiques syntaxiques et sémantiques communes aux différents verbes 82
4.7 Conclusion 83

5 La construction impersonnelle **85**
5.1 Introduction 85
5.2 il + V 86
5.3 il + V + séquence phrastique 86
5.3.1 Les verbes exclusivement impersonnels 86
5.3.2 Les verbes à construction impersonnelle 88
5.3.3 Les verbes auxiliaires 93
5.4 il + V + séquence nominale 95
5.4.1 Les verbes exclusivement impersonnels 95
5.4.2 Les verbes à construction impersonnelle 97
5.4.3 Caractéristiques sémantiques des verbes entrant dans la construction 'il V séquence nominale' 97
5.4.4 Propriétés syntaxiques et sémantiques du constituant nominal . . . 99
5.5 Raisons d'être de la construction impersonnelle 101
5.6 Conclusion 106

Conclusion **109**

Références bibliographiques **113**

Appendice A **119**

Appendice B **129**

Appendice C **139**

Lexique **149**

Introduction

"Le verbe est l'âme d'une langue. C'est, comme on l'a fort bien dit, le mot par excellence",[1] "le verbe est un PIVOT: c'est sur lui que s'appuient sujets et objets pour fixer leurs rapports réciproques; le verbe les *met en rapport*",[2] "régissant de toute la phrase verbale",[3] "élément indispensable à la constitution d'un énoncé assertif fini",[4] "élément générateur de la phrase",[5] "porteur privilégié d'indices qui affectent le sens de la phrase entière",[6] voilà quelques affirmations soulignant la place centrale du verbe à l'intérieur de la phrase. C'est un point sur lequel toutes les théories grammaticales s'accordent, explicitement ou implicitement. Des recherches en psycho-linguistique ont démontré l'importance du verbe pour la compréhension de la phrase.[7] Qu'un mot ayant une fonction aussi importante ait donné lieu à une littérature abondante n'a rien de surprenant. Nombreuses sont les études consacrées au rôle du verbe en général ou à la description de catégories de verbes ou de constructions en particulier. Il y a d'autre part des dictionnaires de verbes. Il s'agit dans pratiquement tous les cas[8] d'études qui décrivent les propriétés syntaxiques et lexicales théoriques de verbes. Nous sommes assez bien au courant des propriétés théoriques de (catégories de) verbes, mais assez mal des propriétés réalisées dans l'usage concret. Une meilleure connaissance des constructions syntaxiques observées aurait un intérêt certain, aussi bien en linguistique descriptive qu'en linguistique appliquée.

En linguistique descriptive, la méthode généralement adoptée pour décrire un phénomène est celle de l'introspection. Quand il s'agit de savoir si un énoncé donné est acceptable ou non, le linguiste se basera sur sa propre intuition et éventuellement sur celle d'autres

---

1 Duhamel (1934:43).

2 *Grammaire Larousse* (1964;1969:75)

3 Tesnière (1959;1976:103).

4 Benveniste (1950:22).

5 von Wartburg & Zumthor (1947;1958:181).

6 Fourquet (1950:78).

7 Voir par exemple Fodor, Garrett & Bever (1968), Fodor (1971), Schank (1972), Moore & Biederman (1979).

8 Nous ne connaissons, pour le français, que deux ouvrages basés entièrement ou en partie sur l'étude d'un corpus: Corbeil (1971) et Roy (1976).

informateurs, le plus souvent d'autres linguistes. Une telle procédure peut être critiquée d'un point de vue méthodologique.[9] D'après Al (1975:11–19), le linguiste qui veut obtenir des jugements d'acceptabilité fiables devrait:

1. se servir d'un grand nombre d'informateurs,

2. contrôler les conditions dans lesquelles l'épreuve d'acceptabilité a lieu, et

3. choisir la méthode adéquate pour obtenir le jugement d'acceptabilité.

Or, il est extrêmement rare que ces conditions soient remplies. Ce qui a sans aucun doute contribué à la popularité de la méthode de l'introspection est sa facilité d'emploi apparente. L'utilisation d'un corpus par contre est laborieuse, surtout celle d'un corpus non enrichi d'informations linguistiques.

Gross (1979) critique vivement la grammaire générative-transformationnelle pour "the anecdotal character of data in GG". Il donne plusieurs exemples de règles transformationnelles qui reposent sur des données dont ses recherches personnelles ont montré qu'elles constituent une exception. Gross est d'avis que, dans l'état actuel des connaissances, l'ensemble des données linguistiques est trop réduit pour permettre la découverte des principes généraux abstraits qui sont à la base de toutes les langues humaines:

> It is virtually an absolute rule in all natural sciences that accumulation of consistent data precedes theoretical advances. Accumulating and classifying facts, and constructing a global image of the language under study, is the major experimental process, all the more in that syntactic problems have never been approached in this way. (p. 568)

L'importance des questions théoriques s'efface devant "la nécessité d'accumuler des données empiriques aussi complètes et détaillées que possible" (Gross 1975:45). Gross et ses collaborateurs du Laboratoire d'Analyse Documentaire et Linguistique visent à rassembler ces données en décrivant de façon systématique les propriétes syntaxiques de l'ensemble du lexique français.

Nous partageons le point de vue de Gross quant à l'importance des activités taxonomiques pour la construction de modèles. Et il ne fait aucun doute que les travaux du LADL constituent une contribution extrêmement importante à une meilleure connaissance de la langue française. Mais les principes méthodologiques mis en oeuvre appellent aussi quelques remarques critiques:

1. Les jugements d'acceptabilité proviennent en général d'une seule personne, le plus souvent le linguiste lui-même. Levelt (1972) montre combien peu les linguistes sont d'accord entre eux quand il s'agit de donner un jugement d'acceptabilité. On devrait donc se méfier d'un jugement d'acceptabilité émis par une seule personne, qui plus est, un linguiste.

---

[9] Voir par exemple Levelt (1972), Levelt (1973,III:13–19), Al (1975:11–19) et Gross (1979).

