Université de Lorraine

THESE DE DOCTORAT en Sciences du Langage

Présentée et soutenue publiquement le 26 novembre 2012 par

Lolita BERARD

Dépendances à distance en français contemporain - Etude sur corpus

« c'est ce qu'on pense qui devrait être fait »

Préparée au laboratoire **ATILF (UMR 7118)**

Dans le cadre de l'Ecole Doctorale **Langages, Temps et Société**

Directrice de thèse : **Jeanne-Marie DEBAISIEUX**
Co-directeur de thèse : **Henri-José DEULOFEU**

JURY

Rapporteurs :

Sylvain KAHANE	Professeur, Université Paris Ouest Nanterre la Défense
Claus PUSCH	Professeur, Université de Freiburg (Allemagne)

Examinateurs :

Jean-Marie PIERREL	Professeur, Université de Lorraine
Catherine SCHNEDECKER	Professeur, Université de Strasbourg
Jeanne-Marie DEBAISIEUX	Professeur, Université Paris 3 Nouvelle Sorbonne
Henri-José DEULOFEU	Professeur Emérite, Université de la Méditerranée

Université de Lorraine

THESE DE DOCTORAT en Sciences du Langage

Présentée et soutenue publiquement le 26 novembre 2012 par

Lolita BERARD

Dépendances à distance en français contemporain - Etude sur corpus

« c'est ce qu'on pense qui devrait être fait »

Préparée au laboratoire **ATILF (UMR 7118)**

Dans le cadre de l'Ecole Doctorale **Langages, Temps et Société**

Directrice de thèse : **Jeanne-Marie DEBAISIEUX**
Co-directeur de thèse : **Henri-José DEULOFEU**

JURY

Rapporteurs :

Sylvain KAHANE	Professeur, Université Paris Ouest Nanterre la Défense
Claus PUSCH	Professeur, Université de Freiburg (Allemagne)

Examinateurs :

Jean-Marie PIERREL	Professeur, Université de Lorraine
Catherine SCHNEDECKER	Professeur, Université de Strasbourg
Jeanne-Marie DEBAISIEUX	Professeur, Université Paris 3 Nouvelle Sorbonne
Henri-José DEULOFEU	Professeur Emérite, Université de la Méditerranée

La vie offre des opportunités qu'il faut savoir saisir

Remerciements

Effectuer ma thèse à Nancy, dans le laboratoire de l'ATILF, a été une opportunité que j'ai saisie avec quelques risques et qui a été une chance.

Les choses ne se seraient sans doute pas aussi bien passées si je n'avais pas eu comme directeurs Jeanne-Marie Debaisieux et José Deulofeu, qui ont toujours su s'entendre pour me guider dans ma recherche. Leur enthousiasme, leur optimisme inébranlable, leurs encouragements et leur confiance en moi m'ont portée pendant les trois ans et m'ont engagée à ne pas les décevoir. A chacune de mes demandes (généralement urgentes), ils ont été disponibles. Ils les ont même parfois anticipées. Non seulement ils m'ont aidée dans cette recherche spécifique, mais ils ont également pris soin de me faire découvrir des projets et de me faire rencontrer des équipes de recherches. Je les remercie pour leur implication dans ma thèse et mon après-thèse.

Bien sûr, mon expérience positive à Nancy est également due à tout un environnement et particulièrement à celui de l'ATILF. J'ai très vite été intégrée aux doctorants, à l'équipe Macrosyntaxe et à l'ensemble du laboratoire. Je souhaite tout particulièrement remercier M. Pierrel, le directeur, qui m'a accueillie personnellement à mon arrivée, et qui a toujours eu un sourire ou un mot gentil pour moi. Son attitude bienveillante et son attention au bien-être des doctorants (dans leur vie professionnelle et personnelle) sont très appréciées. J'ai également eu de très bons rapports avec l'ensemble de mon équipe, menée avec entrain par Annie Kuyumkuyan. Chacun a été attentif à mon travail lors de mes présentations et m'a fourni spontanément des lectures et des annonces de conférences. Les séjours en Suisse avec les équipes de Neuchatel et de Fribourg sont des souvenirs agréables.

Dans le laboratoire, je souhaite remercier énormément Christophe et Mathilde, qui ont pris le relais pour me soutenir sur place, alors que mes directeurs n'étaient pas ou plus à Nancy. Leurs visites dans le bureau et les échanges que nous avons eus ont largement contribué à conserver ma motivation. Travailler sur les projets de corpus avec Christophe, toujours d'attaque, de bonne humeur et bouillonnant d'idées, était un bol d'air dans mes recherches.

D’autres personnes du laboratoire m’ont apporté des coups de pouce précieux dans mon travail (Bertrand, Coralie, Emmanuelle, Evelyne, Pascale, Gilles) et dans la préparation des cours (Me Debaisieux, Christophe, Virginie, Richard, Laurent, Iveta).

Enfin le laboratoire était toujours un lieu chaleureux grâce à la bonne humeur des Atilfiens. Les échanges ont toujours été sympathiques, que ce soit avec Dominique, Simone, Lucia ou Viviane lorsque j’étais perdue dans le classement du centre de documentation, avec Olivier ou William lorsque mon ordinateur faisait des siennes ou que mon disque dur externe était ‘tombé’ par terre, ou tout simplement à la cuisine, dans ces moments de détente et d’humour. Les attentions et les sourires quotidiens (chocolats, bredele, beignets, etc.) n’ont pas manqué, en particulier à mon étage.

Les relations avec les autres doctorantes, notamment les co-arrivantes Cécile, Jinjing et Inga, ont été très amicales, guidées par les rires et le grain de folie des plus anciennes (Sandrine, Delphine, Iveta, Tiphanie, Magali, Aurore et Coralie). Nous avons partagé nos premières expériences de cours, de conférences, d’école d’été, souvent autour d’un repas. Enfin, de notre bureau, je retiendrai les discussions animées, la stimulation et un certain nombre de repas copieux...

Par ailleurs, mes trois ans à Nancy ont été l’occasion de rencontres nombreuses et importantes. Les amis du CIES, du SIUAP, les activités à Jarville et à la MJC Pichon ont comblé mes moments de loisirs. Je voudrais citer en particulier les colocs & co., Maxime, Vincent, Charly et Mathieu, qui m’ont nourrie presque toute la deuxième année, m’ont fait découvrir la ville et offert un cadeau d’anniversaire mémorable ! Je citerai également Aurore, qui, après m’avoir mise à l’aise au bureau et fourni des informations précieuses en début de thèse, est devenue une deuxième famille. Nous avons partagé les trajets quotidiens, les karaokés kitschs, les pizzas de Jamal, les meilleures crêpes (les siennes bien sûr) et quelques moments de ‘tristitude’. Je conserve pour plusieurs années les bases de la langue lorraine qu’elle m’a apprise et la sonnerie de son horloge interne très précise de 11h30 (‘j’ai faim !!’).

Pour finir, je voudrais remercier ma famille, sur qui je peux toujours compter. Elle a participé à la relecture de la thèse mais a aussi été un relais pratique pour le déménagement, la recherche d’appartement, etc. Elle a satisfait le moindre de mes désirs et caprices, jusqu’à me faire écouter le chat au téléphone. Yann, en plus de son aide lors de la rédaction, m’a apporté

son équilibre, sa force tranquille et son réconfort. Je le remercie pour sa patience et sa compréhension. J'ai été touchée par toutes ses délicates attentions. Enfin j'ai une pensée pour le petit béluga qui m'accompagne nuit et jour depuis le début de l'été.

Ce manuscrit symbolise la fin de mes années nancéennes. Avant de laisser la place à l'étude, je remercie Sylvain Kahane, Claus Pusch, Jean-Marie Pierrel et Catherine Schnedecker qui ont accepté de faire partie de mon jury et d'évaluer mon travail.

Sommaire

PRESENTATION GENERALE 19

CHAPITRE 1. INTRODUCTION- DEFINITION DU SUJET ET ETAT DE L'ART 23

1.1 Constructions désignées par ce terme 23

1.1.1 Modèle en constituants avec transformation (Harris, Chomsky) 24

1.1.2 Modèle en dépendances (Tesnière, Kahane) 31

1.1.3 L'approche pronominale 36

1.1.4 Récapitulatif 42

1.2 Délimitation du champ de recherche 44

1.3 Etat de l'art 48

1.3.1 Approches Générativistes transformationnelles : priorité aux contraintes syntaxiques (67, 77) 49

1.3.2 Contraintes sémantico-pragmatiques 52

1.3.2.1. Erteschik-Shir (73) 52

1.3.2.2. Godard (80) 60

1.3.2.3. Ambridge & Goldberg (08) 64

1.3.2.4. Featherston (05) 67

1.3.2.5. Synthèse des contraintes sémantico-pragmatiques / bilan provisoire 69

1.3.3 Hofmeister & Sag : Contraintes cognitives sur la production des énoncés et le calcul du sens (10) 70

1.3.4 De la représentation des possibles à la description des constructions réalisées 73

1.3.5 Contraintes lexicales 75

1.3.5.1 Dąbrowska (08-09) 76

1.3.5.2 Verhagen (05-06) 77

1.4 Synthèse 79

1.4.1 Points retenus et adaptation au français 80

1.4.1.1 Verbes français liés à la présupposition 81
1.4.1.2 Verbes français de mitigation 85
1.4.2 Problématique et plan 87
CHAPITRE 2. CORPUS ET OUTILS 91
2.1 Les données 91
2.1.1 Présentation des corpus originaux 93
2.1.1.1 CERF/CorpAix 93
2.1.1.2 CRFP 94
2.1.1.3 Nancy-TCOF 94
2.1.1.4 C-ORAL-ROM 95
2.1.1.5 CTFP 96
2.1.1.6 CFPP 97
2.1.1.7 PFC 97
2.1.1.8 CPROM 98
2.1.1.9 OTG 98
2.1.1.10 Brassens 99
2.1.1.11 UBS 99
2.1.1.12 Irit 100
2.1.1.13 Synthèse 100
2.1.2 Travail d'uniformisation 101
2.1.2.1 Une raison majeure pour uniformiser les données : le recours aux outils de requêtes 102
2.1.2.2 Uniformisation des conventions de transcriptions 103
2.1.2.3 Uniformisation de l'orthographe 106
2.1.2.4 Suppression des doublons 108
2.1.2.5 Uniformisation du format 113
2.1.2.6 Résultats 114

2.1.3 Synthèse sur le prétraitement des ressources à effectuer avant une analyse 116
2.1.4 Perspectives : annotation morphosyntaxique 118
2.2 Méthodologie de recherche des dépendances à distance 119
2.2.1 Requêtes 120
2.2.1.1 Dépendances à distance en complétive 120
2.2.1.2 Dépendances à distance en infinitive 123
2.2.1.3 Bilan 124
2.2.2 Tri des complétives et des infinitives 126
2.2.3 Evaluation des requêtes automatiques sur les corpus uniformisés 129
2.2.3.1 Requêtes simples, du type « qui \w+ \w+ qu » 130
a) Absence d'apostrophe dans l'empan 132
b) Requête consommante 133
2.2.3.2 Variation de la taille de l'empan (jusqu'à 10 mots) 134
2.2.4 Synthèse des requêtes effectuées 137
2.3 Synthèse des difficultés d'une étude sur corpus 138
2.4 Synthèse sur les occurrences de dépendance à distance repérées 142
2.4.1 Occurrences en complétive 142
2.4.2 Occurrences en infinitive 143
CHAPITRE 3. DESCRIPTION DES CONSTRUCTIONS DE DEPENDANCE A DISTANCE A VERBE FINI .....
..... 149
3.1 Construction prototypique en français ? 150
3.2 Construction libre avec des contraintes syntaxiques (position du gap) ? 152
3.2.1 Contrainte du sujet phrastique 153
3.2.2 Contrainte du SN complexe 154
3.2.3 Contrainte de l'ajout 155
3.3 Sélection dans le lexique des verbes ponts ? 156
3.3.1 Verbes ponts attestés dans les corpus 157

3.3.1.1 Verbes ponts et verbes factifs/implicatifs/semi-factifs 158
3.3.1.2 Verbes ponts et verbes recteurs faibles 164
3.3.1.3 Verbes ponts et verbes à complétive 167
3.3.2 Verbes ponts attestés sur le Web 169
3.3.3 Sélection des verbes ponts parmi les verbes les plus fréquents ? 172
3.3.4 Sélection des verbes ponts parmi une classe syntaxique ? 179
3.3.5 Sélection des verbes ponts parmi une classe sémantique ? 181
3.3.6 Synthèse 188
3.4 Autre source de contrainte ? La construction du « pont » 191
3.4.1 Le sujet du verbe pont 191
3.4.2 Le verbe pont 193
3.4.3 Les objets et les ajouts 196
3.4.4 Le sujet du verbe à complétive 197
3.4.5 Les éléments non-régis et les disfluences 198
3.4.6 De la description de la construction du pont à l'interprétation de contrainte de *processing* 200
3.4.6.1 Récapitulatif 200
3.4.6.2 Absence de poids syntaxique 201
3.4.6.3 Interprétation : contraintes cognitives de distance topologique 201
3.4.6.4 Poids du pronom interrogatif-relatif 202
3.4.6.5 Autre contrainte cognitive : connexion syntaxique 203
3.4.6.6 Corrélation entre les deux contraintes cognitives précédentes : distance topologique et connexion syntaxique 205
3.5 Variation : Les contextes de productions 206
3.5.1 Contexte d'apparition des dépendances à distance 206
3.5.1.1 Echantillonnage 208
a) Critères d'échantillonnage 210
b) Echantillonnage du corpus écrit 216

c) Echantillonnage du corpus oral 217
3.5.1.2 Dépendance à distance et échantillonnage dans le corpus oral 221
3.5.1.3 Dépendance à distance et échantillonnage dans le corpus CERF 226
3.5.2 Adultes-enfants 235
3.6 Conclusion 241
CHAPITRE 4. VERS UNE INTERPRETATION DES RESULTATS : ENCHASSEMENT OU COMPLEXE VERBAL ? 243
4.1 Préalable : Comparaison du pont dans les séquences avec verbe fini et avec verbe infinitif 244
4.1.1 Contexte : pourquoi étudier les infinitives ? 244
4.1.2 Lexique et construction des verbes ponts 247
4.1.2.1 Lexique des verbes ponts 247
4.1.2.2 Construction du verbe pont 254
4.1.3 Lien V1-V2 - Classement syntaxique des verbes ponts 259
4.1.4 Conclusion 265
4.2 Une construction spécifique du français : le complexe verbal fini 268
4.2.1 Une construction complexe comme unité syntaxique 270
4.2.2 Ré-emploi dans le *sluicing* 273
4.2.3 Représentation par le nucléus verbal 275
4.3 La chaîne verbale : unité fonctionnant dans les trois configurations syntaxiques 278
4.3.1 Comparaison des configurations syntaxiques 278
4.3.1.1 Relatives vs interrogatives 279
4.3.1.2 Interrogatives directes vs interrogatives indirectes 282
4.3.2 Cadre syntaxique et lexique des verbes ponts 284
4.3.2.1 Verbes faibles et relatives déterminatives 287
4.3.2.2 « Vouloir » faible 288
CONCLUSION 291
ANNEXES 295

TABLE DES FIGURES 329

TABLE DES TABLEAUX 331

TABLE DES IMAGES 337

BIBLIOGRAPHIE 339

PRESENTATION GENERALE

Notre étude porte très majoritairement sur les constructions « QU- CV1 que CV2 », du type *c'est ce qu'on pense qui devrait être fait.* (exemple tiré du corpus oral CorpAix, de la transcription FTUSPC). L'élément QU- est un interrogatif ou un relatif (ici le relatif « qu' » dont l'antécédent est « ce ») et CV1 est une construction verbale (« on pense qu- ») qui construit la construction verbale complétive CV2 (« devrait être fait »). L'élément QU- dépend de CV2 (*faire_'ça'*). La construction présente donc une 'dépendance à longue distance' (*long distance dependencies*, abrégé en LDD), terme raccourci en 'dépendance à distance'. La question de savoir si la séquence peut comporter plus de deux verbes reste ouverte.

Cette construction a fait l'objet de nombreuses études et toutes concluent que des contraintes en règlent les emplois. Les possibilités ne sont pas aussi étendues que pour les constructions « CV1 que CV2 » : *on pense que ça devrait être fait*. Certains travaux mettent l'accent sur des contraintes syntaxiques, les autres sur des contraintes lexicales, sémantiques, pragmatiques ou processuelles. Les contraintes affectent inégalement chacune des parties de la construction.

Dans la quasi-totalité des travaux, la construction CV2 est considérée comme enchâssée dans celle du V1, ce qui produit une distance hiérarchique entre l'élément antéposé et son recteur V2 enchâssé. Mais depuis quelques années, une nouvelle conception de la construction a été proposée : les deux constructions verbales pourraient former un constructeur complexe, ce qui effacerait la distance structurale entre le V2 et son complément antéposé. Ces deux conceptions découlent largement des propriétés des formalismes qui ont modélisé la question : modèle en constituants et modèle en dépendances.

Que l'objectif soit la définition des contraintes qui pèsent sur la construction de dépendance à distance ou son traitement théorique, les travaux dans le domaine français se sont essentiellement appuyés sur des exemples construits. A notre connaissance, les seules études qui ont été menées sur corpus concernent d'autres langues : l'anglais et le néerlandais.

L'objectif de cette thèse est donc d'étudier les constructions de dépendance à distance en français et sur des données authentiques. La méthodologie a tantôt été *corpus-based*, tantôt *corpus-driven*. L'idéal aurait été de décrire de façon la plus naïve possible les données

attestées afin de découvrir quel était le cadre le plus adapté (*corpus-driven*) mais la question ayant été largement débattue, nous ne pouvions ignorer les résultats les plus célèbres. Nous avons donc choisi d'examiner dans un premier temps si les corpus permettaient d'infirmer ou de confirmer les hypothèses établies sur des données construites (*corpus-based*). Mais l'essentiel du travail aura consisté, dans un deuxième temps, à dégager des données attestées les propriétés de la construction (*corpus-driven*).

Etant donné que l'objectif était de déterminer quelle analyse rendait le mieux compte de la construction (enchâssement ou constructeur complexe) et que, comme nous l'avons mentionné, ces deux conceptions découlent des deux cadres théoriques majeurs en syntaxe, nous nous sommes attachés à décrire le phénomène de la manière la plus neutre possible, sans choisir de cadre théorique. Nous avons utilisé l'Approche pronominale comme un outil descriptif minimal (distinction entre microsyntaxe et macrosyntaxe, importance des pronoms pour définir la rection), qui n'impliquait pas de faire un choix entre constituance et dépendance, les deux représentations étant présentes dans cette approche. Le débat entre enchâssement et constructeur complexe restait donc ouvert. Afin de respecter le caractère descriptif de l'étude, nous avons opté pour un vocabulaire intuitif pour désigner le V1, recteur pour les uns et partie du constructeur pour les autres. Nous avons repris la terminologie de (Erteschik-Shir, 1973): 'verbe pont'. Le verbe est une passerelle entre le V2 et le pronom QU-antéposé.

Afin de mener à bien cette étude descriptive, il était nécessaire de pouvoir s'appuyer sur un certain nombre d'occurrences. Or une étude préalable avait montré que les occurrences de dépendance à distance étaient assez peu fréquentes et qu'elles étaient encore plus rares à l'écrit. Nous avons donc cherché à réunir environ trois millions de mots d'oral et au moins le double d'écrit. Cette tâche n'aurait même pas eu à être mentionnée pour des domaines linguistiques qui disposent de corpus de référence de cette taille (anglais, néerlandais, espagnol, hébreu…). Mais en l'absence d'un tel outil pour le français, il nous a fallu le construire, et ce travail de construction, qui nous a pris pratiquement un an doit être considéré comme partie intégrante de la thèse. Nous avons rassemblé environ neuf millions de mots d'écrit par le corpus CERF et le nombre souhaité d'oral, par accumulation de douze corpus. Une partie d'entre eux a été nettoyée et homogénéisée, afin d'améliorer les performances des outils d'exploitation (notamment les concordanciers). L'absence d'étiqueteurs

morphosyntaxiques et syntaxiques performants sur la langue parlée a rendu la recherche d'occurrences de dépendance à distance fastidieuse. Nous avons donc utilisé une partie du corpus pour entraîner TreeTagger sur l'oral et obtenir un outil d'étiquetage morphosyntaxique de l'oral. Nous souhaitons mettre à la disposition de la communauté les corpus unifiés (proche de 2 300 000 mots) lorsque les problèmes juridiques seront résolus. Les corpus annotés en morphosyntaxe et l'outil d'annotation entraîné sur l'oral sont déjà disponibles sur le CNRTL.

La présentation de l'étude est organisée de la manière suivante.

Le Chapitre 1 présente le contexte de l'étude. Il définit ce qu'est une construction de dépendance à distance, et à quelles constructions le terme fait référence en fonction des différents cadres théoriques (§1.1). Il précise ensuite les constructions spécifiquement examinées dans notre analyse (§1.2) puis révèle les contraintes mises au jour par les précédentes études portant sur les dépendances à distance (§1.3). L'établissement des contraintes sur des données non francophones nous invite à réfléchir à une adaptation et à formuler des hypothèses sur le comportement du phénomène en français (§1.4).

Le Chapitre 2 présente les ressources sélectionnées pour l'étude sur corpus. Après une présentation des données et de leur harmonisation (§2.1), nous exposons la manière dont ont été recherchées les occurrences de dépendance à distance, recherche qui a fait l'objet d'une évaluation (§2.2). Cette expérience nous amène à tirer des leçons sur la méthodologie en linguistique de corpus (§2.3). Pour conclure, nous indiquons sur combien d'occurrences a porté l'étude et pourquoi certaines ont été écartées (§2.4).

Le Chapitre 3 constitue le coeur de l'étude. Il procède à la description des séquences de dépendance à distance avec verbe fini (structure en 'complétive' en grammaire traditionnelle). Par méthode *corpus-based*, nous vérifions si les emplois de dépendance à distance sont formulaires, comme l'ont conclu les études sur corpus anglais et néerlandais (§3.1), si les contraintes syntaxiques définies par les génératives sont respectées (§3.2) et si le lexique du verbe pont est choisi en fonction des critères mentionnés jusqu'alors : factivité, modalité, fréquence, construction syntaxique (§3.3). Par méthode *corpus-driven*, nous examinons l'ensemble de la construction du pont, qui a suscité peu d'intérêt dans les études précédentes et mettons en évidence l'importance des facteurs liés à la performance dans l'usage des dépendances à distance (§3.4). Enfin, nous tentons de définir si la construction varie selon le

contexte de production : variation en fonction des ‘genres’ et en fonction du type d’interlocuteur, adulte ou enfant (§3.5).

Le Chapitre 4 permet alors d’approfondir l’interprétation théorique de ces résultats à travers ces deux questions centrales. Les constructions présentent-elles une dominance de V1 sur V2 avec dépendance à distance ou un complexe verbal sans dépendance à distance ? Quelle relation y a-t-il entre V1 et V2 ? Pour confirmer la tendance fournie par les descriptions du chapitre 3, nous comparons les dépendances à distance avec verbe fini aux dépendances à distance à l’infinitif qui ont la même structure « X2 CV1 CV2 » : *un film que je pensais bien connaître* (§4.1). L’étude ne porte que sur un sous-ensemble de corpus et de formes à l’infinitif. Le but n’est pas de décrire complétement ces structures, mais de fournir des arguments indépendants en faveur de l’une ou l’autre analyse de la séquence avec verbes finis. A la suite de cette comparaison, nous tirons une conclusion sur les implications théoriques de notre analyse (§4.2) et en justifions le choix (§4.3).

CHAPITRE 1. INTRODUCTION-DEFINITION DU SUJET ET ETAT DE L'ART

Après avoir listé les constructions qui peuvent répondre au nom de 'dépendance à distance' dans les différents cadres théoriques, nous limiterons notre champ de recherche aux constructions de structure « X2 CV1 CV2 » dans lesquelles X2 dépend explicitement de CV2. Nous verrons quelles contraintes ont été notées dans les études précédentes analysant le plus souvent des langues étrangères. Nous établirons à partir de là une liste de contraintes qui peuvent peser sur la construction en français dont nous vérifierons la validité dans notre étude sur corpus.

1.1 Constructions désignées par ce terme

Les dépendances à distance désignent de nombreuses séquences dans lesquelles une séquence n'est pas dans une position canonique par rapport à son verbe recteur. On relève par exemple dans la littérature des exemples d'extractions (1), de dislocations ou topicalisations, avec ou sans reprise pronominale (2), d'interrogatives (3), de relatives (4) et de clivées (5).

(1) ***Huit ans****, il avait à l'époque (d'après (Abeillé, Godard, & Sabio, 2008, p. 3))*

(2) ***les livres****, elles* ***les*** *avaient envoyés (Blache, 2001, p. 107)*

(3) ***What*** *did you find / Tell me* ***who*** *you talked to (Bender, Sag, & Wasow, 2003, p. 3)*

(4) la soupe ***que*** *le chef déguste (d'après (Labelle, 1990, p. 20))*

(5) C'est ***huit ans*** *qu'elle avait à l'époque (d'après (Abeillé et al., 2008, p. 3))*

Ces constructions peuvent devenir des dépendances à longue distance ('*long distance dependencies',* ou LDD), lorsqu'un verbe se place entre les éléments distants :

(6) Huit ans, je crois qu'il avait à l'époque (Abeillé et al., 2008, p. 3)

(7) Les livres, je pense qu'elles les avaient envoyés (d'après (Blache, 2001, p. 107))

(8) Je ne sais pas qui Jean a dit qu'il avait vu (d'après (Abeillé, 2007))

(9) La pièce que l'évolution de la situation politique donnait l'impression qu'on pouvait enfin monter n'a pourtant pas été autorisée. (Godard, 1988, p. 38)

(10) C'est huit ans que je crois qu'elle avait à l'époque (d'après (Abeillé et al., 2008, p. 3))

Même si intuitivement on peut noter que tous les exemples ont comme point commun l'éloignement de deux éléments dépendants par un verbe, nous allons montrer que le recouvrement du terme '*long distance dependencies'* dépend de l'approche grammaticale théorique qui est choisie : grammaire en constituants ou grammaire de dépendances. Dans cette section, nous expliquerons quelles structures sont concernées dans chacune des approches et pour quelles raisons.

1.1.1 Modèle en constituants avec transformation (Harris, Chomsky)

D'abord, nous présenterons rapidement les bases du modèle en constituants, développées dans (Chomsky, 1965), c'et à dire le modèle qui sert de base aux approches transformationnelles, pionnières dans l'étude du phénomène. Ensuite, nous détaillerons les constructions qui mettent en jeu une distance entre gouverneur et dépendant et nous distinguerons celles qui appartiennent aux '*long distance dependencies'.*

Définition et principe :

La grammaire en constituants, ou 'syntagmatique', représente la structure des énoncés en se basant sur les catégories grammaticales des groupes de mots. Des règles permettent de définir la composition de chaque groupe, qui porte généralement le nom de sa tête. Les éléments entre parenthèses sont facultatifs. La signification de chaque élément est indiquée dans le TABLEAU 1.

- Ph --> (SP) SN SV (SP)

- SN --> (Dét) N (SP) (SA)
- SN --> (SN) (SP) : complément du nom
- SV --> (Aux) V (SN) (SP) (SA) (SAdv)
- SP --> Prép SN
- SA --> (SAdv) A (SP)
- SAdv --> (SAdv) Adv
- N --> Pro

Abréviation	Signification	Abréviation anglophone
Ph	Phrase	S (sentence)
SN	Syntagme nominal	NP (noun phrase)
SP	Syntagme prépositionnel	PP (prepositional phrase)
SV	Syntagme verbal	VP (verb phrase)
SA	Syntagme adjectival	AdjP (adjectival phrase)
SAdv	Syntagme adverbial	AdvP (adverbial phrase)
Dét	Déterminant	Det (determiner)
N	Nom	N (noun)
Aux	Auxiliaire	Aux (auxiliary)
V	Verbe	V (verb)
A	Adjectif	Adj (adjective)
Adv	Adverbe	Adv (adverb)
Pro	Pronom	Pro (pronoun)
Prép	Préposition	P (preposition)

TABLEAU 1. Liste des principaux constituants et syntagmes utilisés dans une représentation arborescente

Ces règles sont établies à partir d'un ordre des mots canonique pour une phrase simple, active, déclarative et affirmative, ce qui correspond grossièrement en français au schéma « sujet (SN ou Pro), verbe, complément (SN, SP) ». Cette structure est appelée la structure profonde de la phrase. Si l'énoncé ne suit pas l'ordre canonique fixé par les règles de la grammaire, c'est qu'il a subi une transformation. La structure de surface est donc différente de la structure profonde. Les phrases complexes, originellement issues de transformations, peuvent être obtenues directement par des règles de réécriture permettant un usage récursif du symbole Ph

(phrase) ou S (*sentence*) (FIGURE.1). Les phrases enchâssées 'complétives' sont introduites par une conjonction, notée 'C' en anglais pour '*complementizer*'. L'ensemble de la complétive est rassemblé sous le CP ('*complementizer phrase'*).

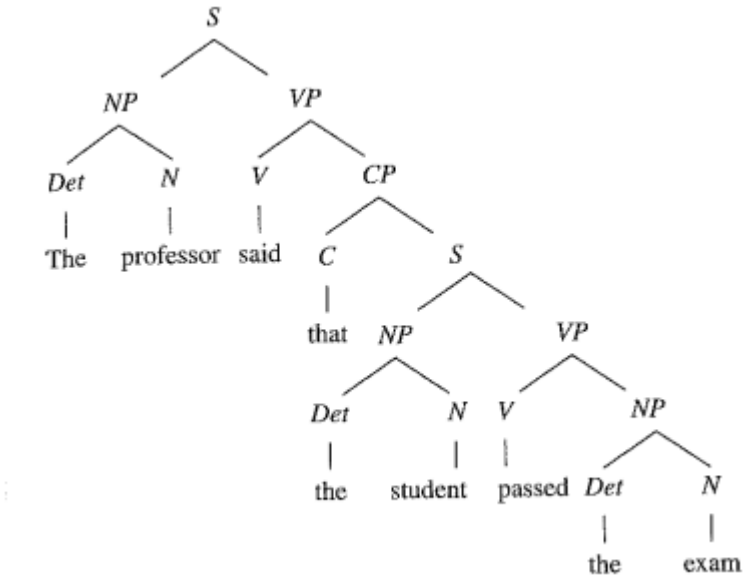

FIGURE.1 Arbre en constituants d'une phrase complexe (Fromkin, Rodman, & Hyams, 2010, p. 133)

Dans les modèles les plus utilisés, les transformations possibles se réduisent au déplacement d'un constituant. Ainsi, les interrogatives sont modélisées comme des transformations de structures déclaratives. En anglais, l'auxiliaire se déplace en tête d'énoncé. Sa position dans l'arbre est hiérarchiquement supérieure. (FIGURE.2)

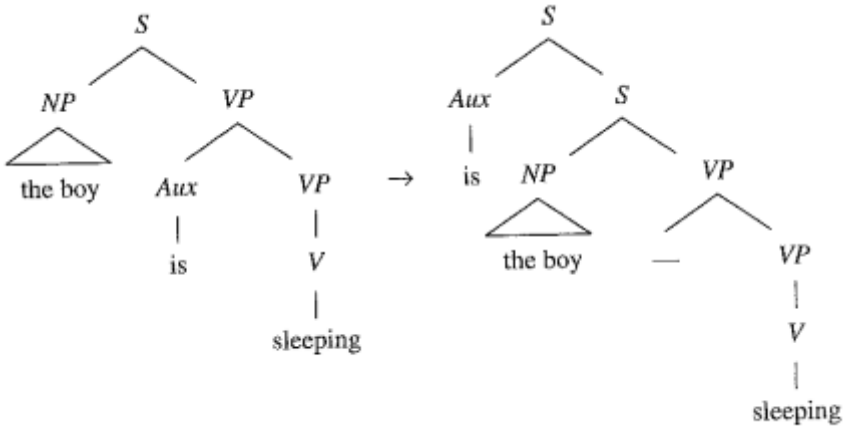

FIGURE.2 Arbre en constituants d'une interrogative (transformée à partir d'une déclarative) (Fromkin et al., 2010, p. 156)

Les mouvements produits par les transformations sont visibles : le nœud laissé vide porte la trace de l'élément qui a été déplacé. C'est la raison pour laquelle le nœud ne peut pas être rempli à nouveau (11) :

*(11) *la soupe que le chef déguste la soupe (Labelle, 1990, p. 20)*

***'Short' vs 'long' distance* :**

La grammaire transformationnelle en constituants distingue deux types de mouvement : les mouvements courts et les mouvements longs.

Tant que l'élément se déplace à l'intérieur d'une clause (C), le mouvement est considéré comme court ('*short distance'*). C'est le cas pour les relatives simples (12), les interrogatives simples (13) et les interrogatives indirectes (14), ainsi que pour la pronominalisation en français (15) :

(12) Who did Mary love ~~who~~?(Boxell, 2012, p. 151)

(13) This is the girl who John loved ~~who~~.(Boxell, 2012, p. 152)

(14) The King wondered who he should buy a castle for ~~who~~.(Boxell, 2012, p. 152)

(15) Il la mange ~~la pomme~~.

Si, en structure de surface, l'élément se déplace au-delà de la clause à laquelle il appartient en structure profonde, le mouvement est long ('*long distance'*). Dans ce cas, il peut y avoir récursivité dans le franchissement des frontières de clauses. En (16), le pronom « *who* », complément de « *to please* » franchit une première clause en se déplaçant en tête de la complétive objet de « *to claim* » (CP) ; ensuite, de façon récursive, il franchit une deuxième clause en se déplaçant en tête de la complétive objet de « *to wonder* » (CP).

(16) The manager wondered [CP who [C the secretary claimed [CP ~~who~~ [C that the new salesman had pleased ~~who~~]]]](Boxell, 2012, p. 152)

Les relatives complexes et les interrogatives complexes (directes et indirectes) sont donc représentées comme (16).

Les extractions sont également le fruit d'un mouvement, qui peut être court ou long. Dans les premiers travaux des générativistes (Ross, 1983)[1], les dislocations n'étaient pas distinguées des extractions. De nombreuses objections ont été soulevées, dont les suivantes : en anglais, la dislocation de compléments prépositionnels ne déplacent pas la préposition et aucun lien syntaxique n'est donc visible, les rares dislocations de pronoms sujet changent de forme (*Me, I*...) (Larsson, 1979)[2]. Les dislocations (ou au moins celles avec reprise pronominale) ont par la suite été placées dans un nœud pré-noyau spécifique, dès la structure profonde. (voir (Gardelle, 2010) pour plus de détails)

Dans ce modèle en constituants, les constructions de '*long distance dependencies'* concernent donc les relatives, les interrogatives et les clivées complexes, ainsi que les extractions et les dislocations sans reprise pronominale dans certains cas.

La seule relation possible entre deux nœuds est que l'un domine l'autre. Il n'existe donc pas de relation spécifique entre deux verbes. Les auxiliaires sont têtes des participes passés, etc. Dans les dépendances à distance, les deux verbes entretiennent également une relation hiérarchique : le premier verbe domine le second. On voit que ces relations hiérarchiques de verbes enchâssés dans d'autres verbes établissent dès la structure profonde un verbe intermédiaire entre le terme régi et le verbe recteur. La distance est en quelque sorte créée par le modèle. Nous allons voir qu'un changement de modèle peut réduire le caractère non canonique de ces dépendances.

Dans les années 70, (Joshi, 1987; Joshi, Levy, & Takahashi, 1975) ont développé le formalisme TAG (*Tree Adjoining Grammars*) dans le but de représenter les dépendances à distance de manière plus intuitive. Cette grammaire en constituants réunit plusieurs arbres, par substitution ou adjonction d'un nœud de catégorie identique. Dans le cas des dépendances à distance, l'arbre initial comporte une dépendance courte et locale tout à fait canonique; le syntagme verbal intermédiaire forme un arbre auxiliaire qui est adjoint au premier, grâce au nœud représentant la complétive (FIGURE.3). Cette représentation présente deux avantages : d'une part, elle montre que le syntagme verbal intermédiaire est second dans l'énoncé, ce qui le rapproche des ajouts adverbiaux par exemple qui sont introduits par la même opération

[1] Ross , J.R. (1973) The Same Side Filter. In C. Corum, T. Smith-Stark et A. Weiser (eds), *Papers from the*

[2] Larsson , E. (1979) *La dislocation en français. Étude de syntaxe générative* . Lund : CWK Gleerup

d'adjonction ; d'autre part, elle donne la possibilité de représenter simplement des structures récursives. Enfin, elle rapproche la structure syntaxique de la structure sémantique, puisque beaucoup d'observateurs ont relevé le caractère sémantiquement 'second' du verbe enchâssant en surface le verbe recteur « offre ».

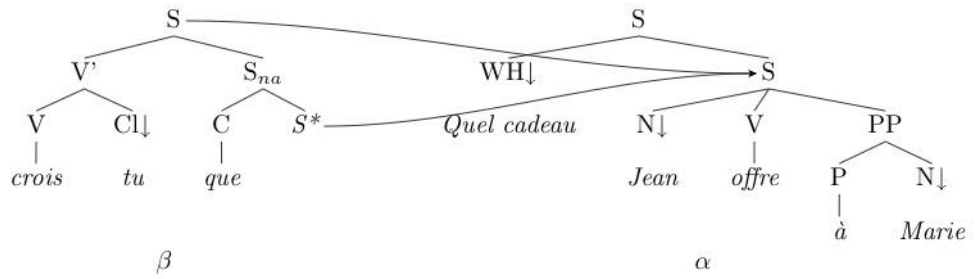

FIGURE.3 Formation d'un arbre de dépendances non bornées en TAG (B. Crabbé)[3]

Néanmoins, l'arbre terminal n'est pas différent de celui du modèle avec mouvement. Les adjonctions et les substitutions qui ont abouti à la formation de l'arbre ne sont visibles que dans le dispositif métagrammatical des 'arbres de dérivation' et finalement, la représentation donne l'impression que le syntagme verbal intermédiaire domine le reste de l'énoncé (FIGURE.4), sans création d'une structure qui permettrait de considérer qu'il n'y a pas de dépendance à distance, mais que « quelles filles » est en relation de dépendance 'locale' avec « regarde ».

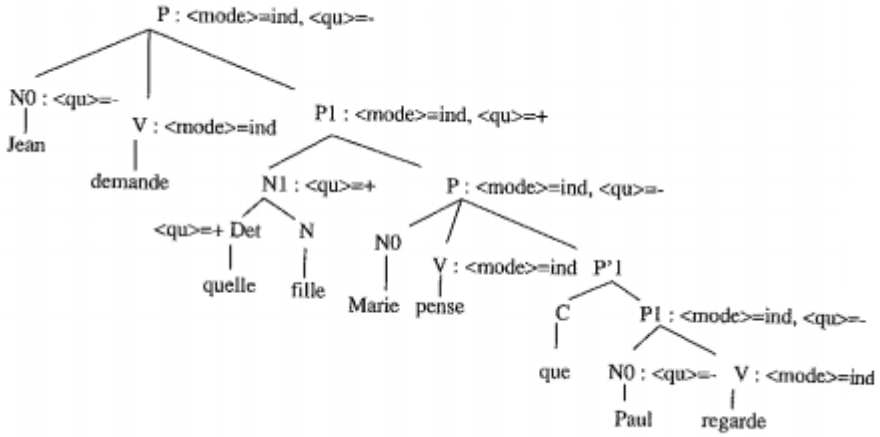

FIGURE.4 Représentation finale des dépendances non bornées en TAG (Abeillé, 1993, p. 248)

[3] Figure issue du Web : http://fr.wikipedia.org/wiki/Grammaire_d%27arbres_adjoints

Parallèlement aux modèles en constituants avec transformation, il existe des modèles en constituants sans transformation. Dans ce courant, se développe à la fin des années 80 le formalisme HPSG (*Head-driven Phrase structure Grammar*) (Pollard & Sag, 1994). Les représentations sont basées sur la contrainte de bonne formation : un énoncé agrammatical ne peut pas être décrit. En outre, ce modèle exclut les catégories vides et les syntagmes vides. Les dépendances à distance ne sont donc pas formalisées par le déplacement d'un élément qui laisse une branche vide. C'est le verbe dont le complément n'est pas à la place canonique qui signale le manque. Il porte un trait *slash* qui indique à la fois le manque et la catégorie souhaitée du syntagme, puis le transmet à sa tête. Tant que la tête ne peut pas saturer le trait *slash* grâce à son autre branche fille, il le donne en héritage à sa propre tête (FIGURE.5).