2. Les jugements portés sont binaires. Un énoncé est considéré comme grammatical ou comme agrammatical, les degrés intermédiaires n'existent pas. Le français étudié est un français standard. Il est possible que des phrases inacceptables à ce niveau de langue, mais acceptables à un autre niveau aient été considérées comme agrammaticales.

3. Les chercheurs donnent peu d'informations au sujet des tests utilisés. Les tests ne peuvent donc pas être reproduits. Cela va à l'encontre des "bonnes procédures scientifiques" exigées par Gross lui-même (1975:13).

4. L'acceptabilité a parfois été forcée pour des raisons théoriques. Ce procédé nous semble discutable d'un point de vue méthodologique: des données réunies dans le but de permettre la construction de modèles théoriques ne devraient pas être manipulées auparavant.

L'utilisation d'un corpus pourra pallier en partie ces inconvénients:

1. Si le corpus a été composé avec soin, les énoncés proviennent d'un grand nombre d'informateurs, variant en âge, en sexe et en position sociale.

2. Les données concernant la fréquence d'un phénomène permettent de distinguer les cas nets et les cas douteux. On pourrait se demander si une construction fréquemment réalisée ne devrait pas avoir un autre statut qu'une construction rarement ou même jamais attestée. Il est vrai qu'un corpus ne contient pas d'énoncés impossibles, argument souvent avancé contre l'utilisation d'un corpus. Un corpus par contre peut contenir des énoncés ne répondant pas à la norme, mais apparemment possibles, car réalisés. Or, nous savons depuis Frei (1929) que les 'fautes' peuvent contribuer à expliquer les phénomènes qui sous-tendent le fonctionnement du langage.

Nous pensons que l'utilisation d'un corpus se justifie par le fait qu'il fournit des informations qu'on ne pourrait jamais ou difficilement obtenir par le biais d'autres procédés méthodologiques. Un corpus permet notamment de déterminer quels sont les phénomènes fréquents et quelles sont les exceptions. Il contribue ainsi à renforcer l'adéquation descriptive d'une théorie grammaticale. Un corpus peut en plus permettre la découverte de faits nouveaux. Pour ce qui est des constructions, Aarts et Meys (1989:9) affirment que la variété des constructions réalisées dans un corpus se révèle toujours plus grande qu'on n'aurait tendance à penser sur la seule base d'ouvrages théoriques. Puis, un corpus constitue un ensemble de textes continus. Une linguistique qui se base sur un corpus de textes ne décrira pas uniquement la structure d'énoncés isolés, mais aussi celle d'énoncés dans leur contexte, ou celle de textes. Pour des exemples précis de ce dernier aspect nous renvoyons au chapitre 5.

Une étude des constructions syntaxiques utilisées se justifie aussi en linguistique appliquée. On pourrait penser à des applications en traduction automatique, ou à l'enseignement du français langue maternelle ou étrangère. De nombreuses publications allemandes ont souligné la pertinence de la valence du verbe pour l'enseignement.[10] Le verbe est pour

[10] Voir par exemple pour l'enseignement du français Happ (1978) et Zöfgen (1982).

une large part responsable de la structure de base d'un énoncé. Son impact sur la phrase est donc plus grand que celui d'autres formes lexicales. Il mérite par conséquent une attention particulière dans l'enseignement. Un produit pratique de la recherche valencielle est le *Französisches Verblexikon* de Busse et Dubost (1983). Ce dictionnaire valenciel bilingue a pour objectif de décrire les constructions courantes d'environ 5 000 verbes français. Il s'agit plutôt d'un ouvrage de référence que d'un manuel. Sur le plan théorique il a certains défauts.[11] En France, les applications didactiques de recherches sur les constructions verbales sont plus rares. L'étude de Le Goffic et Combe McBride, qui se base sur les travaux du LADL, est un exemple d'une application de recherches théoriques. Le livre présente un inventaire des constructions possibles à partir des mots (verbes, adjectifs et substantifs opérateurs) du *Français Fondamental* (1er degré). Les quelque 300 verbes du *FF1* sont classés selon les constructions dans lesquelles ils peuvent entrer. Selon les calculs de Happ (1978:129), le nombre des constructions verbales s'élèverait à 35 environ. Il s'agit d'un inventaire de constructions potentielles. Dans leur 'Introduction', les auteurs exposent pourquoi ils n'ont pas compté les structures et ainsi tenu compte de leur fréquence. Les principales raisons concernent l'imprécision de la notion de structure et l'impossibilité de trouver un corpus représentatif du français en général. Nous sommes d'avis que ce sont là en effet des problèmes, mais que ces problèmes ne sont pas insurmontables. Le dépouillement d'un corpus pourrait d'autre part servir à éclaircir le problème de l'identité et des limites des éléments qui, combinés avec le verbe, forment une unité structurale.

Nous avons choisi de travailler sur un corpus oral.[12] Et ce, pour plusieurs raisons. La première est que la langue parlée reflète peut-être mieux les principes fondamentaux du langage que la langue écrite, plus travaillée. Puis, la langue écrite a été mieux étudiée que la langue parlée.[13] Il est donc plus utile de prendre celle-ci pour objet d'analyse. La troisième raison est d'ordre plus pratique. Il n'existe pas, à notre connaissance, de corpus de textes français enrichis d'informations syntaxiques. Le dépouillement des constructions devra donc entièrement être fait à la main. Pour l'instant, il n'existe pas encore d'analyseur automatique pouvant faire ce travail de façon satisfaisante. Dans quelques années, les programmes

---

[11] Les défauts les plus importants sont les suivants:

1. Les compléments verbaux sont classés tantôt d'après la catégorie grammaticale, tantôt d'après la fonction, tantôt d'après un critère d'ordre sémantique.
2. Certaines informations sont données de façon systématique, alors que d'autres ne le sont pas.
3. Les possibilités d'opération sur le verbe (passivation, extraposition, construction moyenne) ne sont pas indiquées du tout, ou pas de façon systématique.
4. Le traitement des verbes dits essentiellement pronominaux est peu systématique.

Pour une critique détaillée voir Zöfgen (1982).