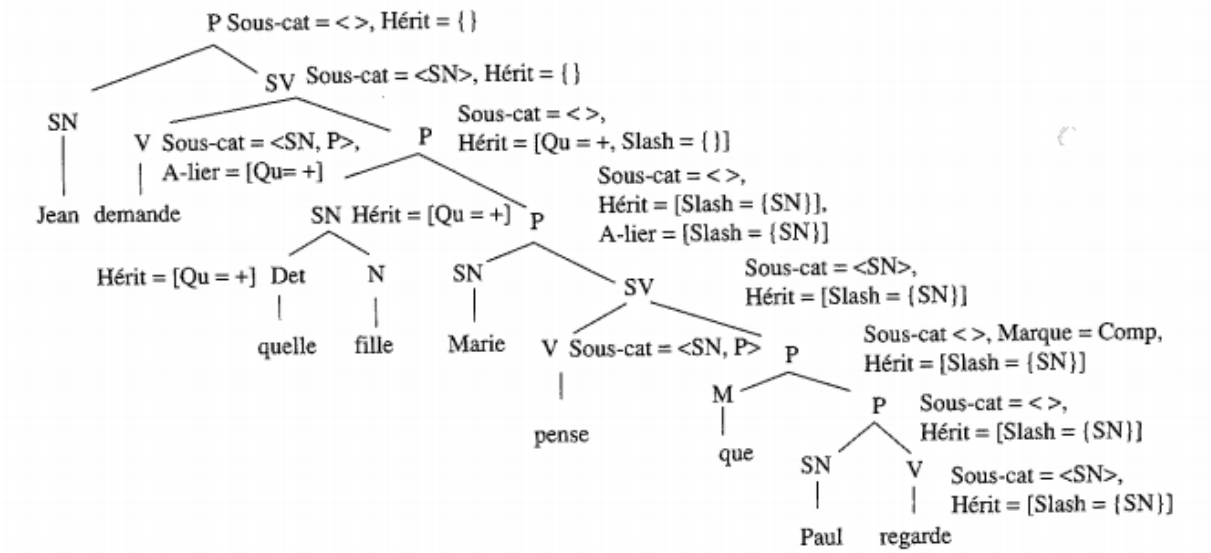

FIGURE.5 Représentation des dépendances non bornées en HPSG (Abeillé, 1993, p. 188)

Même s'il n'y a pas de mouvement, la relation de dépendance entre « quelle fille » et « regarde » se fait à travers la structure englobant « penser » et donc le modèle concourt ici encore à établir l'idée qu'il existe dans la langue des dépendances à distance.

Quelles que soient les variantes, le modèle en constituants interprète de nombreuses constructions comme étant des reconstructions de structure profonde donnant lieu à des dépendances à distance en structure de surface. Voyons maintenant les principes du modèle en dépendances et les constructions qu'il considère comme dépendances à distance.

1.1.2 Modèle en dépendances (Tesnière, Kahane)

Le modèle en dépendances a été développé par Tesnière et publié après sa mort (Tesnière, 1959). Il a donné lieu à différentes adaptations, telles la Théorie Sens-Texte (Mel'čuk, 1988), ou la grammaire à bulle (Kahane, 1997; 2000).

Définition et principe :

La grammaire de dépendance s'attache à représenter la relation dominant-dominé entre les mots. Chaque unité lexicale est liée à une autre, si bien que toute la phrase est représentable par un réseau, semblable à une molécule chimique. La tête, ou 'régissant', sélectionne ses dépendants, 'subordonnés', les domine et leur impose ses traits (genre, nombre, etc.). L'élément central du modèle est le verbe. Tous ses dépendants se trouvent au même niveau (sujet, objets direct, objet indirect), et sont appelés 'actants'.

Cette approche évite autant que possible le recours à un métalangage spécifique. Le modèle se préoccupe uniquement du lexique, et non de la catégorie grammaticale. Les nœuds sont directement les formes lexicales. Les liens qui unissent ces nœuds peuvent être annotés par la fonction du dépendant : sujet, objet, modifieur.

Dans ce modèle, on distingue deux ordres : l'ordre linéaire et l'ordre structural. Dans l'ordre linéaire, des flèches, qui vont de la tête aux dépendants, marquent les relations entre les mots. Ainsi, comme le montre la (FIGURE.6), le verbe « rencontre » pointe sur ses subordonnés sujet « Christine » et objet « collègue ». Le nom tête « collègue » pointe à son tour vers ses subordonnés : les adjectifs « bon », « italien » et le déterminant « un ». Enfin, l'adjectif « bon » pointe vers son subordonné : l'adverbe « très ».

FIGURE.6 Réseau de dépendance – ordre linéaire (Mertens, 2008, p. 89)

Dans l'ordre structural, c'est la hiérarchisation qui est mise en avant (sans préoccupation pour l'ordre des mots). La tête se situe au-dessus de ses dépendants. (FIGURE.7)

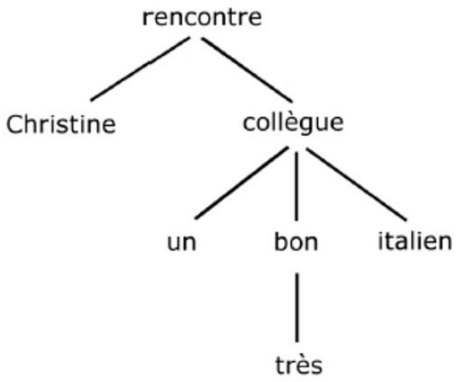

FIGURE.7 Réseau de dépendance – ordre structural (Mertens, 2008, p. 89)

L'ordre structural, de même que l'ordre linéaire, ne semble pas suffisant pour représenter la construction d'un énoncé. Une nouvelle représentation s'est donc développée pour les phrases simples. Cette représentation combine ordre structural et linéaire. (FIGURE.8). L'ordre linéaire est présenté sur l'axe horizontal ; l'ordre structural est présenté sur l'axe vertical.

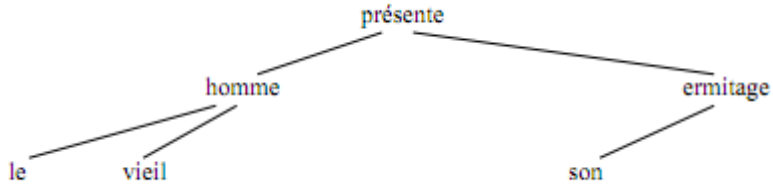

FIGURE.8 Réseau de dépendance – Combinaison des ordres (Brunet-Manquat, 2000, p. 20)

Contrairement à l'approche en constituants, l'approche en dépendances propose donc une représentation hiérarchique identique pour (17) et (18) (FIGURE.9). Seul l'ordre linéaire est modifié. Pour le phénomène des antépositions sans verbe intermédiaire, il n'y a donc pas lieu, à la différence des modèles en constituants, de parler de distance.

(17) Huit ans il avait

(18) Il avait huit ans

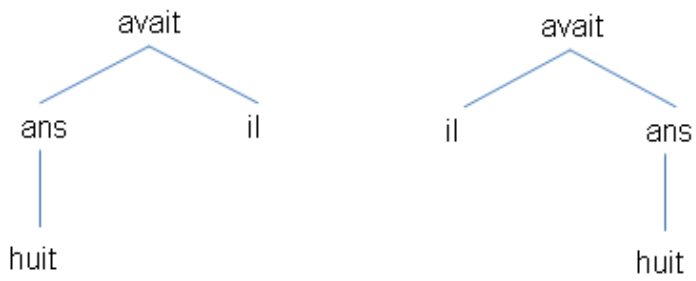

FIGURE.9 Réseau de dépendance des exemples (17)(18)

Il en est de même pour les interrogatives simples. L'interrogatif peut se trouver dans deux positions différentes, mais structurellement, il est toujours le nœud fils du verbe. L'ordre des mots donne lieu à des réflexions que nous ne détaillerons pas ici (Kahane, 2002; Tesnière, 1959).

(19) *Tu demandes à qui ?*

(20) *À qui tu demandes ?*

Les interrogatives complexes fonctionnent de la même manière. La représentation structurale montre simplement les relations de dépendance entre le verbe de la subordonnée et son complément interrogatif.

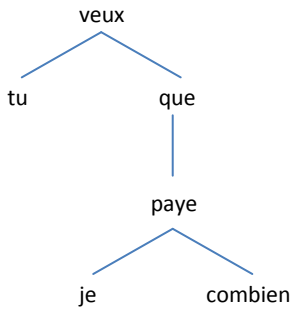

FIGURE.10 Réseau de dépendance des 'dépendances à distance' (Dikovsky & Modina, 2000, p. 87)

Pour les relatives simples, en revanche, il est plus difficile de considérer que seul l'ordre des mots est modifié, puisque le relatif s'ajoute (21)(22).

(21) *Nous voyons l'homme*

(22) l'homme que nous voyons

Dans l'exemple (22), le pronom relatif « que » dépend syntaxiquement de son antécédent (« homme ») mais aussi du verbe de la relative (« voyons ») puisque le pronom est l'objet direct du verbe. D'autre part, le pronom relatif est la tête de la relative et donc en gouverne le verbe (« voyons »). Il y a alors un problème de cycle infini : « que » dépend de « voyons », en tant qu'objet du verbe, qui lui-même dépend de « que », la tête de la relative. (FIGURE.11)

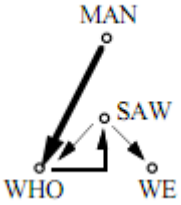

FIGURE.11 Réseau de dépendance des relatives (Mel'čuk, 2003, p. 68)

(Vergne, 1998) propose une représentation intéressante de ces constructions. La connexion s'établit entre l'antécédent et le verbe de la relative, mais l'ensemble de la relative forme un tout. Le verbe étant la tête de la relative, il n'y a qu'un pas pour dire que l'antécédent est finalement connecté à l'ensemble de la relative, par l'intermédiaire du pronom relatif.

Dans (23), l'antécédent « combats » est le gouverneur du verbe de la relative « débuter » ; il est relié à l'ensemble de la relative par l'intermédiaire du pronom relatif « qu-i » (FIGURE.12)[4].

(23) les combats qui ont débuté au mois de décembre (Vergne, 1998, p. 8)

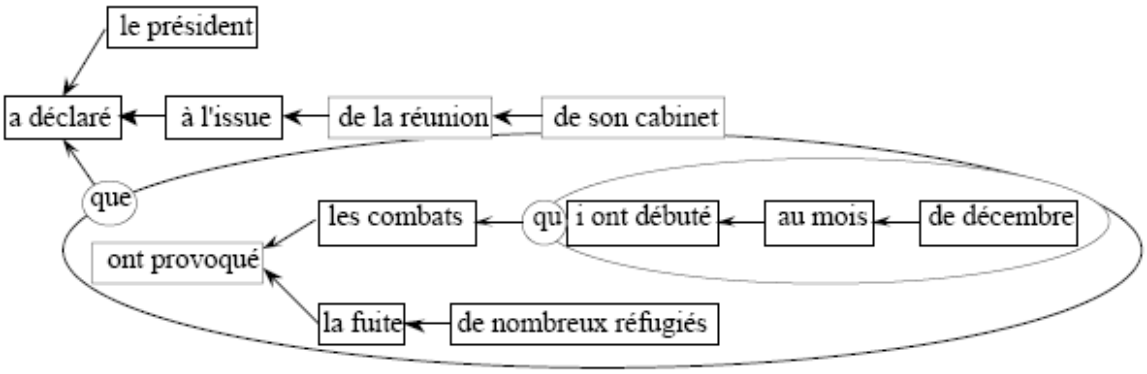

FIGURE.12 Réseau de dépendance d'une phrase avec complétive et relative simple (Vergne, 1998, p. 8)

[4] Dans cette figure, les flèches sont orientées dans le sens inverse de celui présenté dans la FIGURE.6. Ce sont les subordonnés qui pointent vers le gouverneur.

Le problème se complique encore pour décrire les relatives complexes. Pour *l'homme que je pense que nous voyons*, la représentation est la suivante (FIGURE.13) :

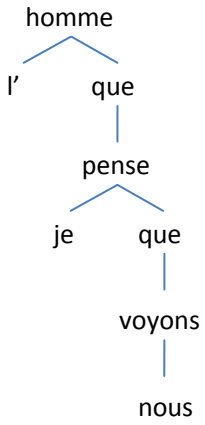

FIGURE.13 Réseau de dépendance d'une relative complexe

Mais dans cette représentation, il manque le lien entre le verbe « voir » et son complément « homme », par l'intermédiaire du relatif « que ».

Dans le modèle de Tesnière, les dépendances sont donc des relations de terme à terme. Dans son extension, (Kahane, 2000) modifie cette relation de un à un en proposant la notion de 'nucléus verbal'. Le gouverneur peut être complexe et se réaliser sous la forme d'une chaîne verbale : auxiliaire + participe, verbe + infinitif (éventuellement avec une préposition), verbe + que + verbe. Il peut également intégrer des prédications complexes formées avec des noms ou des adjectifs : *avoir l'intention de + infinitif, faire mine de + infinitif, être impossible de + infinitif, sembler heureux de + infinitif.*

La représentation par bulle (Kahane, 1997; 2000) met en évidence le lien particulier entre les deux verbes. Le cas des dépendances à distance en infinitive est le plus évident. La bulle formée par la chaîne verbale rapproche le verbe recteur de son complément antéposé (FIGURE.14).

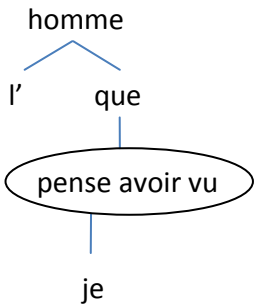

FIGURE.14 Arbre à bulle simplifié d'une dépendance à distance en infinitive

La représentation est identique lorsque la chaîne verbale n'est pas continue, ce qui est forcément le cas des dépendances à distance en complétive (FIGURE.15).

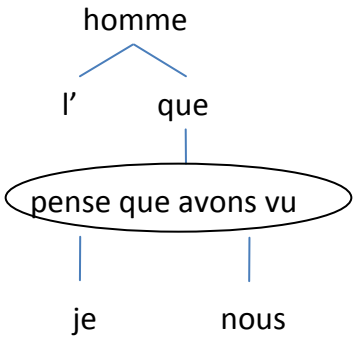

FIGURE.15 Arbre à bulle simplifié d'une dépendance à distance en complétive

La bulle regroupe l'ensemble de la chaîne verbale en un gouverneur unique. A l'intérieur, des dépendances entre les éléments de la chaîne existent. Ce sont les mêmes que celles qui auraient été définies en grammaire de dépendance traditionnelle.

Dans tous les cas précédents, le modèle développé par (Kahane, 1997; 2000) nous montre que l'on n'a pas besoin de parler de dépendance à distance, qu'elle soit courte (interrogatives simples) ou longue. Evidemment cette économie a sa contre partie : la création d'une structure linguistique spécifique : la bulle, que l'on peut interpréter comme un gouverneur complexe. La complexité d'une structure passe de l'empilement de dépendances à la structure du gouverneur.

1.1.3 L'approche pronominale

L'approche pronominale (Blanche-Benveniste et al., 1990; Blanche-Benveniste, Deulofeu, Stéfanini, & Van den Eynde, 1987; Van den Eynde & Mertens, 2003) est basée sur le modèle de la grammaire en dépendances. Elle en a conservé les idées principales : l'élément central est le verbe, qui possède une valence[5], c'est-à-dire des compléments spécifiques à un verbe. Les autres compléments, qui sont non-spécifiques, sont des ajouts. Les objets de la valence et les ajouts forment la rection.

La spécificité de cette approche est d'étudier la dépendance verbale à partir des pronoms.

[5] Terme déjà présent chez Tesnière.

> « C'est le pronom – ou l'unité sous-jacente induite à partir du pronom – qui constitue la base linguistique de l'énoncé. Les autres éléments peuvent être présentés comme étant le résultat du processus de lexicalisation. » (Blanche-Benveniste et al., 1987, p. 26)

Les pronoms sont proportionnels au lexique. Ils symbolisent un paradigme potentiel et mettent en évidence les listes paradigmatiques effectivement produites. Leur référence sémantique est minimale (Van den Eynde & Mertens, 2003, p. 3-article en ligne), ce qui leur permet d'être représentatifs d'un grand nombre d'énoncés. Ils préfigurent le lexique sans toutefois présumer de l'acceptabilité lexicale. Grâce au pronom complément « me » du verbe « dévisager », par exemple, l'acceptabilité du trait '+personnel' est immédiatement identifiée (24). Le fait que « manger » prenne comme pronom complément « la » implique la possibilité d'un complément ayant un trait '+personnel' ou '-personnel', mais le pronom ne permet pas de prédire que le complément lexical « chaise » sera inacceptable, contrairement au complément lexical « pomme » (25)(26).

(24) Il me dévisage

(25) Il la mange -> il mange une pomme

*(26) Il la mange -> *il mange une chaise*

L'absence de proportionnalité lexique/pronom permet d'identifier les locutions ou proverbes. En effet, dans une locution, le complément du verbe est figé. Il n'appartient plus à un paradigme et ne peut donc pas commuter avec l'élément le plus général du paradigme : le pronom (27)(28)

*(27) Il donne sa langue au chat -> *il la donne*

*(28) Il prend la mouche -> *il la prend*

Les pronoms traditionnels ne permettent pas toujours de représenter toute la rection d'un verbe. Afin de représenter les circonstants, la liste des pronoms est étendue aux adverbes interrogatifs (*où, combien, pourquoi, quand, comment*) et à quelques autres adverbes (*ici, là, ainsi*). L'ensemble est regroupé sous le terme 'proforme'.

Le nombre de proformes, ainsi que le nombre de verbes, est limité. Identifier toutes les constructions possibles de la langue devient donc un objectif réalisable.

Cette approche n'est pour autant pas incompatible avec le modèle de la grammaire en constituants. En effet, une proforme est équivalente à un constituant ; elle entre dans le même paradigme et possède la même fonction syntaxique (sujet, objet, etc.)

(29) *Le bébé du [?voisin pleure] toutes les [?nuits à] partir de 2h*

(30) *[Le bébé du voisin]$_{il}$ pleure [toutes les nuits]$_{quand}$ [à partir de 2h]$_{quand}$*

Les proformes montrent également que l'élément constructeur de l'énoncé peut être un verbe, ou un ensemble de catégories :

(31) *J'en (mange)*

(32) *J'en (suis fier)*

(33) *J'en (prends note / fais mention…)*

La pronominalisation révèle les constructeurs complexes (34), mais indique aussi que certains verbes ont plusieurs emplois possibles, en constructeurs complexes et en constructeur à part entière (35).

(34) *Il **prend parti** pour le candidat de gauche / *il le prend*

(35) *Il **est capable** de grimper ce mur / il en **est capable** / il l'est*

Dans la lignée des structures à bulles de (Kahane, 2000), le constructeur peut être une chaîne verbale (Blanche-Benveniste, Non publié; Chu, 2008).

(36) *J'ai pu lui parler ; chaîne = pu parler*

(37) *Je crois devoir vous avertir de ça ; chaîne = crois devoir avertir*

Il est souvent admis que les auxiliaires, le verbe « aller » du futur proche et certains verbes modaux (les verbes à montée) sont transparents à la construction des verbes qu'ils accompagnent (38) et constituent donc avec eux des constructeurs complexes :

(38) *Elle va être furieuse /*elle y va (Blanche-Benveniste, Non publié, p. 3)*

Mais, comme l'indique (Blanche-Benveniste, Non publié),

> « L'accord est moins général pour analyser d'autres types de modaux, des verbes dits « à contrôle », comme *tenter de*, des verbes supports, *mettre fin à*, les causatifs *faire* et *laisser*, ou les verbes de perception, *voir, entendre* » (Blanche-Benveniste, Non publié, p. 2)

Chaque classe verbale a son propre fonctionnement, et l'ensemble des verbes d'une classe ne semble pas homogène. Néanmoins,

> « Quelle que soit l'analyse interne qu'on choisit pour ces séquences, il paraît très utile d'avoir un terme qui permette de désigner l'ensemble, au moins provisoirement, par une saisie externe » (Blanche-Benveniste, Non publié, p. 4)

La notion de chaîne verbale a pour but de délimiter une grande unité, ce qui va permettre notamment de réduire le nombre de structures avec dépendance à distance, comme avec les bulles de (Kahane, 2000).

(39) Qu'est-ce qu'il a cru pouvoir faire ; chaîne = a cru pouvoir faire

Enfin, les proformes rendent compte des limites de la syntaxe de rection ou microsyntaxe : elles ne sont équivalentes qu'aux compléments régis par le verbe. Les éléments non régis (40) sont des associés. La relation d'association a d'abord été décrite en terme négatif : outre l'absence de pronominalisation (41), les associés ne peuvent pas être clivés (42) ni contrastés (43).

(40) [Quand je suis arrivé], il était midi

*(41) *il était midi quand ?*

(42) #c'est quand je suis arrivé qu'il était midi

*(43) * quand je suis arrivé mais pas quand je suis parti il était midi*

L'approche pronominale a été complétée par le module macrosyntaxique (Berrendonner, 1990; Blanche-Benveniste, 2000; Deulofeu, 2003) afin de rendre compte des relations entre ces éléments et le reste de l'énoncé. Les critères définitoires sont devenus positifs : présence de modalités de construction (interrogative, exclamative), liberté de composition des catégories, interprétation sémantique non compositionnelle, etc. (Debaisieux, sous presse).

L'élément central d'un énoncé macrosyntaxique est le noyau, qui possède une autonomie discursive et qui porte la modalité (affirmative ou négative). Il possède des modalités interprétables en termes de force illocutoire. Les autres éléments, appelés satellites, ne peuvent pas être autonomes. Ils sont définis par rapport au noyau, selon des critères distributionnels : les préfixes ou prénoyaux précèdent le noyau (44), les postfixes ou postnoyaux le suivent (45). On parle également parfois de suffixes, pour les éléments non régis analysés en grammaire traditionnelle comme des subordonnées (complément en

puisque, même si, consécutives en *de sorte que, pour*, relatives détachées, compléments détachés en *mais*). (46) Cette distinction de suffixe tend à disparaître (Debaisieux, sous presse).

(44) [comme elle souffrait du ventre $_{préfixe}$] [on avait peur de lui détraqu- de lui déclencher une crise $_{noyau}$](Blanche-Benveniste et al., 1990, p. 130) dans (Benzitoun, 2006, p. 135)

(45) [ils n'ont pas fait ça de leur propre chef $_{noyau}$] [ces C.R.S. $_{postfixe}$](Blanche-Benveniste et al., 1990, p. 141) dans (Benzitoun, 2006, p. 135)

(46) [On attendait donc le coup de sifflet convenu $_{noyau}$], [quand soudain des cris sauvages retentirent $_{suffixe}$]. (Verne, Tour du monde en 80 jours)(Blanche-Benveniste et al., 1990, p. 136) dans (Benzitoun, 2006, p. 135)

La structure prosodique est souvent particulière pour chacun des types de groupes macrosyntaxiques. Les prénoyaux s'achèvent sur un ton haut, à valeur continuative. L'auditeur perçoit que l'énoncé est en suspens ; il attend la production du noyau (IMAGE 1). Les postnoyaux ont une intonation plutôt basse et plate, ou faiblement descendante (IMAGE 2).

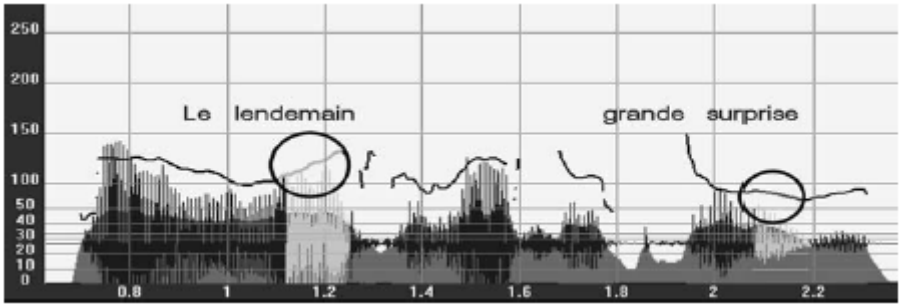

IMAGE 1. Représentation prosodique de « le lendemain grande surprise » : Prénoyau – Noyau (Martin, 2006, p. 155)

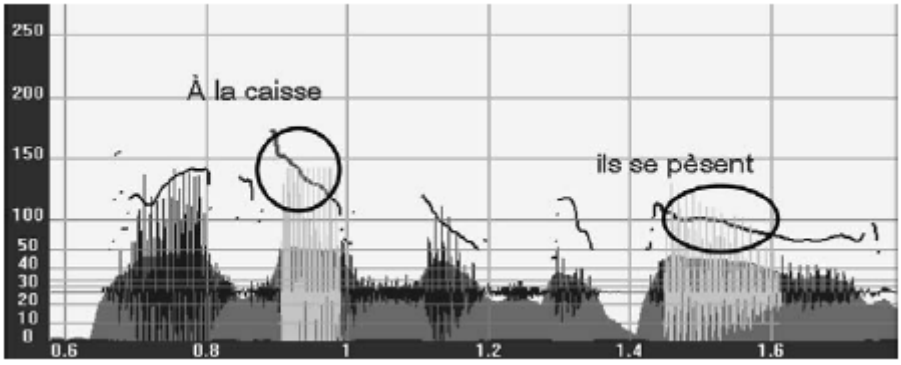

IMAGE 2. Représentation prosodique de « à la caisse ils se pèsent » : Noyau – Postnoyau (Martin, 2006, p. 156)

Les groupes macrosyntaxiques, quelle que soit leur position, n'ont pas de préférences catégorielles, ni de structure hiérarchique. Le noyau peut être occupé par un complément d'objet, tandis que le verbe est relégué en postnoyau (47). Il n'y a pas simplement un déplacement dans l'ordre des mots, mais une réorganisation complète de l'information.

(47) Un franc il manque (J. Pohl, cité par (Blanche-Benveniste, 2000, p. 115))

Ainsi, c'est la prosodie qui permet notamment de choisir entre plusieurs interprétations possibles : prénoyau-noyau ou noyau-postnoyau pour (48).

(48) [Demain] [il vient]

Les répartitions des rôles noyau/satellites touchent également le dispositif de clivée. On distingue les clivées contrastives, réalisées en noyau-postnoyau (49) et les clivées non contrastives, réalisées en prénoyau-noyau (50).

(49) c'est depuis trois ans que j'occupe ce poste [CRFP] (Debaisieux, sous presse)

(50) alors c'est comme ça que j'ai trouvé le moyen de venir [CRFP] (Debaisieux, sous presse)

Les non-régis (qui peuvent également être des intranoyaux) sont insérés dans l'énoncé mais ils ne sont pas enchâssés. On peut en trouver dans toutes les constructions. Ils enrichissent l'énoncé par des procédés autres que les combinaisons de rection. Leur présence dans un énoncé ne rend pas plus complexe la structure de rection à laquelle ils sont associés. Ils n'introduisent donc pas de 'distance grammaticale' entre un régi et son recteur :

(51) A qui (tu penses / à ton avis/ sincèrement) je dois m'adresser ?

Les insertions macrosyntaxiques ne doivent donc pas être prises en compte au même titre que les enchâssements microsyntaxiques dans l'évaluation de la complexité d'un énoncé. Cette remarque nous sera utile plus tard.

Ce cadre théorique, par sa vision micro- et macrosyntaxique, analyse peu de phénomènes comme constructions de dépendance à distance. Les relatives simples et les interrogatives relèvent de l'ordre topologique, comme en grammaire de dépendances. Chaque catégorie (clitique, mots QU-, éléments lexicaux) obéit à une micro-grammaire spécifique qui détermine sa position linéaire dans la chaîne (Blanche-Benveniste et al., 1987, p. 69-75), ainsi que les possibilités de cooccurrences (clitiques). Par exemple, la microgrammaire de la catégorie pronom relatif et interrogatif impose la position en tête de construction verbale,

alors que celle des clitiques les attache étroitement au verbe, sans modification de la relation de dépendance au verbe.

Les dislocations à gauche sont définies comme des prénoyaux, qu'il y ait reprise pronominale ou non (*ça c'est vieux mais de quelle année je saurais pas dire*). La macrosyntaxe invite à considérer qu'il y a un contrôle sémantique et non syntaxique entre des éléments que l'on a envie de rapprocher. On peut par exemple considérer que dans *le chocolat, j'adore* la relation entre les deux termes est de l'ordre de la macrosyntaxe, comme dans *le chocolat, merveilleux !* avec le verbe « adorer » en emploi absolu : il n'a pas de complément. Il n'y a donc pas de relation de dépendance syntaxique entre l'élément détaché et le verbe, mais une simple relation macrosyntaxique interprétable comme '*topic-comment'*. Par conséquent, dans un énoncé comme (52), le modèle ne considèrera pas qu'il y a contrôle à distance.

(52) Ce concert , je crois que je vais adorer

Les extractions du type *Huit ans il avait* sont analysées en terme de noyau-postnoyau. Dans ces constructions, en revanche (53) (avec *à Paris* focalisé), une analyse en dépendance à distance est possible, puisque la relation de rection[6] reste marquée par la préposition en (54), et qu'il est possible d'employer un verbe à complément obligatoire, à la différence d'une structure prénoyau-noyau.

(53) A Paris, j'ai décidé qu'on irait demain (et pas à Lyon)

*(54) A Pierre, je pense qu'il ressemble, Paul / *Pierre, je pense qu'il ressemble, Paul*

1.1.4 Récapitulatif

Les phénomènes nommés 'dépendance à distance' sont en fait conditionnés par le cadre théorique choisi. Certaines constructions, qui au départ étaient considérées comme une dépendance à distance, ne le sont plus car le modèle a évolué. On estime maintenant qu'il y a uniquement un contrôle sémantique. L'élément disloqué possède une place spécifique dans l'arbre 'phrase' des génératives ; dans l'approche pronominale, il est analysé en terme macrosyntaxique.

[6] En effet, une double analyse, micro- et macrosyntaxique, est possible sur un même énoncé.

les livres, elles les avaient envoyés
les livres, je pense qu'elles les avaient envoyés

Les dépendances sans verbe intermédiaire génèrent une complexité minimale pour la grammaire de constituants (mouvement court). Pour les autres modèles, la dépendance est locale (problèmes uniquement topologiques, effet des microgrammaires de catégories ou macrosyntaxe).

What did you find / Tell me who you talked to
La soupe que le chef déguste
Huit ans, il avait à l'époque
C'est huit ans qu'il avait à l'époque

Quant aux dépendances avec verbe intermédiaire, elles sont analysées comme subissant un mouvement long pour la grammaire de constituant. Pour les autres modèles, les constructions posent problème, sauf dans les dernières extensions. En grammaire de dépendance, le nucléus verbal permet de représenter les séquences verbales comme un tout, et la dépendance de l'élément extrait devient locale. Dans l'approche pronominale, la notion de chaîne verbale est équivalente du point de vue de la syntaxe externe.

Huit ans, je crois qu'il avait à l'époque
C'est huit ans que je crois qu'il avait à l'époque
Je ne sais pas qui Jean a dit qu'il avait vu
la pièce que l'évolution de la situation politique donnait l'impression qu'on pouvait enfin monter n'a pourtant pas été autorisée

TABLEAU 2. Phénomènes complexes en grammaire de constituants et en grammaire de dépendances

Quatre phénomènes sont donc interprétés comme ayant une complexité (TABLEAU 2) : soit une complexité dans leur structure (un élément dépend d'un verbe enchâssé : '*extraction from embedded clauses*') soit une complexité dans leur verbe (deux verbes forment un ensemble). Il s'agit des relatives complexes, des interrogatives complexes, des clivées complexes et des extractions complexes.

1.2 Délimitation du champ de recherche

Parmi les énoncés complexes établis par les différents cadres théoriques, voici ceux que nous retenons pour cette étude. Ce que nous appellerons désormais 'dépendance à distance' correspond aux exemples (55), (56) et (57). Ce sont les relatives, les interrogatives (cible de la majorité des études génératives sur les dépendances à distance) et les clivées.

(55) *Je ne sais pas qui Jean a dit qu'il avait vu (d'après (Abeillé, 2007))*

(56) *La pièce que l'évolution de la situation politique donnait l'impression qu'on pouvait enfin monter n'a pourtant pas été autorisée. (Godard, 1988, p. 38)*

(57) *C'est huit ans que je crois qu'elle avait à l'époque (d'après (Abeillé et al., 2008, p. 3))*

Nous écartons donc de cette analyse les extractions complexes marquées seulement par la prosodie et non par la syntaxe.

(58) *Huit ans, je crois qu'il avait à l'époque (Abeillé et al., 2008, p. 3)*

D'une part, ces constructions relèvent en partie de la macrosyntaxe et donc mettent certainement en jeu des paramètres notamment discursifs différents des autres structures. Il est en particulier souvent difficile de distinguer en l'absence d'analyses prosodiques fiables les cas pertinents où l'élément antéposé est noyau (59) et les cas où il est préfixe (60).

(59) *[A la caisse $_{noyau}$] ils se pèsent (Martin, 2006, p. 156)*

(60) *[comme elle souffrait du ventre $_{préfixe}$] on avait peur de lui détraqu- de lui déclencher une crise (Blanche-Benveniste et al., 1990, p. 130) dans (Benzitoun, 2006, p. 135)*

D'autre part, des considérations techniques sont venues appuyer ce choix : les extractions complexes non marquées syntaxiquement sont très difficiles à repérer automatiquement et un repérage manuel est rendu difficile par l'apparente rareté dans nos corpus de ce type de structure (61). Les incises sont certainement plus utilisées dans ce contexte (62).

(61) *? Moi, des pommes, je pense que je vais acheter*

(62) *Des pommes, je pense, je vais acheter*

Nous ne nous inscrivons dans aucun modèle théorique a priori. L'objectif n'est pas de vérifier au moyen du corpus les assertions et la cohérence interne de chaque modèle mais de décrire le fonctionnement et l'usage de ces structures qui ont un certain nombre de paramètres syntaxiques en commun puis de choisir le modèle qui le décrit le mieux. Le cadre de l'Approche Pronominale, qui permet une analyse à la fois en constituant et en dépendance, sera utilisé comme cadre d'analyse aussi léger que possible.

Dans les séquences que nous étudierons, un complément se situe donc hors du cadre de la construction verbale dont il dépend (CV2). Il est placé avant une autre construction verbale (CV1) contenant ce que nous appellerons un 'verbe pont'. L'ordre des élements de la séquence est ainsi « complément CV1 CV2 ». Le complément antéposé peut être un sujet, un objet ou un ajout du verbe. Toutes les combinaisons structure-fonction sont possibles (TABLEAU 3).

Structure	Fonction	Exemple
Relative	Objet	*Ça, c'est des éléments, en tout cas,* ***qu'****on pense que le gouvernement n'a pas* ***évalués***
Relative	Sujet	*il y a trois échographies // d'ailleurs vous dites que c'est ce qui est + couramment admis // c'est pas forcément couramment admis c'est ce* ***qu'****on pense qui* ***devrait être fait***
Relative	Ajout	*c'est un crapaud qui a eu la malchance de tomber dans une famille* ***où*** *il fallait qu'il* ***s'investisse*** *un peu*
Interrogative	Objet	***que*** *veux-tu que j'en* ***fasse***
Interrogative	Sujet	*personne ne voudra aller là-haut //* ***qui*** *voulez-vous qui* ***grimpe*** *comme-ça*
Interrogative	Ajout	***pourquoi*** *tu penses qu'ils* ***ont pas voulu***
Clivée	Objet	*c'est* ***cette thématique là*** *qu'il faut que nous* ***soyons capables*** *(si en tout cas nous estimons que c'est une bonne chose)* ***de populariser*** *auprès des salariés du privé*
Clivée	Sujet	*quand il y avait un client ben la patronne elle se dérangeait pas c'est* ***moi*** *qu'il fallait qui* ***aille servir***
Clivée	Ajout	*C'est* ***dans ce cadre de service public de proximité*** *que je pense qu'il* ***faut réfléchir*** *au rôle, aux missions et aux moyens de la police et de la justice.*

TABLEAU 3. Combinaisons des structures syntaxiques et des fonctions de l'élément extrait dans les constructions de dépendance à distance

Le CV1 peut construire une complétive (63) ou une infinitive (64).

(63) Otto Kranz est loin de tout contrôler, contrairement à ce qu'il souhaite que vous pensiez (CERF, LITC)

(64) c'est la responsable qui qui décide avec les représentants des des maisons d'édition des nouveautés qu'elle souhaite recevoir (CRFP, PRO-NAN-1)

Plusieurs autres constructions auraient pu être considérées comme impliquant des dépendances à distance : les pseudo-dépendances à distance, les comparatives, les exclamatives et les dépendances à distance nominales.

La construction des pseudo-dépendances à distance est très proche de notre objet d'étude. Dans ce type de construction, le pronom interrogatif-relatif réalisé à distance est repris dans la complétive ou l'infinitive. Il y a un double marquage, qui peut être réalisé sous forme pronominale ou lexicale. Etant donné que le complément régi est présent dans la construction enchâssée, on ne peut pas parler de dépendance « à distance ». Il n'y a qu'un contrôle sémantique entre le pronom interrogatif-relatif et le verbe. C'est pourquoi nous les avons écartés de notre analyse. Les pseudo-dépendances à distance sont très fréquentes avec le pronom « dont ». Dans certains exemples, « dont » pourrait être interprété comme le complément du verbe qui le suit immédiatement : « dire de / à propos de », « voir à propos de » etc. (65). Néanmoins, il semble que cette construction soit en fait une construction utilisée à la place d'une dépendance à distance. En effet, on la trouve avec des verbes qui s'emploient très rarement avec un complément pronominalisable en « dont » : « oublier à propos de ? », « nier à propos de ? », « reconnaître à propos de ? » (66)(67). Les pseudo-dépendances à distance en « dont » seraient l'équivalent d'une dépendance à distance du sujet dont la forme est perçue comme fautive (68).

*(65) La formulation vise donc d'abord la simplicité, pour être plus facilement mémorisable, utilisant jusqu'à l'abus les systèmes d'appariements et de doublets symétriques, **dont** on a déjà vu qu'**ils** pouvaient masquer parfois certaines subtilités de raisonnement. (CERF, DIVS)*

*(66) les vérités sont des illusions **dont** on a oublié qu'**elles** le sont (CERF, SCIEN)*

*(67) Je recommande aux écrivains et aux éditeurs de réclamer la part de la part de cette TVA **dont** ils estimeraient qu'**elle** leur revient, parce qu'il y aurait eu valeur ajoutée à notre fond culturel. (CERF, FORM)*

*(68) La formulation vise donc d'abord la simplicité, pour être plus facilement mémorisable, utilisant jusqu'à l'abus les systèmes d'appariements et de doublets symétriques, **qu'**on a déjà vu **qui** pouvaient masquer parfois certaines subtilités de raisonnement (adaptation de CERF, DIVS)*

D'après (Godard, 1988), cette construction de relative en « dont » avec reprise pronominale, qu'elle appelle 'DPR' est beaucoup moins contrainte que les relatives dans les dépendances à distance. Elle n'est pas sensible aux contraintes définies par les génératives (voir §1.3.1) (Godard, 1988, p. 40-42). En revanche, la liste des verbes ponts des DPR est plus restreinte (Godard, 1988, p. 71-73).

Dans les comparatives (69)(70), comme dans les constructions de dépendance à distance, le verbe recteur n'est pas saturé dans la subordonnée : « être [ainsi] ». Un verbe peut venir se placer entre le verbe recteur et le comparatif (*souhaiter*, *penser*, *vouloir*). En revanche, le complément n'est pas réellement antéposé ; il doit être reconstruit à partir du contexte gauche.

(69) Nous comprenons que Dieu nous accepte, que Dieu nous pardonne, que Dieu nous aime comme nous sommes. Pas ***comme nous souhaitons*** *être, pas* ***comme nous pensons qu'il faut que*** *nous soyons, pas* ***comme nous voulons que les autres pensent que*** *nous sommes, mais comme nous sommes. (CERF, DIVS)*

(70) Mais en est-il autrement des Pérec, Zidane, Flessel et autre Bouras ? Ne sont ils pas représentatifs de ces immigrés de la deuxième génération, ***tels que les français voudraient qu****'ils soient ? (CERF, FORM)*

Certains exemples ont les deux interprétations (71) : le comparatif « **le même** rebond **que** la caisse » ou une dépendance à distance dans une configuration de relative : « [le même rebond] **que** la caisse peut **donner** ». Dans la seconde interprétation, il y a bien un pronom relatif, suivi de la CV1 (*pouvoir*) et de la CV2 (*donner*), ce dernier ayant dans sa construction le relatif (*donner que* = donner un rebond).

(71) je m'entraîne avec des substituts de batterie qui sont des espèces de ronds en caoutchouc qui sont censés donner // c'est des pads d'entraînement euh spécialement faits pour ça qui sont censés donner à peu près le même rebond que peut te donner une caisse (CRFP, PRI-PSO-4)

Les exclamatives paraissent tout à fait éloignées de nos constructions. Pourtant, elles leur sont reliées par l'intermédiaire des questions rhétoriques, qui ont une valeur exclamative. Aucune différence formelle n'est visible entre les exclamatives et les questions rhétoriques (72).

(72) c'est c'est la folie on s'est dit mais qu'est-ce qu'on est parti faire chez Tarneaud (PFC, 44aar2)

Pour finir, nous avons limité notre étude aux dépendances à distance verbales. Lorsque le pronom antéposé était régi par un nom de la complétive, nous l'avons écarté, qu'il y ait reprise (pseudo-dépendance) (73) ou non (74). On retrouve dans ces exemples le problème du statut de la forme *dont*.

(73) Et puis, il y a Amina. Parmi ceux ***dont*** *Chevènement braille qu'il ne faut pas avoir de "faiblesse" à* ***leur égard****, elle est bien placée pour lui jeter son histoire dans les gencives. (CERF, PRES)*

(74) ne voyez devant vous qu'un monstre, ***dont*** *je suis heureux que vous ne puissiez apercevoir* ***la figure*** *(CERF, LITA)*

Tous ces types d'énoncés pourront faire l'objet d'études ultérieures.

1.3 Etat de l'art

Quel que soit le traitement théorique que l'on donne de ces séquences, il est avant tout important de décrire le détail de leur organisation syntaxique et lexicale Les constructions définies ci-dessus, désormais nommées 'dépendances à distance', ont d'abord été étudiées empiriquement, notamment par les génératives. Les contraintes limitant leurs formes syntaxiques définies par les génératives ont eu une grande importance dans la recherche. Toutes les études qui ont suivi ont pris appui sur celles-ci, même si le plus souvent, il s'agissait de tenter de les remettre en cause. Les premières conclusions étaient en effet basées sur des données issues de l'intuition des chercheurs, qui inventaient des exemples selon leur connaissance de la langue. Les études suivantes ont d'une part fabriqué des contre-exemples, et d'autre part soumis les données au jugement des locuteurs. Les expériences sur le jugement de grammaticalité ont permis aux chercheurs d'avoir un contrôle sur leurs intuitions. Enfin, l'accès à la parole spontanée, par les corpus, a fourni une preuve de l'existence de certaines données et démontré quels étaient les usages les plus courants. Néanmoins, les corpus ne sont qu'un échantillon de la langue et il est indispensable de combiner les résultats des corpus et des données expérimentales. Nous présenterons les résultats des précédentes études en tentant de tenir compte de l'évolution des données (intuition, expériences, corpus) et en les confrontant aux généralisations descriptives que l'on peut formuler à propos de ces séquences en termes de contraintes syntaxiques, sémantico-pragmatiques, cognitives et lexicales. Notre objectif n'est pas d'entrer dans les discussions internes aux modèles, et d'évaluer leurs

conclusions théoriques, mais de dégager leur contribution à notre connaissance du phénomène étudié.

1.3.1 Approches Générativistes transformationnelles : priorité aux contraintes syntaxiques (67, 77)

La théorie générative (Chomsky, 1977; Ross, 1967) présume qu'il existe un système commun à toutes les langues, appelé la 'Grammaire universelle'. Chaque langue instancie ce système en le paramétrant d'une certaine façon. Lors de l'acquisition de sa langue maternelle, l'enfant, à partir d'une généralisation de ce qu'il entend, met à jour les paramètres. C'est pourquoi il ne répète pas seulement ce qu'il entend, mais peut créer rapidement de nouveaux énoncés. L'objectif des générativistes est donc de définir ce qui appartient au système minimal et de trouver les règles les plus générales possibles qui permettent de passer du système à la langue telle qu'elle est produite. Les règles sont regroupées en domaines :

- les règles de base,
- les règles de transformation,
- les règles phonologiques,
- les règles d'interprétation sémantique. (Chomsky, 1977, p. 71)

Ce sont les règles de transformation, dont font partie les mouvements, qui permettent de passer de la structure profonde à la structure de surface (cf §1.1.1). Le fait que le mouvement soit une règle de la Grammaire universelle permet de réduire et de simplifier considérablement le système de base. Le principe général est représenté par la formule 'Move α', où α peut être n'importe quel nœud. Ce principe général doit ensuite être soumis à des contraintes, afin de décrire les réelles possibilités de la langue. Dans cette modélisation, le mouvement se produit entre la position canonique du complément désignée par '*gap*' (marquée par '_') et la position réalisée, le '*filler*'. Le *gap* garde la trace ('t') du complément déplacé.

En grammaire générative, l'analyse des dépendances à distance s'inscrit donc dans une étude bien plus large qui englobe notamment celle des relatives et des interrogatives simples, dans

lesquelles un élément est antéposé. La construction de dépendance à distance n'est pas étudiée comme construction en soi. Elle apparait uniquement comme résultat des tests de mouvement entre la structure profonde et la structure de surface d'énoncés qui impliquent plusieurs catégories enchâssées les unes dans les autres. Ces tests montrent qu'il existe des barrières qui bloquent le mouvement des constituants. Les génératives considèrent dans leur modèle de base qu'a priori le mouvement d'un constituant est possible à partir de n'importe quelle configuration CV1 CV2. Par exemple CV2 peut être objet ou ajout de V1, elle peut être objet d'un nom lui même objet de V1 etc. Le résultat des tests de déplacement montre qu'il n'en est rien et que certaines configurations de départ interdisent le mouvement. Le chercheur va interpréter ces configurations interdites comme des 'contraintes' sur le mouvement et donc sur les possibilités de créer des dépendances à distance. Comme le modèle donne une place centrale à la syntaxe, les génératives ont mis l'accent sur une formulation syntaxique de ces contraintes. La formulation des contraintes syntaxiques a évolué avec le développement du modèle pour s'adapter à certains contre-exemples. Elles ont en outre été articulées avec des contraintes lexicales et sémantico pragmatiques pour rendre la description plus précise. Nous citerons ici uniquement les règles les plus célèbres, et de manière simplifiée, en nous appuyant sur (Ross, 1967). Nous verrons successivement la contrainte des syntagmes nominaux complexes, la contrainte du sujet phrastique et la contrainte de l'ajout.

Les SN complexes (Complex Noun Phrase Constraint - CNPC)

Les noms font partie des 'nœuds cycliques'. Ils sont donc des barrières pour leurs dépendants, que ces derniers soient des adjectifs, des syntagmes prépositionnels, des verbes en –ing (première citation, exemplifiée par (75)), des complétives ou des relatives (deuxième citation, exemplifiée par (76) pour la complétive). On ne peut donc pas extraire une partie d'un SN (SN dont la tête est *girl* dans (75) et *claim* dans (76)) :

> « No element of a constituent of an NP which modifies the head noun may be questioned or relativized » (Ross, 1967, p. 125)

(75) **Who does Phineas know **a girl jealous of**?(Ross, 1967, p. 125)*

Who does Phineas know **a girl behind?*

Who does Phineas know **a girl working with?*

> « No element contained in a sentence dominated by a noun phrase with a lexical head noun may be moved out of that noun phrase by a transformation » (Ross, 1967, p. 127)

(76) **The hat which I believed **the claim that Otto was wearing** is red.(Ross, 1967, p. 126)*

Le sujet phrastique

On ne peut pas déplacer une partie du sujet phrastique ; le noeud 'S' (pour *sentence*), qui regroupe l'ensemble du sujet, étant également un noeud cyclique. Dans l'exemple suivant (77), l'élément *the teacher* ne peut pas être extrait par relativisation car il est gouverné par la phrase (S) *the principal would fire {the teacher}*, elle-même gouvernée par un noeud assimilable à un SN (*that the principal would fire {the teacher}*), sujet de l'ensemble de la phrase S :

> « No element dominated by an S may be moved out of that S if that node S is dominated by an NP which itself is immediately dominated by S » (Ross, 1967, p. 243)

(77) *That the principal would fire some teacher was expected by the reporters.*

The teacher who **that the principal would fire was expected by the reporters is a crusty old battleax.(Ross, 1967, p. 241)*

L'ajout

On ne peut pas non plus déplacer une partie d'un ajout, c'est-à-dire un complément non sélectionné par le verbe (*theta-governed*) :

(78) **What did she leave the movie **because they were eating**? ((Ross, 1967) dans (Ambridge & Goldberg, 2008, p. 359))*

Les trois environnements cités précédemment (le nom complexe, le sujet phrastique et l'ajout) empêchent donc le mouvement, et par conséquent la construction de dépendance à distance. Lorsque le mouvement est autorisé, il faut noter que le mouvement peut être récursif, comme on le voit en (79).

> « The rule of wh-movement applies successive cyclically » (Chomsky, 1977, p. 74)

(79) *Who did Mary hope that Tom would tell Bill that he should visit? (Chomsky, 1977, p. 74)*

Les contraintes syntaxiques sont ainsi mises au jour mais les génératives eux-mêmes reconnaissent que la syntaxe n'explique pas tout.

> « For some reason that is obscure to me, I find sentences like those in (2.23a), where the embedded question consists of a wh-word followed by an infinitive, by and large more acceptable than corresponding sentences, like those in (2.23b), where the wh-word is followed by a clause with a finite verb.
>
> 2.23a *he told me about a book which I can't figure out whether to buy or not / how to read / where to obtain / what to do about*
> 2.23b *he told me about a book which I can't figure out why he read / ?wheter I should read / ??when I should read »* (Ross, 1967, p. 27)

D'autres pistes ont alors été suivies afin de décrire les contraintes qui régissaient les constructions de dépendance à distance. D'abord, ce sont les contraintes sémantico-pragmatiques qui ont été explorées.

1.3.2 Contraintes sémantico-pragmatiques

Les trois contraintes syntaxiques majeures ne permettent pas d'expliquer les différences de grammaticalité qui existent entre des énoncés qui répondent à un même schéma syntaxique. L'hypothèse développée ensuite est que des contraintes sémantico-pragmatiques s'ajoutent aux contraintes syntaxiques, voire les remplacent.

1.3.2.1. Erteschik-Shir (73)

Dès 1973, (Erteschik-Shir, 1973) tente de trouver une alternative aux contraintes syntaxiques définies par les génératives. En effet, toutes les séquences que permettent les contraintes de (Ross, 1967) ne sont pas acceptables. Ainsi ces contraintes, formulées sur des variables, autorisent en principe sans limitation le mouvement à partir d'une *clause* objet d'un verbe or cette configuration en principe autorisée produit des énoncés inacceptables :

(80) **We discussed the issue that he **exulted** that you had understood (Erteschik-Shir, 1973, p. 94)*

(81) *We discussed the issue that he **was happy** that you had understood (Erteschik-Shir, 1973, p. 94)*

Selon elle, l'extraction est soumise à des contraintes sémantiques basées sur la notion de dominance. Elle définit ainsi la dominance :

> « A clause or phrase is semantically dominant if it is not presupposed and does not have contextual reference. » (Erteschik-Shir, 1973, p. 22)

Il s'agit d'un terme proche de celui de l''assertion', ou du 'focus'. Un élément dominant sémantiquement est une information nouvelle, assertée. Pour définir quel élément est sémantiquement dominant ou subordonné dans une phrase, (Erteschik-Shir, 1973) propose quatre tests : la réfutation, les *question-tag*, l'interprétation de l'adverbe et l'interjection.

1) Réfutation

Le test de réfutation, emprunté à (Ross, 1967), consiste à enchâsser la phrase, puis à la nier. A partir de (82), on obtient (83). L'élément interprété comme étant réfuté est l'élément dominant sémantiquement. Dans notre exemple, « *Bill saw* » est dominant (84), et « *the picture of his mother* » est subordonné, parce qu'il ne peut pas être réfuté (85).

(82) Bill saw the picture of his mother

(83) Tom said « Bill saw the picture of his mother », which is a lie.

(84) Tom said « Bill saw the picture of his mother », which is a lie. He didn't.

*(85) *Tom said « Bill saw the picture of his mother », which is a lie. Such a picture doesn't exist.(Erteschik-Shir, 1973, p. 12)*

Certaines phrases, comme (86), sont ambiguës. Dans ce cas, les deux parties peuvent être interprétées comme étant dominantes (87)(88).

(86) Bill thinks that they're gonna win

(87) Tom said « Bill thinks that they're gonna win », which is a lie. He doesn't.

(88) Tom said « Bill thinks that they're gonna win », which is a lie. They're not.(Erteschik-Shir, 1973, p. 11)

2) Question-tag

Le deuxième test qui permet de trouver l'élément sémantiquement dominant d'un énoncé est le test de la *question-tag*. La *question-tag* est une construction anglaise qui est généralement

juxtaposée à une phrase. Sa fonction est de demander une confirmation à un interlocuteur. La *question-tag* reprend sous la forme d'un auxiliaire et d'un pronom sujet la construction verbale qui demande confirmation. L'auxiliaire porte la modalité contraire à celle de la construction verbale, mais garde le même temps et la même conjugaison. La *question-tag* peut être traduite par « n'est-ce pas » en français.

En anglais, ce qui peut donner lieu à une *question-tag* est sémantiquement dominant. Dans l'exemple suivant (89), « *He mumble* » est dominant. Il autorise la production de la *question-tag* « *didn't he* ». En revanche, « *John was a fool* » est sémantiquement subordonné, puisque la *question-tag* « *wasn't he* » rend l'énoncé agrammatical (90).

(89) He mumble that John was a fool, didn't he ?

*(90) *He mumble that John was a fool, wasn't he ? (Erteschik-Shir, 1973, p. 12)*

Là-encore, certaines phrases peuvent donner lieu à deux interprétations ; les deux *question-tag* sont alors correctes (91)(92).

(91) I believe that Dr. Spock is innocent. Don't I ?

(92) I believe that Dr. Spock is innocent. Isn't it ? (Erteschik-Shir, 1973, p. 12)

3) Interprétation de l'adverbe

Le troisième test s'appuie sur l'interprétation de l'adverbe. D'après Brecht (1972)[7], certains adverbes sont rattachés à la construction verbale subordonnée, mais portent sur l'opinion du locuteur. Le fait énoncé étant donc simplement un point du vue du locuteur, il peut être contredit par un second locuteur. Ainsi, dans (93) le locuteur John a fait part de son sentiment de certitude à l'égard de l'affirmation « *be dead by midnight* » grâce au recours à l'adverbe « *doubtlessly* ». Le second locuteur exprime son désaccord ; son point de vue sur l'affirmation « *be dead by midnight* » est totalement différent.

*(93) John assumes that he **doubtlessly** will be dead by midnight, but I for one have my doubts (Erteschik-Shir, 1973, p. 13)*[8]

[7] Brecht, R.D. (1972) "Modal Reference and Presupposition", North-East Linguistic Society III.

[8] John suppose qu'il sera sans doute mort avant minuit, mais pour ma part, j'ai des doutes

D'autres adverbes, au contraire, portent sur la situation. Dans (94), l'affirmation est « *doubtlessly be dead by midnight* ». Elle ne dépend pas du point de vue de John. Dans ce cas, l'énoncé n'est pas réfutable.

*(94) *John understand that he **doubtlessly** will be dead by midnight, but I for one have my doubts (Erteschik-Shir, 1973, p. 13)*[9]

Dans l'exemple (94), seule la CV1, également appelée 'matrice', est considérée comme dominante, puisque le contenu de la CV2 ne peut pas être nié. En revanche, dans l'exemple (93), la matrice et la CV2 enchâssée peuvent chacune être interprétée comme dominante. Pour Brecht, ce test est lié au caractère factif/non factif des verbes[10]. Lorsque le verbe est non factif (« *assume* »*)*, la subordonnée peut être assertée, et réfutée. Les compléments des verbes factifs (« *understand* »*)*, eux, sont présupposés ; leur contenu ne peut donc pas être contesté. Voici un tableau récapitulatif qui met en valeur l'opposition entre les deux exemples étudiés.

	Exemple (93)	Exemple (94)
Adverbe	Opinion du locuteur	Situation
Verbe	Non-factif	Factif
Subordonnée	Assertée – réfutable	Présupposée - *réfutable
Dominance	Matrice / subordonnée	Matrice

TABLEAU 4. Interprétation de l'adverbe et conséquences sur le verbe de la matrice, la subordonnée et la dominance sémantique

Cette position est contestée par (Erteschik-Shir, 1973), pour qui seule la dominance sémantique est en jeu. Dans l'exemple (95) « m*umble* » est un verbe non factif, mais il se comporte comme les verbes factifs pour ce test : la subordonnée ne peut pas être niée.

*(95) ?Fred mumbled that Bess would **certainly** marry him, but it is not at all certain (Erteschik-Shir, 1973, p. 14)*

Etant donné que le verbe est non-factif, la matrice et la CV2 enchâssée devraient chacune pouvoir être interprétée comme dominante. Pourtant, on ne peut pas considérer que la subordonnée soit dominante, comme le prouve le test de réfutation.

[9] *John comprend qu'il sera sans doute mort avant minuit, mais pour ma part, j'ai des doutes

[10] Un verbe factif possède un complément présupposé, dont la valeur de vérité ne dépend pas de l'assertion ou de la négation du verbe.

(96) Bill said "Fred mumbled that Mary is a fool", which is a lie -- he didn't.

(97) ?? Bill said "Fred mumbled that Mary is a fool", which is a lie -- She isn't. (Erteschik-Shir, 1973, p. 14)

Le fait de pouvoir nier la subordonnée ne serait donc pas lié à la factivité mais à la dominance sémantique.

4) Interjection

Le test de l'interjection, qui est le dernier test de (Erteschik-Shir, 1973), est emprunté à James (1972)[11]. Il consiste à ajouter une interjection, marquant la surprise. Cette marque porte sur l'élément nouveau de l'énoncé, qui est donc sémantiquement dominant. Dans les exemples suivants, l'élément dominant est souligné.

(98) Ah it's reported by <u>Newsweek</u> that Kissinger is a vegetarian !

(99) Ah it's reported by Newsweek that <u>Kissinger is a vegetarian</u> !

(100) Ah it's editorialized by <u>today's Times</u> that Kissinger is a vegetarian ! (Erteschik-Shir, 1973, p. 15)

En revanche, l'énoncé (101) est impossible.

*(101) *Ah it's editorialized by today's Times that <u>Kissinger is a vegetarian</u> !*

Le test de l'interjection aboutirait aux mêmes conclusions que celui de la réfutation. Il renforce l'argumentation de (Erteschik-Shir, 1973).

Les quatre tests permettent d'identifier la partie de l'énoncé qui est dominante. (Erteschik-Shir, 1973) tente ensuite de démontrer que l'extraction est liée à la dominance sémantique, en s'appuyant sur l'Anglais et le Danois :

> « Extraction can occur only out of clauses or phrases which can be considered dominant in some context. » (Erteschik-Shir, 1973, p. 27)

Les transformations s'appliquent donc aux éléments sémantiquement dominants. Or dans le discours, ce sont les parties interrogées, relativisées ou clivées qui portent l'information centrale, si bien qu'il n'est pas étonnant que l'ensemble de la CV2 enchâssée à laquelle elles

[11] James, B. (1972) "Some Aspects of the Syntax and Semantics of Interjections", Papers from the Eighth Regional Meeting of the Chicago Linguistic Society

appartiennent soit sémantiquement dominant. Cela signifie en outre que la matrice est sémantiquement subordonnée. Ainsi, les CV1 sémantiquement subordonnées autorisent l'extraction. (Erteschik-Shir, 1973) est à l'origine de la terminologie de 'pont', reprise dans les études ultérieures, et que nous-mêmes utiliserons (ainsi que 'verbe pont').

> « The following more picturesque terms will also be used: A matrix which is subordinate (i.e., where the embedded clause is dominant and allows extraction) will be called a bridge. Extraction out of an island is therefore possible only across a bridge. » (Erteschik-Shir, 1973, p. 83)

Les verbes factifs, dont les compléments ont un contenu présupposé, et qui sont donc sémantiquement subordonnés, ne devraient pas permettre l'extraction. Pourtant, de nombreux exemples l'attestent. Pour pallier cette difficulté, (Erteschik-Shir, 1973) s'appuie sur la proposition de Karttunen (1973)[12] selon qui certains verbes factifs ont un complément qui n'est pas vraiment présupposé ; le complément apporte une nouvelle information, ce que confirme le test de réfutation (102)(103).

(102) Jack has children and it is strange that all of them are bald.(Karttunen, dans (Erteschik-Shir, 1973, p. 91))

(103) Harvard regrets that children cannot be accommodated (Karttunen, dans (Erteschik-Shir, 1973, p. 91))

L'exemple (103) est très clair et on imagine facilement une brochure dans laquelle cette phrase aurait pour but d'informer que « les enfants ne peuvent pas être logés ». Ainsi, dans certains contextes, le verbe (factif) « *regret* » prend des compléments qui sont interprétés comme sémantiquement dominants. Il peut alors être verbe pont, même dans des énoncés dans lesquels son complément est présupposé ((104) construit à partir de (105)).

(104) We discussed the issue that I regretted that I had ever thought of (Erteschik-Shir, 1973, p. 92)

(105) I regretted that I had ever thought of the issue (Erteschik-Shir, 1973, p. 92)

C'est la raison pour laquelle il est précisé dans la définition « *in some context* » :

> « Extraction can occur only out of clauses or phrases which can be considered dominant in some context. » (Erteschik-Shir, 1973, p. 27)

L'auteur remarque en outre que la complexité sémantique du verbe a un impact sur la possibilité d'extraction. Les verbes factifs complexes sont toujours jugés inacceptables en

[12] Karttunen L. (1973) "Remarks on Presupposition", presented at the Texas Conference on Performances, Conversational Implicature, and Presuppositions.

verbe pont, et ce, par tous les locuteurs interrogés : « *rejoice* » (106), « *exulte* » (107), par opposition au verbe plus simple « *be happy* » (108).

(106) **We discussed the issue that he **rejoiced** that you had understood (Erteschik-Shir, 1973, p. 94)*

(107) **We discussed the issue that he **exulted** that you had understood (Erteschik-Shir, 1973, p. 94)*

(108) *We discussed the issue that he **was happy** that you had understood (Erteschik-Shir, 1973, p. 94)*

Bien que la notion de complexité soit difficile à définir[13], elle semble assez intuitive, en tout cas lorsque l'on compare des verbes appartenant au même champ sémantique. La difficulté d'extraire un complément d'un verbe complexe ne se limite pas aux verbes factifs. Dans tous les cas, les verbes avec lesquels l'extraction est la mieux acceptée sont les verbes simples. Dans le champ sémantique de DIRE, (Erteschik-Shir, 1973, p. 84-86) distingue trois classes de verbes, pour lesquels l'extraction est acceptable, douteuse et mauvaise :

- La première classe rassemble les verbes qui signifient uniquement « dire »
- La seconde rassemble les verbes qui décrivent plus ou moins précisément la façon dont cela a été dit
- La troisième rassemble les verbes qui sont rarement associés au sens « dire » et les verbes qui mentionnent le contexte de la situation

Voici la répartition des verbes dans les trois classes, et leur traduction française (TABLEAU 5).

Acceptable	*say, tell, report, announce* Dire, dire, déclarer, annoncer
Discutable/douteux ??	*grunt, holler, murmur, mumble, mutter,* *roar, scream, shout, sigh, snort,* *stammer, wail, whine, tell, exclaim* Grogner, brailler, murmurer, marmonner, bredouiller, rugir_vociférer, crier, crier, soupirer, grogner_mépris/pouffer de rire, balbutier, gémir, se plaindre, dire, s'exclamer
Mauvais *	*purr, snarl, editorialize,* *eulogize, coo, jeer, rumble,* *simper, lisp, quip, croak, dictate,* *transcribe, ululate, animadvert*

[13] La possibilité de placer le participe passé avant le nom serait une indication de complexité : « the muttered words » vs « *the said words » (Erteschik-Shir, 1973, p. 86). Ce test n'est pas applicable au Français.

	Ronronner, gronder_rugir, exprimer une opinion (publier ?), faire l'éloge, roucouler_s'extasier, se moquer_huer, gronder_ ronfler, minauder, zozoter, railler, grommeler_coasser, dicter_ordonner, transcrire, hululer, critiquer_blamer

TABLEAU 5. Degré d'acceptabilité des verbes du champ sémantique de DIRE en verbe pont

Deux contraintes sont donc définies jusque-là : la dominance sémantique et la complexité sémantique du verbe pont.

Enfin, la négation a une influence sur la dominance, et donc l'extraction. Lorsque la matrice est niée, elle est sémantiquement dominante et la CV2 est sémantiquement subordonnée. Cela est prouvé par le test de réfutation, ici renversé en test de confirmation (109).

*(109) Bill said « It is unlikely that he enjoys eating that stuff, which is true – it is / ?he doesn't / *he does] (Erteschik-Shir, 1973, p. 95)*

Il est possible de réaffirmer la matrice (« *it is* »), mais pas la complétive (« **he does* »), sauf si on fait porter à cette dernière le trait de négativité (« *?he doesn't* »).

De même, dans (110) il est possible de nier la matrice mais pas la complétive.

(110) Otto: When we were talking about it before, I didn't believed that I had read that book
*Francine: That's a lie – you did / you had / *you hadn't (Erteschik-Shir, 1973, p. 97)*

Sans la négation, les deux parties de la phrase peuvent être interprétées comme dominantes. La confirmation est possible pour les deux verbes de la phrase affirmative (111), et la réfutation prend la modalité opposée à celle de chacun des verbes (112).

(111) Bill said « It is likely that he enjoys eating that stuff, which is true – it is / he does (Erteschik-Shir, 1973, p. 95)

(112) Otto: When we were talking about it before, I believed that I hadn't read that book
Francine: That's a lie – you did not/ you had so(Erteschik-Shir, 1973, p. 96)

Nous pouvons donc retenir que l'extraction est possible lorsque l'élément est extrait d'une subordonnée sémantiquement dominante. Il peut alors être réfuté, repris par une question-tag, modalisé par un adverbe d'opinion ou de certitude et appuyé par une interjection. Dans ce cas, la matrice (ou principale) est sémantiquement subordonnée et son verbe, le verbe pont, doit

être simple. Si le verbe pont est sémantiquement complexe ou bien s'il est nié, il tend à empêcher l'extraction.

L'argumentation présentée par (Erteschik-Shir, 1973) est complète et elle sera citée à de nombreuses reprises. Néanmoins, quelques problèmes se posent : certaines notions sont considérées comme floues mais intuitives (la complexité sémantique) et d'autres s'appuient sur des tests spécifiques à une langue : l'anglais (certains tests d'identification de la dominance). Ces conclusions demandent donc à être vérifiées pour le français, comme l'a remarqué (Godard, 1980).

1.3.2.2. Godard (80)

(Godard, 1980) reprend les tests de (Erteschik-Shir, 1973) pour les appliquer au français, sur les relatives. Concernant la dominance, elle observe les tests 1) et 3), c'est-à-dire la réfutation et l'interprétation de l'adverbe. Outre le problème d'ambiguïté dans le cas où la principale et la complétive peuvent toutes deux être dominantes, il semble que le test de l'adverbe ne débouche pas sur les résultats attendus : l'interprétation de l'adverbe comme description du point de vue du locuteur ne garantirait pas la possibilité de l'extraction en français. Rappelons l'ensemble de l'argumentation : lorsque l'adverbe décrit la façon de voir du sujet, la complétive peut être niée par un locuteur différent, qui n'a pas le même point de vue, comme l'illustre (113).

> *(113) "Astérix faisait l'hypothèse que les villageois aimeraient **certainement** aller combattre les Romains, mais moi, lecteur, je n'en suis pas si sûr." (Godard, 1980, p. 69)*[14]

En cela, le français est identique à l'anglais. Si l'on poursuit l'argumentation, la matrice et la CV2 enchâssée peuvent chacune être interprétée comme dominante ; l'extraction devrait donc être possible. Pourtant, d'après (Godard, 1980), « faire l'hypothèse » ne peut pas être verbe pont (114).

> *(114) *les Romains contre lesquels Astérix faisait l'hypothèse que les villageois aimeraient combattre se dérobaient aussi souvent qu'ils le pouvaient" (Godard, 1980, p. 69)*

[14] Le caractère gras n'est pas dans la citation originale.

En revanche, (Godard, 1980) reprend l'idée que les verbes qui autorisent l'extraction sont des verbes plus 'simples'. Pour définir la notion de complexité et de simplicité, elle reprend le champ sémantique de DIRE, et analyse une partie des verbes présents dans les tables de (Gross, 1975). Elle qualifie la complexité par la notion de « trait sémantique indépendant » et répartit les verbes en deux groupes (Godard, 1980, p. 72-74). Dans le premier groupe, se trouvent les verbes complexes, qui contiennent DIRE + un trait indépendant. Il peut s'agir :

- De la condition physique, soit la force ou le débit de la parole : « crier », « s'écrier », « s'esclaffer », « s'exclamer », « expectorer », « exploser », « tempêter », « tonitruer », « vociférer », « ânonner », « bafouiller », « balbutier », « bégayer », « chuchoter », « crépiter », « débiter », « déclamer », « éructer », « haleter », « murmurer », « susurrer », ainsi que les cris d'animaux employés métaphoriquement.
- De l'état d'esprit du locuteur (qui peut influencer son débit) : « bougonner », « grommeler », « marmonner », « maugréer », « ronchonner », « chanter », « fredonner », « pleurer », « pleurnicher », « larmoyer », « sangloter »
- Du moyen de transmission : « écrire », « téléphoner », « télégraphier »

Dans le second groupe (verbes simples), le trait ajouté n'est pas indépendant. Il ajoute une information qui porte également sur la parole, sur ce qui est dit :

- L'acte de parole : « avouer », « confesser », « jurer », « promettre »
- La volonté de faire savoir : « annoncer », « apprendre », « clamer », « colporter », « communiquer », « dévoiler », « diffuser », « divulguer », « ébruiter », « enseigner », « insinuer », « mentionner », « préciser », « proclamer », « propager », « publier », « révéler », « souligner », « spécifier », « suggérer », « taire »
- Un lien avec ce qui vient d'être dit (confirme, nie, ajoute au contenu) : « ajouter », « contredire », « corroborer », « confirmer », « démentir », « nier », « rappeler », « répéter »

Les deux groupes se comportent différemment : les verbes du premier groupe rendent l'extraction difficile, contrairement aux verbes du second. Si un verbe porte plusieurs traits qui ne vont pas dans la même direction, le degré d'acceptabilité dépend de l'aspect sur lequel l'accent est mis.

Selon (Godard, 1980), d'autres facteurs influencent la difficulté de former une extraction, parmi lesquels :

- Le fait que le verbe ne soit pas employé « naturellement » avec une complétive, et que l'on attende plutôt un GN («formuler », « conter », « ébruiter », « colporter », « aboyer », etc.) (Godard, 1980, p. 71-72,73)
- La présence d'un second complément, hormis les clitiques et le complément de « faire » factitif introduit par « à » (lui aussi attendu) : « avertir quelqu'un que », « jurer à quelqu'un que », « suggérer à quelqu'un que », etc. Les énoncés dans lesquels le verbe pont n'a pas d'autres compléments (« avertir que », « jurer que », « suggérer que », « faire jurer que ») sont jugés plus acceptables par les locuteurs que les mêmes verbes avec un complément.

(Godard, 1980) conteste en outre la contrainte syntaxique CNPC des générativistes, selon laquelle on ne peut pas extraire une partie d'un groupe nominal, soit en français N+complétive et N+relative. Elle cite des exemples dans lesquels un élément est extrait du GN[15] et qui semblent acceptables.

(115) Nommez donc cet endroit où vous affirmez posséder [la preuve que Gilles se trouvait _ ce soir-là] ! (Godard, 1980, p. 49)

(116) Cette pièce que la situation politique donnait [l'impression qu'on pouvait enfin monter _] n'a pourtant pas été autorisée. (Godard, 1980, p. 49)

Selon l'auteur, l'acceptabilité de ces exemples ne vient pas du fait que les noms tête n'ont plus les propriétés d'un GN (d'après les tests de passivation : **l'impression est donnée que...*, de clivage ou pseudo-clivage : *?ce qu'elle donnait, c'était l'impression que...*, et de pronominalisation : **elle la donnait*) car d'une part, d'autres locutions verbales n'autoriseraient pas les extractions (« faire l'hypothèse que ») et d'autre part, de vrais GN l'autoriseraient (« éprouver l'espoir », « donner l'espoir »)[16].

[15] Les crochets et les marques de la position initiale (_) ont été ajoutés aux exemples originaux.

[16] J'éprouve quelques réserves sur ces affirmations. Selon moi, *faire l'hypothèse* pourrait permettre l'extraction, contrairement à *éprouver l'espoir, donner l'espoir* :

« C'est avec elle que je fais l'hypothèse que tu aimerais danser. »

Par ailleurs, l'expression *éprouver l'espoir* est en soi peu usité (3 occurrences de *éprouve l'espoir que* dans *google* le 26mai12).

La possibilité pour un GN d'autoriser l'extraction serait plutôt liée à la valeur sémantique du GN :

> « Si ce SN décrit une réalité (présentée comme) objective, la parenthétisation est bloquée, au contraire, si ce SN vise à caractériser, à qualifier un autre SN présent dans la phrase, (c'est ordinairement le sujet) alors la parenthétisation est possible. » (Godard, 1980, p. 77)

Elle définit deux classes de noms. La première classe ne s'emploie pas avec le verbe « avoir ». La seconde classe, contrairement à la première, s'emploie avec le verbe « avoir » ; elle regroupe les noms qui décrivent les sentiments, les désirs, mais aussi des noms qui appartiennent à d'autres catégories sémantiques (TABLEAU 6):

CLASSE	VERBES
N qui décrit une réalité objective (*extraction)	l'annonce, le bruit, le fait, l'hypothèse, la nouvelle, la prédiction, la prévision, la supposition.
N qui caractérise le sujet (extraction)	l'audace, la bonté, la correction, la certitude, le courage, le désir, l'impression, l'intention, l'intelligence, l'espoir, la négligence, le souci, la volonté. l'autorisation, l'occasion, l'ordre, la preuve, le temps.

TABLEAU 6. Classe des noms qui facilitent ou non l'extraction

Pour résumer, (Godard, 1980) remet en cause une partie des tests d'identification de la dominance proposés par (Erteschik-Shir, 1973). Seul le test de réfutation semble convenir au français. Elle précise la notion de complexité sémantique d'un verbe de (Erteschik-Shir, 1973), en la définissant par l'ajout d'un trait sémantique indépendant, et confirme la difficulté d'extraire un complément d'un verbe pont complexe. Elle ajoute d'autres paramètres qui rendent l'extraction difficile. Nous retiendrons particulièrement deux contextes : le fait pour un verbe pont de s'employer « naturellement » avec une complétive, ce qui oriente vers des contraintes de performance, et la réalisation du verbe sans autre complément que la complétive (ou éventuellement des compléments réalisés en pronoms clitiques), ce qui oriente vers un autre type de contrainte syntaxique que celle des barrières limitées à une catégorie. Parallèlement, elle conteste la contrainte syntaxique du nom complexe définie par les générativistes. Selon elle, en français, l'extraction du complément du nom est possible si ce nom ne décrit pas une réalité objective. N'ayant pas retenu les dépendances à distance

nominales dans notre étude, nous ne pourrons pas confirmer ces dires. En revanche, nous pourrons observer si la même distinction se produit pour les verbes ponts.

Si (Godard, 1980) s'est attachée à reprendre l'ensemble des contraintes sémantico-pragmatique de (Erteschik-Shir, 1973) pour tester leur application au français, (Ambridge & Goldberg, 2008) se sont concentrés à affiner et à vérifier le point de vue pragmatique par des expériences.

1.3.2.3. Ambridge & Goldberg (08)

(Ambridge & Goldberg, 2008) s'inscrivent dans la lignée de (Erteschik-Shir, 1973). Elles relèvent à leur tour que les contraintes syntaxiques sont insuffisantes pour décrire la construction de dépendance à distance, sur la base des exemples suivants (117), qui ont la même construction syntaxique (interrogative directe avec extraction du sujet) et qui ne violent aucune des contraintes définies par les génératives (extraction d'un nom complexe, d'un sujet phrastique ou d'un ajout).

(117)	*a) Who did she think that he saw _?*	*Qui pense-t-elle qu'il a vu ?*
	b) ??Who did she mumble that he saw_?	*?? Qui marmonne-t-elle qu'il a vu ?*
	c) ??Who did she realize that he saw_?	*?? Qui réalise-t-elle qu'il a vu ?*

Mais elles ne se contentent pas de raisonner sur des exemples fabriqués : par des expériences, elles prouvent que ces énoncés n'ont pas le même degré d'acceptabilité pour les locuteurs. Alors que (117)(a) est tout à fait courant et accepté, (117)(b) et (117)(c), qui n'ont comme différence que le lexique du verbe pont, sont difficilement acceptables (Ambridge & Goldberg, 2008, p. 260).

Pour expliquer la différence d'acceptabilité entre (a), (b) et (c), deux hypothèses sont proposées : soit l'acceptabilité est liée à la fréquence d'emploi de la construction, liant fortement grammaire et lexique (Bybee, 2006; Sinclair, 1991), soit elle est déterminée par des propriétés discursives, mettant en jeu la notion d'arrière-plan.

(Ambridge & Goldberg, 2008) défendent l'hypothèse que la différence d'acceptabilité dépend de paramètres discursifs. La contrainte qui pèserait sur le phénomène des dépendances à distance serait que le gap, étant le focus de l'énoncé, doit se trouver dans une partie assertée.

L'extraction serait plus difficile si la CV2 est en « arrière plan ». Deux cas de figure sont possibles pour que le complément soit en arrière-plan :

- le complément est présupposé : que l'énoncé soit positif (118)(a) ou négatif (118)(b)[17], l'information donnée par le complément est garantie. Son verbe recteur est alors un verbe factif qui ne serait donc pas un candidat favorable en V1 (118)(c).

(118) *a) Elle a réalisé qu'il avait vu les roses* *-> il avait vu les roses*
b) Elle n'a pas réalisé qu'il avait vu les roses *-> il avait vu les roses*
c) ?les roses qu'elle (n') a (pas) réalisé qu'il avait vues

- plus généralement, le complément n'est pas asserté : on ne sait pas si l'information qu'il donne est vraie ou fausse, quelle que soit la modalité de l'énoncé (119)(a, b). Les verbes qui introduisent ce type de compléments sont les verbes de parole spécifiés. Ces verbes ne seraient pas non plus des candidats idéaux pour être V1 (119)(c).

(119) *a) Elle a crié qu'il avait vu les roses* *-> il avait vu les roses ?*
b) Elle n'a pas crié qu'il avait vu les roses *-> il avait vu les roses ?*
c) ?les roses qu'elle (n') a (pas) crié qu'il avait vues

Le statut discursif de « premier plan » et d'« arrière plan » est révélé par le test de négation. Ce test consiste à vérifier la portée de la négation sur l'ensemble de l'énoncé. Si la négation ne touche pas l'affirmation CV2, ce dernier est présupposé.

(120) *Elle ne regrette pas qu'il soit parti / il est parti est vrai => présupposition de CV2*
Elle ne pense pas qu'il soit parti / il est parti ? => pas de présupposition de CV2

(Ambridge & Goldberg, 2008) mettent en place des expériences basées sur le jugement d'acceptabilité afin de tester cette hypothèse. Trois types de verbes (des trois classes) sont pris en compte : des verbes qui régissent des compléments assertés (« *say* », « *decide* », « *think* », « *believe* »), des verbes qui régissent des compléments présupposés (« *realize* », « *remember* », « *notice* », « *know* ») et des verbes qui régissent des compléments non-assertés (« *whisper* », « *stammer* », « *mumble* », « *mutter* »). Deux tests sont convoqués :

- le test de négation. Il s'agit pour le locuteur interrogé de dire si l'énoncé nié implique la négation du complément, sur une échelle de 1 à 7.

[17] Exemple traduit de (Ambridge & Goldberg, 2008, p. 364)

- le score de différence. Dans un premier temps, le locuteur interrogé doit juger de l'acceptabilité d'une phrase affirmative avec le complément en place canonique, sur une échelle de 1 à 7. Dans un deuxième temps, le locuteur interrogé doit juger de l'acceptabilité de la phrase interrogative avec antéposition du complément, également sur une échelle de 1 à 7. Le score de différence calcule la différence entre le chiffre donné pour l'affirmative et le chiffre donné pour l'interrogative.

Les résultats montrent que les scores des deux tests sont corrélés : si l'un est élevé, l'autre est bas (TABLEAU 7). Les énoncés avec extraction les plus acceptables sont ceux dans lesquels *think, say, believe, decide* sont verbes ponts. Ensuite viennent les verbes de parole : « *whisper* », « *stammer* », « *mumble* », « *mutter* ». Enfin, les énoncés avec extraction les moins acceptables sont ceux dont le verbe pont est un verbe factif : « *realize* », « *remember* », « *notice* », « *know* ».

Type de verbe	**'dire, penser'**	**manière de parler**	**Verbes factifs**
Exemples choisis	*think, say, believe, decide*	*whisper, stammer, mumble, mutter*	*realize, remember, notice, know*
Score du Test de négation	Haut (3.35) donc assertion du complément	Moyen (2.75)	Bas (1.90) donc non-assertion du complément
Score de différence	Bas (0.96) interrogation aussi acceptable que l'affirmation	Moyen (1.74)	Haut (2.06) interrogation loin d'être aussi acceptable que l'affirmation
Verbe pont	+	+/-	-

TABLEAU 7. Résultat de l'expérience de (Ambridge & Goldberg, 2008)

Cette contrainte de présupposition expliquerait aussi pourquoi il est impossible d'extraire une partie d'un GN complexe (nom tête d'une relative, d'une complétive, etc.), d'un sujet phrastique ou d'un ajout.

D'après (Ambridge & Goldberg, 2008), la contrainte la plus forte concerne le verbe pont. La possibilité d'extraire un complément est liée à son degré d'assertion : plus le complément est asserté, plus l'extraction est acceptable. Ces résultats fournis par les expériences sur l'anglais sont confirmés pour l'allemand par les travaux de (Featherston, 2005).