[12] Il s'agit du corpus d'Orléans. Ce corpus, qui sera présenté en 1.2, répond aux conditions formulées plus haut: les informateurs ont été sélectionnés en fonction de leur âge, de leur sexe et de leur position sociale.

[13] Voir Blanche-Benveniste et Jeanjean (1987) pour une description de l'attitude des linguistes et des grammairiens à l'égard du français parlé.

seront probablement suffisamment sophistiqués pour permettre l'analyse automatique de la langue écrite. Pour la langue parlée ce sera beaucoup plus difficile à réaliser. C'est pourquoi nous préférons prendre un corpus de textes parlés.

Nous nous proposons d'étudier les constructions des verbes français les plus fréquents. Les verbes fréquents sont des verbes très riches en constructions. Nous supposons qu'en décrivant leurs constructions, nous couvrons à peu de choses près les constructions possibles de tous les verbes français. L'approche par le biais d'un certain nombre de verbes permet de concilier deux points de vue complémentaires: le point de vue syntaxique (étant donné une construction, quels sont les verbes pour lesquels elle se réalise?), et le point de vue lexicographique (étant donné un verbe, quelles sont les constructions dans lesquelles il entre?).

Les trois premiers chapitres de l'étude présentée ici sont consacrés à l'exposition de la démarche adoptée lors du dépouillement du corpus. Le choix des verbes à analyser constitue l'objet du premier chapitre. Nous comparons à cet effet les 250 verbes les plus fréquents de cinq listes de fréquence. Ce chapitre contient en outre une présentation du corpus d'Orléans. Le second et le troisième chapitre ont pour sujet la question de savoir quoi et comment compter. Les résultats globaux du dépouillement sont présentés dans le quatrième chapitre. Le cinquième et dernier chapitre donne une analyse détaillée d'une partie de nos matériaux. Il a pour objet la construction impersonnelle. Nous aimerions souligner ici que le dépouillement d'un corpus, tel que nous l'avons entrepris, procure un grand fonds de données. Une petite partie seulement de toutes ces données pourra être exploitée dans cet ouvrage. Le lexique, qui constitue la dernière partie de ce livre, présente les données de chaque verbe étudié à la manière d'un dictionnaire. Chaque entrée mentionne le nombre total des occurrences d'un verbe, les constructions dans lesquelles il entre et, pour chaque construction, le nombre des occurrences, un ou plusieurs exemples et, s'il y a lieu, des remarques.

Chapitre 1

Matériaux analysés

1.1 Verbes étudiés

1.1.1 Introduction

Notre objectif est d'étudier les constructions des verbes les plus fréquents en français. On peut se demander de combien de verbes se compose ce groupe des verbes les plus fréquents et quels sont les verbes qui en font partie. Mais il est impossible de donner une réponse absolue à ces questions. En effet, le groupe des verbes les plus fréquents en tant que tel n'existe pas. Quand on range les verbes sur une échelle de fréquence, on ne constate aucune rupture entre verbes fréquents et verbes non fréquents. Il faudra déterminer quantitativement, en fonction de critères externes, l'étendue du groupe visé. Une fois le nombre de l'ensemble visé déterminé, il faut savoir quels sont ces verbes. Car cela ne va pas de soi: le groupe des 20, 50 ou 100 verbes les plus fréquents en français selon telle liste de fréquence n'est pas identique à celui d'une autre liste. Dans ce qui suit nous essaierons donc de nous doter de critères pour répondre à ces deux questions. Les recherches faites pour répondre à la seconde question (quels verbes?) nous ont fourni des matériaux qui nous permettent de faire des remarques de nature générale concernant les listes de fréquence. Nous en ferons état dans les différentes sections.

1.1.2 Nombre de verbes à étudier

Nous avons d'abord cherché à déterminer le nombre de verbes à étudier. Nous nous sommes posé la question suivante: si nous décidons de décrire les verbes faisant partie du groupe des 1 000 mots les plus fréquents du français, combien de verbes obtenons-nous, et est-il possible de déterminer le pourcentage que représentent ces verbes parmi tous les verbes utilisés dans un texte français moyen?

Parmi les 1 000 mots les plus fréquents du français, combien y a-t-il de verbes? Ou plutôt, les listes de fréquence intéressantes pour notre objectif combien de verbes contiennent-elles parmi les 1 000 mots les plus fréquents? Le tableau 1.1 rassemble les données concernant quatre listes de fréquence: la *Table des fréquences décroissantes* du

Dictionnaire des fréquences du *Trésor de la Langue Française* de Nancy (XIXe et XXe siècles) (désormais *TLF*), la liste des fréquences décroissantes de Juilland, Brodin et Davidovitch, *Frequency Dictionary of French Words* (désormais *FDFW*), l''Index hiérarchique des cinq mille lemmes majeurs' de Engwall, *Vocabulaire du roman français (1962–1968)* (désormais *VRF*) et 'La liste des mots par fréquences décroissantes' de 'L'Enquête statistique sur la langue parlée', donnée par Gougenheim *et al.* (1967) (désormais *ESLP*). Nous donnons, pour chaque liste, le nombre de verbes par tranche de 500 mots.[1]

	1e tranche de 500 mots	2e tranche de 500 mots	total sur 1 000 mots
TLF	84	108	192
FDFW	110	129	239
VRF	135	159	294
ESLP	101	114	215

Tableau 1.1: Nombre de verbes par tranche de 500 mots dans quatre listes de fréquence