1.3.2.4. Featherston (05)

(Featherston, 2005) présente une expérience qui s'appuie sur huit verbes allemands différents. Ces huit verbes forment des paires sémantiques et peuvent présenter d'autres particularités : « *sagen/behaupten* » (« dire/affirmer » - verbes de parole, mais qui présentent un contraste de fréquence et de longueur), « *erklären/erzählen* » (« expliquer/raconter » - verbes de parole spécifiques), « *hoffen/fürchten* » (« espérer/craindre » - contraste de sens : positif/négatif) et « *glauben/bezweifeln* » (« croire/douter » - contraste de sens : positif/négatif et de factivité : non factif/factif).

Il affirme que l'acceptabilité de ces verbes en extraction n'est pas seulement liée à leur fréquence d'utilisation (dans les corpus écrits et oraux) ou à leur longueur (en nombre de syllabes). L'une des hypothèses qu'il valide est qu'il y a une corrélation entre les verbes ponts et les verbes qui acceptent les complétives sans la conjonction « *dass* », ce qui implique que le verbe de la subordonnée revienne en deuxième position. Un verbe permettrait d'abord un complément sans conjonction dans les déclaratives (121), puis une extraction sans conjonction (122), puis une extraction avec conjonction et verbe en position finale (123).

(121) *Er sagt, der Lehrer* ***hat*** *den Schüler ausgeschimpft*[18] *(Featherston, 2005, p. 8)*

(122) *Wen sagt er,* ***hat*** *der Lehrer ausgeschimpft?*[19] *(Featherston, 2005, p. 8)*

(123) *Wen sagt er,* ***dass*** *der Lehrer ausgeschimpft* ***hat****?*[20] *(Featherston, 2005, p. 8)*

La structure sans conjonction n'est pas possible en français. Nous ne développerons donc pas la présentation des résultats en allemand concernant la comparaison les constructions avec et sans « *dass* ». Nous pouvons néanmoins retenir deux faits de son étude : d'une part, il existe des relations entre les verbes ponts et les verbes appartenant à une classe syntaxique particulière ; d'autre part, l'acceptabilité des huit verbes choisis en pont correspond à celle des verbes anglais.

Ses résultats sont les suivants : le verbe apparaissant dans les structures en extraction les plus acceptables est « *sagen* » (« dire ») ; ceux présents dans les moins acceptables sont les verbes de parole avec un sens spécifique « *erzählen* », « *erklären* » (« raconter », « expliquer ») et, en dernier lieu, le verbe factif « *bezweifeln* » (« douter »). Les verbes négatifs sont moins

[18] Il dit que l'enseignant a grondé l'élève [la page indiquée est celle de l'article en ligne : http://www.sfb441.uni-tuebingen.de/~sam/papers/bridge.pdf]
[19] Qui dit-il (que) l'enseignant a grondé ? [id]
[20] Qui dit-il que l'enseignant a grondé ? [id]

acceptables que les positifs en pont. Les résultats sont donc totalement convergents avec ceux de (Ambridge & Goldberg, 2008). Voici ses résultats (TABLEAU 8) :

Critères	Ordre des verbes
Fréquence	*sagen* > *glauben* > *behaupten* > *hoffen* > *fürchten* > *erzählen* > *erklären* > *bezweifeln*
Longueur	Monosyllabes : *sagen, glauben, hoffen, fürchten,* Bisyllabes : *behaupten, erzählen, erklären, bezweifeln*
Degré d'acceptabilité en verbe pont	*sagen* > *behaupten* > *glauben* > *hoffen* > *fürchten* > *erzählen* > *erklären* > *bezweifeln* dire > affirmer > croire > espérer > craindre > raconter > expliquer > douter

TABLEAU 8. Observation de huit verbes allemands selon différents critères

Nous pouvons remarquer qu'hormis l'échange de place entre « *glauben* » et « *behaupten* », l'acceptabilité des verbes ponts correspond à leur fréquence. Mais pour l'auteur, c'est la combinaison de la fréquence, de la longueur et des traits sémantiques (notamment la complexité sémantique) qui ont un rôle dans l'acceptabilité du verbe pont :

> « This data would therefore tend to support the suggestion that it is a combination of these factors which determines bridge and V2 quality. » (Featherston, 2005, p. 13)

Enfin, l'auteur fournit une conclusion intéressante, affirmée par peu de linguistes : les verbes formeraient un continuum et non deux groupes bien distincts : verbes ponts / verbes non ponts.

> « our data underlines that the bridge feature is a continuum and not a categorial distinction; there is no absolute group of bridge verbs, only better and worse ones. [...]The best we might manage would be to distinguish a group of structures which are good enough for them to be used, and another group, which are so bad that they in practice never occur. » (Featherston, 2005, p. 12-13)

Pour résumer, (Featherston, 2005), comme (Ambridge & Goldberg, 2008), reprend les grandes classes de (Erteschik-Shir, 1973) (verbes simples, verbes de parole spécifiques, verbes factifs) mais sans la notion de dominance. Les expériences menées sur l'allemand et l'anglais aboutissent au même résultat mais ne permettent pas de prévoir totalement le degré d'acceptabilité du verbe en verbe pont. Il leur faudrait pour cela tester tous les verbes à complétive, comme nous tenterons de le faire et pas seulement opposer des sous classes déterminées a priori.

1.3.2.5. Synthèse des contraintes sémantico-pragmatiques / bilan provisoire

L'extraction semble avoir un lien fort avec les éléments sémantiquement dominants, se trouvant dans le focus de l'énoncé. C'est pourquoi certains verbes dont le complément est asserté, sont de bons candidats au verbe pont (les verbes de paroles) et d'autres, dont le complément est présupposé sont de mauvais candidats (les verbes factifs). Les verbes qui attirent l'attention, par un fort développement syntaxique ou une complexité sémantique, sont également de moins bons candidats (les verbes de parole spécifiques). En revanche, les verbes fréquents et courts sont de bons candidats.

Le raisonnement global semble commun à l'ensemble des études basées sur des contraintes sémantico-pragmatiques : tous s'appuient sur la distinction entre la partie 'assertée' ou 'dominante' et la partie 'présupposée' ou 'subordonnée' de l'énoncé (notamment grâce à la négation de la principale et à la réfutation, qui montre la portée de l'assertion). Pourtant, les résultats divergent sur un point : les verbes factifs. (Erteschik-Shir, 1973) affirme que l'extraction n'est pas liée à la factivité du verbe pont. (Ambridge & Goldberg, 2008) et (Featherston, 2005) affirment le contraire en s'appuyant sur leurs expériences. Cette divergence pourrait venir des données analysées. (Ambridge & Goldberg, 2008) et (Featherston, 2005) ne traitent que des constructions interrogatives. (Erteschik-Shir, 1973) étudie à la fois les interrogatives et les relatives. Or le contre-exemple (Erteschik-Shir, 1973) est tiré d'une construction relative (124).

(124) We discussed the issue that I regretted that I had ever thought of (Erteschik-Shir, 1973, p. 92)

Il est possible que les contraintes diffèrent en fonction de la configuration syntaxique de l'énoncé. L'autre hypothèse est que « *regret* » fasse exception dans les verbes factifs. Il est pour le moment impossible de trancher entre ces deux hypothèses. En effet, l'acceptabilité de ce verbe en verbe pont n'est pas testée dans les expériences de (Ambridge & Goldberg, 2008) et de (Featherston, 2005).

Les contraintes sémantico-pragmatiques ne sont pas la seule alternative aux contraintes syntaxiques. (Hofmeister & Sag, 2010) se sont penchés sur une explication cognitive du phénomène.

1.3.3 Hofmeister & Sag : Contraintes cognitives sur la production des énoncés et le calcul du sens (10)

D'autres paramètres entrent en compte pour décrire la construction. Ceux-ci ont été repérés très tôt dans les études mais ils ont été considérés comme secondaires et ne remettant pas en question la théorie générativiste, comme l'indique (Hofmeister & Sag, 2010).

> « Beyond the graded nature of acceptability judgments observed in the primary literature on islands, a number of other sources have observed manipulations that affect the 'naturalness' of island constructions. Notably, these manipulations do not change or eliminate the syntactic configurations that form the basis of the island constraints in question. As we see below, however, these observations have not been interpreted as evidence against configurationally defined syntactic islands; instead, they have spawned a series of proposals that proliferate mechanisms within the theory of phrasal movement. » (Hofmeister & Sag, 2010, p. 374)

Selon (Hofmeister & Sag, 2010), les dépendances à distance réellement observées dans des productions ou des expériences ne répondent pas tant aux contraintes syntaxiques, auxquelles on peut déroger, qu'aux contraintes cognitives liées à la production et reconnaissance effective de ces séquences (contraintes de performance plutôt que de compétence). Toutes les contraintes syntaxiques peuvent être violées si d'autres facteurs compensent. Le jugement d'acceptabilité des dépendances à distance dépend du matériel (notamment lexical) qui intervient tout au long de la séquence, ce qui n'est pas du tout pris en considération dans le modèle générativiste.

L'élément à la base de la non-acceptabilité d'un énoncé est la difficulté de '*processing*' (production et reconnaissance en situation des énoncés). Cette difficulté a été notée dans de nombreux travaux, notamment ceux de (Hawkins, 1999[21]).

> « Filler-gap dependencies are difficult structures to process. [⋯] Identifying the gap is not easy. It is an empty element with no surface manifestation and its presence must be inferred from its immediate environment. At the same time, the filler must be held in working memory, and all other material on the path from filler to gap must be processed simultaneously, and the gap

[21] Hawkins, J. (1999) Processing complexity and filler-gap dependencies across grammars. *Language* 75, p.244–85.

must be correctly identified and filled. » (Hawkins 1999:246−47) dans (Hofmeister & Sag, 2010, p. 380)

Le principe supposé du fonctionnement de l'encodage et du décodage est que si une partie d'un énoncé nécessite des ressources cognitives importantes, la disponibilité de ces ressources est réduite pour les autres parties. Par exemple, plus l'attention est portée sur la recherche du référent mentionné, moins la représentation de la structure syntaxique est mémorisée. Le processus de rattachement de l'élément antéposé (*filler*) et de son verbe est donc rendu plus couteux par les conditions suivantes notamment :

- si les deux éléments sont éloignés. Les interrogatives sujets, par exemple, seraient jugées plus acceptables que les interrogatives objets, parce que l'élément antéposé reste moins longtemps en mémoire.

(125) *a. Which man __ saw the girl in the bar on California Avenue? ≥*

b. Which man did the girl in the bar on California Avenue see __ ? (Hofmeister & Sag, 2010, p. 381)

- si l'énoncé implique beaucoup d'éléments spécifiques ou référentiels. Les groupes nominaux définis et les noms propres, par exemple, demandent plus d'attention que les pronoms, parce qu'ils sont plus saillants et impliquent que la mémoire effectue une recherche plus précise.

(126) *The consultant who (Donald Trump/the chairman/a chairman/we) called advised wealthy companies (d'après (Hofmeister & Sag, 2010, p. 382)*

- si l'introducteur de subordonnée est un interrogatif (par exemple « où ») ou la conjonction « si », plutôt que la conjonction « que ». Les premiers ouvriraient un champ de réponse possible plus grand, et ce même en interrogative simple.

(127) *a. Has she forgotten that he dragged her to a movie on Christmas Eve? ≥*

b. Has she forgotten if he dragged her to a movie on Christmas Eve? ≥

c. Has she forgotten who he dragged _ to a movie on Christmas Eve? (Hofmeister & Sag, 2010, p. 383)

- si le filler est simple plutôt que complexe. Un pronom complexe syntaxiquement et sémantiquement serait plus facilement identifiable comme appartenant à la valence du V2. Dans les interrogatives, il réduirait en outre le champ des réponses possibles.

(128) *a. Which article don't you remember who wrote?* ≥

b. What don't you remember who wrote? (Maling and Zaenen, 1982[22]*) dans* (Hofmeister & Sag, 2010, p. 374)

- si les éléments sont peu fréquents dans la langue ou en collocation.

C'est la combinaison de plusieurs facteurs qui rendrait un énoncé de dépendance à distance agrammatical. Pour étayer cette analyse, (Hofmeister & Sag, 2010) ont recours à trois expériences. L'une des expériences montrent qu'un énoncé a priori agrammatical avec un interrogatif très spécifié (129)(b) est lu plus rapidement qu'un énoncé agrammatical avec un interrogatif générique (129)(c), au point de réduire la différence entre énoncé grammatical (129)(a) et agrammatical (129)(b).

(129) *a. Who did Albert learn **that** they dismissed after the annual performance review?*

*b. ***Which employee** did Albert learn **whether** they dismissed after the annual performance review?*

*c. ***Who** did Albert learn **whether** they dismissed after the annual performance review?(Hofmeister & Sag, 2010, p. 394)*

La conclusion proposée par (Hofmeister & Sag, 2010) est la suivante : lorsque les difficultés de 'process' augmentent, l'acceptabilité diminue.

> « These empirical investigations into the processing of island constraint violations converge on the conclusion that where processing difficulty increases, acceptability decreases » (Hofmeister & Sag, 2010, p. 401)

Ce serait la complexité du process qui influencerait le jugement de grammaticalité et non l'inverse.

[22] Maling, J. & Zaenen, A. (1982) A phrase structure account of Scandinavian extraction phenomena. *The nature of syntactic representation*, ed. by Pauline Jacobson and Geoff Pullum, 229–82. Dordrecht: Reidel.

> « One interpretation of these facts -the one that we adopt here- is that processing differences are influencing perceptions of acceptability ». (Hofmeister & Sag, 2010, p. 401)

Ainsi, d'après les expériences, il est probable que l'énoncé soit plus accessible si l'interrogatif est très spécifié et que les éléments qui le séparent de sa position canonique sont simplifiés. Cela implique par exemple que le V1 fasse partie des verbes fréquents, qui ont un sens générique, et que le sujet soit un référent déjà présent dans le discours. D'autres paramètres peuvent avoir une influence sur le degré d'acceptabilité : le temps du verbe, le type du verbe (verbe support, verbe faible), la classe du verbe (verbe de communication du type « expliquer » plutôt qu'un verbe décrivant la façon de parler du type « chuchoter »), l'introducteur de la complétive (« que » plutôt que « si »), le type de subordonnée (infinitive plutôt que complétive), le type de référent des noms (indéfini plutôt que défini), etc. Ces paramètres sont très nombreux et nous aurions souhaité qu'un classement soit établi, pour définir, au moins de manière tendancielle, quels sont les critères ou les combinaisons de critères qui ont le plus d'impact sur l'acceptabilité d'un énoncé en extraction. Nous pourrons tout de même observer nos données à la lumière de ces résultats.

1.3.4 De la représentation des possibles à la description des constructions réalisées

Toutes les études présentées jusqu'à présent (hormis les extensions) ont un point commun : elles s'appuient sur un modèle syntaxique qui, par construction, établit une longue distance structurale entre le verbe V2 et son dépendant, distance considérée comme potentiellement infinie. C'est ce long chemin structural que doivent parcourir les constituants qui se déplacent dans le modèle génératif ou les traits qui se propagent dans les grammaires d'unification. Les contraintes syntaxiques restent formulées en termes de contraintes sur des configurations complexes. Ces contraintes sont sans doute plus simples dans les modèles d'unification. Ainsi, dans la version d'HPSG, une contrainte est définie afin d'interdire au trait *slash* de se propager hors d'un groupe sujet ou ajout, ce qui serait agrammatical (130).

*(130) *Je me demande quels articles tu es parti [sans lire]. (Abeillé, 1993, p. 187)*

(Abeillé, 1993) formule la contrainte ainsi :

> « Chaque élément valeur du trait Hérit[Slash] d'un syntagme doit également appartenir au trait Hérit[Slash] de la branche Tête ou d'un de ses compléments « stricts » (ce qui exclut le sujet et les ajouts). » (Abeillé, 1993, p. 189)

HPSG utilise donc une part de syntaxe et une part de sémantique lexicale pour représenter un phénomène qui se situe à l'interface des deux domaines. Le *slash* ne marque pas un constituant (vide). Il est introduit par une règle lexicale et se propage en tant que trait sémantico-syntaxique par les têtes (critère syntaxique). En quelque sorte, ce formalisme conserve le mouvement des génératives et élimine la trace.

Pour (Culicover & Jackendoff, 2005), le principe reste néanmoins le même pour les deux courants : il s'agit de définir des contraintes sur le mouvement de la trace ou sur la formation de la chaine du *slash*.

> « Still, a treatment of constraints as prohibitions on the formation of chains of slash-category is not so much better than a treatment in terms of prohibitions on movement. » (Culicover & Jackendoff, 2005, p. 332)

Ces représentations par contraintes découlent d'une conception commune de la langue selon laquelle les possibilités seraient au départ maximales, à représenter par les règles les plus générales possible (*move a*) puis réduites dans certains contextes. Selon (Culicover & Jackendoff, 2005), il serait avantageux de partir du principe inverse et de formaliser en termes positifs les structures possibles, en s'appuyant sur le principe que les locuteurs généraliseraient uniquement ce à quoi ils sont confrontés, et non l'inconnu. Ils citent pour preuve des travaux sur les relatives[23] indiquant que les positions relativisables sont, dans l'ordre, sujet > objet direct > objet indirect > objet prépositionnel (obliques) > ajout et les structures dont proviennent les éléments relativisés sont principales > complétives à l'infinitif > complétives finies (enchâssées). Les généralisations des locuteurs se baseraient sur ces deux observations. Les extractions complexes particulières seraient donc complètement absentes du système.

La tendance à décrire « ce qui n'est pas possible » plutôt que « ce qui l'est » est ancrée dans la tradition linguistique. La majorité des études recherche les contextes et les verbes qui rendent

[23] Keenan, E. & Comrie, B. (1977) "Noun Phrase Acceptability and Universal Grammar", Linguistic Inquiry 8, 63-99.
Hawkins, J.A. (1994) A Performance Theory of Order and Constituency. Cambridge, Cambridge University Press

les énoncés agrammaticaux. Les études sur corpus remarquent que les emplois sont dans l'ensemble limités et qu'il est plus judicieux de décrire les contextes et les verbes ponts attestés. Mais se pose alors la question des données observées et du recours au corpus

Toutes les descriptions des dépendances à distance présentées ci-dessus se basent sur des exemples construits par les linguistes. Même dans les expériences, les exemples soumis aux locuteurs sont tirés de leur imagination et de leur représentation de la langue. Dans l'ensemble, ces exemples sont assez différents des données attestées, comme le montre (Dąbrowska, 2010). Ils présentent souvent un enchâssement récursif (131), deux interrogatifs (132) ou une complétive introduite par « *whether* » (133). Ces exemples semblent peu naturels.

(131) What was it obvious that Harry said you had forced Bill to buy? (Ross, 1967)

(132) How do you wonder what John could fix? (Sabel, 2002), dans (Hofmeister & Sag, 2010)

(133) Which employee did Albert learn whether they dismissed after the annual performance review? (Hofmeister & Sag, 2010)

Nous allons maintenant exposer les conclusions des études sur corpus. Les analyses des données attestées mettent en avant une conclusion tout à fait inédite : les dépendances à distance auraient un aspect formulaire. Les énoncés feraient intervenir des constructions totalement lexicalisées.

1.3.5 Contraintes lexicales

Les études sur corpus vont dans le sens de la première hypothèse de (Ambridge & Goldberg, 2008) : l'acceptabilité d'un verbe par rapport à un autre serait liée à sa fréquence d'emploi dans la construction. Les verbes employés de manière extrêmement fréquente dans cette construction avec verbe pont amèneraient le locuteur à lier fortement grammaire et lexique (Sinclair, 1991). Cela pousse même (Dąbrowska, 2008) et (Verhagen, 2006) à parler d'aspect formulaire de la construction de dépendance à distance.

1.3.5.1 Dąbrowska (08-09)

Les études de (Dąbrowska, 2008; Dąbrowska, Rowland, & Theakston, 2009), sont menées sur des corpus de langue anglaise : la partie orale du British National Corpus (BNC), un corpus adultes-enfants Manchester et un corpus enfants ABE. Leurs recherches portent sur les dépendances à distance dans les structures interrogatives, dans lesquelles le complément est extrait d'une complétive.

(134) What did they say it meant? (Dąbrowska, 2008)

Dans le corpus oral du BNC (10 millions de mots), (Dąbrowska, 2008) constate que 67% des occurrences de dépendance à distance ont la forme « WH do you think S » ou « WH did you say S », c'est-à-dire un mot interrogatif suivi de « *do you think/say* » suivi immédiatement de la subordonnée, sans « *that* ». Les autres occurrences montrent une variation minimale par rapport à cette forme prototypique. Un seul élément est modifié : soit l'auxiliaire (« *did* », « *may* » *(135)*), soit le sujet (« *he* », « *they* » *(136)*), soit le verbe (« *hope* », « *expect* » *(137)*), soit le *complementizer* (présence de « *that* » *(138)*), soit un autre élément (ajout d'un adverbe (139)).

(135) Wh- did you think ...

(136) Wh- do they think

(137) Wh- do you hope ...

(138) Wh- do you think that ...

(139) Wh- do you really think ...

Seuls 6% des occurrences présentent des constructions dans lesquelles plusieurs éléments diffèrents des deux formes les plus fréquentes.

La conclusion que l'auteur en tire est que la construction n'est pas libre ni exclusivement soumise à des règles syntaxiques. Il s'agit plutôt d'un idiome très spécifié lexicalement. Cette conclusion s'appuie notamment sur les résultats d'une expérience, qui montrent que si l'on modifie tous les éléments (sujet, V1, auxiliaire, présence de « *that* », récursivité de V1) la construction est jugée agrammaticale, comme si l'idiome n'était plus reconnu (Dąbrowska, 2008).

Les études sur un corpus de type adultes-enfants, le corpus Manchester (1 450 000 mots d'adultes), aboutissent à des résultats encore plus probants. Sur 326 énoncés de dépendances à

distance produits par des adultes, 85% des occurrences ont la forme « WH do you think S », 9% « WH did you say S » et 73% des occurrences restantes varient de façon minimale (variation du sujet ou de V1, présence de « *that* »).

Enfin, dans le corpus Abe (environ 350 000 mots), sur 44 énoncés de type dépendances à distance produits par des enfants de 3-4 ans, 42 occurrences sont du type « WH do you think S », 1 occurrence correspond à « WH do you think that S » et la dernière est du type « WH did he say S ».

Outre leur caractère formulaire, la construction de dépendance à distance se caractérise par une différence de fréquence selon la répartition des locuteurs en adultes-enfants. Il y a cinq fois plus d'occurrences dans le corpus Manchester que dans le corpus BNC-oral. Voici la fréquence correspondante dans les deux corpus pour un million de mots (TABLEAU 9).

Corpus	Nombre de dépendances à distance
Manchester	225
BNC_oral	42

TABLEAU 9. Proportion d'occurrences de dépendances à distance pour 1 million de mots dans un corpus adultes-enfants et dans un corpus adultes

Un enfant entendrait une construction de dépendance à distance environ une fois toutes les soixante-dix minutes, et aussi souvent que les termes « *bus* », « *old* » et « *able* ». La formule prototypique serait donc largement diffusée.

Au travers de recherches dans le Brown Corpus, (Verhagen, 2006) confirment l'hypothèse de la structure formulaire de la construction en anglais. Ses résultats sont également similaires pour le néerlandais. Pour lui la construction n'est pas seulement formulaire ; elle répond également à des paramètres pragmatiques.

1.3.5.2 Verhagen (05-06)

(Verhagen, 2005, 2006) étudie les constructions de dépendance à distance en anglais et en néerlandais. Ses travaux portent également sur les complétives et les interrogatives. Il montre que la structure a tendance à être très spécifiée lexicalement. Dans le corpus oral Eindhoven (800 000 mots), toutes les occurrences sont construites avec le V1 « *denken* » (« penser »). Le

pronom de la deuxième personne est majoritairement employé comme sujet du V1. Dans le corpus de presse du quotidien Volkskrant (année 1995), 80% des occurrences ont pour V1 « *denken* » et 84% le pronom de deuxième personne en sujet de V1. 65% des occurrences présentent une structure avec les deux paramètres : un pronom de deuxième personne et le verbe pont « *denken* ».

La forme prototypique est « WH- denk- P2 dat ». Elle est similaire à la forme prototypique anglaise. La fréquence de cette construction n'est pas due à la fréquence générale de chaque élément. Dans le corpus de presse par exemple, « *denken* » n'est pas le verbe le plus fréquent. D'autres verbes sont beaucoup plus fréquents : « *zeggen* » (« dire »), « *vinden* » (« trouver/penser »). De même, les groupes nominaux ou les pronoms de troisième personne sont beaucoup plus employés que le pronom de deuxième personne.

(Verhagen, 2006) explique la présence de la deuxième personne dans l'interrogative en ayant recours à la notion d'intersubjectivité. La construction pont ne constitue pas une description objective. Elle permet au locuteur d'interroger le point de vue de l'interlocuteur. Le principe de l'intersubjectivité est justement d'émettre un point de vue sur l'objet (la complétive), la notion de 'point de vue' étant opposée à celle de 'description. Cette valeur de point de vue a pour effet d'affaiblir l'assertion dans le cas des déclaratives. Le locuteur ne juge pas de la valeur de vérité du fait, mais fait part de sa croyance (première personne) ou interroge le point de vue de l'interlocuteur (deuxième personne), ce qui atténue la brutalité de la question. (Verhagen, 2006) s'oppose à (Benveniste, 1966) en posant que la troisième personne est également une manifestation de la subjectivité, tout en reconnaissant qu'elle est exprimée avec moins de force. En effet, la troisième personne porte une subjectivité affaiblie parce que la subjectivité ne se situe pas au niveau des interlocuteurs mais au niveau d'une tierce personne et qu'elle est restituée par le locuteur. C'est ce que représente la (FIGURE.16). Pour (Verhagen, 2006), la moins grande force de la subjectivité dans la troisième personne ne signifie pas qu'il y a une plus grande part d'objectivité.

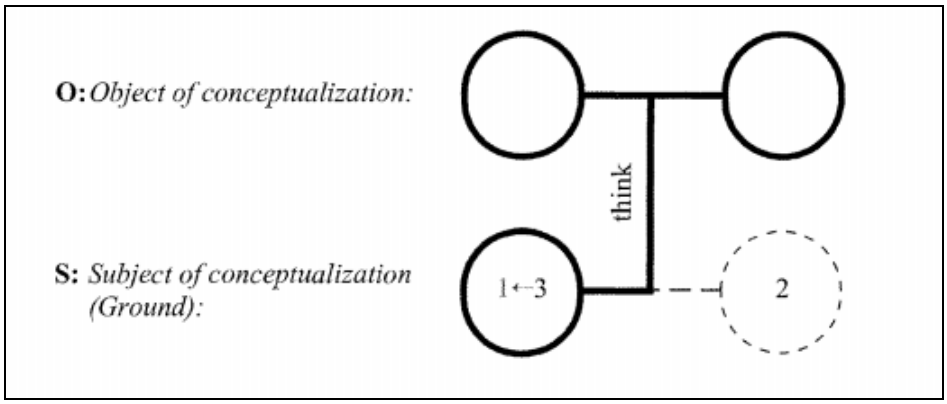

FIGURE.16 Représentation de l'intersubjectivité de la troisième personne (Verhagen, 2006, p. 330)

La présence de verbes ponts à la troisième personne ne remet donc pas en cause la théorie de l'intersubjectivité. Il y a d'ailleurs quelques occurrences de troisième personne dans les corpus :

*(140) What **did she think** he could do? (Brown Corpus) (Verhagen, 2006, p. 334)*

Pour résumer, les corpus mettent en évidence un emploi formulaire de la construction de dépendance à distance dans les structures interrogatives. Que ce soit en anglais ou en néerlandais, la formule se présente sous la forme « Wh- P2 penser/dire (que) subordonnée ». Cette formule correspond à un paramètre pragmatique : l'intersubjectivité. Nous observerons si la même formule existe en français et si les autres structures syntaxiques ont également leur propre formule. Cela signifierait que la construction est apprise avec le lexique et que les contraintes (syntaxiques, sémantiques ou cognitives) relèvent d'une réflexion métalinguistique à partir d'une structure que les modèles supposent à tort libre.

Nous présenterons dans la section suivante un récapitulatif de toutes les conclusions émises par les linguistes cités dans le paragraphe 1.3 afin de poser nos hypothèses pour le français.

1.4 Synthèse

Dans un premier temps, nous rappellerons les analyses qui ont été apportées dans les précédentes études. Nous verrons comment elles peuvent être adaptées au français. Ensuite, nous formulerons les questions et les hypothèses auxquelles nous tenterons de répondre.

1.4.1 Points retenus et adaptation au français

Dans la section précédente (1.3), nous avons noté que différentes perspectives de recherche ont été choisies afin de définir les propriétés des constructions de dépendance à distance. Après la description syntaxique des générativistes, les orientations lexicales, sémantiques, pragmatiques et cognitives ont apporté de nouveaux éléments de réponse.

Afin de synthétiser ces conclusions, tirées d'expériences, de descriptions de corpus ou tout simplement d'analyses introspectives, nous adopterons un point de vue tout à fait descriptif. Nous regrouperons les contraintes en fonction de l'objet sur lequel elles pèsent : la position canonique non réalisée de l'élément antéposé, le verbe pont, d'autres éléments ponctuels de la construction (interrogatif-relatif, groupes nominaux, conjonction, etc.) et enfin, l'ensemble de la construction. Pour une meilleure lisibilité, les résultats sont présentés dans le (TABLEAU 10) :

Domaine	**Règle exprimant la contrainte**
Place du gap	La place du gap est contrainte : on ne peut pas déplacer une partie d'un GN, une partie d'un sujet complexe, ni une partie d'un ajout. Le groupe duquel est extrait l'élément doit être dominant/dans le focus
Verbe pont	Sont difficilement acceptables en verbe pont : - Les verbes complexes, avec des traits indépendants du sens principal - Les verbes niés - Les verbes qui régissent un autre complément (non clitique) - Les verbes factifs - Les verbes de parole spécifiques Sont plus acceptables en verbes ponts : - Les verbes fréquents - Les verbes naturels avec une complétive - Les verbes courts - Les verbes simples / à sens générique - Les verbes de communication - Les verbes supports / faibles
Filler	- filler spécifié (à quel endroit > où) - référent du filler déjà dans le discours (pronom > GN)
Construction pont	- pont simple : seule expansion du verbe = complément clitique

conjonction	- type de conjonction (que > si)
GV2	- type de subordonnée (infinitive > complétive)
Formule	Dans les interrogatives, on trouve habituellement : « WH- do you think S » / « WH- denk- P2 dat » « WH- did you say S »

TABLEAU 10. Récapitulatif des contraintes pouvant avoir cours dans les constructions de dépendance à distance

Comme l'illustre le tableau, l'intérêt des études, qui portait à l'origine sur la structure de l'ensemble de la construction, s'est rapidement déplacé sur la forme du verbe pont. L'objectif était de déterminer si les verbes ponts formaient une classe particulière et si cette dernière pouvait recouper d'autres classes. L'observation la plus courante a concerné la classe des verbes factifs. Cette classe combine le lexique et la structure informationnelle de l'énoncé. Elle se place donc en continuité avec les analyses syntaxiques.

L'observation des données présentes dans les corpus a néanmoins ouvert une autre piste. Les verbes ponts attestés dans les interrogatives seraient des marques de l'intersubjectivité des participants. Ils appartiendraient donc à une classe de verbes à valeur modale dont la fonction est de donner une information sur le degré d'assertion de l'énoncé.

Ces deux conclusions nous amènent à rechercher quels peuvent être les verbes français qui appartiennent à la classe des verbes de présupposition et ceux qui appartiennent à la classe des verbes modaux. Pour définir ces deux groupes, nous nous appuierons respectivement sur les travaux de (Mørdrup, 1975) et de (Blanche-Benveniste & Willems, 2009).

1.4.1.1 Verbes français liés à la présupposition

(Mørdrup, 1975) a établi une typologie des verbes français liés à la présupposition et à l'implication. En s'appuyant sur l'analyse d'environ 500 verbes, elle distingue trois groupes, en fonction de la réaction du complément (B) face à la négation du verbe de la principale (A).

Si la valeur de vérité du complément est indépendante de la modalité du verbe de la principale, celui-ci est factif. Si la valeur de vérité du complément est indépendante de la modalité du verbe de la principale uniquement dans certains environnements, celui-ci est

semi-factif. Si la valeur de vérité du complément est systématiquement parallèle ou systématiquement opposée à la modalité du verbe de la principale, celui-ci est implicatif.

Les trois groupes de verbes sont donc les verbes factifs, les verbes semi-factifs et les verbes implicatifs. Ils sont ensuite distingués en nombreux sous-groupes (voir Annexe 1). Voyons plus en détail chacun des groupes.

Les verbes factifs régissent un complément qui n'est pas affecté par sa modalité. Si le complément est vrai quelle que soit la modalité du verbe, celui-ci est factif positif. Si le complément est faux quelle que soit la modalité du verbe, celui-ci est factif négatif.

- Factifs positifs : **que A soit positif ou négatif, B est vrai**

  (141) Pierre regrette que Sheila soit enceinte -> Sheila est enceinte

  (142) Pierre ne regrette pas que Sheila soit enceinte -> Sheila est enceinte (Mørdrup, 1975, p. 126)

- Factifs négatifs : **que A soit positif ou négatif, B est faux**

  (143) Pierre s'est imaginé que Sheila est partie -> ~~Sheila est partie~~

  (144) Pierre ne s'est pas imaginé que Sheila est partie -> ~~Sheila est partie~~ (Mørdrup, 1975, p. 137)

Les verbes semi-factifs sont tantôt factifs, tantôt non factifs. Ils perdent leur factivité selon l'environnement dans lequel ils apparaissent. On peut citer plusieurs cas de figure. Les verbes, par exemple, ne sont pas factifs lorsqu'ils sont conjugués à la première personne, dans une structure hypothétique portant sur l'avenir, le présent ou le passé. La complétive n'est pas présupposée être vraie (*Je n'ai pas dit la vérité ?*, *Marie a acheté une voiture ?*).

(145) Si j'apprends un jour que je n'ai pas dit la vérité, je l'avouerai à tout le monde (Mørdrup, 1975, p. 129)

(146) Si j'apprenais que Marie avait acheté une voiture, je serais désolé (Mørdrup, 1975, p. 130)

(147) Si j'avais appris que Marie avait acheté une voiture, j'aurais été désolé (Mørdrup, 1975, p. 130)

En revanche, dans l'exemple suivant, dans lequel la construction n'est pas hypothétique, la complétive est présupposée :

(148) J'ai appris que Marie avait acheté une voiture -> Marie a acheté une voiture

Toutefois d'autres paramètres peuvent facilement modifier cette règle. Il suffit qu'un adverbe insistant sur l'action ou le moment de l'énonciation soit présent pour que le verbe retrouve sa factivité. Dans les deux exemples suivants, la construction *Marie a acheté une voiture* est présupposée.

(149) Si j'avais appris que Marie avait acheté une voiture de cette manière-là, j'aurais été désolé

(150) Si j'avais appris hier que Marie avait acheté une voiture, j'aurais été désolé (Mørdrup, 1975, p. 130)

D'autres environnements font perdre sa factivité au verbe. D'après (Mørdrup, 1975), si le verbe est régi par un verbe factif négatif (comme « s'imaginer »), il perd sa factivité ; son complément n'est pas présupposé (151). Mais il regagne sa factivité en étant nié (152).

(151) Pierre s'est imaginé que j'apprendrais que je n'avais pas dit la vérité. -> je n'ai pas dit la vérité ? (Mørdrup, 1975, p. 131)

(152) Pierre s'est imaginé que je n'apprendrais pas que je n'avais pas dit la vérité. -> je n'ai pas dit la vérité (Mørdrup, 1975, p. 131)

L'exemple proposé ne nous semble cependant pas vraiment convaincant. Notre interprétation de (151) implique la présupposition de *je n'ai pas dit la vérité.*

Les verbes du troisième groupe, les verbes implicatifs, ont un complément dont la modalité dépend de celle du verbe. (Mørdrup, 1975) distingue les implicatifs positifs, les implicatifs négatifs et les implicatifs simples. Pour les implicatifs positifs (153)(154), la modalité du complément est identique à celle du verbe.

- Implicatifs positifs : **si A est positif, B est vrai ; si A est négatif, B est faux**

(153) Pierre a commencé à parler -> Pierre parle

(154) Pierre n'a pas commencé à parler -> ~~Pierre parle~~ (Mørdrup, 1975, p. 139)

Pour les implicatifs négatifs, la modalité du complément est contraire à celle du verbe (155)(156).

- Implicatifs négatifs : **si A est positif, B est faux ; si A est négatif, B est vrai**

(155) Pierre a cessé de parler -> ~~Pierre parle~~

(156) Pierre n'a pas cessé de parler -> Pierre parle (d'après (Mørdrup, 1975, p. 149))

Les implicatifs simples sont particuliers parce que l'implication ne se fait que pour l'une des modalités ((157) à (162)). Un seul postulat est vérifié, son contraire ne l'est pas. Il y a donc quatre possibilités pour un verbe d'être étiqueté implicatif simple.

- Implicatifs simples :

si A est positif, B est vrai [mais si A est négatif, B n'est pas faux]:

(157) Pierre a prouvé que la terre tourne autour du soleil -> la terre tourne autour du soleil.

(158) [« Pierre n'a pas prouvé que la terre tourne autour du soleil » n'implique pas « ~~la terre tourne autour du soleil~~ »] (Mørdrup, 1975, p. 151)

si A est négatif, B est faux [mais si A est positif, B n'est pas vrai] : aucun verbe français n'entre dans cette classe.

si A est positif, B est faux [mais si A est négatif, B n'est pas vrai]:

(159) Pierre n'a pas pu partir -> ~~Pierre est parti~~.

(160) [« Pierre a pu partir » n'implique pas « Pierre est parti »] (Mørdrup, 1975, p. 151)

si A est négatif, B est vrai [mais si A est positif, B n'est pas faux] :

(161) Pierre n'a pas hésité à sortir -> Pierre est sorti

(162) [« Pierre a hésité à sortir » n'implique pas « ~~Pierre est sorti~~ »] (Mørdrup, 1975, p. 152)

Les recherches précédentes sur les constructions de dépendance à distance ont conclu qu'il était difficile pour le verbe pont d'être un verbe factif, sans chercher à faire de différences entre les classes de verbes qui mettent en jeu la factivité. La distinction fine établie par (Mørdrup, 1975) entre tous ces verbes liés à la présupposition et à l'implication mérite d'être utilisée dans le cadre d'une étude sur les verbes ponts. Quelles sont les classes les moins acceptées dans cette construction ? Y a-t-il de grandes différences d'appréciation entre les classes et les sous-classes définies ? Nous tenterons de répondre à ces questions dans notre analyse des constructions de dépendance à distance en français.

La classe des verbes de présupposition étant bien établie, nous avons cherché une classe de verbes modaux qui pourraient s'inscrire dans le cadre de l'intersubjectivité.

1.4.1.2 Verbes français de mitigation

En français, il existe une classe de verbes qui ont un emploi dans lequel leur fonction est de diminuer la force d'assertion de l'énoncé. On parle de 'verbes de mitigation', de 'verbes recteurs faibles' ou 'verbes faibles' (Blanche-Benveniste & Willems, 2009).

> « Leur rôle fondamental est de permettre au locuteur de prendre ses distances par rapport à une assertion simple, souvent ressentie comme trop forte. » (Blanche-Benveniste & Willems, 2009)

Parmi eux se trouvent par exemple « je pense », « je crois », « je trouve », « il me semble », « j'ai l'impression », « on dirait » (Blanche-Benveniste, 1989; Blanche-Benveniste & Willems, 2007, 2009). Ces verbes ont été regroupés d'après un certain nombre de critères.

En emploi de recteur faible, ces verbes possèdent trois types de réalisation : une réalisation avec un complément en 'que-phrase' (163), une réalisation en parenthétique, en milieu (incise) ou en fin d'énoncé (164) (qui ne suit pas un patron prosodique spécifique d'après (Gachet & Avanzi, 2009)) et une réalisation autonome, en réponse ou dans le cours du tour de parole (165).

(163) On dirait qu'il va pleuvoir (Blanche-Benveniste & Willems, 2007, p. 225)

(164) Il va pleuvoir, on dirait (Blanche-Benveniste & Willems, 2007, p. 223)

(165) Je sais pas s'il va pleuvoir mais on dirait

Dans les trois réalisations, le lien entre le verbe et la proforme est faible[24]. Dans la première, il n'y a pas d'équivalence entre la complétive et un pronom : **on le dirait, *on dirait ça*. Dans les deux suivantes, aucune proforme complément n'est produite. La seule proforme éventuellement acceptable est celle en « qu'est-ce que » (166) ; le dispositif en « ce que » (pseudo-clivage) est également usuel (167).

(166) ?Qu'est-ce qu'on dirait ?

(167) Ce qu'on dirait, c'est qu'il va pleuvoir

Ce groupe de verbes se distingue des verbes comme « jurer » qui, dans un environnement similaire (les trois réalisations), ont un lien fort avec la proforme (168)(169)(170).

[24] Pour (Gachet, 2012), la mitigation se situe au niveau pragmatique, et non syntaxique.

(168) Je te jure que je viendrai = je te le jure (d'après (Blanche-Benveniste & Willems, 2007, p. 223))

(169) Je viendrai, je te le jure

(170) Tu viendras ? Je te le jure.

Il se distingue par ailleurs des verbes qui n'ont qu'une partie de ces réalisations, voire aucune d'elles, tels que les groupes suivants :

> « Les verbes de « dire » ne présentent pas la disjonction (ils ont *que* + V et l'incise). Les verbes comme *regretter* ont les trois constructions mais avec la pro-forme *le*. Les verbes modaux comme *je dois*, *je peux* ont les constructions en incise et les disjonctions mais pas la construction en *que* + V. » (Blanche-Benveniste & Willems, 2007, p. 226)

Du point de vue du statut discursif, on considère que dans les énoncés avec verbes faibles, l'information principale est donnée dans la complétive. Le verbe faible n'apporte pas une information descriptive mais une information qui s'ajoute à celle fournie par un autre verbe en modulant l'attitude du locuteur à l'égard de son propos. L'analyse de l'oral spontané, qui laisse voir pas à pas la construction de l'énoncé, met en évidence l'ajout du verbe à complétive : dans l'exemple (171), le locuteur amorce son énoncé puis l'interrompt ; il le réorganise ensuite et l'enrichit en ajoutant le verbe faible.

(171) il faut (//) je pense qu'il faut ou avoir fait des études ou bien s'intéresser à la littérature (CorpAix, BEC)

Enfin, l'interprétation sémantique du verbe faible est comparable à celle des adverbes épistémiques « selon moi », « à mon avis » (Apothéloz, 2003). Dans les corpus, les verbes faibles sont souvent réalisés à proximité de termes exprimant le non-savoir (*je sais pas*) ou l'engagement personnel (« moi », « personnellement »). Ils reflètent une précaution prise par le locuteur pour limiter la validation de son assertion. Une assertion simple, en revanche, véhicule à la fois une affirmation et une validation.

> « Une analyse plus détaillée des lexèmes verbaux […] nous a permis de découvrir (1) qu'il y a une relation syntaxique et sémantique stable à travers les 3 emplois faibles (+P, incise, disjoint) des verbes étudiés; (2) que cette valeur sémantique stable porte sur la validation du contenu de l'énoncé P : ce contenu est présenté comme ayant un degré moindre de validation; [...]. » (Blanche-Benveniste & Willems, 2007, p. 248)

La relation syntaxiquement affaiblie entre le verbe et ses compléments, la réorganisation informationnelle de l'énoncé et le sémantisme des verbes caractérisent particulièrement les verbes faibles. Ils correspondent sur de nombreux points aux verbes parenthétiques décrits par (Urmson, 1952).

> « A verb which, in the first person present, can be used, as in the example above, followed by 'that' and an indicative clause, or else can be inserted at the middle or end of the indicative sentence, is a parenthetical verb. » (Urmson, 1952, p. 481)
> « The claim to truth need not be very strong, we shall indeed find that the whole point of some parenthetical verbs is to modify or to weaken the claim to truth which would be implied by a simple assertion. » (Urmson, 1952, p. 484)

La conclusion de (Verhagen, 2006) pour le néerlandais est que les verbes ponts ont pour fonction d'atténuer l'assertion, en donnant le point de vue d'un locuteur. Si cette conclusion se révèle être identique en français, les verbes faibles pourraient être de bons candidats pour remplir la fonction de verbe pont dans les constructions de dépendance à distance. L'une des difficultés à étayer cette hypothèse est liée au fait que tous les verbes recteurs faibles ont par ailleurs un emploi de verbes pleins. La classe n'est donc pas facile à définir.

1.4.2 Problématique et plan

Les différentes perspectives des travaux antérieurs soulèvent de nombreuses questions, d'autant que ceux-ci ont presque toujours porté sur d'autres langues que le français. Voici d'abord une série de questions très ponctuelles, qui permettront de répondre à la question plus générale suivante : Comment les locuteurs font-ils usage de la construction de dépendance à distance ?

- Y a-t-il une formule, comme en anglais ('do you think') et en néerlandais ('denk- P2 dat') ? (Dąbrowska, 2008) Ou la construction fait-elle simplement intervenir des verbes spécifiques, comme les verbes recteurs faibles, afin d'atténuer l'assertion ? (Verhagen, 2006)
- Les verbes factifs, semi-factifs et implicatifs sont-ils exclus des données, ainsi que les verbes de parole spécifiques ? (Ambridge & Goldberg, 2008; Featherston, 2005; Godard, 1980)

- Les verbes ponts sont-ils des verbes fréquents, courts, de sens générique ? L'ensemble de la séquence située entre l'interrogatif-relatif et la place canonique est-il réalisé de manière à faciliter le processing ? (Hofmeister & Sag, 2010)
- Y a-t-il beaucoup d'occurrences qui transgressent les contraintes syntaxiques de (Ross, 1967) ? (comme le sous-entendent de nombreux auteurs désireux d'imposer une autre théorie)
- Quel est l'impact du type d'interlocuteur sur ce comportement et sur la fréquence de la construction : adultes-enfants, adultes-adultes ? (Dąbrowska et al., 2009)
- Y a-t-il de façon similaire une importance du genre communicatif : presse, littérature, oral spontané, oral préparé, écrits scientifiques, etc. ?

Nous répondrons à ces questions en nous appuyant sur des données issues de corpus oraux et écrits. Dans un premier temps, l'analyse sera corpus-based et nous vérifierons simplement les hypothèses soumises ci-dessus (contraintes syntaxiques, aspect formulaire, fréquence des verbes sélectionnés en verbes ponts, traits de factivité des verbes ponts ou appartenance à la classe des verbes faibles). Dans un second temps, nous opterons pour une méthode corpus-driven afin de décrire de façon la plus précise possible l'ensemble de la construction pont. Nous pourrons alors croiser nos résultats et ceux de (Hofmeister & Sag, 2010). L'analyse se veut la plus descriptive possible et éloignée de tout cadre théorique. Pour cette raison, le vocabulaire neutre et imagé de 'verbe pont' et de 'construction pont', repris de (Erteschik-Shir, 1973), sera utilisé. Le détail des méthodes d'analyse et des résultats sera présenté dans le Chapitre 3.

Après cette partie proprement descriptive, nous reviendrons sur les questions théoriques (Chapitre 4). La description des énoncés de dépendance à distance permettra de choisir l'interprétation théorique qui s'adapte le mieux : Y a-t-il un enchâssement de V2 dans V1 qui éloigne le complément antéposé (par mouvement ou un autre processus) ou bien les deux verbes forment-ils un constructeur complexe duquel l'élément antéposé dépend (dans ce cas sans distance) ? Pour nous donner une base empirique plus large, nous présenterons une étude complémentaire sur corpus des cas où V2 est à l'infinitif.

Avant de commencer la description sur corpus, nous souhaitons insister sur la nécessité à laquelle nous avons été confronté de mener un long travail préparatoire. A la différence du

domaine Anglo-saxon que nous avons cité, le chercheur français ne dispose pas d'un corpus de référence de taille suffisante sur lequel il puisse immédiatement procéder à des requêtes. La constitution et l'exploitation des ressources nécessaires à une étude sur corpus se heurte en outre pour le Français à de nombreuses difficultés de mise en oeuvre : difficulté d'accès aux corpus existants, manque d'enrichissements des corpus (peu d'étiquetage morphosyntaxique et syntaxique), immense nécessité d'uniformiser les transcriptions et les formats, difficulté à trouver des outils permettant notamment d'effectuer des requêtes complexes. Nous détaillerons notre expérience dans le Chapitre 2. L'établissement de notre corpus d'étude nous a en fait occupé pendant un an lors des trois ans consacrés à la thèse.

CHAPITRE 2. CORPUS ET OUTILS

L'étude des dépendances à distance que nous allons développer s'appuie sur des données attestées. Comme nous l'avons noté dans le chapitre précédent, les études théoriques ne s'attachent pas à décrire les usages de la langue tels qu'ils sont produits par les locuteurs. Elles tentent de définir l'ensemble des possibilités, qui sont bien plus vastes que les emplois réels. Pour décrire la langue parlée et écrite, les corpus sont indispensables. Il est dangereux de s'appuyer sur sa propre intuition, d'autant que l'intuition des linguistes peut aussi s'avérer être très différente de celle d'un échantillon de locuteurs.

> « Because we are looking for typical pattern, analyses cannot rely on intuitions or anecdotal evidence. In many cases, humans tend to notice unusual occurrences more than typical occurrences, and therefore conclusions based on intuition can be unreliable. » (Biber, Conrad, & Reppen, 2002, p. 3)

Par ailleurs, de nombreux paramètres entrent en jeu pour rendre un énoncé acceptable, ce qui rend la création d'exemples compliquée :

> « Moreover, expressions that linguists have sometimes categorised as "ungrammatical" have been found to be accepted by people [...] or found to be used by speakers and to sound good compared to the examples contrived by linguists [...]. Many researchers today thus see the need for placing less emphasis on linguists' judgments of grammaticality and more emphasis on usage and experimental data [...]. » (Ford & Bresnan, 2010, p. 1)

2.1 Les données

Une étude préalable sur un corpus limité (en Master 1[25]) avait révélé que le phénomène des constructions de dépendance à distance est peu fréquent. Pour obtenir un nombre suffisamment représentatif d'occurrences, il était donc nécessaire de travailler sur un corpus

[25] Ce travail a donné lieu à une présentation lors de la journée Rhapsodie du 1er juillet 2008 à Aix en Provence.

d'étude de taille suffisante. Les standards internationaux fixent par exemple la taille minimale d'un corpus d'étude oral à environ trois millions de mots (Biber et al., 2002). Dans certains domaines linguistiques (anglais, espagnol, néerlandais…), le linguiste dispose de corpus de cette dimension directement utilisables pour des requêtes en raison du fait qu'ils ont été constitués par une seule équipe et transcrits selon des conventions uniques. Le français ne dispose pas à ce jour d'un tel corpus de référence, particulièrement pour la partie orale. Comme il était exclu de le construire en vue de notre travail, la seule solution possible pour obtenir une masse de données de taille raisonnable était de rassembler plusieurs corpus existants. Le partage des données, qui étaient une richesse jalousement gardée, commence à se mettre en place : des corpus sont maintenant accessibles en ligne (CFPP, PFC), sous leur forme la plus complète (son, transcription, métadonnées) et non plus simplement grâce à une interface d'interrogation. Néanmoins, de nombreux corpus sont encore stockés dans des laboratoires, éparpillés dans toute la France et il n'est pas facile d'y avoir accès. Actuellement, des projets de banque de données se mettent en place pour dépasser cette difficulté[26].

Rassembler des corpus existants est a priori une bonne solution aussi bien quantitativement que qualitativement : ils permettent en particulier de constituer un ensemble de données orales voisin de trois millions de mots et leur contenu est diversifié : on y trouve du français écrit aussi bien que du français parlé, avec dans chaque groupe un continuum allant du spontané au préparé et de l'informel au formel.

Les corpus sélectionnés comme corpus de travail sont :

- Le Corpus Evolutif de Référence du Français (CERF/CorpAix)
- Le Corpus de Référence du Français Parlé (CRFP)
- Nancy
- C-Oral-Rom
- Choix de Textes du Français Parlé (CTFP)
- Le Corpus de Français Parlé Parisien (CFPP)
- Phonologie du Français Contemporain (PFC)
- C-PROM

[26] On peut notamment citer les anciens CRDO : COllections de COrpus Oraux Numériques (CoCoON) (http://crdo.risc.cnrs.fr/exist/crdo/) et Corpus (http://www.corpus-ir.fr/) ainsi que le projet ANR « Outils et Recherche pour le Français Ecrit et Oral » (ORFEO).

- IRIT itinéraire
- Office de Tourisme de Grenoble (OTG),
- Brassens,
- Université Bretagne Sud (UBS)

Mais, comme l'indique (Luzzati, 2009), un rassemblement de corpus différents ne devient un corpus d'étude qu'au prix d'un considérable travail de préparation :

> « La linguistique, pour accéder aux données, pour adhérer au réel, devient une linguistique de corpus, quelque chose comme une linguistique "du code barre", dont les principaux mots clés deviennent "harmonisation", "normalisation", "codage"… » (Luzzati, 2009, p. 109)

Nous présentons donc ci-dessous les corpus de départ et le travail d'unification des conventions de transcription et d'unification des formats que nous avons dû réaliser pour obtenir un corpus de travail de langue parlée. Ce corpus d'étude compte au final près de trois millions de mots, ce qui était notre objectif.

2.1.1 Présentation des corpus originaux

2.1.1.1 CERF/CorpAix

Le CERF (Corpus Evolutif de Référence du Français) est un corpus réalisé au début des années 2000 par l'équipe Description Linguistique Informatisée sur Corpus (DELIC) d'Aix en Provence (terminé en 2004). Il est composé de 10 tranches. Parmi elles, neuf attestent de variétés d'écrit : « Littérature/Fiction ancienne, Littérature/Fiction contemporaine, Institutions, Divers, Forum internet, Politique, Loisirs et vie pratique, Presse d'informations, Sciences et techniques ». Chaque tranche d'écrit est divisée en 10 sous-tranches, qui précisent la source journalistique, le genre de littérature (théâtre, conte, poésie, roman), etc. Chaque sous-tranche donne lieu à un fichier de format texte. Il n'y a pas vraiment de possibilité de repérer la source à l'intérieur de ce fichier.

La dernière tranche représente le français parlé. Cette tranche nommée Oral est une partie de CorpAix, corpus oral constitué par le Groupe Aixois de Recherche en Syntaxe (GARS) dans les années 70. A l'origine, CorpAix était disponible en format papier. Une partie a été

transcrite entre 1990 et 2000 à l'Université d'Aix en Provence. Elle comporte 265 transcriptions, disponibles en format texte. Il n'y a pas d'alignement avec les enregistrements qui par ailleurs ne sont plus accessibles, de même que les métadonnées. La majorité du corpus est constituée de récits de vie de personnes habitant la Provence. D'autres types de situations apparaissent également : conversations, cours, prises de parole professionnelles en public, émissions de radio et de télévision, etc. Le CERF compte environ 10 millions de mots, répartis équitablement entre les 10 tranches.

2.1.1.2 CRFP

Le CRFP (Corpus de Référence du Français Parlé) (DELIC, 2004) a été constitué au début des années 2000 (terminé en 2003) par l'équipe DELIC, à Aix en Provence, à la demande de la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF). Les 134 enregistrements, effectués dans 38 villes, sont le reflet du français parlé dans toute la France métropolitaine et la Corse. Les 357 locuteurs sont âgés de 18 à 60 ans, ont reçu un enseignement scolaire plus ou moins long et exercent une soixantaine de métiers différents. Dans un enregistrement, entre 1 et 16 locuteurs peuvent intervenir, mais le plus souvent, seules deux ou trois personnes sont réunies. Les discussions sont privées, publiques ou professionnelles. Les enregistrements provoqués se présentent sous la forme d'entretiens ou de conversations ; les autres sont des conférences, des réunions, des visites guidées, des journaux d'informations ou des émissions de radio/tv. Les transcriptions sont accessibles en format .txt et les enregistrements en .wma. Le CRFP compte environ 430 000 mots pour plus de 36 heures de parole.

2.1.1.3 Nancy-TCOF

Nancy est un corpus constitué entre 2000 et 2005 sous la direction de Jeanne-Marie Debaisieux, grâce à la collaboration des étudiants de l'université Nancy 2. Il comporte 73 fichiers au format texte, d'environ 360 000 mots. Les contenus sont divers : conversation spontanée, entretien, émission de radio, réunions professionnelles, etc. Ce corpus est en partie à l'origine du corpus TCOF, pour lequel une partie des trancriptions a été revérifiée et anonymisée. A partir de 2005, tous les corpus constitués ont été versés dans le corpus TCOF.

Le corpus TCOF[27] (Canut, André, & Gaiffe, 2010), réalisé à l'ATILF (Analyse et Traitement Informatisé de la Langue Française), est divisé en deux grandes catégories : les corpus d'interactions d'adultes et les corpus d'interactions adultes-enfants. Il réunit des enregistrements constitués dans les années 80-90 ainsi que des enregistrements récents, effectués entre autres par les étudiants de Sciences du langage de l'université de Nancy. Le corpus adultes-enfants (20h d'enregistrement, environ 175 000 mots) est constitué de conversations et de narrations dialoguées avec le support d'un livre illustré lu préalablement par l'adulte à l'enfant. Le corpus TCOF est diffusé librement sur le site du CNRTL[28]. Les transcriptions et les métadonnées sont disponibles au format .xml et les fichiers audio sont au format .wav.

Nous utiliserons le corpus Nancy et les corpus adultes-enfants de TCOF.

2.1.1.4 C-ORAL-ROM

Ayant pour but de permettre des comparaisons entre les langues romanes, C-Oral-Rom (Deulofeu & Blanche-Benveniste, 2006) est constitué de 4 sous-corpus, représentatifs des langues espagnole, française, italienne et portugaise. Le projet, proposé par la Commission européenne, a débuté en 2001 et s'est terminé en 2004. Il était dirigé par Emanuela Cresti. Pour chaque langue, les transcriptions sont réparties selon la structure suivante : Catégorie, Contexte et Domaine. Le TABLEAU 11 présente les champs présents dans ces trois partitions.

Catégorie	Contexte	Domaine
INFORMEL	Privé - familial	Monologue / dialogue / conversation
	Public	Monologue / dialogue / conversation
FORMEL	Contexte naturel	Affaire / conférence / loi / débat politique / discours politique / prière / explication professionnelle / enseignement
	Média	Interview / météo / informations / reportages / reportage scientifique / sport / débat télévisé
	Téléphone	Conversation privée / dialogue homme-machine

TABLEAU 11. Structure du corpus C-Oral-Rom, d'après (Moneglia, s. d., p. 12)

[27] Voir le site http://www.cnrtl.fr/corpus/tcof/
[28] Voir le site du CNRTL : http://www.cnrtl.fr/

La répartition du corpus est en principe équilibrée pour chaque langue en termes de catégories (50% Formel, 50% Informel), mais les proportions sont inégales en termes de contextes et de domaines. La provenance des locuteurs, leur sexe et leur âge, ainsi que la longueur des enregistrements sont également variés mais disproportionnés. L'ensemble du corpus contient 1 200 000 mots, soit environ 300 000 mots par langue.

La partie française, dont était responsable l'équipe du GARS, comporte 206 transcriptions, enregistrées entre 1980 et 2002. Nous avons eu accès à 175 d'entre elles.

Le corpus a été étiqueté automatiquement en morphosyntaxe, par Cordial Analyzer (Laurent, 2008). Les ressources ont été modifiées et des modules de pré- et post-traitement ont été ajoutés. L'évaluation, effectuée sur un échantillon, indique un taux de rappel de 99,9% et un taux de précision de 98,2% (pour un intervalle de confiance de 95%).

Le corpus est délivré par ELDA (European Language Distribution Agency) sous forme de 9 DVDs pour un prix variant entre 1500€ et 20000€. Les enregistrements sont disponibles en format .wav, les transcriptions en CHAT, .txt et .xml. Le fichier d'alignement texte-son est en .xml.

2.1.1.5 CTFP

L'équipe du GARS a réalisé un ouvrage, publié en 2002 chez Champion, qui s'intitule *Choix de Textes du Français Parlé : trente-six extraits* (Blanche-Benveniste, Rouget, & Sabio, 2002). Il réunit 36 transcriptions, enregistrées entre 1989 et 2000. Les locuteurs, âgés de 5 à 80 ans, viennent de plusieurs régions de France et ont différents niveaux d'éducation. Le corpus est divisé en quatre rubriques : Métiers-Activités, Récits, Témoignages, Enfants. Dans la dernière rubrique, on trouve des récits et des parodies produits par des enfants.

L'ouvrage est vendu avec CD qui contient les transcriptions informatisées en format .txt et les fichiers audio en format .mp3. CTFP contient environ 67 000 mots.

2.1.1.6 CFPP

Le CFPP (Corpus de Français Parlé Parisien) (Branca-Rosoff, Fleury, Lefeuvre, & Pires, 2009a, 2009b) a été élaboré par l'équipe de l'université Paris III pour rendre compte du français parlé à Paris et dans sa banlieue. Il contient 24 enregistrements, qui sont récents (2007-2010) et qui sont généralement faits au domicile des enquêtés. Les locuteurs, âgés de 15 à 88 ans, ont été sélectionnés parce qu'ils ont vécu toute leur vie ou presque dans une zone de Paris. Tous les entretiens suivent plus ou moins un questionnaire portant sur la vie à Paris et plus spécifiquement sur la vie dans le quartier dans lequel habitent les locuteurs.

Les fichiers audio sont disponibles au format .wav et .mp3 ; les transcriptions sont en .trs (c'est-à-dire xml). Les enregistrements sont relativement longs. Ainsi, le corpus comporte environ 340 000 mots.

2.1.1.7 PFC

Le projet PFC (Phonologie du Français Contemporain) (Durand, Laks, & Lyche, 2003, 2009; Durand & Lyche, 2003; Durand & Tarrier, 2006; Lyche, 2005, 2007), codirigé par Marie-Hélène Côté (Université d'Ottawa), Jacques Durand (ERSS, Université Toulouse-Le Mirail), Bernard Laks (MoDyCo, Université Paris X) et Chantal Lyche (Universités d'Oslo et de Tromsø) est un projet international en cours de réalisation, depuis 2002[29]. L'objectif est de constituer une base de données du français parlé dans toutes les régions de France métropolitaine, dans les DOM-TOM et dans les pays francophones. La base de données s'étendra même à d'autres pays pour enregistrer des locuteurs dont le français n'est pas la langue maternelle.

Un protocole a été établi afin que les enquêtes puissent être comparables quel que soit le pays dans lequel elles sont réalisées. Chaque locuteur participe à quatre enregistrements : une lecture de 94 mots, une lecture d'un texte, un entretien guidé avec un enquêteur qu'il ne connaît pas, une discussion libre avec des amis.

Nous avons utilisé les transcriptions disponibles sur le site du projet fin mars 2010, en sélectionnant uniquement les discussions libres enregistrées en France métropolitaine, soit

[29] Il semble que certains enregistrements antérieurs aient été intégrés au corpus. Le plus ancien date de 1997.

163 fichiers, d'une taille approximative de 270 000 mots. Les transcriptions sont au format .trs et les fichiers audio sont en .wav.

2.1.1.8 CPROM

Le Corpus CPROM (Auchlin, Avanzi, Goldman, & Simon, 2012) est un corpus récent, développé par Antoine Auchlin (Université Genève), Mathieu Avanzi (Université Neuchâtel / Paris X), Jean-Philippe Goldman (Université Genève) et Anne Catherine Simon (Université Catholique Louvain). Il inclut 24 enregistrements, qui fonctionnent souvent par trios : français de France, de Belgique et de Suisse. Les sept genres présents dans le corpus sont « journaux radiophoniques, lectures, discours politiques, conférences, interviews radiophoniques, itinéraires, et récits de vie ». Le corpus, d'environ 13 000 mots, est diffusé sur le Web. Les fichiers audio sont au format .wav et .mp3. Les transcriptions, annotées en phonétique, syllabes, mots et syllabes proéminentes, sont au format textgrid. Il existe également un fichier Pitch, qui contient la détection de la fréquence fondamentale par Praat corrigée manuellement. Le corpus étant restreint, nous avons conservé les enregistrements belges et suisses ainsi que les lectures.

2.1.1.9 OTG

OTG[30] (Office de Tourisme de Grenoble) (Antoine, Letellier-Zarshenas, Nicolas, Schadle, & Caelen, 2002), développé en 2002 par les laboratoires VALORIA [Laboratoire de Recherche en Informatique et ses Applications de Vannes et Lorient] et CLIPS-IMAG [Communication Langagière et Interaction Personne-Système - Informatique et Mathématiques Appliquées de Grenoble], est un corpus qui regroupe des échanges entre les employés de l'office de tourisme de Grenoble et les personnes en quête d'informations. Les dialogues sont non seulement spontanés mais aussi non provoqués par les chercheurs, ce qui est assez rare. Seuls les réceptionnistes sont au courant de l'enregistrement au moment de la discussion. Les interlocuteurs ont décidé par la suite s'ils permettaient ou non que l'enregistrement soit conservé. Le corpus totalise ainsi 7h d'enregistrements. Deux heures ont été transcrites, soit

[30] Voir le site http://www.info.univ-tours.fr/~antoine/parole_publique/OTG/index.html

environ 26 000 mots pour 315 fichiers. Les entretiens, effectués entre 5 réceptionnistes et 315 « touristes », sont généralement courts (moins de 3 minutes).

Le corpus est distribué librement par l'université de Tours dans le cadre du projet Parole Publique. Les fichiers audio sont envoyés par CD pour 15€. Les transcriptions sont en .xml, .txt et .pdf ; les fichiers audio sont en .wav.

2.1.1.10 Brassens

Brassens[31] est un corpus enfants-adulte, développé par le laboratoire LiCoRN de l'université de Bretagne Sud [Equipe de recherche en Lexicographie, Corpus et Ressources Numériques] lors du projet ANR EmotiRob, réalisé de 2007 à 2009. Il a été enregistré dans une classe d'école primaire. Les élèves, approximativement âgés de 7 ans, tentent avec l'aide de l'institutrice d'inventer une histoire ensemble puis par groupe en partant de la même base. Lorsque leur histoire est achevée, ils la répètent pour la mémoriser puis la racontent.

Le corpus inclut 138 dialogues, qui sont plutôt courts (moins de 3 minutes) et qui comptent environ 34 000 mots. Cela correspond à environ 4h d'enregistrement. Seuls 120 dialogues sont disponibles sur Internet. (version 0.1 du 4 juin 2008)

Le corpus est disponible en format .txt, .trs, .xml (tei), .pdf pour les transcriptions et .wav pour les fichiers audio. Comme pour OTG, le corpus Brassens est distribué librement par l'université de Tours dans le cadre du projet Parole Publique. Les fichiers audio sont envoyés par CD pour 15€.

2.1.1.11 UBS

Le corpus UBS[32] (Université Bretagne Sud) a été réalisé en 2008 par le laboratoire VALORIA de l'université Bretagne Sud [Laboratoire de Recherche en Informatique et ses Applications de Vannes et Lorient] dans le cadre du projet AGILE-OURAL du programme TECHNOLANGUE du Ministère de la Recherche. Il est uniquement constitué d'entretiens téléphoniques, passés entre les secrétaires de l'université Bretagne Sud et les étudiants. Les

[31] Voir le site http://www.info.univ-tours.fr/~antoine/parole_publique/Brassens/index.html

[32] Voir le site http://www.info.univ-tours.fr/~antoine/parole_publique/Accueil_UBS/index.html

appels sont réels et les autorisations d'utilisation ont été demandées aux appelants après enregistrements. Les secrétaires, elles, étaient au courant de l'enregistrement. Trois heures d'enregistrement ont ainsi bénéficié de l'accord des participants.

L'université de Tours distribue librement un tiers du corpus, soit une heure et environ 10 000 mots, sélectionné notamment pour sa qualité sonore. Cela correspond à 40 dialogues, produits par 40 appelants et deux secrétaires des UFR Sciences et Droit.

Les transcriptions, aux formats .txt, .doc, .odt, .pdf et .xml (.trs), sont disponibles sur le site de l'université de Tours (projet Parole Publique). Les fichiers audio, en .wav, sont envoyés par CD pour 15€.

2.1.1.12 Irit

Le laboratoire IRIT a enregistré en 2002 des explications d'itinéraires réalisées par téléphone. Les deux locuteurs sont à deux endroits différents et le premier doit expliquer au second comment le rejoindre. Les appels ne sont pas réalisés en situation réelle, mais l'appelant ne connaît pas à l'avance le lieu de départ. En outre, les participants ne se connaissent pas, ce qui écarte la possibilité de s'appuyer sur des connaissances communes (adresses d'amis, lieux publics peu célèbres, etc.). Tout ou presque doit être explicité. 21 situations ont été enregistrées, avec 23 participants. Les participants ne jouent jamais deux fois le même rôle. Ils sont tantôt l'appelant, tantôt l'appelé. Les dialogues sont disponibles en ligne sur la page personnelle de Laurent Prévot[33]. Ils sont courts et ont une taille totale d'environ 9000 mots.

2.1.1.13 Synthèse

Le tableau suivant récapitule les principales informations sur les corpus sélectionnés, soit le type de données (écrit/oral), la date de constitution (plus ou moins contemporaine), le nombre de transcriptions, le nombre approximatif de mots et les formats des fichiers disponibles (TABLEAU 12) :

[33] Voir le site http://www.irit.fr/ACTIVITES/LILaC/Pers/Prevot/Corpus.html.

Corpus	Source	Date	Nombre de transcriptions	Taille	Format
CERF/CorpAix	Ecrit / Oral	2004	? / 265	10 000 000 mots	txt
CRFP	Oral	98-2003	134	430 000 mots	txt - wma
C-Oral-Rom (fr)	Oral	80-2004	175	300 000 mots	txt/xml - wav
Nancy	Oral	2009	73	360 000 mots*[34]	txt
TCOF (adultes-enfants)	Oral	2011	177	175 000 mots*	trs/xml-wav
CTFP	Oral	1989-2000	36	67 000 mots	txt - mp3
CFPP	Oral	2007-2010	24	340 000 mots	trs - mp3/wav
PFC (France, discussions libres)	Oral	(2010)	163	270 000 mots*	txt/xml - wav
C-PROM	Oral	07-2010	24	13 000 mots*	trs - mp3/wav
OTG	Oral	2002	315	26 000 mots	xml/txt
Brassens	Oral	2009	120	34 000 mots*	trs/tei/txt/pdf
UBS	Oral	2008	40	10 000 mots	xml/txt
IRIT	Oral	2002	21	9 000 mots	txt

TABLEAU 12. Présentation globale des corpus sources

Au total, le corpus oral compte environ 3 034 000 mots, et le corpus écrit 9 000 000 mots. Le fait de réunir plusieurs corpus dont les formats de support et les formats de transcription diffèrent pose des problèmes de compatibilité. Pour rendre possible l'exploitation de l'ensemble, il est nécessaire d'uniformiser les formats.

2.1.2 Travail d'uniformisation[35]

Au départ, nous n'avions pas prévu d'uniformiser les données. Certes les corpus ne présentaient pas tout à fait les mêmes conventions de transcription, mais cela ne semblait pas être un obstacle à l'exploitation des corpus. En effet, s'il est désormais couramment admis

[34] Les nombres de mots indiqués avec un astérisque sont le résultat de nos propres estimations. Ces dernières concernent les corpus qui ne fournissent pas l'information ou les corpus toujours en cours de constitution, pour lesquels nous n'avons donc sélectionné qu'une partie des données qui existeront à terme.

[35] Ce travail a donné lieu à plusieurs communications et à une publication :

Benzitoun C. & Bérard L. (2011) *Leçon à tirer de la mutualisation de corpus oraux*. Communication présentée à la réunion de préparation au projet ANR ORFEO, Nancy, 25 janvier.

Bérard L. & Benzitoun C. (2011) *Mutualisation de corpus oraux.* Communication présentée au Colloque Jeunes Chercheurs « Corpus, Données, Modèles : approches qualitatives et quantitatives », Montpellier, 10 juin.

Bérard L. & Benzitoun C. (à paraître) Mutualisation et uniformisation de ressources de français parlé, *Cahiers de Praxématique 54*

qu'il existe de nombreux corpus français (Cappeau & Seijido, 2005) souvent peu accessibles, leur hétérogénéité est rarement mentionnée, et personne ne semble se soucier des recommandations d'EAGLES (EAGLES, 1996; Sinclair, 1996). Ces dernières années, la communauté a pris de conscience du problème que pose cette diversité :

> « On peut poser qu'il y a sans doute entre quatre ou cinq millions de mots effectivement disponibles mais l'absence de coordination rend l'exploitation de l'ensemble impossible » (Debaisieux, 2005, p. 8)

L'observation des résultats obtenus par les concordanciers révèlent plusieurs problèmes dus à cette hétérogénéité : présence de doublons, occurrences non repérées à cause de la présence de caractères spéciaux, des conventions de transcription, et de la spécificité du concordancier.

2.1.2.1 Une raison majeure pour uniformiser les données : le recours aux outils de requêtes

Nous avons choisi d'utiliser pour nos recherches le concordancier Contextes (Véronis, 1998). Développé par Jean Véronis, ce concordancier possède un grand atout : il est conçu non seulement pour les corpus écrits, mais aussi pour les corpus oraux. Il permet en effet d'écouter les extraits sonores correspondants aux occurrences sélectionnées.

Les fichiers d'entrée sont des fichiers textes. Les fichiers audio lisibles sont wav, mp3 et wma. Le fichier d'alignement texte-son est en format alg. Il doit indiquer pour chaque fragment de texte la position du premier et du dernier caractère et les deux temps correspondants, en seconde. Les fichiers de sortie sont au choix en txt et en xls. On peut rechercher par forme ou par expression régulière.

Notre première requête pour repérer les dépendances à distance était une expression régulière du type « qui \w+{2,5} que ». Cela signifie « qui », suivi de 2 à 5 mots, puis de « que », la forme « qui » étant potentiellement un interrogatif-relatif et la forme « que » potentiellement une conjonction de subordination introduisant la complétive (*à **qui** crois-tu **que** Jean donnera le livre ?*). Mais les résultats obtenus ne sont pas ceux attendus, car ils n'appliquent pas la restriction « entre 2 et 5 mots ». Or les occurrences avec un seul mot intermédiaire génèrent beaucoup de bruit.

La seconde requête a été développée pour expliciter le {2,5} en, par exemple, « qui \w+ \w+ que ». Le problème de cette requête est qu'elle écarte les énoncés du type *qui voulez-vous que*, qui présentent un tiret.

Une autre possibilité de l'outil pour expliciter le {2,5} est la fonction « contextes additionnel » de l'outil. On recherche un élément (« qui ») et on indique dans son contexte droit (ou gauche) un élément filtreur (par exemple « que ») à une distance choisie (jusqu'à 6 mots). Mais il semble qu'il y ait des problèmes avec cette fonction : certains fichiers faisaient échouer la requête ; pour l'interrogatif-relatif « qu », le nombre de résultats était plus faible avec « qu » qu'avec « qu' » ce qui est incohérent.

Nous avons alors essayé une autre méthode. Nous avons effectué les requêtes les plus basiques possible, en ne cherchant que les interrogatifs-relatifs. La restriction du contexte droit (présence de « que ») pouvait être réalisée automatiquement à l'aide d'un programme écrit en langage python. A ce moment-là, des problèmes d'encodage sont survenus et ont perturbé le programme. En effet, les corpus (surtout CorpAix, mais aussi Nancy et CTFP) comportent de nombreux caractères qui perturbent les traitements : la lettre « œ », le bullet « • », le dash « – », l'em dash : « — », etc. Certains caractères faussent également les résultats : dans CorpAix et CTFP coexistent deux types d'apostrophe : « ' » et « ' ».

Pour cette raison, nous avons entrepris le nettoyage des corpus, avec l'aide de Christophe Benzitoun.

2.1.2.2 Uniformisation des conventions de transcriptions

En plus de la suppression des caractères spéciaux, il nous a paru utile d'uniformiser l'ensemble des conventions de transcription. Dans les divers corpus, certains phénomènes sont annotés avec des symboles différents. Par ailleurs, certains symboles se révèlent polysémiques. Les annotations deviennent donc inutilisables. Nous avions besoin d'un système qui soit pertinent pour les traitements et analyses ultérieurs. Il fallait donc que les symboles soient acceptés par les outils, repérables automatiquement, non ambigus et si possible explicites.

L'uniformisation a été appliquée aux six corpus suivants : CRFP, CTFP, PFC, C-Oral-Rom, CFPP, CorpAix. Nous n'avons pas traité Nancy et TCOF car les transcriptions étaient en cours de relecture et les conventions étaient déjà proches des nôtres. Les autres corpus n'ont

pas été normalisés car nous y avons eu accès alors que l'uniformisation était déjà bien entamée.

Le tableau suivant présente les conventions que nous avons repérées le plus régulièrement dans chaque corpus et les conventions que nous avons adoptées (TABLEAU 13).

	CTFP	CFPP	PFC	CRFP	CorpAix	C-oral-rom	**CHOIX**
Pauses	- - -	+ ++ +++	, .	+	+ ++ +++	+	
Intonations		?	, . ?		//	? / // [/] #	
Allongements		un: un:: un:::			un:		
Discours rapporté		"x"	'x'				
Chevauchement		spk1 spk3	< >	< >	< >	< >	**Balises**
Multi-transcript°	/x, y/	(x, y) (x ; y)		/x, y/	(x, y)	/x, y/	**/x, y/**
Anonymisation				*P1*		yyy	***P1***
Pb écoute	XXX	XXX	(XXX)	XXX	XXX	xxx	**XXX**
Répétitions						&le &le	
Hésitations						&euh	
Amorces	c-	c-	c/	c-	c-	&c	**c-**
Liaisons	-z-	z-		=z=	=		**-z-**
Commentaires		[soupir] (rires)	(rires)	{rires}	(bruit)	hhh	**Balises**

TABLEAU 13. Synthèse des conventions adoptées par chaque corpus

En général, nous avons conservé et étendu les conventions les plus fréquentes (multi-transcription, amorce). Nous avons le plus possible évité de surcharger la transcription notamment par le marquage des répétitions et des hésitations. Le signe « & » a donc été supprimé ; les répétitions et les hésitations sont restées « le le » et « euh ». Enfin, les commentaires (rires, bruits ambiants et événements ponctuels) ont été placés entre balises car ils ne correspondaient pas à la parole des locuteurs et perturbaient les recherches. Ainsi, si l'on recherchait « rire(s) » dans les corpus originaux, on obtenait 1980 résultats alors qu'après le balisage, on en obtient plus que 69 occurrences. Il était auparavant impossible de distinguer de manière systématique la langue de la métalangue.

Nous avons décidé de supprimer toutes les marques de ponctuation (point, point d'interrogation, point d'exclamation, virgules). La plupart des transcriptions n'étant pas ponctuées, il était plus simple de supprimer des informations que d'en ajouter. De plus, il a été souligné que la ponctuation de l'écrit n'était pas forcément adaptée à l'oral et qu'elle

imposait une analyse préalable (Blanche-Benveniste, 1998; Blanche-Benveniste & Jeanjean, 1987). Il est donc gênant d'avoir des transcriptions ponctuées alors qu'elles doivent être la base des analyses ultérieures. Nous avons également supprimé les majuscules suivant ces ponctuations.

De même, nous avons supprimé toutes les marques d'intonations, de pauses et d'allongements. Ce choix est radical et il aurait sans doute été possible de conserver l'intégralité des informations marquées à l'aide de symboles sous forme de balises, puis de les supprimer et de les réinjecter à volonté. Mais nos compétences en informatique rendaient les résultats peu sûrs et certaines informations n'étaient pas directement utiles pour nous. Cette manipulation aurait également eu l'inconvénient de gêner certains logiciels. En outre, l'annotation de ces phénomènes n'était pas systématique mais liée à la subjectivité de chaque transcripteur.

Nous avons essayé d'effectuer les changements de façon automatique, mais le plus souvent, la transformation automatique s'est heurtée à de nombreuses difficultés.

La conversion des majuscules en minuscules n'était pas toujours transparente. Par exemple, le « E » majuscule pouvait correspondre à « e,é,è,ê », etc.

Même la suppression des marques (pause, intonation, allongement, discours rapporté) a nécessité de nombreuses vérifications et corrections. Les pauses, par exemple, étaient marquées dans CTFP par un ou plusieurs tirets « - ». Les amorces, elles, étaient notées également avec un tiret, mais collées à l'amorce, dans CorpAix, CFPP, CTFP et CRFP. Il ne devrait donc pas y avoir d'ambiguïté, comme le montre l'exemple suivant :

> *(172) elle a pas du tout réalisé euh - que que je continuais à courir et que - que je n'avais pas vu la voiture donc - euh - - b- bon moi j'ai quand je l'ai vue la voiture c'était déjà un peu tard quoi (CTFP, L'accident)*

Pourtant, il était fréquent que les espaces ne soient pas respectés et on trouvait aussi bien des pauses collées aux mots que des tirets marquant des amorces mais décollés de celles-ci. Ayant décidé de supprimer les pauses mais de garder les amorces, nous avons dû vérifier manuellement l'utilisation des tirets. Cette étape, qui aurait dû être rapide et automatique, a donc nécessité une longue vérification manuelle.

Certaines marques étaient utilisées pour noter différents phénomènes selon les corpus. Les parenthèses, par exemple, pouvaient indiquer une transcription multiple comme (ces ; ses) ou

bien un commentaire métalinguistique tel que (raclement de gorge), (marque d'acquiescement : hum), etc. Pour pouvoir les transformer, il était donc impossible d'utiliser des expressions régulières.

Parfois, les marques avaient même plusieurs significations dans le même corpus. Ainsi, « & » pointe à la fois les amorces, les hésitations et les répétitions. Souhaitant conserver seulement les amorces, et les transformer en tiret (a-), nous n'avons pas pu automatiser cette uniformisation.