Ces chiffres doivent être considérés avec une certaine prudence. C'est que les listes ne sont pas tout à fait comparables, étant donné que le regroupement des formes est différent pour chaque liste (phénomène qui rend épineuse toute comparaison de listes de fréquence), et que le *FDFW* contient des formes homographes. Nous reviendrons plus loin sur ce problème. On constate d'assez gros écarts entre les listes pour ce qui est du nombre de verbes parmi les 1 000 mots les plus fréquents. Le *TLF*, qui a le moins de verbes, et le *VRF*, qui en a le plus, forment les extrêmes, alors que les deux autres listes ont un nombre de verbes qui tient le milieu entre ces deux extrêmes. Que le *TLF* ait moins de verbes que les trois autres listes s'explique par le fait qu'il regroupe relativement peu les mots grammaticaux et les formes verbales. Il a donc, parmi les 1 000 mots les plus fréquents, moins de lemmes différents que les trois autres listes, qui regroupent davantage. Le *FDFW* a des verbes homographes, ce qui gonfle évidemment l'effectif total. Cette liste contient en réalité 230 verbes différents parmi les 1 000 mots les plus fréquents. Le nombre élevé de verbes dans le *VRF* est difficile à expliquer. Il est vrai qu'il regroupe beaucoup, mais il n'est pas sûr que cette donnée à elle seule puisse expliquer l'écart dans sa totalité. Il est possible qu'il faille chercher dans la nature du corpus du *VRF* l'explication de ce phénomène. L'ampleur du corpus est de toute façon un facteur qui ne peut pas jouer de rôle.[2] Nous ne pouvons examiner ici cette question, qui dépasse le cadre de notre étude. La liste de l'*ESLP* a ceci de commun avec celle du *VRF* que les formes verbales ont toutes été regroupées sous l'infinitif et qu'elle ne connaît pas d'homographes. Pour les quatre listes

[1] (1) Quant au *TLF*, il s'agit en réalité, non des verbes, mais des infinitifs de cette liste.
(2) La seconde tranche du *VRF* contient 493 mots, et non 500. Ceci pour des raisons techniques.

[2] Pour être fiable, une liste de fréquence doit être basée sur un corpus suffisamment grand. Sciarone (1979:50) pose qu'un mot devrait avoir une fréquence absolue égale ou supérieure à 40 pour avoir une fiabilité satisfaisante. Les derniers verbes de la liste des 294 verbes du *VRF* ont une fréquence de 53, et satisfont donc largement au critère de Sciarone.

on constate que la première tranche contient moins de verbes que la seconde. Cela est dû au nombre relativement élevé des mots grammaticaux dans la première tranche de 500 mots. La moyenne des trois listes où toutes les formes verbales ont été lemmatisées sous l'infinitif (*FDFW*, *VRF* et *ESLP*) est de 246 verbes sur les 1 000 mots les plus fréquents.

Ces 1 000 mots les plus fréquents, combien de mots d'un texte français moyen couvrent-ils? Le *FDFW* fournit les chiffres suivants pour sa liste de fréquence et le corpus utilisé: "the first ten words account for 30.52% of the total occurrences, the first hundred for 62.66%, the first thousand for 85.58%"(p.LXI). Ces chiffres (qui se rapportent au corpus du *FDFW*, non à un texte quelconque) correspondent avec ceux donnés par Guiraud (cf. Sciarone 1979:51) d'après qui les 100 mots les plus fréquents couvrent 60% d'un texte quelconque, les 1 000 mots les plus fréquents 85% et les 4 000 mots les plus fréquents 97.5%, tandis que le reste du vocabulaire ne couvre que 2.5%. Sciarone est d'avis que ces estimations sont trop optimistes.[3] Lui-même (1977:91) arrive, pour l'italien, à un pourcentage de 80% pour 1 000 mots. Il attribue ces différences à l'étendue des corpus qui ont servi de base aux calculs: 500 000 occurrences pour le *FDFW* et un million et demi pour le sien. Nous pensons qu'il est prudent de nous en tenir à une estimation conservatrice et de poser que les 1 000 mots les plus fréquents ont un pourcentage de couverture de 80%.

On peut se demander s'il est permis d'extrapoler ce chiffre et de l'appliquer à la catégorie des verbes, et d'affirmer que les verbes de ces 1 000 mots les plus fréquents couvrent 80% des verbes utilisés dans un texte français moyen. Pareille généralisation serait fausse à notre avis. C'est que les premières tranches de fréquence contiennent plus de verbes que les tranches qui viennent après. Le *FDFW* a 239 verbes sur les 1 000 premiers mots (= 24%) et 956 verbes sur les 5 000 mots les plus fréquents (= 19%). Le *VRF* en a respectivement 294 (= 29%) et 1 222 (= 24%). Il est frappant de constater que l'écart entre les deux pourcentages est le même pour les deux listes de fréquence, c'est-à-dire 5%. *FDFW* et *VRF* s'arrêtent aux 5 000 mots les plus fréquents. Nous ne pouvons donc pas nous prononcer sur le pourcentage des verbes dans les tranches au-delà de ce seuil, mais il est probable qu'il diminuera. Sciarone (communication personnelle) a calculé que les verbes du *Dictionnaire du français contemporain* (environ 25 000 entrées) représentent 16% environ de toutes les entrées. Les 1 000 mots les plus fréquents ont donc un pourcentage de verbes plus élevé que le vocabulaire total. Les verbes de ces 1 000 premiers mots doivent par conséquent couvrir plus de 80% de tous les verbes utilisés dans un texte moyen. Un autre calcul nous fournit un résultat similaire. Sciarone (communication personnelle) a calculé, pour les verbes des 10 tranches de 500 mots du *FDFW*, le pourcentage de couverture sur l'ensemble du corpus du *FDFW*. Le tableau 1.2 présente ces chiffres. On constate que le nombre de verbes diminue lentement après la deuxième tranche. Nous voyons d'autre part que le pourcentage de couverture diminue rapidement. Les verbes des deux premières tranches (239) couvrent 12.31% de tous les mots du corpus, les 117 verbes de la troisième tranche y ajoutent 0.9% seulement, et les tranches suivantes encore moins.

---

[3] D'autres chiffres peuvent confirmer cette opinion. Selon Engwall (1984:XL) les cinq mille vocables les plus fréquents couvrent "environ 90 pour cent du texte courant."

	nombre de verbes	% de couverture
1re tranche de 500 mots	110	10.6802
2e	129	1.6284
3e	117	.8948
4e	86	.4554
5e	103	.4042
6e	93	.2732
7e	83	.1920
8e	82	.1544
9e	79	.1168
10e	74	.0860
Total	956	14.8854

Tableau 1.2: Pourcentages de couverture des verbes du *FDFW* sur l'ensemble du corpus

Si nous divisons le pourcentage de couverture des verbes des deux premières tranches par celui de tous les verbes du *FDFW*, nous obtenons un résultat de 82.7%.

Ces chiffres nous confirment dans notre opinion que le groupe des verbes rencontrés parmi les 1 000 mots les plus fréquents représente un choix tout à fait raisonnable pour ce qui est du nombre des verbes à étudier.

Nous avons trouvé plus haut une moyenne de 246 verbes sur les 1 000 mots les plus fréquents. Si nous décidons d'étudier les 250 verbes les plus fréquents, nous couvrons un pourcentage des formes verbales d'un texte quelconque qui se situe au-delà de 80%. Cela nous semble satisfaisant.

Les verbes les plus fréquents sont aussi les verbes les plus 'polyvalents': ils entrent dans beaucoup de constructions. En examinant leurs propriétés syntaxiques, nous saisissons une catégorie de verbes difficiles à étudier. En outre, les constructions de ces verbes représentent selon toute probabilité une très grande partie de toutes les constructions verbales possibles. L'intérêt de ce groupe de verbes est donc multiple.

1.1.3 Quels verbes?

Muller (1974) cherche à savoir quels sont les vingt verbes les plus fréquents en français, et compare à cet effet plusieurs listes de fréquence. Il arrive à la conclusion que

> il n'y a pas de réponse absolue à une question comme: "Quels sont les 20 verbes les plus fréquents en français?"; mais les données statistiques dont on dispose maintenant permettent de dresser des listes dont seules les dernières unités laisseraient place à l'hésitation. (1974:17)

Il est évident que cette hésitation s'accroîtra à mesure qu'on avancera dans la liste. Une autre conclusion de Muller est que

> les fréquences, dont l'intérêt n'est pas discutable, sont fortement dépendantes du type de corpus qui a servi à leur calcul. Elles n'ont de valeur que par

> rapport à un certain type de textes, ou à un ensemble de textes divers dont elles reflètent une moyenne. [...] C'est pourquoi un corpus très vaste et très diversifié comme celui du T.L.F. nous propose des données plus solides que des relevés partiels; c'est pourquoi aussi la répartition des vocables à l'intérieur d'un corpus est un élément important. (1974:17)

Il nous paraît absolument nécessaire d'asseoir le choix des verbes à étudier sur une base solide. C'est pourquoi nous partirons, non d'une seule liste de fréquence, mais de cinq (celles nommées sous 1.1.2 et la liste de fréquence du corpus d'Orléans), que nous comparerons. Ces listes sont toutes plus ou moins différentes par leur lemmatisation. Le *FDFW*, par exemple, a des verbes homographes, contrairement aux autres listes. Le *VRF*, pour donner un autre exemple, regroupe sous le lemme *un* les formes masculine et féminine, le singulier et le pluriel, alors que le *Dictionnaire des fréquences* a quatre entrées. Si l'on veut savoir quels sont les verbes les plus fréquents parmi les 1 000 premiers mots, il faudra éliminer toutes ces différences. C'est un véritable travail de Sisyphe, possible, mais ingrat et tout à fait disproportionné à l'objectif de cette étude. C'est pourquoi nous avons décidé de sélectionner, pour chaque liste, les 250 verbes les plus fréquents. Les données concernant ces verbes, elles, ont été, autant que possible, rendues comparables.

1.1.4 Listes de fréquence utilisées. Procédures de sélection appliquées

Nous donnerons dans cette section un descriptif succinct des cinq listes de fréquence utilisées et des corpus sous-jacents et préciserons les opérations effectuées pour rendre les chiffres comparables. Les listes du *VRF* et de l'*ESLP* sont les listes les plus lemmatisées: toutes les formes verbales sont regroupées sous l'infinitif; aucun verbe n'apparaît plus d'une fois sur les listes. Nous avons donc aligné les trois autres listes sur ces deux-là.

1.1.4.1 Le Dictionnaire des fréquences

Le *Dictionnaire des fréquences* de Nancy repose sur un corpus de 71 millions d'occurrences environ. Le corpus dépouillé se compose de textes littéraires du XIXe et du XXe siècle (entre 1789 et 1964). Le *Dictionnaire des fréquences* donne, pour chaque vocable, des informations sur (1) sa fréquence absolue dans le corpus des textes du XIXe siècle et dans celui du XXe siècle, (2) le rang qu'occupe le vocable dans l'ensemble des vocables, et (3) des indications sur sa fréquence relative dans les deux corpus du XIXe et du XXe siècle et dans dix sous-ensembles. Les verbes apparaissent dans le *Dictionnaire des fréquences* sous trois formes: l'infinitif (qui regroupe toutes les formes conjuguées au mode personnel), le participe présent et le participe passé. Ce regroupement, particulier au *TLF*, constitue un grand problème quand on veut comparer différentes listes de fréquence. Il s'y en ajoute un autre. Par suite d'une lemmatisation très incomplète, les trois formes verbales peuvent regrouper des formes non verbales: des substantifs homographes avec un infinitif (*dîner, devoir* par exemple), et des adjectifs, des substantifs et des prépositions homographes avec des participes présents ou passés (*amusant, habitant, devant* par exemple). Nous aurions pu totaliser pour chaque verbe les données concernant les trois formes sous lesquelles il

se manifeste dans le *Dictionnaire des fréquences*. Seulement, ce chiffre aurait contenu beaucoup de bruit. C'est pourquoi il nous a paru préférable de nous baser uniquement sur les chiffres concernant les infinitifs, catégorie "relativement pure" d'après Brunet (1981, I:297). Nous sommes partie des données fournies par la *Table des fréquences décroissantes* du *Dictionnaire des fréquences* et les avons ensuite vérifiées dans Brunet (1981,II), qui a corrigé les erreurs numériques qui s'étaient glissées dans les Tables. Les chiffres sélectionnés représentent la fréquence absolue des verbes dans l'ensemble du corpus.