Enfin, l'intervention humaine était indispensable car il existait de nombreuses variations, indépendamment des conventions fournies aux transcripteurs. Pour les multi-transcriptions, nous n'aurions pas imaginé qu'il puisse y avoir tant d'annotations. Entre les marques de début et de fin (parenthèse, crochet, slash) [celle de fin n'étant pas forcément identique à celle du début], la marque de séparation des deux transcriptions possibles (virgule, point virgule) et la présence ou non d'espace(s) avant et après chaque marque, le nombre d'annotations différentes était impressionnant.

2.1.2.3 Uniformisation de l'orthographe

Tous les corpus que nous avons traités utilisent l'orthographe standard (Blanche-Benveniste & Jeanjean, 1987). Mais il y a cependant quelques exceptions tels *d'la*, *t'as*, *y'a*. Ces conventions posent au moins deux problèmes. Le premier est que les deux versions existent dans les corpus (*de la / d'la*). Il y a donc fort à parier que les utilisateurs ne pensent pas à rechercher les deux formes et le résultat de leur requête doit être biaisé. Le deuxième est que les programmes d'annotation automatique (parseurs, taggueurs) sont généralement élaborés pour l'écrit et ne peuvent traiter de telles suites.

Nous avons donc décidé de ne pas faire d'exception à l'orthographe standard. Pour le phénomène du 'e' élidé devant une consonne (*d'la*), nous avons repéré automatiquement toutes les suites du type 'consonne + apostrophe + consonne' (hormis 'h') et avons remplacé l'apostrophe par un 'e' et un espace. Une fois de plus, un traitement totalement automatisé n'a pas été possible car il aurait généré de nombreuses erreurs. Nous avons dû prendre soin de mettre à part les cas particuliers comme « qu » qui peut correspondre à « que » ou « qui », « s'te plait », « à c't'époque », etc. Les suites 'consonne + apostrophe + voyelle' ont nécessité

un traitement en grande partie manuel. En effet, « t'es » peut par exemple correspondre à « te » dans « tu t'es fait mal » ou à « tu » dans « t'es parti ».

En ce qui concerne les chevauchements de parole, nous avons rencontré deux types de problème. Il arrive que les transcripteurs, dans le but d'indiquer scrupuleusement à quel endroit les chevauchements de parole commençaient, aient coupé un mot en deux. Mais cette manière de procéder possède l'inconvénient que le mot ainsi transcrit ne peut pas être repéré lorsque l'on recherche la chaîne de caractères complète. Lorsque nous les avons repérés, nous les avons rassemblés.

De même, à l'intérieur d'un chevauchement, les paroles d'un même locuteur peuvent être fragmentées en plusieurs tours. Cela rend le texte peu lisible et pour les requêtes ultérieures, cela limite grandement les contextes des concordances. Lorsque le chevauchement avait lieu entre un locuteur qui produisait un énoncé et un autre qui acquiesçait de temps à autre, nous avons supprimé les acquiescements et reconstitué linéairement l'énoncé. Il serait utile pour la recherche de structure de coller tous les tours de parole d'un même locuteur afin de réunir les syntagmes, comme la relative et sa tête dans le passage suivant :

(173) *<Turn speaker="spk2 spk1 spk2 spk1 spk2 spk1" startTime="164.05053637270618"*
endTime="166.69460469827209">
<Sync time="164.05053637270618"/>
<Who nb="1"/>
j'ai que Anne
<Who nb="2"/>
tu as
<Who nb="3"/>
j'ai que Anne
<Who nb="4"/>
que Anne
<Who nb="5"/>
qui a deux enfants
<Who nb="6"/>
deux enfants
</Turn> (PFC, 11aml1)

Nous souhaiterions obtenir (174):

(174) *<Turn speaker="spk2 spk1 spk2 spk1 spk2 spk1" startTime="164.05053637270618"*
endTime="166.69460469827209">

<Sync time="164.05053637270618"/>

<Who nb="1"/>

j'ai que Anne

<Who nb="2"/>

L1 tu as que Anne

<Who nb="1"/>

L2 j'ai que Anne qui a deux enfants

<Who nb="2"/>

L1 deux enfants

</Turn>

Mais l'idéal serait également de conserver une trace du fait que les dépendances ont été réalisées de manière découpée. Ces informations seraient utiles dans des perspectives macrosyntaxiques et discursives.

2.1.2.4 Suppression des doublons

La fusion des corpus n'a pas engendré que des problèmes de conventions, elle a également créé des doublons. En effet, comme la constitution de corpus est longue et coûteuse, il arrive que le même enregistrement soit réutilisé dans la réalisation d'un nouveau corpus. Bien évidemment, ces doublons doivent être éliminés en vue d'études statistiques mais leur repérage n'est pas toujours évident car les conventions de transcriptions peuvent changer d'un corpus à un autre.

Ex : C-Oral-Rom, ffamdl02

*EDO: il en faut hein / il faut un [/] il faut un gestionnaire / il faut un comptable / il faut [/] donc moi je remplis ce rôle-là // après c' est avec mon frère que je tiens le magasin aussi / et lui est magicien et jongleur // # *SAN: d' accord // alors c' est de famille hhh //

Ex : CRFP, PRI-CAE-2

L1 il en faut hein il faut un il faut un gestionnaire il faut un comptable il faut donc moi je remplis ce rôle-là après c'est avec mon frère que je tiens le magasin aussi et lui est magicien et jongleur L2 d'accord alors c'est de famille {rire}

Nous avons donc décidé de rechercher pour chaque transcription d'un corpus donné un mot clé du début et un mot clé de la fin de la transcription. Grâce au concordancier, nous pouvions voir les répétitions (IMAGE 3).

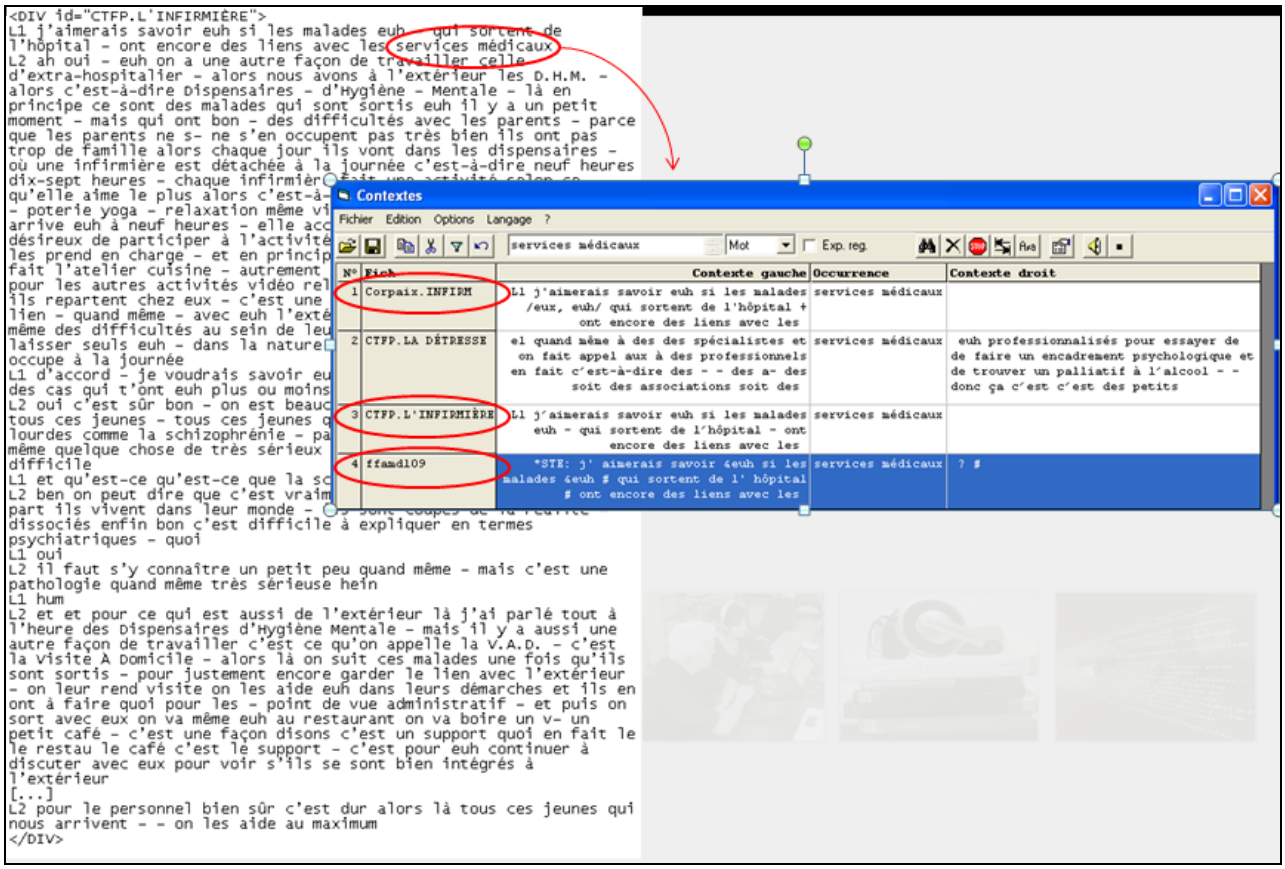

IMAGE 3. Recherche des doublons dans les corpus

Cette technique a permis d'identifier 49 transcriptions en double. Généralement, l'un des fichiers représentait un extrait de l'autre. Nous avons conservé les transcriptions les plus longues, sauf indication contraire (id = identique). Le tableau suivant présente les doubles supprimés dans chaque corpus (TABLEAU 14) :

CorpAix		CRFP		CTFP	
Avocat	1	PUB-LIM-1	1	Biologiste	1
Infirm (+lg)	2	PUB-CLE-1	2	La_dame du Finistère	2
Sapho	3	PUB-PNE-1	3	Secrétaire	3
		PRI-LYO-1	4	Agriculteur	4
		PRI-ROU-1	5	Cours	5
		PUB-PER-1(+lg)	6	Réflexion (ou Professeur)	6
		PUB_PSE_1	7	Editeur	7
				Poulets	8
				Vendeurs	9
				Infirmière	10

C-Oral-Rom	
ffamdl02	1
ffamdl05 (id)	2
ffamdl12	3
ffamdl14	4
ffamdl18	5
ffamdl19	6
ffamdl20	7
ffamdl21	8
ffamdl22	9
ffammn02 (id)	10
ffammn06	11
ffammn08	12
ffammn09	13
ffammn10	14
ffammn11	15
ffammn24	16
fmednw01	17
fnatco02	18
fnatpe02 (id)	19
fpubcv01	20
fpubdl06	21
fpubdl07	22
fpubdl08	23
fpubmn03	24
fpubmn04	25
ffamdl04	26
fnatpr02	27
fmedts02	28
ffamdl13	29

TABLEAU 14. Liste des fichiers supprimés dans chaque corpus

Dans deux cas, nous avons conservé les extraits qui étaient pourtant plus courts. Pour « CorpAix.Infirm » nous avons décidé de conserver le double car celui-ci était aligné texte-son.

Le tableau suivant indique les noms des fichiers dont le contenu était partiellement identique à ceux des fichiers supprimés :

CorpAix = C-Oral-Rom		CRFP = C-Oral-Rom		CTFP = CorpAix	
Avocat	fnatla01-02-03	PUB-LIM-1	fnatte01	Biologiste	Claire
Infirm	ffamdl09	PUB-CLE-1	fnatte03	Dame_Finistère	Barian
Sapho	fmedin02	PNE-PUB-1	fnatco01	Secrétaire	Jeanne
		PRI-LYO-1	ffammn01	Agriculteur	Merl3
		PRI-ROU-1	fpubdl01	Cours	Posie
		PUB-PER-1	fmedsp02	Réflexion	Corpus
		PUB-PSE-1	fnatte02	Editeur	Casane
				Poulets	Volatil
				Vendeurs	Aidouni
				Infirmière	Infirm

C-Oral-Rom	
ffamdl02	CRFP.PRI-CAE-2
ffamdl05	CRFP.PRI-BAY-2
ffamdl12	CRFP.PRI-NAR-1
ffamdl14	CRFP.PRI-PNO-1
ffamdl18	CRFP.PRI-NIC-1
ffamdl19	CRFP.PRI-PCR-3
ffamdl20	CRFP.PRI-AUX-1
ffamdl21	CRFP.PRI-AUX-1
ffamdl22	CRFP.PRI-ROU-2
ffammn02	CRFP.PRI-COR-1
ffammn06	CRFP.PRI-PSO-3
ffammn08	CRFP.PRI-MON-1
ffammn09	CRFP.PRI-CLE-2
ffammn10	CRFP.PRI-PNO-4
ffammn11	CRFP.PRI-AIX-1
ffammn24	CRFP.PRI-NIC-1
fmednw01	CRFP.PUB-BAY-1

fnatco02	CRFP.PUB-STR-1
fnatpe02	CRFP.PUB-NIC-1
fpubcv01	CRFP.PUB-AMI-1
fpubdl06	CRFP.PRO-NAR-1
fpubdl07	CRFP.PRO-NAR-1
fpubdl08	CRFP.PRO-NAR-1
fpubmn03	CRFP.PUB-DIJ-1

fpubmn04	CRFP.PUB-PNO-1
ffamdl04	Nancy.Chant
fnatpr02	Nancy. Homelie 1, 2, 3
fmedts02	Nancy. Messier
ffamdl13	corpAix.AZ

TABLEAU 15. Correspondance entre les corpus : fichiers au contenu identique

Les transcriptions des corpus réutilisés ont la plupart du temps été corrigées et adaptées avant d'être intégrées au nouveau corpus. C'est pourquoi nous avons décidé de copier les extraits plus courts mais mieux transcrits dans leur doublon (plus long), notamment ceux de CTFP dans CorpAix. Il faut noter que « CTFP.Agriculteur » comporte un passage non transcrit, qui en revanche est transcrit dans « CorpAix.Merl3 ». Dans ce cas, nous avons donc copié la transcription révisée de CTFP en y ajoutant le passage interne manquant de CorpAix.

Etrangement, dans un même corpus, CorpAix, on trouve également des doublons, qui portent des noms différents. Arbitrairement, nous avons gardé le premier fichier cité ci-dessous :

- 27sos-11fragra
- 3cachou-cachou
- adrien-corpusad
- barian-justif
- dur58b-mmeduran
- pr54ab-pr55ab
- pavrayst-pr61ab
- morin-sourd1 (morin +lg)
- stan-stan2

Nous avons rencontré un cas particulier. Le fichier « Le cordonnier » de CTFP avait une grande partie de son texte dupliqué. Nous avons supprimé les lignes répliquées.

Enfin, les doublons étaient quelquefois partiels dans les deux fichiers : la fin d'une transcription correspondait au début de l'autre, mais le début du premier était unique et la fin du deuxième également. Cela peut être représenté par l'image suivante :

IMAGE 4. Doublons partiel dans les corpus

Ce cas de figure est arrivé à trois reprises :

- La fin de CRFP.PUB-PER-1 était identique au début de C-Oral-Rom.fmedsp02. Nous avons complètement supprimé PUB-PER-1, après avoir injecté la fin dans fmedsp02. Les fichiers audio ont également été fusionnés. Les noms des deux fichiers ont été suffixés par '-v2'. Cette opération n'était pas la solution la plus simple et nous avons changé de méthode pour les autres cas.
- La fin de PFC.75csb1 était identique au début de PFC.75cvl1. Nous n'avons pas modifié le second fichier. Pour 75csb1, seul le début a été conservé et nous avons coupé le fichier audio. Les deux fichiers (texte et audio) ont été renommés 75csb1-v2.
- La fin de PFC.13bfa1 était identique au début de PFC.13bsa1. Nous n'avons pas modifié le premier fichier. La fin de 13bsa1 a été sauvegardée sous le nom 13bsa1-v2. Le fichier audio a subi le même traitement.

Voici l'illustration de la sélection des données :

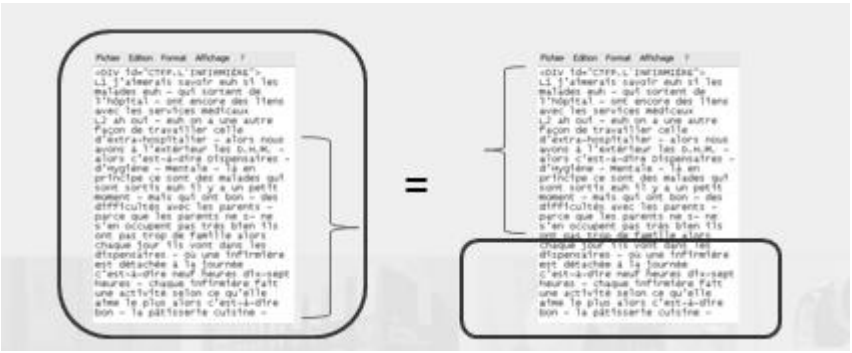

IMAGE 5. Solution adoptée pour les doublons partiels

Le corpus PFC a posé un problème spécifique. Celui-ci est mis en ligne mais il est téléchargeable par locuteur. Chaque locuteur ayant participé au projet est relié à quatre transcriptions, qui portent dans leur code les initiales du locuteur : discussion libre, entretien guidé, texte lu, mots lus. Lorsque plusieurs locuteurs se retrouvent dans la discussion libre, on obtient autant de fichiers, contenant la même discussion. Il ne s'agit pas à proprement parler de doublons mais si l'on télécharge l'ensemble des fichiers, les discussions se répètent et il n'y a pas de liens visibles entre les fichiers pour repérer cette répétition.

Réunir des corpus nécessite donc quelques ajustements. Mais ce travail préalable est largement compensé par un gain de temps lors de l'exploitation des données. En outre, la démarche peut aboutir à une réévaluation de la taille de chaque corpus, finalement moins important qu'il n'y paraît. Les éventuels calculs statistiques doivent prendre en considération, qu'un corpus fusionné comporte beaucoup moins de mots que la somme de chaque corpus.

2.1.2.5 Uniformisation du format

Les transcriptions des différents corpus étaient stockées dans trois formats distincts : texte brut (ASCII), *Transcriber*[36] et *Praat*[37]. Nous avons harmonisé l'ensemble au format *Transcriber*. Ce dernier est en effet un langage XML. Or l'utilisation du format XML est de plus en plus répandue à l'heure actuelle car ce format permet l'échange de données. Les balises facilitent la distinction entre langue et métalangue, ce qui, nous l'avons vu, est nécessaire à l'analyse.

Le passage au format Transcriber a nécessité l'ajout des balises <Turn> et <Sync> à chaque tour de parole et le renseignement des attributs « startTime », « endTime » et « speaker ». Dans les cas de chevauchements de parole, la balise <Who> a également dû être ajoutée afin d'identifier le locuteur courant.

Certains corpus ont nécessité un traitement plus important. Par exemple, pour C-ORAL-ROM, nous disposions d'un côté d'un fichier d'alignement texte-son et de l'autre de la transcription orthographique. Il a donc fallu élaborer un programme en Python afin de

[36] Voir le site http://trans.sourceforge.net/en/presentation.php
[37] Voir le site www.praat.org

convertir le corpus en format *Transcriber* valide[38]. La conversion n'a pas été nécessaire pour CorpAix qui ne dispose pas d'alignement texte-son.

Une fois l'uniformisation des corpus effectuée et vérifiée, nous avons entamé la projection vers le format d'entrée de certains logiciels d'exploitation. Dans un premier temps, nous avons effectué la conversion pour le concordancier *Contextes* (txt, wma, alg).

2.1.2.6 Résultats

La suppression des doublons et la normalisation des corpus fait apparaître les résultats suivants que nous présentons dans le (TABLEAU 16). Le nombre de mots est calculé sans les mentions des locuteurs, grâce à un programme python qui prend comme séparateur les fins de ligne, les espaces et les apostrophes.

Corpus oraux	**Date**	**Nb transcriptions**	**Taille**	**Format**[39]
CorpAix	2004	262	970 000 mots	txt
CRFP	98-2003	127	420 000 mots	trs – mp3
C-Oral-Rom (fr)	80-2004	146	233 000 mots	trs – mp3
Nancy	2009	73	360 000 mots	txt
TCOF adultes-enfants	2011	177	175 000 mots	trs – wav
CTFP	1989-2000	26	55 000 mots	trs – mp3
CFPP	2007-2010	24	340 000 mots	trs – mp3
PFC (France, discussions libres)	(2010)	163	270 000 mots	trs – mp3
C-PROM	07-2010	24	13 000 mots	trs – mp3/wav
OTG	2002	315	27 000 mots	txt/xml
Brassens	2009	120	34 000 mots	txt/xml/tei/pdf
UBS	2008	40	8 000 mots	txt/xml
IRIT	2002	21	7 000 mots	txt
TOTAL		**1518**	**2 912 000 mots**	

TABLEAU 16. Présentation globale des corpus oraux, après uniformisation d'une partie d'entre eux

[38] Nous remercions B. Gaiffe pour avoir effectué ce programme de conversion.

[39] Pour simplifier le tableau, nous n'indiquons pour les corpus uniformisés que le format que nous avons choisi : Transcriber pour les transcriptions et mp3 pour les fichiers audio.

La comparaison du nombre de caractères des deux versions de corpus (TABLEAU 17) donne une idée du nombre de marques qui étaient initialement ajoutées aux paroles des locuteurs (1 609 252 caractères en tout)[40]. Le corpus le plus affecté par la standardisation est C-Oral-Rom.

	Nombre de caractères (avec les espaces)	
	Corpus source	Corpus cible
PFC[41]	1 681 465	1 354 510
CFPP[42]	1 793 298	1 683 782
CTFP	337 649	271 284
CRFP	2 313 466	2 151 694
CorpAix	5 338 200	4 921 096
C-Oral-Rom	1 772 859	1 245 319
TOTAL	13 236 937	11 627 685

TABLEAU 17. Comparaison du nombre de caractères de chaque corpus avant et après standardisation

Au total, le corpus uniformisé, issu de la fusion et de l'homogénéisation de CRFP, CTFP, PFC, C-Oral-Rom, CFPP et CorpAix, représente plus de deux millions de mots (2 288 000 mots). Il constitue donc un corpus oral de parole adulte-adulte important pour la langue française. Pour une grande partie, il est vrai, la partie sonore est inexistante. Nous avons tout de même conservé l'intégralité car la dimension prosodique n'était pas capitale dans notre étude qui porte sur des segments courts. Les corpus avec partie sonore ont pu permettre de vérifier la faible variation de réalisation du schéma prosodique de la séquence (sauf cas d'insertion d'adverbes). La séquence forme un tout prosodique.

Nous verrons dans les chapitres suivants que nous avons fréquemment utilisé le corpus uniformisé pour vérifier des hypothèses. Un certain nombre de requête a été fait uniquement sur cette partie. La recherche des occurrences de dépendance à distance, bien sûr, a été menée dans l'ensemble des corpus, atteignant presque trois millions de mots.

[40] Le nombre de tokens et de caractères reproduit ici est celui fourni par Notepad++. La mention des locuteurs est cette fois prise en compte, et ce dans les deux versions des corpus.

[41] PFC est un corpus actuellement toujours en évolution. La version considérée ici, récupérée en mars 2010, comporte 163 sans doubles.

[42] Le compte est effectué pour une version antérieure de CFPP, qui comporte 16 textes.

2.1.3 Synthèse sur le prétraitement des ressources à effectuer avant une analyse

Rassembler plusieurs corpus pose des difficultés. Généralement, les corpus ne sont pas disponibles sous le même format, ce qui peut poser problème à certains outils d'interrogation. Au-delà du format du fichier, la structuration et la typologie des informations diffèrent. Certaines indications sur la prosodie, la prononciation, les pauses, etc. peuvent être notées par des symboles, symboles eux-mêmes arbitraires et donc variés. Ces indications se trouvent soit dans le fil du texte de la transcription, soit dans un balisage. La transcription est généralement orthographique, de manière stricte ou approximative. Pour tirer le meilleur parti des corpus, il faut uniformiser ces informations (à moins de faire une recherche d'un simple token, qui ne subit jamais de variation de prononciation). L'uniformisation est longue, souvent manuelle, mais elle permet d'éviter à la fois le silence et le bruit dans les données extraites par les requêtes.

A l'issue du travail de traitement, on peut aboutir à la conclusion que la qualité des ressources importe autant, ou plus, que leur taille. Constituer un corpus n'est pas un but en soi. Les corpus ont vocation à être analysés en tant qu'échantillon représentatif d'une langue ou à servir de ressource de base pour divers outils. Certains éléments doivent donc être impérativement pris en compte lors des étapes de la constitution de corpus :

- Il faut éviter les caractères spéciaux afin de pouvoir effectuer des traitements automatiques ultérieurs.
- Comme la constitution de corpus est coûteuse et que l'objectif de tous les chercheurs est de pouvoir travailler à partir d'une grande masse de données, il est nécessaire de réfléchir à un format commun, tel que le XML ou la TEI, qui permettent l'échange et le partage des données.
- Le téléchargement des corpus doit être conçu de façon à éviter à l'utilisateur de devoir télécharger successivement chaque fichier de transcription et chaque fichier audio. Par ailleurs, lorsque l'accès au corpus est donné par locuteur, il serait bon de permettre de repérer les enregistrements à multi-locuteurs, fournis plusieurs fois.
- Enfin, le plus important concerne les conventions de transcription.

- La transcription d'une forme doit être systématique, afin d'être retrouvée automatiquement par un concordancier[43]. L'uniformisation du corpus constitue un gain de temps pour la recherche: il n'est plus nécessaire d'imaginer toutes les possibilités de transcription d'une forme (ex : je suis, j'suis, ch'suis, chuis, chu, etc.). Ainsi, les résultats obtenus sont plus pertinents et le tri en est accéléré.
- Pour faciliter la recherche d'expressions complexes, notamment syntaxiques, il est important que les commentaires métalinguistiques (pause, rires, bruits, précision de l'annotateur sur la prononciation d'une forme, etc.) ne soient pas inscrits sur le même niveau. Cela permet d'éviter le biais de résultats d'une requête prévoyant un empan entre deux formes ou même ceux d'une requête cherchant des collocations ou des expressions figées. Les balises XML facilitent l'annotation d'éléments à différents niveaux. Elles permettent non seulement de commenter les bruits extérieurs et les rires mais aussi les gestes des locuteurs, leur regards, ce par quoi ils peuvent être distraits, ainsi que des particularités prosodiques, phonétiques, morphosyntaxiques ou syntaxiques. En outre, la linéarité dans la parole des locuteurs n'est pas seulement utile pour la recherche mais également pour enrichir automatiquement les données. Suite au travail d'uniformisation, nous avons par exemple pu commencer à annoter les transcriptions en catégories morphosyntaxiques (POS tagging).

[43] Il semble que le laboratoire Clapi ait développé un concordancier qui prenne en compte la variation graphique. Malheureusement, ce concordancier n'est pas diffusé. Il peut être utilisé seulement en ligne et sur les corpus de Clapi. (*CLAPI-concordancier [Computer Software]*, s. d.; ICOR, s. d.)

2.1.4 Perspectives : annotation morphosyntaxique [44]

Certaines requêtes seraient bien plus précises si l'on pouvait formuler les catégories grammaticales attendues. Or les corpus oraux présentent des spécificités qui ne sont pas prises en compte par les étiqueteurs morphosyntaxiques, qui ont été élaborés à partir de données écrites et pour des données écrites. La solution envisagée est généralement d'adapter les outils existants a postériori (Blanc, Constant, Dister, & Watrin, 2008; Dister, 2007) ou de masquer certains phénomènes tels que les « disfluences » (Valli & Véronis, 1999).

Pour notre part, nous[45] avons décidé de travailler directement à partir des transcriptions brutes avec un minimum d'aménagement pour ne pas dénaturer les données. Nous avons uniquement supprimé les commentaires. Ainsi, au lieu d'adapter les sorties des outils existants par des post-traitements, nous avons préféré développer une méthode spécifique, à l'image de ce que proposent Eshkol et al. (Eshkol, Tellier, Taalab, & Billot, 2010; Tellier, Eshkol, Taalab, & Prost, 2010). Nous avons souhaité former l'outil *Treetagger* (Schmid, 2009) à l'étiquetage de corpus oraux. TreeTagger est un outil libre, qui dispose d'un module d'apprentissage. Son fonctionnement est basé sur les statistiques de n-grams, ce qui le rend facilement modifiable. D'après ses développeurs, il est très vite performant, même avec un petit corpus d'apprentissage (Schmid, 1994; Stein & Schmid, 1995) : il atteindrait une précision de 95,55% avec un corpus d'apprentissage de 43 834 mots[46]. Le module d'apprentissage, TrainTreeTagger, demande en entrée plusieurs ressources : un lexique forme-étiquette (POS), un corpus d'apprentissage étiqueté (pour désambigüiser les formes qui ont plusieurs étiquettes d'après leur contexte), une liste de tokens à ne jamais segmenter (du type « pomme de terre ») et la liste des étiquettes possibles pour un mot inconnu.

Après avoir réfléchi à un jeu d'étiquettes qui prendrait en compte les spécificités de l'oral (mention du locuteur, amorces, multi-transcriptions), nous avons construit un corpus de

[44] Ce travail a donné lieu à plusieurs communications :

Bérard L. & Benzitoun C. (2011) *Réflexion sur l'annotation syntaxique de corpus oraux*. Communication présentée à la réunion de préparation au projet ANR ORFEO, Paris, 1 avril

Benzitoun C. & Bérard L. (2011) *Annotation morphosyntaxique de corpus oraux : présentation d'une expérience en cours.* Communication présentée à la journée ATALA « Annoter les corpus oraux », Paris, 9 avril 2011.

Benzitoun C. & Bérard L. (2011) *Premiers tests en vue de l'étiquetage d'un corpus oral par apprentissage automatique exclusivement endogène*. Communication présentée au colloque AFLS 2011 « Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française », Nancy, 8-10 septembre.

[45] Christophe Benzitoun et moi-même.

[46] Voir le fichier en ligne ftp://ftp.ims.uni-stuttgart.de/pub/corpora/french-par-linux.info

base étiqueté pour servir de corpus d'apprentissage. Ce corpus de base est tiré d'une partie du corpus TCOF. Il est constitué d'environ 22 000 mots. Il a été pré-étiqueté avec l'outil d'origine TreeTagger puis les étiquettes ont été corrigées, manuellement, par une lecture intégrale des fichiers étiquetés.

A partir de ce corpus de base étiqueté, nous avons élaboré un lexique Forme/POS, qui présentait l'avantage d'avoir exactement les étiquettes voulues. Sa couverture lexicale n'était pas très importante mais les premiers tests effectués montraient que plus de 60% des mots inconnus recevaient une bonne étiquette.

La première évaluation, portant à la fois sur les lemmes et les étiquettes, a donné un taux d'erreurs de 10,4%, l'outil étant entraîné sur 10 000 mots. La précision des étiquettes morphosyntaxiques seulement (et non de la lemmatisation) s'élevait à 91,7%. Cela était très encourageant vu la taille du lexique et du corpus de référence. Néanmoins, les résultats ont été améliorés par l'ajout d'un lexique extérieur, Morphalou (Romary, Salmon-Alt, & Francopoulo, 2004), et ils ont atteint 94,2% de précision. Cette dernière solution a donc été conservée[47].

Un second étiqueteur a également été évalué : MElt (P. Denis & Sagot, 2009), utilisant comme lexique extérieur le Le*fff* (Sagot, 2010). Avec le même corpus d'apprentissage de 10 000 mots, ses résultats étaient plus intéressants : il obtenait une précision de 95,7%. Avec un apprentissage de 20 000 mots, la précision atteignait 96,9%[48].

2.2 Méthodologie de recherche des dépendances à distance

Le corpus étant relativement important, nous n'avons généralement pas recherché manuellement les attestations de constructions présentant des dépendances à distance. Nous avons tenté d'en repérer les occurrences de manière automatique, en utilisant un

[47] Toutes les ressources qui ont permis de former TreeTagger à l'étiquetage de corpus oraux sont disponibles sur le site du CNRTL, à l'adresse http://www.cnrtl.fr/corpus/perceo/. On y trouve le corpus de base étiqueté, ainsi qu'un corpus additif de 80 000 mots corrigés manuellement et le lexique Forme-POS obtenu par fusion des tokens du corpus et du lexique Morphalou. Le fichier paramètre obtenu à l'aide du module d'entraînement est également disponible. Il est donc possible d'annoter automatiquement de nouveaux corpus oraux avec ce Tree Tagger entraîné sur et pour le français parlé.

[48] Pour plus de précision sur la méthodologie adoptée et les résultats, consulter (Benzitoun, Fort, & Sagot, 2012).

concordancier. La lecture intégrale d'une partie des corpus, avec relevé manuel des occurrences de dépendance à distance, a attesté de la fiabilité des requêtes.

2.2.1 Requêtes

Les requêtes soumises au concordancier doivent couvrir la variété de réalisation de la construction. Nous avons recherché d'abord les constructions dans lesquelles le verbe recteur est fini, enchâssé dans une subordonnée complétive (175) puis les constructions dans lesquelles le verbe recteur est à l'infinitif (176).

(175) Qu'est-ce que tu veux que je fasse ?

(176) Qu'est-ce que tu veux faire ?

2.2.1.1 Dépendances à distance en complétive

Les corpus Brassens, OTG, UBS, CPROM et Irit ont été lus en intégralité. La recherche a donc été entièrement manuelle. Ce traitement particulier tient au fait que le format et le type de transcription de ces corpus ne facilitaient pas le traitement automatique. Le corpus Brassens, en particulier, qui présente des transcriptions de paroles d'enfants, est difficilement exploitable automatiquement. La lecture du corpus a également permis de connaître le contexte de production et le lien entre les termes employés par l'institutrice et ceux (parfois repris) des enfants.

Pour les autres corpus, nous avons récolté les occurrences de dépendances à distance à l'aide d'outils.

Les dépendances à distance en complétive en « que » ont été repérées par le concordancier Contextes. Compte tenu des problèmes rencontrés avec cet outil (voir §2.1.2.1), et notamment la fonctionnalité 'restriction du contexte additionnel', nous avons dû poser de nombreuses requêtes (374 requêtes pour les complétives).

La première série de requêtes était du type « qui \w+ \w+ qu » « Mot+/-1 », soit un pronom interrogatif-relatif suivi de 2 à 5 mots puis de la forme « qu » à laquelle peut être ajoutée une lettre. L'option « mot+/-1 » s'applique uniquement à « qu » et reconnaît les mots qui sont exactement constitués de « qu » ou de « qu + une lettre », soit « qui, que, qu' ». Toutes ces

formes peuvent être annonciatrices d'une complétive. Ces requêtes ont été faites sur les corpus dans leur état d'origine, ce qui a pu biaiser les réponses obtenues.

Ce type de requête ne renvoie pas les énoncés qui ont un tiret entre les mots de l'empan. Cela élimine donc toutes les interrogatives directes avec inversion du sujet (« qui voulez-vous qu »). Nous avons explicité la présence d'un tiret dans la seconde série de requêtes. Nous avons cependant restreint la recherche à un seul tiret par énoncé, afin de ne pas multiplier davantage les requêtes. Les corpus oraux interrogés avaient à ce moment-là déjà subi une première phase d'unification (traitement des doublons et uniformisation d'une grande partie des conventions) (voir §2.1.2) afin d'obtenir moins de silence que dans la requête précédente.

Les dépendances à distance en complétive en « si » ont également été recherchées par le concordancier Contextes. Les corpus oraux utilisés avaient subi une deuxième phase d'uniformisation (uniformisation des multi-transcriptions). Nous avons renoncé à la requête permettant d'obtenir les constructions présentant des tirets, un tri préalable ayant montré qu'il n'y avait aucune occurrence de dépendance à distance. Il était donc peu probable que la requête sur les tirets en fasse apparaître.

Le tableau suivant présente les requêtes que nous avons soumises pour les dépendances à distance en complétive (TABLEAU 18) :

Requête des dépendances à distance en complétive (374 requêtes)		
Interrogatifs - relatifs	Empan (sans et avec tiret)	Indice d'une complétive potentielle
qui	\w+ \w+	qu « Mot+/-1 »
que	\w+ \w+ \w+	si
qu'	\w+ \w+ \w+ \w+	s'
dont	\w+ \w+ \w+ \w+ \w+	
où		
pourquoi	\w+-\w+	
quand	\w+-\w+ \w+	
comment	\w+ \w+-\w+	
combien	\w+-\w+\w+ \w+	
lequel	\w+ \w+-\w+ \w+	
laquelle	\w+ \w+ \w+-\w+	
lesquels	\w+-\w+ \w+ \w+ \w+	
lesquelles	\w+ \w+-\w+ \w+ \w+	
quel	\w+ \w+ \w+-\w+ \w+	
quelle	\w+ \w+ \w+ \w+-\w+	
quels		
quelles		

TABLEAU 18. Requête des dépendances à distance en complétive

Avec AntConc, découvert plus tard, il aurait été possible, en une requête, de définir l'empan « 2-5 mots », ainsi que les pronoms interrogatifs-relatifs souhaités (IMAGE 6). Mais d'autres problèmes se seraient posés, les expressions régulières et les apostrophes n'étant pas autorisées dans le champ « Context Words ».

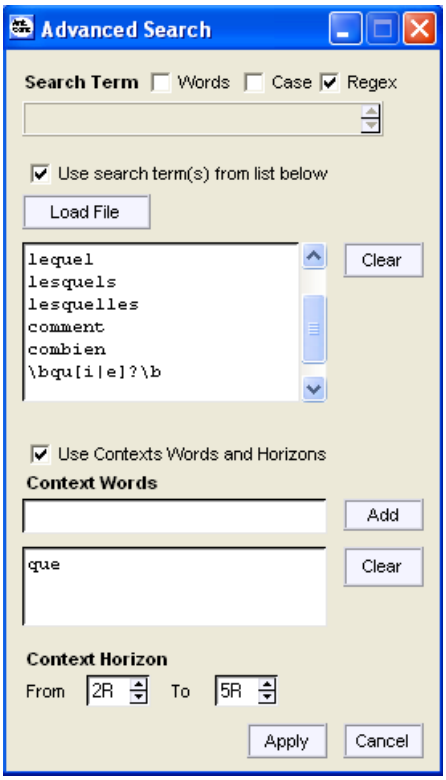

IMAGE 6. Interface de recherche avancée d'AntConc

Pour le corpus adultes-enfants TCOF, les requêtes ont été effectuées d'abord sur le corpus de narration, puis sur le corpus de conversation (ces derniers n'étant pas disponibles au départ). Pour le corpus de narration, la méthode appliquée était similaire à celle utilisée pour les corpus adultes. Nous avons recherché avec Contextes les séquences suivantes : un interrogatif relatif « que », « qu' » (incluant « qu'est-ce que » et « qu'est-ce qu' »), « où » suivi de 2 ou 3 mots et de « qu ». La requête a donc concerné les paroles de l'ensemble des locuteurs. Pour le corpus de conversation, en revanche, nous avons utilisé le concordancier JConc (Canut et al., 2010), développé dans le cadre du projet TCOF[49]. Ce concordancier permet de poser des requêtes élaborées, avec restriction des contextes gauche et droit, et d'écouter les séquences recherchées. Il possède également une fonctionnalité qui permet de sélectionner une partie des locuteurs. Nous avons donc ciblé les locuteurs adultes et fourni la requête suivante : un interrogatif-relatif « qui », « qu'est-ce qu' », « que », (incluant « qu'est-ce que »), suivi de 0 à

[49] Cet outil sera prochainement disponible librement sur le CNRTL.

5 mots, puis de « qu ». Le contexte gauche a été restreint afin d'écarter quelques complétives en « que » : la séquence ne devait pas être précédée immédiatement des mots « parce », « crois », « sais », « veux » et « faudra ».

2.2.1.2 Dépendances à distance en infinitive

Les dépendances à distance en infinitive, tels que « qu'est-ce que tu veux y faire », sont très nombreuses. C'est pourquoi nous avons choisi de restreindre notre relevé à cinq corpus oraux : CFPP, C-Oral-Rom, CRFP, CTFP et PFC, dans une version quasiment identique à celle utilisée pour les requêtes en « si » (voir p.121. La seule modification était la suppression des espaces après les apostrophes dans C-Oral-Rom).

Pour rendre les résultats plus pertinents, il était souhaitable de pouvoir exprimer dans les requêtes la volonté de trouver dans la construction un verbe fini et un verbe à l'infinitif. Pour ce faire, l'étiquetage des corpus en morphosyntaxe était indispensable. Comme le paramétrage de TreeTagger pour l'oral était loin d'être terminé (voir §2.1.4), nous avons utilisé un autre outil : Cordial Analyseur (Laurent, 2008). Développé par la société Synapse Développement, c'est un « logiciel d'analyse syntaxique et sémantique », jugé comme l'un des étiqueteurs morphosyntaxiques les plus fiables. Il comporte neuf catégories (nom, verbe, adjectif, pronom, déterminant, adverbe, préposition, conjonction et ponctuation). Celles-ci sont divisées en sous-catégories (nom : commun / propre, pronom : personnel / démonstratif / indéfini / relatif / possessif, etc) et précisent certains traits morphologiques (mode, temps, personne, genre, nombre) et phonologiques (débute par un h aspiré), ce qui donne un jeu de 128 étiquettes (voir Annexe 2). Cordial est orienté vers la langue écrite : son corpus de référence, notamment, est un corpus de presse et de littérature. Son analyse se base sur la « phrase », notion déjà problématique à l'écrit, et totalement inutilisable pour l'oral (Blanche-Benveniste, 1998). Néanmoins, la catégorie verbale étant souvent repérable immédiatement, sans ambiguïté, de par sa forme, nous avons estimé que l'étiquetage devait convenir pour la requête spécifique de dépendances à distance en infinitive.