1.1.4.2 Le Frequency Dictionary of French Words

Le *Frequency Dictionary of French Words* est basé sur un corpus de textes en prose de 500 000 mots, réparti en cinq sous-corpus: pièces de théâtre, romans et nouvelles, essais, périodiques, textes scientifiques et techniques. Les textes dépouillés couvrent une période allant de 1920 à 1940. Nous avons utilisé les chiffres concernant la fréquence[4] de la Table décroissante du *FDFW* pour sélectionner les 250 verbes les plus fréquents. Le *FDFW* a comme particularité que certains mots ont plusieurs entrées. En ce qui concerne les verbes, il s'agit d'une part des verbes *avoir*, *être*, *aller* et *faire* qui ont une entrée comme verbe et une autre comme verbe auxiliaire. Les verbes homographes d'autre part ont plusieurs entrées, d'après leur régime syntaxique. Nous avons totalisé les nombres d'occurrences pour ces verbes.

1.1.4.3 Le Vocabulaire du roman français 1962–1968

Le corpus sur lequel se base le *VRF* comprend un demi-million d'occurrences et est tiré de 25 romans best-sellers, publiés entre 1962 et 1968. Pour faire la liste des 250 verbes les plus fréquents du corpus du *VRF*, nous avons utilisé la liste 6 ('Classement grammatical des cinq mille lemmes majeurs') qui donne une liste, par fréquences décroissantes, des verbes faisant partie de cet ensemble. Pour des raisons techniques (huit verbes ayant la même fréquence occupent les places 246 à 253), notre sélection comprend 253 verbes.

1.1.4.4 L'Enquête statistique sur la langue parlée

Nous ne disposons actuellement que d'une seule liste de fréquence du français parlé, celle qui a servi à l'élaboration du *Français Fondamental (1er 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 enregistrement direct au début des années 50, en vue du projet d'élaboration du *Français Fondamental*. Le nombre total de mots dépouillés est de 312 135. Les mots d'une fréquence égale ou supérieure à 20, 1 063 au total, ont été classés dans deux listes de fréquence, la première par

[4] La Table décroissante contient en fait trois listes, établies d'après l'usage, la fréquence et la dispersion. Cf. *FDFW*, p. XLVIII – LXI.

fréquences décroissantes, la seconde par ordre alphabétique. Nous avons utilisé la première liste. Elle comprend 228 verbes.

1.1.4.5 L'Etude sociolinguistique sur Orléans

Nous avons cherché à compléter les données fournies par la liste de l'*ESLP* par des données sur la fréquence de verbes en langue parlée provenant d'une autre source, à savoir l'*Etude sociolinguistique sur Orléans*. Ce corpus d'enregistrements du français parlé 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. Les témoins des quatre premières catégories ont été choisis selon des critères sociologiques explicites. Pour une description plus détaillée du corpus d'Orléans nous renvoyons à la section 1.2.

Des 157[5] interviews face-à-face sur des questionnaires standardisés, 60 ont été mises sur ordinateur à l'Université Libre d'Amsterdam. Le nombre total des occurrences est de 494 028. La liste de fréquence des 60 textes dont nous disposons actuellement est une liste provisoire, car la lemmatisation n'est pas encore achevée. En ce qui concerne les verbes, la situation est comparable à celle du *Dictionnaire des fréquences*. Les formes verbales y apparaissent sous l'infinitif, le participe présent et le participe passé. Ces trois formes peuvent regrouper des formes non verbales, tout comme pour le *Dictionnaire des fréquences*. Nous nous sommes basée sur les occurrences regroupées sous l'infinitif, pour les raisons explicitées ci-dessus. Nous avons suivi l''Enquête statistique sur la langue parlée' pour ce qui est du seuil adopté: nous n'avons admis dans notre sélection que les infinitifs ayant une fréquence égale ou supérieure à 20. Ce critère nous a fourni une liste de 195 verbes.

1.1.4.6 Principales caractéristiques des cinq corpus

Le tableau 1.3 résume les caractéristiques des cinq corpus.

1.1.5 Remarques sur les cinq listes de verbes. Sélection des verbes retenus pour notre recherche

Les verbes de chaque liste ont été classés par ordre de fréquence décroissante et pourvus d'un numéro de rang.[6] Ces données ont été regroupées par ordre alphabétique pour les cinq

---

[5] Blanc et Biggs (1971) et Robach (1974) mentionnent un nombre de 147 interviews sur questionnaires standardisés, mais il y a, en réalité, 157.

[6] Engwall (1984) contient déjà une telle liste (pp. 406–411). Les verbes qui ont la même fréquence partagent le même numéro de rang. Il est d'usage de prendre la moyenne des rangs occupés. Ce chiffre est entier si les verbes ayant la même fréquence sont en nombre impair et contient un demi si ces verbes sont en nombre pair. Or, Engwall a toujours des chiffres entiers. Nous avons modifié sa liste dans ce sens que nous avons toujours pris la moyenne exacte pour ces verbes.

corpus	étendue du corpus	période	langue écrite ou parlée	type de textes	nombre de verbes sél.
TLF	71 mill. d'occ.	1789 - 1964	langue écrite	litt. (prose et poésie)	250
FDFW	500 000 occ.	1920 - 1940	langue écrite	prose (litt. et non-litt.)	250
VRF	500 000 occ.	1962 - 1968	langue écrite	25 romans best-sellers	253
ESLP	312 135 occ.	1950 (princ.)	langue parlée	interviews libres	228
Orléans	494 028 occ.	1969	langue parlée	interviews*	195

* 60 interviews sur questionnaires standardisés

Tableau 1.3: Principales caractéristiques des cinq corpus

listes. Elles forment l'Appendice A. Cette liste contient 396 verbes au total. Cet ensemble a été soumis à un certain nombre d'analyses, qui seront traitées dans cette section, et sur lesquelles nous nous baserons pour sélectionner les verbes à étudier.

Une première question qu'on peut se poser est celle de savoir combien de verbes sont communs aux différentes listes. Le tableau 1.4 présente les chiffres. Le pourcentage des

nombre total de verbes	396 verbes	
sont présents sur 5 listes	118 verbes	(= 30%)
sont présents sur 4 listes	43 verbes	(= 11%)
sont présents sur 3 listes	55 verbes	(= 14%)
sont présents sur 2 listes	69 verbes	(= 17%)
sont présents sur 1 liste	111 verbes	(= 28%)

Tableau 1.4: Répartition des verbes sur les cinq listes

verbes ne figurant que sur une seule liste est presque aussi élevé que celui des verbes figurant sur toutes les listes. Ces chiffres invitent à une analyse plus détaillée. Une telle analyse n'est pas seulement intéressante pour notre objectif, elle pourra avoir aussi un intérêt intrinsèque.

Examinons d'abord les verbes qui n'apparaissent que sur une seule liste, les verbes 'solitaires'. Leur distribution sur les cinq listes n'est pas égale. Le TLF contient 13 verbes solitaires (soit 5.2% sur 250 verbes), le $FDFW$ en a 27 (soit 10.8% sur 250 verbes), le VRF 27 (soit 10.7% sur 253 verbes), l'$ESLP$ 25 (soit 11% sur 228 verbes) et Orléans 19 (soit 9.7% sur 195 verbes). Nous présenterons brièvement ces verbes à l'aide de tableaux

commentés.

Le *TLF* a relativement peu de verbes solitaires, comme le montre le tableau 1.5. *Dîner*

achever	admirer	attacher	dîner
écrier	fuir	mêler	nommer
pardonner	rapporter	remplir	reposer
soutenir			

Tableau 1.5: Verbes figurant uniquement sur la liste du *TLF*

peut devoir sa place sur la liste au fait que le *TLF* ne distingue pas entre emploi verbal et emploi substantif d'une forme. Les autres verbes n'ont rien de bien étonnant. Leur numéro de rang est assez élevé, au-dessus de 180. Seul *écrier* a un numéro de rang assez peu élevé (r=104). Le *TLF* est le seul corpus dont les textes s'étalent sur deux siècles. A première vue, aucun des verbes solitaires ne paraît appartenir à un vocabulaire daté.[7] Cependant, quand on consulte les données statistiques procurées par Brunet (1981), on constate que sept de ces verbes sont surreprésentés dans une ou plusieurs tranches chronologiques du 19e siècle. Il s'agit des verbes suivants: *attacher*, *écrier*, *mêler*, *nommer*, *pardonner*, *remplir* et *soutenir*.

Le *FDFW* a beaucoup de verbes solitaires, comme le montre le tableau 1.6. Beaucoup

accorder	appliquer	citer	comporter
conclure	constater	constituer	décrire
déterminer	développer	examiner	figurer
fournir	imposer	indiquer	maintenir
mériter	procéder	promettre	publier
réaliser	réduire	résoudre	révéler
signaler	souhaiter	subir	

Tableau 1.6: Verbes figurant uniquement sur la liste du *FDFW*

de ces verbes appartiennent au vocabulaire abstrait et/ou scientifique. Quand on consulte les données numériques concernant les sous-corpus, on constate en effet que plus d'une vingtaine de ces verbes sont surreprésentés dans un ou plusieurs sous-corpus (le plus souvent dans le sous-corpus des textes techniques et scientifiques (14 verbes), ou celui des journaux et des magazines (7 verbes)).

La liste du *VRF* compte également de nombreux verbes solitaires (voir le tableau 1.7). Certains mots surprennent (*accrocher*, *allonger*, *allumer* par exemple). La présence

[7] Cf. Muller (1974:15) qui trouve, pour le corpus du *TLF*, que la nature et le classement des 20 verbes les plus fréquents sont chronologiquement stables, mais stylistiquement instables, et Brunet (1981,I:113–114, 807, 846) qui conclut, pour différents groupes de verbes du même corpus, parmi lesquels les 255 verbes les plus fréquents, que la marque 'genre' est cause de beaucoup plus de variance que la marque 'temps'.

d'autres verbes (*caresser, chérir, épouser, mentir*) pourrait s'expliquer par le caractère romanesque de ce corpus.

accrocher	agiter	allonger	allumer
appuyer	briller	caressser	chérir
croiser	dresser	écarter	endormir
épouser	fixer	hésiter	inventer
mentir	pencher	peser	presser
rassurer	refermer	rejoindre	réveiller
secourir	surprendre	traîner	

Tableau 1.7: Verbes figurant uniquement sur la liste du *VRF*

La liste de l'*ESLP* (voir le tableau 1.8) contient à peu près autant de verbes solitaires que celles du *FDFW* et du *VRF*: beaucoup de verbes concrets, reflétant la vie de tous les

attraper	chauffer	coller	danser
dépenser	engager	enlever	ennuyer
enregistrer	entretenir	loger	louer
nager	nettoyer	opérer	organiser
pleuvoir	prévoir	ramasser	remercier
réparer	repasser	risquer	soigner
surveiller			

Tableau 1.8: Verbes figurant uniquement sur la liste de l'*ESLP*

jours; pas de termes appartenant typiquement à la langue parlée (à l'exception peut-être de *ramasser*). Rien de vraiment surprenant, sauf *enregistrer*, qui doit sa présence sur la liste aux conditions de prélèvement des textes, raison pour laquelle la Commission chargée de la composition du *Français Fondamental* a décidé de l'exclure de sa liste.

Comme le montre le tableau 1.9, la liste d'Orléans a un peu moins de verbes solitaires que le *FDFW*, le *VRF* et l'*ESLP*. Il ne faut cependant pas perdre de vue que la liste est moins longue que les trois autres (195 verbes). Les chances d'avoir des verbes solitaires sont évidemment moins grande que pour une liste plus longue. Dix des verbes solitaires de cette liste (*augmenter, consister, déplacer, déplaire, enseigner, fréquenter, regretter, relire, utiliser* et *voter*) peuvent s'expliquer par les questionnaires[8] utilisés lors de la constitution du corpus. Un verbe utilisé dans une question a évidemment beaucoup de chances de revenir dans la réponse. Quant aux neuf verbes ne figurant pas dans une question, on peut cependant dire pour *choquer, contrôler, corriger, cuire* et *participer* qu'ils peuvent avoir été amenés par une question.[9]

[8] Ces questionnaires forment l'Appendice C.