Les fichiers étiquetés par Cordial peuvent être interrogés à l'aide de CooLoX (Audibert, 2001). Basée sur la bibliothèque[50] LoX de C++, WinLoX (la version Windows de CooLoX) permet de faire des recherches simples de formes sur des textes bruts ou de combiner des recherches de formes, de lemmes et de POS sur des textes étiquetés se présentant dans un

[50] Une bibliothèque fournit des codes utilisables par divers programmes.

format tabulaire[51]. Le langage de recherche de CooLoX s'appuie sur celui des expressions régulières ; au lieu de viser des chaînes de caractères, il se situe au niveau des suites de mots.

Nous avons soumis des requêtes du type « [POS~"^V" & POS != "VINF"] » avec un filtrage à gauche : « [lemme="qui"] [] {0,5} » et à droite : « [] {0,5} [POS="VINF"] ». Si l'on décompose la requête, cela signifie que l'on recherche un mot dont l'étiquette commence par 'V', c'est-à-dire un verbe conjugué à n'importe quel temps et à n'importe quelle personne mais qui ne soit pas à l'infinitif (POS != "VINF"). L'outil doit ensuite sélectionner les verbes précédés du lemme « qui », qui dans certaines occurrences sera l'interrogatif-relatif que nous cherchons, et ce dans les six mots précédents. Enfin, dans les six mots qui suivent le verbe, doit être présent un mot dont l'étiquette est VINF, c'est-à-dire un verbe à l'infinitif. Le (TABLEAU 19) explicite l'ensemble des requêtes posées pour repérer les dépendances à distance en infinitive.

Requête des dépendances à distance en infinitive (10 requêtes)		
Gauche =	**Recherche =**	**Droite =**
[lemme="qui"] [] {0,5} [lemme="que"] [] {0,5} [lemme="dont"] [] {0,5} [lemme="où"] [] {0,5} [lemme="pourquoi"] [] {0,5} [lemme="quand"] [] {0,5} [lemme="comment"] [] {0,5} [lemme="combien"] [] {0,5} [lemme="lequel"] [] {0,5} [lemme="quel"] [] {0,5}	[POS~"^V" & POS != "VINF"]	[] {0,5} [POS="VINF"]

TABLEAU 19. Requête des dépendances à distance en infinitive

2.2.1.3 Bilan

Le tableau suivant récapitule l'ensemble des requêtes automatiques qui ont été soumises aux concordanciers. Nous indiquons précisément sur quels corpus les requêtes ont été appliquées.

[51] Les requêtes peuvent porter sur de nombreux autres enrichissements du moment que les informations sont séparées d'une tabulation.

Nous précisons également les requêtes manquantes afin de donner une vue globale sur le travail effectué.

Lorsque nous avons recherché la totalité des interrogatifs-relatifs, le tableau mentionne « QU ». Ce sigle correspond à la liste suivante : « qui, que|qu' (qu'est-ce que|qu' inclus), dont, où, pourquoi, quand, comment, combien, lequel, laquelle, lesquels, lesquelles, quel, quelle, quels, quelles ». Dans le cas où nous avons limité la requête, les pronoms recherchés sont précisés.

Type	Corpus	Requête	Manque
Complétives	Brassens, OTG, UBS, CPROM, Irit	Lu	
	CERF, CorpAix, CRFP, Nancy, C-Oral-Rom, CTFP, CFPP, PFC	-QU {2,5} qu -QU {2,5} si | s'	
	CERF, CRFP, Nancy, C-Oral-Rom, CTFP, CFPP, PFC	QU {2,5 tiret} qu	CorpAix QU {2,5 tiret} si
	TCOF_narration	Que|qu', où, qu'est-ce que|qu' {2,3} qu	Autres pronoms, Intervalle {4,5}
	TCOF_ conversation	Qui, que, qu'est-ce que|qu' {0,5} qu	Autres pronoms, Locuteurs enfants
Infinitives	CFPP, C-Oral-Rom, CRFP, CTFP, PFC	QU{0,5}[POS~"^V"& POS!="VINF"] {0,5} POS="VINF"]	Autres corpus

TABLEAU 20. Synthèse des requêtes effectuées dans un premier temps

De nombreuses requêtes ont été nécessaires pour repérer les énoncés de dépendances à distance. Pour simplifier ces requêtes, il aurait été intéressant de pouvoir interroger les corpus au moyen de parties du discours et de lemmes. Cela implique bien sûr que les corpus soient étiquetés en catégories morphosyntaxiques et que les catégories correspondent à ce que l'on veut chercher. Nous aurions par exemple souhaité préciser dans la requête la présence

- d'un interrogatif-relatif : 'PRO:Rel' et 'PRO:Int'[52]
- d'un verbe fini : 'V'
- d'une complétive : lemme « que » ou « si » ayant comme étiquette 'KON'
- d'un verbe infinitif : 'Vinf'

[52] En espérant que les concepteurs des étiquettes ne distinguent pas pronom et adverbe interrogatif.

Malheureusement, nous sommes actuellement loin de pouvoir réaliser ce type de requête. Peu de corpus sont étiquetés à l'heure actuelle ; la catégorie « pronom interrogatif » existe rarement ; et le lemme « que », qui peut recevoir plusieurs d'étiquettes (adverbe restrictif, pronom relatif, pronom interrogatif, conjonction de subordination), est sujet à de nombreuses erreurs.

2.2.2 Tri des complétives et des infinitives

Les requêtes destinées à repérer les énoncés de dépendances à distance en complétive, qui étaient très larges, ont renvoyé 9 665[53] résultats pour l'oral sur environ 2 800 000 mots [54] et 11 622 résultats pour l'écrit, sur environ 9 millions de mots (CERF)[55].

Les requêtes destinées à repérer les énoncés de dépendances à distance en infinitive, qui utilisaient les informations de lemmatisation et de catégories syntaxiques, ont renvoyé 9 415 résultats pour environ 1 318 000 mots (correspondant à cinq corpus oraux : CFPP, C-Oral-Rom, CRFP, CTFP et PFC)[56].

Le tri de ces 30 702 résultats a été effectué manuellement. Nous avons écarté assez rapidement les occurrences qui ne présentaient pas de verbe entre les deux bornes recherchées : l'interrogatif-relatif et l'introducteur de complétive. Nous avons également rejeté les énoncés dans lesquels « que », qu'il soit en borne gauche ou droite, n'était pas une forme autonome, mais une partie d'une locution : « parce que, alors que, bien que, tandis que, dès que, pour que, quand même, etc. » ou un comparatif : « pire que, meilleur que, mieux que ». Nous n'avons pas retenu les énoncés dans lesquels le lemme « que » de borne gauche était une conjonction, c'est-à-dire soit coordonné à une autre conjonction « si, quand, parce que, etc. », soit précédé d'un verbe à complétive : « penser, dire/disons, falloir, demander, sentir, vouloir, affirmer, admettre, savoir, montrer, paraître, c'est vrai, c'est sûr, etc. ». Enfin,

[53] Il manque dans ce total les résultats de la requête spécifiant la présence d'un tiret pour CorpAix (requête de complétives)

[54] Nombre de mots des corpus interrogés, soit C-Oral-Rom, CRFP, Nancy, CorpAix, CFPP, CTFP, PFC et TCOF adultes-enfants

[55] Nous percevons nettement grâce à ces chiffres l'utilité qu'aurait un marquage des corpus oraux en grandes unités syntaxiques.

[56] Preuve de l'utilité de l'annotation morphosyntaxique et de la lemmatisation.

nous avons considéré que la présence d'une conjonction de coordination « mais, et, ou, donc » avant la borne gauche ou avant la borne droite était un indice de rejet, même si les dépendances à distance ne sont pas incompatibles avec la coordination.

Dans la suite du tri, plusieurs questions ont été soulevées :

- Par quel verbe l'élément extrait est-il régi ?
- Le pronom est-il un interrogatif-relatif ?

Alors que les objets peuvent être rattachés sans ambiguïté à un des deux verbes en vertu du principe de sous catégorisation, les ajouts ne sont pas sous-catégorisés par le verbe et posent donc des problèmes de rattachement. Dans certains énoncés, le rattachement est clair : rattachement au verbe enchâssé (177), rattachement au verbe pont (178).

*(177) il y a des jours **où** je trouve que les mots en français me **viennent** plus facilement (CorpAix, PORTUGA)*

*(178) il y a des moments où c'est aussi tu tu sens que tu vas vendre il y a des moments **où tu sens** aussi que tu vas p- que tu as perdu l'affaire ou tu tu es en train de la perdre (CorpAix, INTERWY)*

On peut confirmer cette analyse en réorganisant l'énoncé, c'est à dire en déplaçant le verbe pont en incise si le verbe le permet, ou en première position. Si le déplacement du verbe est possible, on peut considérer que l'ajout dépend du verbe enchâssé (179)(180). Dans le cas contraire, il dépend du verbe pont (181).

(179) il y a des jours où les mots en français me viennent plus facilement, je trouve

(180) je trouve qu'il y a des jours où les mots en français me viennent plus facilement

(181) #tu sens qu'il y a des moments où tu as perdu l'affaire

Mais il est parfois difficile d'identifier le verbe recteur et de s'assurer que l'ajout est régi par le verbe enchâssé et non par le verbe pont. La portée semble pouvoir s'étendre sur les deux verbes, notamment dans le cas des clivées en « c'est pour ça que » et « c'est là que » (182).

(182) la langue de la rue est déstructurée par rapport n'a pas de structures par rapport à la langue académique en tout cas /elle,Ø/ en a vachement moins mais c'est justement l'intérêt de la langue de la rue c'est pour ça que je pense qu'elle est euh elle est dans un sens supérieure à l'autre (CorpAix, CYRIL)

La cause (langue déstructurée) implique deux conséquences (fait penser que la langue est supérieure, la langue est supérieure), qui s'influencent (si le locuteur le pense, il est probable qu'elle le soit ; si elle est supérieure, il est logique qu'il le pense).

L'ambiguïté est plus forte encore avec les dépendances à distance en infinitive (183). L'ajout modifie l'ensemble du groupe verbal, qui est peu dissociable. L'action est représentée sous la forme d'un processus et l'ajout porte aussi bien sur l'action (écrire) que sur le moment particulier précisé par l'aspectuel (commencer à écrire).

(183) à la fin du seizième siècle époque à laquelle Lope De Vega commence à écrire (CRFP, PUB-PAU-1)

Il existe un autre cas où l'on peut se demander de quel verbe dépend l'interrogatif. Il s'agit des énoncés en « comment ça se fait que », qui sont assez productifs dans les corpus (184)(185).

(184) comment ça se fait que votre papa est venu (CorpAix, SCA59AB)

(185) comment ça se fait que tu t'es retrouvé sur Nancy (Nancy, activpro)

L'adverbe n'a pas un sens de manière qui pourrait être attribué au premier verbe (*faire comment*) ou au second (*venir comment*). La tournure semble figée. On peut la considérer comme un ensemble qui « interroge sur les conditions dans lesquelles le procès peut avoir lieu » (TLFi, consulté le 29/03/12). Par ailleurs, la formule n'est pas modifiable : on ne peut pas supprimer « comment » (186), ni le remplacer par un élément censé être dans son paradigme, tel que « comme ça » (187) ou « d'un coup » (188). Il n'y a donc pas de dépendance à distance.

*(186) *ça se fait que papa est venu*

(187) #ça se fait comme ça

(188) #ça se fait d'un coup.

La deuxième source d'ambiguïté dans l'identification des constructions de dépendance à distance se situe au niveau des interrogatifs-relatifs. Certaines formes peuvent appartenir à des locutions qui ne sont pas interrogatives ni relatives. Le pronom « où », par exemple, est un élément des locutions conjonctives « à partir du moment où », « dans la mesure où », etc., qui sont dans le paradigme de « quand » ou « comme » (189)(190). Mais on peut se demander

dans quelle mesure la locution « au moment où » est figée (191). Le pronom « où » ne reste-t-il pas en partie un relatif ayant pour antécédent « moment » ?

(189) *je me suis dit putain à partir du moment où ils disent qu'ils appellent ap- qui pourquoi ils ont pas appelé quoi (C-ORAL-ROM, ftelpv12)*

(190) *ça t'apporte des choses dans la mesure où où tu te rends compte que bon euh que malheureusement que malheureusement dans cette re- [inachèvement] (CorpAix, CITU)*

(191) *au moment où je trouvais que c'était euh où ça me prenait vraiment tout mon temps je dépensais trop d'argent dans euh des livres euh des photos et ben je me suis dit quand il[inachèvement] (Nancy, Marais)*

De même, « quand » peut être employé seul ou régir un complément. Voici deux exemples de cette forme :

(192) *vers quand tu penses que ce sera (PFC, 11agm1)*

(193) *Quand il est arrivé je crois qu'il avait une chemise blanche*

Dans le premier cas, il est interrogatif et nous pouvons valider l'énoncé comme construction de dépendance à distance (192). Dans le second, il est conjonction (193). Bien que le groupe conjonctif puisse être interprété comme ajout du V2, nous n'avons pas souhaité conserver l'occurrence afin de garder une certaine cohérence dans l'ensemble des énoncés dans lesquels l'élément antéposé est soit interrogé, soit relativisé, soit clivé. En revanche, lorsque l'ensemble du groupe conjonctif était clivé, nous avons inclus l'occurrence dans notre corpus d'exemples.

Les limites entre les différentes constructions sont donc floues. Nous avons conservé uniquement les occurrences dans lesquelles la construction de dépendance à distance était la plus probable.

2.2.3 Evaluation des requêtes automatiques sur les corpus uniformisés

La première série de requêtes, portant sur les complétives en « que », a été effectuée sur les corpus bruts. L'uniformisation des conventions de transcription a mis en lumière non seulement des différences de conventions (transcription orthographique ou non, codage

métalinguistique) mais également un grand nombre d'erreurs (de transcription, du locuteur mentionné, de découpage en tours, etc.) Nous avons souhaité estimer l'écart entre les données initiales et les données nettoyées et vérifier l'importance du silence des premières requêtes.

L'évaluation a eu lieu sur les trois corpus oraux suivants : CRFP, CFPP et PFC, soit environ 970 000 mots[57]. Nous avons sélectionné ces trois corpus parce qu'ils ont été récoltés par trois équipes de recherche différentes (Aix en Provence, Paris, Toulouse) dans différentes optiques de recherche (phonologie, syntaxe) et qu'ils présentent donc des formats et des conventions de transcriptions variés. En revanche, ils présentent tous trois des enregistrements d'entretiens ou de conversations. Les données de PFC et de CFPP sont en outre comparables par la taille.

Les premières requêtes avaient été lancées pour CFPP sur la version qui ne contenait que 16 transcriptions. L'évaluation a été l'occasion d'effectuer les recherches sur les textes supplémentaires. Néanmoins, les résultats présentés pour l'évaluation compareront uniquement les textes initiaux.

Dans un premier temps, nous avons procédé aux mêmes requêtes (du type « qui \w+ \w+ qu ») avec le même concordancier, Contextes, sur les données nettoyées. Les nouvelles occurrences apparues grâce à la standardisation des données ont révélé plusieurs failles, à la fois dans la requête et dans l'outil. Ces failles ont à leur tour fait l'objet d'une évaluation, lorsque cela était possible.

Dans un deuxième temps, nous avons élargi les requêtes à un empan plus grand. Nous avions considéré jusqu'alors un espace entre l'interrogatif-relatif et la complétive de cinq tokens maximum ; nous l'avons étendu à dix tokens maximum.

2.2.3.1 Requêtes simples, du type « qui \w+ \w+ qu »

Si l'on compare le nombre de résultats correspondant à la requête : « forme + empan de 2 à 5 mots + qu » dans les deux versions des corpus (TABLEAU 21), on voit bien l'importance de l'uniformisation des données. En effet, sur les 4043 chaînes qui répondent au motif dans le corpus uniformisé, seules 64% étaient repérées dans le corpus brut. Des trois corpus, celui qui est le plus modifié par la standardisation est PFC : la requête obtient le double de résultats une fois le corpus nettoyé de toute marque métalinguistique et transféré en orthographe standard.

[57] CFPP 16txt = 280 000 (arrondi de 276 077) + CRFP = 420 000 + PFC = 270 000 mots

Nombre de réponses renvoyées interro {2,5} qu +/-1	Corpus brut	Corpus nettoyé	% d'occurrences supplémentaires
Total	**2590**	**4043**	**36%**
CRFP	630	969	35%
CFPP	1510	2168	30%
PFC	450	906	50%

TABLEAU 21. Comparaison des résultats fournis par le concordancier sur les corpus bruts et standardisés

L'interrogatif-relatif le moins touché par la standardisation est « combien », avec 82% de chaînes repérées dans le corpus brut ; le plus touché est « pourquoi », avec 44% de chaînes repérées (TABLEAU 22).

Nombre de réponses renvoyées interro {2,5} qu +/-1	Corpus brut	Corpus nettoyé
Total	**2590**	**4043**
Qui	831	1419
Que	779	1287
Qu'	563	728
Quand	264	366
Où	53	75
Comment	32	57
Quel-	29	42
Dont	13	21
Pourquoi	11	25
Combien	9	11
l-quel-	6	12

TABLEAU 22. Répartition des résultats par interrogatif-relatif, avant et après standardisation

Sur tous ces résultats supplémentaires, trouvés après standardisation, on relève seulement deux exemples qui présentent une occurrence de dépendance à distance. Les occurrences supplémentaires sont :

(194) je pense qu'effectivement c'étaient des gens croyants et c'étaient des gens qui avaient les moyens d'élever euh sept ou huit enfants et qui avaient des appartements ***dans lequel*** *il faut dire que ces enfants* ***s'entassaient*** *un peu mais c'était à cette époque là c'était normal (CFPP, Audin_7e)*

(195) c'était un quartier qui est bon qui est très vivant très connu mais c'est un quartier ***dans lequel*** *euh je pense pas que* ***j'apprécierais autant de vivre*** *que ici (CFPP, daSilva_7e)*

Le premier exemple (194) n'avait pas été repéré car dans le corpus brut, il comprenait une pause, notée '+', non spécifiée par la requête : *dans lequel + il faut dire que ces*. Ce phénomène aurait pu produire beaucoup de silence dans le relevé des structures mais cela n'a pas été le cas. On peut donc en conclure que les dépendances à distance ne facilitent pas la présence de pauses. Nous pouvons y voir un indice de la cohésion de la construction qui corrobore l'unité prosodique.

Le deuxième exemple (195) était transcrit sous la forme *dans lequel euh j'pense pas que j'apprécierais* dans le corpus brut. Il n'y avait pas de marque, ni de commentaire métalinguistique interrompant les paroles du locuteur. Il semble que ce soit la transformation de « j'pense » en « je pense » qui ait permis de retrouver l'occurrence. Cela a certes été rendu possible par l'uniformisation des corpus mais cela soulève un problème plus large : les chaînes ayant comme séparateur de mots une apostrophe ne sont pas repérées par la requête « \w+ \w+ ». Nous avons rencontré le même problème avec les tirets (voir p.121). Cela signifie que nous n'avons récupéré aucun exemple contenant dans l'empan les formes « j' » (1ère personne sujet), « m' » (1ère personne objet), « n' » (négation), ni « c' », « d' », « l' ».

a) Absence d'apostrophe dans l'empan

Pour pallier ce problème, nous avons utilisé la fonctionnalité de Contextes 'restriction du contexte additionnel', malgré les problèmes rencontrés (voir §2.1.2.1). Nous avons recherché les interrogatifs-relatifs et contraint la présence de « que », « qui » ou « qu' » dans les cinq mots suivants. Cette recherche a été effectuée sur la version la plus aboutie de l'uniformisation. Le TABLEAU 23 présente les requêtes formulées pour évaluer les pertes subies avec « \w+ \w+ ».

Mots recherchés (interrogatifs – relatifs)	Qui, que, qu', dont, où, pourquoi, quand, comment, combien, lequel, laquelle, lesquels, lesquelles, quel, quelle, quels, quelles
Contexte additionnel	qu « Mot+/-1 », dans les 5 mots suivants

TABLEAU 23. Requête souple, permettant notamment de repérer dans l'empan des tokens séparés d'une apostrophe

Ces requêtes ont fourni 1035 résultats dont 154 avec des apostrophes. Parmi ces derniers, deux occurrences présentaient des constructions de dépendance à distance.

corpus	Contexte gauche	Cible de la requête	Contexte droit
CFPP Teixeira_11e	Julie Teixeira non mais c'est ça oui c'est c'est une période	où c'est vrai qu'	on se prend on se prend pas mal la tête pour des choses qu'en valent peut-être pas la peine
CRFP PRI-PER-1	et après on ouvrait les cochons en deux et on découpait les cochons les euh les jambons euh les ventrèches les lards les costello les pieds les têtes tout était découpé et chaque morceau allait là	où c'était prévu qu'	il devait aller on faisait des euh des confits de porc voyez

b) Requête consommante

Le tri de ces résultats a révélé un autre défaut du concordancier : les requêtes sont 'consommantes'. Cela signifie qu'il ne peut pas y avoir de recouvrement dans les chaînes extraites. Si « qu » est repéré dans un premier temps comme marque de complétive (2ème forme du motif dans la requête), il ne peut pas ensuite être repéré comme interrogatif-relatif (1ère forme du motif). Ainsi, dans (196), trois chaînes devraient correspondre au motif de la requête. Elles sont présentées dans le tableau suivant (TABLEAU 24).

(196) c'est là où on construit vraiment des enfin des amitiés qu'on enfin ***qu'****on sait qu'on* ***gardera*** *après ou qu'on fait des rencontres (CFPP, Teixeira_11^e^)*

Contexte gauche	Cible attendue	Contexte droit
c'est là où on construit vraiment des enfin des amitiés	qu'on enfin qu'	on sait qu'on gardera après ou qu'on fait des rencontres
c'est là où on construit vraiment des enfin des amitiés qu'on enfin	qu'on sait qu'	on gardera après ou qu'on fait des rencontres
c'est là où on construit vraiment des enfin des amitiés qu'on enfin qu'on sait	qu'on gardera après ou qu'	on fait des rencontres

TABLEAU 24. Résultats attendus pour la requête « qu' {2 mots} qu » sur l'exemple (196)

Mais la deuxième chaîne n'est pas repérée car le « qu' » précédant « on sait » est déjà consommé par la première chaîne. Sur ces trois résultats, seuls le premier et le troisième sont proposés par Contextes. L'existence de cette faille dans le concordancier implique que toutes les dépendances à distance présentes dans les corpus n'ont certainement pas été repérées. Il nous était difficile de combler ce problème et donc de l'évaluer. Néanmoins, s'il est probable

que certaines occurrences nous aient échappé, il arrive aussi que la deuxième chaîne, dans le contexte de la première, ait été remarquée lors du tri.

L'évaluation a permis de trouver 5 occurrences supplémentaires, ce qui est peu en soi. Ces occurrences constituent tout de même 12% des résultats. Après considération de la part de gain par rapport au temps demandé par les requêtes et le tri, nous avons décidé de ne pas mettre à jour les requêtes manquantes dans l'ensemble des corpus (apostrophe-si), ni les requêtes manquantes pour les corpus non évalués (requête_que, apostrophe_que). Rappelons que, de même, nous n'avons pas recherché les tirets pour toutes les requêtes et tous les corpus (voir le TABLEAU 20).

2.2.3.2 Variation de la taille de l'empan (jusqu'à 10 mots)

Dans un premier temps, nous avons limité l'empan entre l'interrogatif-relatif et la conjonction de subordination à cinq mots maximum (toutes les requêtes présentées jusqu'à présent). Ce choix va de pair avec l'idée développée par (Hofmeister & Sag, 2010) selon laquelle la production est limitée par la mémoire et par la complexité du *process*. La construction étant complexe, il est probable que la distance entre le complément antéposé et le verbe recteur soit courte afin de permettre à l'auditeur de garder en mémoire le premier élément et de le relier avec le second. Les premiers résultats ont semblé corroborer cette hypothèse : plus la distance entre le pronom et la complétive est grande, plus les occurrences de dépendance à distance se raréfient, comme l'illustre la (FIGURE.17).

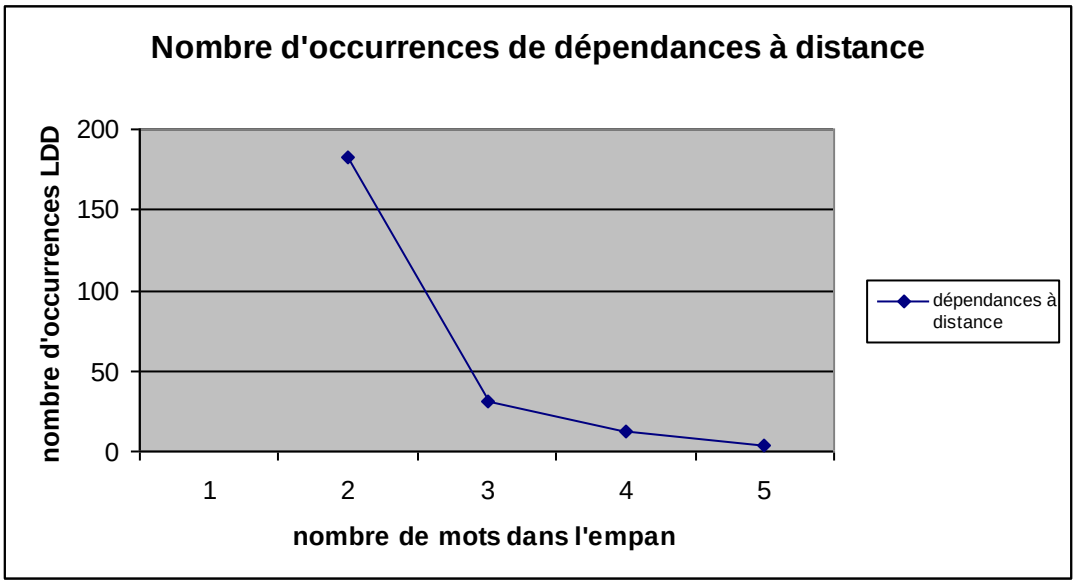

FIGURE.17 Nombre de dépendances à distance repérées avec un empan de n mots entre le pronom et la complétive

Néanmoins, on aurait pu imaginer un second pic lors d'un empan beaucoup plus grand, correspondant par exemple aux occurrences avec une incise, ou avec un ajout de taille importante[58].

(197) Il viendra te rendre visite à Paris où il a pourtant regretté la semaine dernière, dans une lettre émouvante, que tu déménages.

(198) On avait empoisonné la pomme que, fort heureusement, la veille, il avait vérifié que tu n'avais pas mangé.

(199) Il rédigea une lettre que, parce que le chaos continuait dans le pays, il espérait qu'elle recevrait avant qu'il ne soit trop tard.

Pour vérifier cette hypothèse et découvrir un éventuel biais à la requête, nous avons recherché les occurrences de dépendances à distance avec un empan de dix mots, en utilisant le concordancier CooLox sur six corpus étiquetés avec Cordial (corpus préalablement uniformisés) : CFPP, C-Oral-Rom, CorpAix, CRFP, CTFP, PFC et sur quatre tranches d'écrit du CERF : Littérature, Presse, Politique et Divers.

Etant donné le grand nombre de résultats, nous avons restreint la vérification au pronom « que », incluant « qu'est-ce que ».

[58] Les exemples suivants ont été soumis au jugement de quelques locuteurs natifs. Voici une partie de leurs commentaires : « c'est pas trop français je trouve. Disons carrément lourd », « Je les trouve toutes horriblement pas françaises !!! »

Le classement des trois exemples, du plus acceptable au moins acceptable, est également variable selon les locuteurs sondés. [1, autres / 2,1,3 / 1,3,2 / 1,2,3]

La requête effectuée, donnée ci-dessous (TABLEAU 25), signifie que nous avons recherché :

- le lemme « que », suivi d'
- un empan allant de zéro à dix mots quelconques, suivi d'
- un verbe conjugué, suivi d'
- un empan allant de zéro à dix mots, suivi du
- lemme « que ».

La recherche est stoppée lorsque l'ensemble de la séquence qui correspond à la requête dépasse 10 mots. Les chevrons donnent l'indication à l'outil de rechercher en priorité ce qui se trouve à l'intérieur.

Pour les corpus écrits, on peut spécifier l'absence de ponctuation forte (point, point d'interrogation, point d'exclamation, etc.) dans les empans.

Requêtes	Résultats	Corpus
[lemme="que"] []{0,10} <[POS~"^V" & POS != "VINF"]> []{0,10} [lemme="que"] stop[(end.index-begin.index)>10]	5264	Oral
[lemme="que"] [POS != "PCTFORTE"]{0,10} <[POS~"^V" & POS != "VINF"]> [POS != "PCTFORTE"]{0,10} [lemme="que"] stop[(end.index-begin.index)>10]	4189	Ecrit
	9453	**Total**

TABLEAU 25. Requêtes fournies à CooLoX et nombre de résultats obtenus

Les résultats étant encore très nombreux, nous nous sommes limités à la vérification de deux lemmes verbaux fréquents dans les occurrences de dépendance à distance : « vouloir » et « penser ».

Les requêtes soumises ont été du même ordre que celles exposées ci-dessus, si ce n'est que nous avons recherché un lemme particulier et plus seulement un verbe conjugué. Le tableau suivant indique le nombre de résultats obtenus pour chaque verbe, pour les corpus écrits et oraux réunis (TABLEAU 26) :

Requêtes	Résultats
[lemme="que"] []{0,10} <[lemme="penser"]> []{0,10} [lemme="que"] stop[(end.index-begin.index)>10]	**271**
[lemme="que"] []{0,10} <[lemme="vouloir"]> []{0,10} [lemme="que"] stop[(end.index-begin.index)>10]	**294**

TABLEAU 26. Requêtes avec un verbe spécifique fournies à CooLoX et nombre de résultats obtenus

Pour *vouloir*, on ne relève aucune occurrence de dépendance à distance avec un empan plus grand que 5 mots, ni à l'écrit, ni à l'oral. Pour *penser*, on relève une seule occurrence qui pourrait correspondre à une construction de dépendance à distance à condition de poser que c'est la transcription en « ait » qui corresponde à ce qui a été dit[59] :

> *(200) L1 ouais parce que la confiance euh elle y est mais ils peuvent pas savoir réellement si l'enseignant euh détient réellement tous le(s) savoir(s) qui qu'ils pensent euh L2 hum hum L1 qu'il /est,ait/ (CorpAix, FLE)*

Nous pouvons observer que la distance est produite par l'intervention de l'interlocuteur, qui acquiesce. Il n'y a donc pas réellement d'occurrences de dépendance à distance avec une distance très importante entre le pronom et la complétive. Cela confirme les appréciations des locuteurs sondés sur les exemples construits.

Le test, effectué sur une partie des corpus et limité à quelques lemmes (pronominal et verbal), indique fortement que la requête initiale semble satisfaisante. La restriction de l'empan à cinq mots n'a pas dû occasionner beaucoup de perte d'occurrences de dépendance à distance.

2.2.4 Synthèse des requêtes effectuées

Le tableau suivant récapitule l'ensemble des requêtes automatiques qui ont été soumises aux concordanciers, avant et après évaluations des premières requêtes (TABLEAU 27). Comme dans le §2.2.1.3, le symbole « QU » indique que l'ensemble des interrogatifs-relatifs a été recherché : « qui, que|qu' (qu'est-ce que|qu' inclus), dont, où, pourquoi, quand, comment, combien, lequel, laquelle, lesquels, lesquelles, quel, quelle, quels, quelles ».

Type	Corpus	Requêtes	Manque
Complétives	Brassens, OTG, UBS, CPROM, Irit	Lu	
	CERF, CorpAix, CRFP, Nancy, C-Oral-Rom, CTFP, CFPP, PFC	QU {2,5} qu QU {2,5} si | s'	
	CERF, CRFP, Nancy, C-Oral-Rom, CTFP, CFPP, PFC	QU {2,5 tiret} qu	CorpAix QU {2,5 tiret} si
	CRFP, CFPP, PFC (uniformisé)	Que|qu', qui {0,5} qu => apostrophe et autres signes	Autres corpus Autres pronoms (Apostrophe) + si

[59] Dans le doute, nous n'avons pas conservé cet exemple.

	CFPP, C-Oral-Rom, CorpAix, CRFP, CTFP, PFC, CERF : Littérature, Presse, Politique, Divers	Qu'est-ce que|qu' + vouloir|penser + qu (max 10 mots)	Autres corpus Autres pronoms
	TCOF_narration	Qu'est-ce que|qu', que|qu', où, {2,3} qu	Autres pronoms Intervalle {4,5}
	TCOF_ conversation	Qu'est-ce que|qu', que, qui {0,5} qu	Autres pronoms Locuteurs enfants
Infinitives	CFPP, C-Oral-Rom, CRFP, CTFP et PFC	QU{2,5}[POS~"^V"& POS!="VINF"] {2,5}[POS ="VINF"]	Autres corpus

TABLEAU 27. Synthèse de la totalité des requêtes effectuées

Ce sont ces requêtes qui ont permis de relever dans les corpus les occurrences de dépendances à distance à l'origine de la description proposée dans les chapitres suivants.

2.3 Synthèse des difficultés d'une étude sur corpus

Pour établir une liste d'attestations d'un phénomène, il ne suffit pas de télécharger des corpus et de les interroger avec une expression régulière. La tâche est beaucoup moins triviale. Elle demande du temps et de la réflexion. On peut résumer les problèmes rencontrés de la manière suivante.

Une fois l'accès aux corpus obtenu, la diversité des corpus peut être source de difficultés, car elle n'est pas prévue par les outils informatiques. Nous avons vu précédemment (§2.1.3) les obstacles liés à la diversité des conventions (polysémie des symboles, variation orthographique d'une forme, utilisation ou non de balises). Il faut y ajouter les problèmes posés par les outils d'interrogation et de traitement des données.

Les concordanciers ne sont pas toujours adaptés aux données : ils ne sont pas capables de lire tous les formats de fichiers ni d'interroger la totalité des informations fournies[60]. Peu d'outils tirent profit du balisage, par exemple, alors qu'il serait intéressant de définir des couches d'annotation, chacune balisée, et de pouvoir les repérer ou les interroger. Il faut également se méfier du langage de requête, et penser à toutes les possibilités avant de déterminer quelle requête soumettre. Dans les expressions régulières, le symbole « \w+ » correspond à une lettre

[60] L'objectif du projet ANR ORFEO est de répondre à ces deux problématiques : donner accès à de multiples corpus dans leur version originale et dans une version homogénéisée ; proposer des outils adaptés aux données.

(accentuée ou non), un chiffre ou un tiret bas « _ », présent une ou plusieurs fois. Il est donc pratique pour repérer un mot. Mais repérer une suite de mots avec la séquence « \w+ \w+ » est trop restrictif, car cette dernière élimine deux mots séparés par une apostrophe ou un tiret. Il faudrait insérer un « \W », qui correspond à tous les caractères non compris dans « \w » et soumettre la requête suivante : « \w+\W\w+ ».

Enfin, l'annotation morphosyntaxique des corpus permettrait de cibler de façon beaucoup plus précise les constructions recherchées. S'il existe encore peu d'outils fiables permettant d'étiqueter les corpus oraux, plusieurs sont en cours de développement, avec des apprentissages également basés sur l'oral (Benzitoun, Fort, & Sagot, 2012; Eshkol et al., 2010; Tellier et al., 2010). Leur diffusion devrait améliorer les conditions de travail de la recherche en linguistique de corpus.

Pour ce travail, toutes ces étapes ont été réalisées au fur et à mesure de la rencontre des problèmes. Elles ont débuté après les requêtes et le tri des données obtenues, ce dernier ayant mis en évidence certains problèmes (l'absence de tiret par exemple). Les étapes de l'uniformisation n'ont pas été réalisées dans l'ordre présenté ci-dessus. Le format des données (Transcriber) a été harmonisé de manière parallèle à l'uniformisation des conventions de transcritption. De plus, l'exécution a parfois été expérimentale, opérée avec peu de moyens et peu d'outils. Le résultat n'est sans doute pas idéal. Il est donc difficile d'estimer le temps qui serait gagné par la réalisation du processus avant d'entamer les requêtes. Néanmoins s'appuyer sur des données que l'on sait propres est rassurant.

Je présenterai brièvement ici une expérience qui prouve l'intérêt d'avoir des corpus annotés en morphosyntaxe. Cette expérience s'appuie sur le corpus construit pour servir de référence à l'apprentissage de TreeTagger (voir §2.1.4). Ce corpus, étiqueté par TreeTagger puis entièrement relu et corrigé de façon manuelle est constitué de plus de 100 000 mots, extraits du corpus adulte de TCOF. Son format tabulaire est lisible par le concordancier CooLoX. Grâce à l'étiquetage morphosyntaxique, une seule requête permet de rechercher à droite du verbe (éventuellement verbe pont) toutes les complétives, en « que » et « si », et à gauche du verbe la présence d'un interrogatif ou d'un relatif. Le concordancier renvoie 161 occurrences. Si l'on compare ces résultats avec les réponses obtenues pour des corpus de même genre, les corpus oraux, mais non étiquetés (CorpAix, CRFP, Nancy, C-Oral-Rom, CTFP, CFPP, PFC),

le gain est vraiment important. En effet, nous avions obtenu 9 393 résultats pour 2 700 000 mots, soit environ 355 résultats pour le même nombre de mots que le corpus annoté. Le nombre de résultats renvoyés est donc diminué par deux grâce à l'étiquetage morphosyntaxique[61].

Nous avons poussé plus loin l'expérience, en effectuant une recherche sur un corpus annoté non seulement en morphosyntaxe mais également en dépendance : le corpus arboré, également nommé « French Treebank » (Abeillé, Clément, & Toussenel, 2003). Dans ce corpus écrit contenant environ 700 000 mots, tiré d'extraits du quotidien *Le Monde*, publié entre 1989 et 1993, chaque token est annoté en lemme, en partie du discours, en traits morphologiques (genre, nombre, etc.), en dépendance et en fonction. Les dépendances à distance ne sont pas annotées dans le corpus, comme l'atteste Abeillé[62].

Les relatives, qu'elles soient simples ou complexes, sont analysées de la façon suivante : l'antécédent (nom ou pronom) gouverne le verbe de la relative, duquel dépendent tous les compléments, soit, dans le cas d'une relative objet, le sujet, le pronom relatif objet, et éventuellement la conjonction de subordination qui introduit la complétive. La représentation de « l'homme que je pense que j'ai vu » est donc la suivante (FIGURE.18).

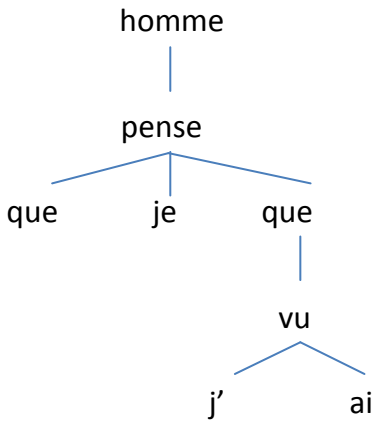

FIGURE.18 Réseau de dépendance d'une relative complexe dans le French TreeBank : *homme que je pense que j'ai vu*

[61] La comparaison n'est pas tout à fait exacte car le résultat des requêtes effectuées sur les corpus non annotés comporte les occurrences de clivées, contrairement à l'autre résultat.

[62] Voir le site http://www.llf.cnrs.fr/Gens/Abeille/French-Treebank-fr.php

Pour trouver les relatives complexes, il faut rechercher un verbe qui gouverne à la fois un pronom relatif (noté « PROREL »)[63] et une conjonction de subordination (dont les lemmes sont « que » et « si »). Pour les interrogatives, le principe est le même pour les adverbes et les pronoms (notés ADVWH et PROWH) : *où tu penses que*[64]. En ce qui concerne les déterminants et les adjectifs (notés respectivement DETWH et ADJWH), la tâche est plus complexe puisque plusieurs liens les séparent du verbe : nom, éventuellement préposition, puis enfin le verbe ou le participe passé. Dans l'exemple (201), « quel » dépend du nom (« sens »), qui dépend de la préposition (« dans »), qui permet à son tour d'atteindre le verbe (« fonctionnent »).

(201) Dans quel sens fonctionnent …

Mais un adjectif interrogatif peut également être directement relié au verbe (202) tout comme il peut uniquement dépendre d'un nom (203).

(202) Quelle sera votre attitude ?

(203) Quels partenaires ?

Un programme en langage python a permis d'effectuer les recherches appropriées. Il a sélectionné 66 phrases pour les relatives et 3 pour les interrogatives, soit 69 occurrences pour environ 360 000 mots.

La même recherche (augmentée de celle des clivées) effectuée sur le corpus brut écrit du CERF avait renvoyé 11 722 occurrences pour 9 millions mots. Le nombre de résultats est donc diminué par plus de six grâce à l'annotation en parties du discours, en lemme et en dépendance, et ce bien que les dépendances à distance ne soient pas annotées[65].