[9] *Choquer* par le terme 'agacer' utilisé dans les questions 32, 33 et 35 du Questionnaire socio-

améliorer	augmenter	choquer	consister
contrôler	corriger	cuire	déplacer
déplaire	diriger	enseigner	envisager
fréquenter	participer	regretter	relire
ressentir	utiliser	voter	

Tableau 1.9: Verbes figurant uniquement sur la liste d'Orléans

Ce phénomène qu'un vocable doit sa fréquence au procédé de recueil des textes ne se limite évidemment pas aux verbes solitaires. Il arrive qu'un verbe figurant sur notre liste du corpus d'Orléans ainsi que sur d'autres listes ait un numéro de rang beaucoup plus bas (et donc une fréquence plus élevée) dans la liste d'Orléans que dans les autres listes. Cela concerne entre autres les verbes *apprendre*, *casser*, *écrire*, *plaire*, *préférer*, *réussir* et *travailler*. Cette fréquence élevée peut s'expliquer de la même façon que pour les verbes solitaires: le verbe figure dans une ou plusieurs questions ou est amené par une question. Il est clair que la liste de fréquence d'un corpus recueilli à l'aide de questionnaires standardisés est quelque peu faussée par ce procédé. On devrait donc s'en servir avec circonspection.

L'analyse de ces verbes solitaires fait nettement ressortir que les listes reflètent le caractère des textes du corpus ou les conditions de recueil de ces textes. Les trois listes basées sur des corpus écrits se différencient nettement. La liste du *TLF* est sans aucun doute la plus neutre. On ne peut que souscrire à la conclusion de Muller (1974:17) citée plus haut.

Ensuite, nous avons soumis nos données à des analyses statistiques. Le choix du test à appliquer à ce type de données ne va pas de soi. Nous avons d'abord calculé la corrélation des rangs des verbes de l'Appendice A. Le tableau 1.10 présente les résultats. Nous constatons des corrélations élevées ou assez élevées entre les listes basées sur des corpus écrits d'une part et les deux listes basées sur des corpus oraux d'autre part. Les corrélations entre les listes basées sur des corpus écrits et celles basées sur des corpus oraux sont moins élevées.

Ce type de test, qui s'applique à des données de niveau ordinal, fixe à 1 la distance entre un verbe d'une liste et celui qui le suit et ne tient pas compte des fréquences réelles des verbes. Puis, l'attribution des rangs est relativement aléatoire dans le domaine des verbes de fréquence basse. C'est pourquoi nous avons calculé également les coefficients de corrélation (coefficients de Pearson) entre les verbes des cinq listes, c'est-à-dire entre les occurrences de *tous* les verbes dans *tous* les corpus, pour autant que ces données soient disponibles.[10]

---

linguistique; *contrôler* par le mot 'contrôle' utilisé deux fois dans la question 12 du même questionnaire; *corriger* par le synonyme 'reprendre' utilisé dans les questions 32, 33, 34, 35 et 36 de ce questionnaire; *cuire* par la question sur la préparation d'une omelette (Questionnaire ouvert, Branche Langue et culture); *participer* par le terme 'participation' de la question E12 du même questionnaire.

[10] Le *FDFW* a sélectionné les entrées d'après le coefficient d'usage (formule qui tient compte de la fréquence et de la dispersion): tous les mots ayant un coefficient de 3.00 ou plus apparaissent sur la liste. Ce critère a donné un nombre total de 5 082 entrées. Le nombre d'occurrences dans le corpus des verbes qui n'atteignent pas ce seuil n'est pas mentionné.

	TLF	FDFW	VRF	ESLP	Orléans
TLF					
FDFW	.8296				
	(209)				
VRF	.7999	.7314			
	(201)	(183)			
ESLP	.6685	.6854	.6544		
	(163)	(156)	(171)		
Orléans	.6040	.5966	.5892	.7337	
	(145)	(144)	(140)	(160)	

Entre parenthèses le nombre de verbes impliqués pour chaque paire
($p < .000$)

Tableau 1.10: Corrélations des rangs des verbes communs à chaque paire de listes

L'Appendice B présente les données concernant le nombre d'occurrences des 396 verbes de l'Appendice A dans les différents corpus. Pour pouvoir effectuer un tel calcul, il faut au préalable que les conditions techniques d'une analyse des corrélations soient satisfaites. Une de ces conditions est que la distribution des données soit normale. Or, cela n'est pas le cas. La distribution des données est fortement dissymétrique vers la droite. Nous avons donc procédé comme suit: (a) élimination des verbes *avoir* et *être* à cause de leurs fréquences extrêmement élevées, (b) transformation des données en valeur logarithmique (base = 10). Le tableau 1.11 regroupe les résultats obtenus de cette façon. Nous voyons que les corrélations vont dans le même sens que les corrélations des rangs, à cette différence près que les premières sont plus élevées. Nous constatons que les corrélations entre *TLF*, *FDFW*, *VRF* et *ESLP*[11] sont élevées ou assez élevées. Orléans a une corrélation élevée avec *ESLP*, liste basée elle aussi sur un corpus oral, mais des corrélations nettement moins élevées avec les autres listes. Ce résultat concorde avec les remarques faites plus tôt dans cette section à propos du procédé de collecte du corpus d'Orléans.

Si nous résumons les analyses traitées dans cette section, nous pouvons dire que:

- La liste du *TLF* a le moins de verbes solitaires.
- Le test de corrélation des rangs révèle une scission entre les listes basées sur des corpus écrits et les listes basées sur des corpus oraux.
- Le test des coefficients de Pearson à base de données transformées fait ressortir des corrélations élevées entre *TLF*, *FDFW*, *VRF* et *ESLP* d'une part et entre *ESLP* et Orléans d'autre part.

La liste de fréquence de l'*ESLP* s'arrête à la fréquence 20. Gougenheim *et al.* (1967) ne fait pas mention des données concernant les vocables ayant un nombre d'occurrences inférieures à ce seuil.

[11] F. Debyser a donc raison quand il affirme que: " [...] français fondamental, dont la partie la plus sûre reste le lexique issu des relevés de fréquence et, en particulier, les verbes."