Cette étude montre donc que l'aide à la recherche de séquences pertinentes par des logiciels n'est efficace que si l'on conjugue plusieurs modes de requêtes : expressions régulières, analyse en catégories et analyse en dépendances. Tout chercheur devra cependant déterminer

[63] Ou bien un élément qui lui-même gouverne le pronom. Dans *l'homme à qui il parle*, 'qui' dépend de 'à' qui dépend de 'parler'.
Le relatif et la conjonction sont gouvernés par le verbe ou par le participe passé, si le verbe a une forme composée.

[64] La représentation des interrogatives complexes s'éloigne du modèle en dépendances qui ne voit dans ces structures qu'un problème topologique (voir §1.1.2).

[65] Parmi les résultats obtenus dans le French Treebank, nous n'avons relevé qu'une occurrence de dépendance à distance, dans le fichier flmf7ab2ep.xml : *Le public plébiscite la télévision de son choix. Qui n'est pas toujours celle que l'on croit, ni celle que l'on voudrait qu'elle soit*

la part de chaque outil en fonction de son objet d'étude. Il ne semble pas y avoir de procédure unique applicable dans tous les cas.

2.4 Synthèse sur les occurrences de dépendance à distance repérées

A la suite de la fusion et de l'harmonisation minutieuse des corpus, des nombreuses requêtes, et d'une sélection longue et parfois difficile, nous avons réuni les occurrences comportant des dépendances à distance, en structure finie et infinitive. Ces occurrences sont la base de notre travail de description.

2.4.1 Occurrences en complétive

En intégrant les occurrences repérées grâce aux évaluations de qualité des requêtes, nous avons obtenu un corpus d'attestations de 279 occurrences de constructions de dépendance à distance en complétive.

Dans ces données, certaines occurrences ne sont pas du français contemporain. Dans le corpus CERF, l'une des tranches de l'écrit est en effet consacrée à la littérature ancienne (LITA). Celle-ci comporte 45 occurrences de notre construction. Dans d'autres tranches, on note l'apparition d'occurrences qui appartiennent à d'autres variétés du français contemporain : français de Suisse, d'Afrique, du Québec. Ces occurrences, même si elles ressemblaient beaucoup aux occurrences du français de France, ont été écartées pour ne pas ajouter de complexité à la description. En outre, étant peu nombreuses, elles ne permettaient pas de décrire l'ensemble du phénomène dans l'espace francophone. Voici reproduits les quelques exemples francophones écartés :

(204) *Rosalie : **Qu'est-ce que** tu veux que je te **dise** ? (CERF, LITC) [théâtre, Québec]*

(205) *Émile : **Qu'est-ce que** tu veux que je **fasse**, y'est deux fois plus gros que moi (CERF, LITC)[théâtre, Québec]*

*(206) En France, on milite pour l'affichage. Mais **comment** voulez-vous que le commerçant qui **vend** des corn-flakes sache si le producteur a mis dans sa boîte du maïs modifié (CERF, PRES)[Suisse]*

*(207) Et, le lièvre sortit fièrement de sa cachette : "Je vous l'avais dit ; **comment** voulez-vous qu'un candidat incapable de se retenir **puisse valablement défendre** vos intérêts ? " (CERF, PRES)[Afrique]*

Enfin, nous avons voulu éviter de décrire des occurrences traduites, dont la structure et le lexique étaient certainement influencés par la langue source. Par sécurité, nous avons écarté un énoncé qui pourrait bien être le fruit d'une traduction (208).

*(208) Max van den Berg, Pays-Bas : Nous devons expliquer **pourquoi** nous pensons que la politique étrangère, les politiques liées à l'immigration, à la justice **doivent être harmonisées**. Il faut montrer que l'intégration européenne n'est pas un but en soi, mais sert l'intérêt général (CERF, POLI)*

Toutes les occurrences écartées sont présentées Annexe 3.

Ainsi, nos analyses, sauf mention contraire, se sont appuyées sur 229 occurrences de dépendance à distance en complétive. Le paragraphe suivant présente la constitution du corpus d'exemples des dépendances à distance en infinitive.

2.4.2 Occurrences en infinitive

Les occurrences en infinitive sont très nombreuses, et comme nous l'avons mentionné (voir §2.2.2), assez ambiguës.

La première source d'ambiguïté vient des ajouts. Puisque le complément n'est pas sélectionné lexicalement et sémantiquement par le verbe, il est difficile d'affirmer s'il y a une dépendance à distance ou non. Dans le premier cas, il faudrait certifier que l'ajout dépende du V2 ; dans le second, que l'ajout dépende du V1. Or généralement, on considère que l'ajout a une portée large et qu'il modifie l'ensemble de l'énoncé. Cette difficulté a été soulignée par (Hofmeister & Sag, 2010) :

> « The theoretical syntactic and semantic literatures suggest that adjuncts, including temporal adjuncts, may modify entire CPs or IPs (Hitzeman 1997 ;

> Ernst 2002[66]). Such research supports the idea that a dislocated adjunct could be retrieve at the boundary of a clause to modulate its entire meaning. [...] Nevertheless, additional research is clearly needed to determine when stored information related to adjuncts is generally retrieved, particularly when the adjunct modifies an embedded constituent. » (Hofmeister & Sag, 2010, p. 52)

Afin de ne conserver que les occurrences avec une dépendance à distance clairement identifiable, nous n'avons retenu que les occurrences dans lesquelles l'élément mis à distance était dans la valence du verbe. Même avec cette restriction importante, les résultats restent nombreux et soulèvent d'autres questions.

Le second problème pour définir les dépendances à distance en infinitive vient des relatives en « qui » (209). Les énoncés de ce type soulèvent les questions suivantes : peut-on extraire le sujet d'un infinitif ? L'infinitif possède-t-il un sujet ?

(209) Laurent Blanc pour Zidane maintenant dans le rond central Zidane ***qui tente de pousser*** *oui bon tacle là (C-ORAL-ROM, fmedsp03)*

Selon le cadre théorique adopté, ces occurrences peuvent être considérées comme des dépendances à distance. En effet, pour les génératives, les constructions verbales sont toutes identiques, qu'elles soient à l'infinitif ou finies. Les constructions verbales à l'infinitif comportent un noeud sujet, qui est vide, mais marqué d'une coréférence à un élément précédemment exprimé. Ainsi, *Pierre veut travailler* serait la transformation de *Pierre veut que Pierre travaille* (Ruwet, 1972, p. 292[67] dans (Lablanche, 2007, p. 36)). La preuve de l'existence du sujet de l'infinitif serait l'accord produit avec l'agent de l'infinitif :

(210) Il demande à Marie d'être ***parfaite***

(211) Il la remercie d'être ***venue***

L'observation des différentes constructions à l'infinitif invite (Lablanche, 2007, p. 38-39) à distinguer trois types de sujets de l'infinitif : le sujet propre (212)(213), le sujet coréférentiel (214) et le sujet extralinguistique (215). Le sujet propre est différent du sujet du verbe fini. Il se situe avant l'infinitif s'il est réalisé sous forme lexicale et avant le verbe fini s'il est réalisé

[66] Hitzeman, J. (1997) Semantic partition and the ambiguity of sentences containing temporal adverbials. *Natural Language Semantics* 5, 87–100

Ernst, T. B. (2002) *The syntax of adjuncts*. Cambridge: Cambridge University Press.

[67] Ruwet N., 1972, Théorie syntaxique et syntaxe du français, Seuil : Paris.

sous forme pronominale. Il peut également se trouver après l'infinitif si aucun complément n'est produit. Le sujet coréférentiel, comme son nom l'indique, est identique au sujet du verbe fini. Il n'est pas répété. Le sujet extralinguistique n'est pas présent dans l'énoncé, mais il n'est pas non plus coréférent à un autre élément.

(212) Jean regarde les avions atterrir

(213) Je regarde tomber la pluie

(214) Jean veut partir / Jean aime manger du chocolat (Le Bidois et Le Bidois, 1967[68], p. 307, dans (Lablanche, 2007, p. 35))

(215) Jean a entendu sonner / Jean fait repasser sa chemise

Le sujet extralinguistique rend l'argumentation du sujet de l'infinitif peu crédible. Pour (Riegel et al., 1994[69], dans (Lablanche, 2007, p. 36)) il ne s'agit pas dans ce cas d'une 'proposition subordonnée infinitive' mais d'une 'construction infinitive'. Outre le problème du sujet extralinguistique, qui ne fournit pas un sujet à l'infinitif, d'autres arguments remettent en cause le fait qu'un infinitif ait un sujet. D'abord, l'infinitif ne possède pas de marque de personne. Ensuite, il existe en français des énoncés dans lesquels l'analyse la plus logique est qu'un accord se produise entre deux objets associés (216). La même analyse pourrait être proposée pour les exemples (210)(211).

(216) Je la trouve intéressante

D'autres linguistes arrivent à une conclusion différente en observant les énoncés avec infinitive. Pour (D. Denis & Sancier-Chateau, 1994), l'infinitif est un fait, un procès non actualisé, qui n'a donc pas de sujet :

> « S'il appartient bien à la conjugaison du verbe, il ne présente du procès que sa pure image virtuelle, sans le situer dans le monde actuel, c'est-à-dire sans le rattacher explicitement à un support sujet (puisqu'il ne connaît pas la flexion personnelle) ni à la temporalité. On dira donc que l'infinitif n'actualise pas. Il laisse au contraire le procès verbal dans sa plus grande virtualité, un peu à l'image du nom employé sans déterminant » (D. Denis & Sancier-Chateau, 1994)

[68] Le Bidois G., Le Bidois R. (1967) Syntaxe du français moderne. Ses fondements historiques et psychologiques... 1, Prolégomènes. Les articles. Les pronoms. Théorie générale du verbe. Les voix. Le temps. Les modes, 2e éd. rev. et compl., Paris : A. Picard

[69] Riegel M., Pellat J.C, et Rioul R. (1994) *Grammaire Méthodique du français*, PUF, Paris.

Cette analyse est confortée par des exemples qui ne supportent aucune reformulation avec un sujet (217). L'infinitive serait plutôt à mettre sur le même plan qu'un groupe nominal (218).

*(217) nous voulons des hommes politiques responsables bien formés compétents honnêtes et qui sachent travailler en équipe en relation avec leurs électeurs (C-ORAL-ROM, fnatps01) -> *des hommes qui sachent qu'ils travaillent en équipe*

(218) il y en a qui n'accepteront pas de le faire (CRFP, PRO-LYO-1) // y en a qui n'accepteront pas la critique

Pour nous, le problème est plutôt ici celui de l'interprétation de l'agent du verbe à l'infinitif qu'un problème de structure syntaxique.

Enfin, même si l'on accepte l'idée d'un sujet de l'infinitif, qui serait extrait dans les relatives en « qui », doit-on considérer qu'il y a une dépendance à distance alors que le pronom est sujet du verbe pont (219) ? Et doit-on considérer qu'il y a une dépendance à distance lorsque le verbe pont est un verbe syntaxiquement transparent, un verbe à montée (220) ? Doit-on dans un premier temps définir la classe syntaxique du V1 afin de déterminer s'il y a une dépendance à distance ?

(219) c'est Bixente Lizarazu qui s'apprête à effectuer cette remise en jeu (C-ORAL-ROM, fmedsp03)

(220) la pluie qui semble tomber

Au vu de tous les arguments exposés, nous considérons que l'infinitif ne comporte pas de sujet et qu'il ne peut donc pas y avoir de dépendance à distance sujet. Nous avons donc écarté tous les énoncés avec relativisation sujets en « qui ».

Parmi les occurrences en infinitive repérées et triées, se trouve une occurrence particulière : la dépendance à distance est construite par une interrogative indirecte (221).

(221) on va les évacuer pour les remplacer par des professionnels qu'on ne sait pas où trouver (CRFP, PUB-POI-1)

C'est la seule occurrence qui a été relevée dans l'ensemble du corpus. Nous l'avons mise de coté, pour plusieurs raisons. D'une part, du fait de sa construction typique d'interrogative indirecte, il était délicat de la mélanger aux autres occurrences. D'autre part, puisque l'occurrence est unique, il n'aurait pas été possible d'effectuer des comparaisons. Enfin, les

interrogatives indirectes n'ont pas été recherchées pour les complétives. Par souci d'homogénéité, nous l'avons donc écartée. Mais nous sommes conscients de l'intérêt de la structure et nous nous proposons de l'étudier spécifiquement par la suite.

Ainsi, les occurrences sur lesquelles ont porté les analyses ont été restreintes aux corpus oraux, aux antépositions des objets du V2 et aux infinitives directes ou prépositionnelles. Malgré ces limitations, elles sont beaucoup plus nombreuses que celles en complétive et représentent 1234 occurrences.

CHAPITRE 3. DESCRIPTION DES CONSTRUCTIONS DE DEPENDANCE A DISTANCE A VERBE FINI[70]

Ce chapitre est consacré à l'analyse des 229 occurrences de dépendance à distance relevées dans les corpus écrits et oraux dont le verbe pont construit une complétive (Annexe 4). Compte tenu du faible nombre d'exemples, nous nous appuierons dans un premier temps sur les observations précédentes : ainsi nous examinerons si le français possède une construction prototypique de dépendance à distance, comme il semble qu'il y en ait une dans les interrogatives anglaises et néerlandaises : « Wh- do you think/say that S », « Wh- denk- P2 dat » (voir §1.3.5.1 et §1.3.5.2). Nous pourrions ainsi d'emblée conclure que la faible fréquence d'usage découlerait d'un statut structural de construction contrainte lexicalement et syntaxiquement. Si la construction se révèle structurellement ouverte, c'est-à-dire si le sujet et le verbe pont sont des variables dans le schéma « wh-do N V that S », nous vérifierons les contraintes proposées dans la littérature pour régler un processus général : les contraintes syntaxiques, sémantiques et pragmatiques. Nous essayerons de déterminer comment s'établissent la liste des verbes ponts et la liste des verbes non-ponts.

Dans un deuxième temps, nous décrirons la syntaxe des verbes ponts de nos occurrences, de la manière la plus détachée possible de tout cadre théorique. Cette description sans a priori nous permettra de croiser nos résultats avec les contraintes cognitives établies dans la littérature et de déterminer si le verbe pont se comporte comme n'importe quel verbe. Enfin, nous tenterons de définir si un contexte favorise l'apparition de dépendance à distance.

[70] Une partie de ce chapitre a fait l'objet de communications :

Bérard L. (2011) *Les contrôles à distance : Les conclusions formulées pour l'anglais et le néerlandais sont-elles valables en français ?* Communication présentée à la journée d'étude « Langage » de l'ED LTS, Nancy, 19 janvier.

Bérard L. (2011) *Corpus et relation lexique-grammaire dans les contrôles à distance*. Communication présentée au colloque AFLS 2011 « Regards nouveaux sur les liens entre théories, méthodes et données en linguistique française », Nancy, 8-10 septembre.

3.1 Construction prototypique en français ?

Sur les 229 occurrences de dépendance à distance repérées dans les corpus oraux et écrits, 92 occurrences présentent une structure en interrogative directe (structure étudiée par (Dąbrowska, 2008) et (Verhagen, 2006)). La construction de dépendance à distance la plus fréquente en interrogative directe est composée d'un pronom interrogatif, d'un pronom personnel sujet à la deuxième personne, du verbe « vouloir » au présent de l'indicatif et de la conjonction « que ».

(222) Qu'est-ce que tu veux que je t'explique ?

L'ordre sujet-verbe peut être canonique ou inversé dans ces interrogatives. Les énoncés qui présentent l'ensemble de ces éléments représentent 72% des occurrences de dépendance à distance en interrogative directe.

WH- P2 'veux/voulez' que	66 occurrences	**72%**

Dans environ 14% des occurrences, la construction la plus fréquente subit une seule modification : celle-ci touche soit le mode du verbe (223), soit la forme du sujet (224), soit le lexique verbal (225).

*(223) Quand vous allez chez le coiffeur, vous espérez tous que ce sera un professionnel qui vous coupera les cheveux... Pourquoi voud**riez**-vous que ce soit des citoyens, la plupart du temps désinvestis politiquement, qui décident de l'avenir d'une nation ? ! ; -) (CERF, FORM)*

*(224) que veut-**il** que je fasse (Nancy, Séminaire)*

*(225) quand tu **penses** que cela a permis à la langue française de se développer par la suite (CorpAix, 32VERLAN)*

Néanmoins, cette construction dominante subit dans presque autant d'occurrences plusieurs modifications simultanées. Il peut s'agir de deux modifications qui touchent le même mot : lexique et morphologie du verbe (226), ou bien de deux modifications qui touchent plusieurs éléments de l'énoncé : soit la forme du sujet et le lexique verbal (227), soit la forme du sujet

et la morphologie verbale (228). Enfin, il arrive qu'il y ait trois modifications par rapport à la forme la plus fréquente : celle du sujet, du lexique verbal et du temps du verbe (229).

*(226) pourquoi tu **as dit** qu'il n'a pas eu une mort extraordinaire (Nancy, Marais)*

*(227) qui **il vaut mieux** que je voie (Nancy, assoc)*

*(228) Qu'est-ce qu'**il** voul**ait** que je fasse (CERF, LITC)*

*(229) dans quel temps **va falloir** que je mette de cette phrase (CorpAix, 6CELINE)*

Le TABLEAU 28 présente la répartition des énoncés par type de modifications :

WH- P2 'veux/voulez' que	Présent, indicatif, pas de négation	66 occ **72%**
Une modification	conditionnel : 1 occ autre sujet: 2 occ **autre verbe : 10 occ**	13 occ **14%**
Plusieurs modifications	temps + verbe + sujet : 2 occ **verbe + sujet : 9 occ** verbe + temps : 1 occ temps + sujet: 1 occ	13 occ **14%**

TABLEAU 28. Répartition des occurrences d'interrogatives directes entre la forme la plus fréquente et les formes subissant une ou plusieurs modifications

La modification la plus fréquente touche le lexique verbal (22 occurrences sur 26) c'est à dire l'élément qui constitue le cœur de la construction. On peut donc se demander s'il y a réellement une construction prototypique en français, surtout lorsqu'on compare la répartition des formes en français et en anglais[71] (TABLEAU 29). En français, la construction la plus fréquente est moins dominante et il y a beaucoup plus de modifications multiples.

	Français	**Anglais**
Construction dominante, Patron ?	72%	82%
Une modification	14%	15%
Plusieurs modifications	14%	3%

TABLEAU 29. Comparaison des interrogatives directes en français et en anglais

[71] Les calculs pour l'anglais sont le résultat d'une moyenne effectuée sur les corpus BNC_oral, Manchester et ABE d'après les chiffres fournis par E. Dąbrowska dans ses articles.

Dans ces conditions, il est difficile d'affirmer l'existence d'une formule prototypique. En français, la construction aurait plutôt tendance à être libre en tout cas en ce qui concerne le lexème verbal, qui ne se réduit pas à un modèle d'un ou deux items, et son sujet. L'emploi majoritaire confirme néanmoins l'idée de Verhagen : dans les interrogatives, les constructions de dépendance à distance répondent au principe d'intersubjectivité.

Puisque les dépendances à distance ne sont pas vraiment figées en français, comme nous pouvions le supposer au vu des conclusions établies sur les données attestées anglaises et néerlandaises, toutes les interrogations posées par les études empiriques et expérimentales se manifestent. Il est possible que la construction soit en partie libre, mais soumise à des contraintes syntaxiques, pragmatiques, et cognitives.

Nous allons dans un premier temps vérifier si les contraintes syntaxiques posées par les génératives (voir §1.3.1) sont respectées dans les données produites par les locuteurs français enregistrés dans nos corpus. Ensuite, nous tenterons de comprendre pourquoi certains verbes apparaissent en position de ponts et d'autres pas.

3.2 Construction libre avec des contraintes syntaxiques (position du gap) ?

En observant les données attestées, nous ne pouvons aucunement certifier de ce qui est impossible. D'une part parce que les corpus sont un échantillon de la langue produite ; ils ne permettent donc pas de connaître tous les énoncés attestés. D'autre part parce que les corpus sont uniquement le reflet de la performance des locuteurs, et non de leur compétence. Une formule peut être possible et comprise par l'ensemble des natifs (compétence) mais ne jamais être produite (performance). En aucun cas, la description des corpus ne remplace donc les études théoriques qui prédisent qu'un énoncé est impossible pour des raisons structurelles.

En revanche, les données attestées peuvent appuyer une théorie ou la remettre en cause selon qu'elles s'y conforment ou non. Elles peuvent aussi nous permettre de choisir entre deux

analyses théoriques. Nous allons examiner si les contraintes syntaxiques définies par les génératives sont respectées dans nos corpus. Nous rappelons que l'hypothèse de ces modèles est que, en vertu du principe de récursivité, le contrôle à distance peut s'établir à travers un nombre illimité (*unbounded dépendency*) de structures 'ponts' enchâssées les unes dans les autres. Pour atteindre l'adéquation descriptive, il s'agit donc de repérer en premier lieu les configurations qui interdisent la dépendance à distance et qui opèrent donc comme des contraintes sur les configurations a priori possibles.

3.2.1 Contrainte du sujet phrastique

La contrainte du sujet phrastique interdit à la complétive du GN sujet d'être antéposée. Un énoncé qui infirmerait la contrainte du sujet phrastique serait du type (230) :

*(230) *Qui ils croient que le fait qu'elle connaissait _ l'a dérangé ?*

Le sujet complexe de « déranger » *[le fait qu'elle connaissait x]* est scindé en deux ; le complément de « connaître » est antéposé et interrogé. Aucun exemple de ce type n'est présent dans les corpus étudiés, ce qui semble confirmer la validité de la contrainte.

Néanmoins, l'emploi de sujets phrastiques n'est pas très fréquent. Il n'est pas habituel qu'un sujet soit composé d'un adjectif ou d'un nom suivi d'une complétive (*fiers que, la certitude que, l'idée que*). La forme la plus fréquente est sans doute celle en « le fait que ». Or dans les corpus oraux uniformisés, elle n'apparaît que 3 fois en position sujet (et toujours en double marquage avec un pronom) (231) ; dans les corpus écrits, on relève 17 occurrences présentant une phrase qui commence par « Le fait que » (232). Les attestations sont donc très rares.

(231) alors le fait qu'on travaille à l'étranger ça c'est euh un peu particulier (CRFP, PRI-VAL-2)

(232) Le fait qu'Ariel Sharon ait empéché Yasser Arafat de célébrer noël à Bethléem montre bien qu'il ne met aucune volonté pour régler le conflit. (CERF, FORM)

Par ailleurs, l'extraction de sujets, même pris dans leur totalité, est rare. Nous n'en avons relevé que six. En voici deux exemples (233)(234).

(233) qu'est-ce que tu trouves qui est super (Nancy, Marais)

(234) quand il y avait un client ben la patronne elle se dérangeait pas c'est moi qu'il fallait qui aille servir (CTFP, La fleuriste)

(Godard, 1988, p. 40) note que la relativisation du sujet implique d'autres contraintes, notamment celle du choix du verbe pont. Une poignée est employée sans difficulté dans ce contexte : « affirmer », « croire », « dire », « penser », « savoir », et quelques autres s'ajoutent, avec une certaine hésitation : « estimer », « ignorer », « imaginer », « juger », « prétendre », « supposer », « trouver ». (Sportiche, 2011) donne une liste approximativement identique et indique que certains verbes sont acceptables en pont avec une dépendance à distance objet (235) mais pas sujet (236).

*(235) *la fille qu'il est important qui vienne (Sportiche, 2011, p. 111)*

(236) la fille qu'il est important que tu voies (Sportiche, 2011, p. 111)

Les verbes qui autoriseraient l'extraction du sujet seraient des verbes de parole, des épistémiques et des désidératifs. Ceux qui rendraient l'extraction difficile ou exclue seraient des verbes sans lien avec l'attitude et des verbes de parole spécifiques.

Il est donc possible que d'autres contraintes, non syntaxiques (telles que la fréquence de la séquence ou le type de verbe pont), soient en jeu dans les énoncés à sujets phrastiques et qu'elles empêchent l'extraction dans des contextes qui englobent ceux des sujets phrastiques. Enfin, si la question des limites de l'extraction du sujet dans son ensemble se pose pour tous les cadres d'analyse, la question de l'extraction d'une partie de sujet phrastique peut être un problème créé par le modèle lui même puisqu'il n'existe pas par exemple dans le traitement par adjonction proposé par le modèle TAG (Abeillé, 2007, p. 241). Le faible nombre d'exemples attestés ne nous permet pas de discuter cette question plus avant.

3.2.2 Contrainte du SN complexe

La seconde contrainte syntaxique notée par les génératives interdit l'extraction d'une partie d'un syntagme nominal complexe, c'est-à-dire une partie de la relative (237) ou de la complétive (238) qui dépend du nom.

(237) l'homme [qui a un chapeau]

(238) l'envie [qu'il parte d'ici]

Si les relatives en « que » et en « qui » sont effectivement absentes des occurrences de dépendance à distance (239), les relatives en « dont », en revanche, sont fréquentes (240).

*(239) *Qu'est-ce que je pense que tu connais [quelqu'un [qui a _]] ?*

(240) Je mets immédiatement cette lettre non datée dans une bouteille dont vous remarquerez que j'ai soigneusement gratté [l'étiquette _] (CERF, DIVS)

Ces exemples, dans lesquels l'élément antéposé dépend d'un nom, ont été exclus lors de la délimitation du sujet. Nous en avons relevé une quinzaine lors de l'analyse des segments proposés par le concordanciers.

Quant aux noms complexes formés par la construction 'nom + complétive', ils donnent lieu à des dépendances à distance dans notre corpus. La complétive peut être antéposée dans le cas où le nom est employé avec un verbe support. La séquence 'verbe support + nom' forme en quelque sorte une tournure verbale : « avoir le sentiment », « prendre conscience ».

*(241) Le présent même existe-t-il ? A peine l'a-t-on évoqué qu'il a disparu. Mais il est cependant des moments, et je crois que nous en vivons un, où on **a le sentiment** qu'une porte s'ouvre sur l'avenir. (CERF, POLI)*

*(242) quand on aura atteint cette perspective où euh ben on sera tous salariés de notre société entre guillemets idéale où chacun **aura pris conscience** que son travail c'est plus un travail subordonné (Nancy, Retraites)*

La contrainte du nom complexe n'est donc pas totalement applicable en français. Elle ne concerne que les cas où le nom est bien tête d'un groupe nominal autonome, comme le montre la pronominalisation (243) et pas les cas qui peuvent être analysés comme des tournures verbales ou des constructeurs complexes

*(243) J'approuve ça, l'idée qu'il faut renoncer à ça => * à quoi approuves-tu l'idée qu'il faut renoncer*

3.2.3 Contrainte de l'ajout

La troisième contrainte interdit l'extraction d'une partie d'un ajout. Aucun contre-exemple n'est attesté dans les corpus mais il semble que ce type d'extraction ne soit pas totalement interdit, du moins avec une infinitive. Les ajouts en « pour », notamment, donnent lieu à discussion.

(244) il te manque juste le tampon que tu vas t'arranger pour obtenir bientôt (oral, observation personnelle)

(245) ?un livre que j'ai tout fait pour me procurer

On peut évidemment se demander si dans (244) le complément infinitif introduit par « pour » est toujours ajout : une analyse en objet du verbe « s'arranger pour » n'est pas exclue, le verbe portant un trait sémantique de but. De façon générale, nous avons vu que certains modèles HPSG ou TAG considèrent que de telles séquences ne sont pas prévues par les règles syntaxiques de base.

Les contraintes dites 'syntaxiques' relevées par les génératives sont donc globalement respectées. Elles reviennent à dire que les seules structures source de dépendance à distance sont des objets de verbes (complétives ou infinitives). Il reste alors à expliquer pourquoi deux énoncés syntaxiquement identiques sur ce point sont l'objet de jugements de grammaticalité très différents lors des enquêtes menées auprès des locuteurs (246).

(246) le livre que tu dis / ?déclares / ??murmures qu'il a acheté

Le jugement est-il non conforme à la réalité de la production ou y a-t-il d'autres paramètres qui contraignent la construction ? Quel rôle joue le verbe pont dans l'acceptabilité des énoncés ?

3.3 Sélection dans le lexique des verbes ponts ?

Le lexique des verbes ponts est l'élément qui suscite le plus de questions dans les analyses des constructions de dépendance à distance. Nous avons vu que c'est le lexique qui varie le plus dans les énoncés qui s'écartent de la construction la plus fréquente (voir §3.1) et le fait que certains verbes autorisent l'extraction et d'autres non n'est pas justifiée par les contraintes syntaxiques qui pèsent sur la construction. L'hypothèse est qu'il existe une autre contrainte, qui différencie les verbes ponts des verbes non-ponts. Pour répondre à cette question, nous allons examiner quels verbes apparaissent dans cette position et tenter de trouver une explication sur la sélection qui s'opère.

3.3.1 Verbes ponts attestés dans les corpus

Dans les 229 occurrences de dépendance à distance relevées dans les corpus oraux et écrits, 34 formes verbales sont présentes. Ces formes sont des verbes simples ou des locutions verbales. Voici la liste des lexèmes attestés, dans l'ordre alphabétique.

1	attendre
2	avoir le sentiment
3	comprendre
4	croire
5	désirer
6	dire
7	entendre
8	estimer
9	être
10	être clair
11	être évident
12	être intéressant

13	être normal
14	être vrai
15	falloir
16	imaginer
17	penser
18	prendre conscience
19	présumer
20	prévoir
21	rappeler
22	remarquer
23	se rendre compte
24	ressentir

25	savoir
26	sembler
27	sembler important
28	sentir
29	souhaiter
30	supposer
31	trouver
32	valoir mieux
33	voir
34	vouloir

Chacun de ces verbes apparaît une ou plusieurs fois (jusqu'à 88 occurrences). A deux reprises, deux verbes se succèdent dans la même occurrence. Au total, il y a donc 231 attestations des verbes cités.

Dans les paragraphes suivants, nous allons tenter d'expliquer pourquoi ces verbes sont présents et comment s'est opérée la sélection. D'abord, nous vérifierons les hypothèses formulées précédemment : les verbes mettant en jeu un contenu présupposé sont-ils exclus des verbes ponts ? Les verbes recteurs faibles, qui permettent d'atténuer la valeur de vérité de l'assertion, sont-ils favorisés ? Ensuite, nous comparerons le lexique des verbes ponts à celui des verbes à complétive.

3.3.1.1 Verbes ponts et verbes factifs/implicatifs/semi-factifs

Une idée très présente dans la littérature est que les éléments qui font partie du *background*, c'est à dire qui sont présupposés, ne peuvent pas être sujets à une mise en valeur ou à une focalisation. L'extraction est un type de focalisation. Les verbes qui ont pour objets des compléments présupposés, tels que les verbes factifs, ne devraient donc pas être attestés en verbes ponts. Cette théorie a été réfutée par (Erteschik-Shir, 1973) dès 1973 mais elle a tout de même traversé les décennies (Godard, 1980; Featherston, 2005) et a reçu l'appui des expériences de (Ambridge & Goldberg, 2008) (voir §1.3.2).

(Mørdrup, 1975) a défini une typologie des verbes concernés par la présupposition en français (voir §1.4.1.1). Elle distingue trois classes et plusieurs sous-classes en fonction de la modalité de la principale et de la valeur de vérité de la complétive. Le tableau suivant rappelle les principales catégories liées à la présupposition (TABLEAU 30).

Factifs	positifs	**que A soit positif ou négatif, B est vrai**	*Pierre regrette que Sheila soit enceinte => Sheila est enceinte* *Pierre ne regrette pas que Sheila soit enceinte => Sheila est enceinte*
	négatifs	**que A soit positif ou négatif, B est faux**	*Pierre s'est imaginé que Sheila est partie => ~~Sheila est partie~~* *Pierre ne s'est pas imaginé que Sheila est partie => ~~Sheila est partie~~*
Semi-factifs	non-factif factif	si conjugués à la P1 dans une structure hypothétique autre structure	*Si j'apprends un jour que je n'ai pas dit la vérité, je l'avouerai à tout le monde (pas dit vérité ?)* *J'ai appris que Marie avait acheté une voiture => Marie a acheté une voiture*
Implicatifs	positifs	**si A est positif, B est vrai ;** **si A est négatif, B est faux**	*Pierre a commencé à parler -> Pierre parle* *Pierre n'a pas commencé à parler -> ~~Pierre parle~~*
	négatifs	**si A est positif, B est faux ;** **si A est négatif, B est vrai**	*Pierre a cessé de parler => ~~Pierre parle~~* *Pierre n'a pas cessé de parler => Pierre parle*
	simples	**si A est positif, B est vrai** [mais si A est négatif, B n'est pas faux]	*Pierre a prouvé que la terre tourne autour du soleil -> la terre tourne autour du soleil.* [« *Pierre n'a pas prouvé que la terre tourne autour du soleil* » n'implique pas « *~~la terre tourne autour du soleil~~* »]

		si A est négatif, B est faux [mais si A est positif, B n'est pas vrai]	aucun verbe français
		si A est positif, B est faux [mais si A est négatif, B n'est pas vrai]	*Pierre n'a pas pu partir => ~~Pierre est parti~~.* [« *Pierre a pu partir* » n'implique pas « *Pierre est parti* »]
		si A est négatif, B est vrai [mais si A est positif, B n'est pas faux]	*Pierre n'a pas hésité à sortir => Pierre est sorti* [« *Pierre a hésité à sortir* » n'implique pas « ~~*Pierre est sorti*~~ »]

TABLEAU 30. Typologie des verbes liés à la présupposition (Mørdrup, 1975)

Une partie des verbes ponts attestés dans les corpus se retrouve dans le classement de (Mørdrup, 1975). Parmi eux, certains sont des verbes assertifs, ce qui est attendu, mais d'autres sont factifs ou semi-factifs (TABLEAU 31)[72].

Types de verbes	Lexèmes verbaux
Assertifs	Attendre, croire, désirer, dire, estimer, imaginer, penser, présumer, prévoir, sembler, souhaiter, supposer, trouver, valoir mieux, vouloir
Factifs	Comprendre ?, ressentir, savoir
Semi-factifs	Remarquer, sentir, voir, (se rendre compte), Entendre (sauf s'il exprime la volonté)

TABLEAU 31. Verbes ponts du corpus classés par (Mørdrup, 1975) en fonction de leur trait de factivité

Nous présentons ici une attestation relevée dans les corpus pour chaque verbe factif classé par (Mørdrup, 1975) :

*(247) Paris est une cité, on s'y entasse également, c'est pour cela qu'il faut bien **comprendre** qu'en mai 68, le beans n'était pas tant une question de confort au sens qu'on lui connaît. (CERF, FORM)*

*(248) on n'aurait pas pu euh participer à cette évolution dont on **ressentait** qu'on avait vraiment besoin (PFC, 81arc1)*

*(249) mais c'est euh non c'est génial quoi mais faire un truc que tu aimes quoi et que tu **sais** que tu maîtrises (PFC, 12afl1)*

[72] Nous ne sommes pas toujours convaincus du classement de certains verbes, notamment « ressentir » et « entendre » mais nous respectons le classement de l'auteur.

Et ici une attestation relevée dans les corpus pour chaque verbe semi-factif classé par (Mørdrup, 1975) :

(250) *tu as été dans d'autres euh enfin d'autres euh théâtres spectacles où euh tu tu* ***entendais*** *que c'était un français un peu euh (PFC, 75ccm1)*

(251) *il y a des périodes où vous* ***remarquez*** *que vous vendez plus ce produit que ce produit (CorpAix, AZ)*

(252) *c'est pour ça d'ailleurs qu'on* ***se rend compte*** *ben que le livret A de la Caisse d'Epargne qui avait été complètement euh laissé à l'abandon dans les années soixante-et-dix quatre-vingts revient très très fort (CorpAix, PRVOYAN)*

(253) *vous voyez je recherche toujours un petit peu ce qui - pas ce qui est facile enfin c'est pas c'est pas que c'est pas facile - mais - où on* ***sent*** *que les gens ont plus besoin de vous les familles ont besoin de vous – hein (CRFP, PRI-BOR-2)*

(254) *tout ça s'intègre quand même dans un contexte où nous* ***avons vu*** *que euh aussi bien les frais d'inscription que les frais de sécurité sociale que le prix de ticket de restaurant universitaire augmentent bien plus vite que le montant des bourses (Nancy, grève)*

Les autres verbes ponts, « avoir le sentiment, être, être clair/évident/intéressant/normal/vrai, falloir, prendre conscience, rappeler, sembler important », n'ont pas été cités par (Mørdrup, 1975). Voici les énoncés dans lesquels ils sont attestés dans les corpus.

(255) *Le présent même existe-t-il ? A peine l'a-t-on évoqué qu'il a disparu. Mais il est cependant des moments, et je crois que nous en vivons un, où on* ***a le sentiment*** *qu'une porte s'ouvre sur l'avenir. Nous sommes en train de l'ouvrir nous-mêmes, nous sommes donc des ouvriers privilégiés de l'histoire. (CERF, POLI)*

(256) *Il suffit de mesurer à cet égard ce que suscite comme amertume ou laisse comme frustration dans la conscience collective telle ou telle grande affaire où le sentiment commun* ***est*** *que la lumière n'a pas été faite et que la justice a été tenue en échec.(CERF, POLI)*

(257) *"C'est une révolution silencieuse", commente Naqvi. Une révolution où l'armée ne renoncera pas entièrement à son pouvoir, et où il* ***est clair*** *que Musharraf ne compte pas se retirer de son poste à la tête du pays. (CERF, PRES)*

(258) *Et bien c'était, comme tout ce qu'a toujours fait le Général de GAULLE, quelque chose de visionnaire car nous arrivons aujourd'hui à l'époque où il* ***est évident*** *que l'Europe et l'Amérique Latine doivent marcher la main dans la main. (CERF, POLI)*

(259) *En application du paragraphe 11 de la résolution 53/102 de l'Assemblée générale en date du 8 décembre 1998, la Commission indique ci-après, pour chaque sujet, les points sur lesquels il*

*pourrait **être particulièrement intéressant** que les gouvernements expriment leurs vues, soit à la Sixième Commission, soit par écrit, afin de la guider utilement dans la poursuite de ses travaux. (CERF, INST)*

*(260) "Monsieur le Président, Mesdames et messieurs les députés, Monsieur le député, La réunion informelle des chefs d'Etat et de gouvernement qui s'est tenue vendredi dernier à Petersberg, et dont je pense qu'il **est normal** que je rendre compte à la représentation nationale, a été positive à plus d'un titre. (CERF, POLI)*

*(261) il y a des périodes où c'**est vrai** qu'il y a des produits qu'on vend pas et que coup sur coup on vous demande (CorpAix, AZ)*

*(262) Après l'Euro, les mêmes trouvent cette équipe banale. Mais que **faut**-il qu'elle fasse : gagner pendant 10 ans toutes les coupes ? Gagner tous ses matches 10-0 ? (CERF, FORM)*

*(263) quand on aura atteint cette perspective où euh ben on sera tous salariés de notre société entre guillemets idéale où chacun aura **pris conscience** que son travail c'est plus un travail subordonné mais il a une rémunération euh bon bref euh je sais pas j'ai j'ai l'impression que c'est un peu ça que vous dites mais sans euh sans dire le mot (Nancy, Retraites)*

*(264) La seule solution d'accueil pour la titularisation à laquelle je **rappelle** qu'il était déjà procédé précédemment - est l'entrée dans le corps des professeurs certifiés. (CERF, INST)*

*(265) Ce sont là des questions, M. le Président, qu'il **semble très important** qu'un gouvernement responsable doit se poser et surtout auxquelles nous devons avoir les réponses afin de prendre une décision tout à fait éclairée pour l'ensemble des personnes touchées (CERF, INST)*

En appliquant à ces verbes le test de (Mørdrup, 1975), qui vérifie la valeur de vérité de la complétive en fonction de la modalité de la principale, nous pouvons définir leur statut de verbe factif, implicatif ou assertif (TABLEAU 32).

Verbe	**Exemple**	**Classement**
être	Le sentiment commun est qu'elle est riche -> Elle est riche ? Le sentiment commun n'est pas qu'elle est riche -> Elle est riche ?	assertif
falloir	Il faut qu'elle soit riche -> Elle est riche ? Il ne faut pas qu'elle soit riche -> Elle est riche ?	assertif
prendre conscience	Il a pris conscience qu'elle était riche -> Elle est riche Il n'a pas pris conscience qu'elle était riche -> Elle est riche	Factif positif
rappeler	Il rappelle qu'elle est riche -> Elle est riche Il ne rappelle pas qu'elle est riche -> Elle est riche	Factif positif

avoir le sentiment	Il a le sentiment qu'elle est riche -> elle est riche ? Il n'a pas le sentiment qu'elle est riche -> Elle est riche Il n'a pas le sentiment qu'elle soit riche -> Elle est riche ?	assertif / implicatif simple ?
être clair	Il est clair qu'elle est riche -> Elle est riche Il n'est pas clair qu'elle est riche -> Elle est riche Il n'est pas clair qu'elle soit riche -> Elle est riche ?	factif positif / implicatif simple ?
être évident	Il est évident qu'elle est riche -> Elle est riche Il n'est pas évident qu'elle soit riche -> Elle est riche?	implicatif simple A positif, B vrai
être intéressant	Il est intéressant qu'elle soit riche -> Elle est riche Il n'est pas intéressant qu'elle soit riche -> Elle est riche ?	implicatif simple A positif, B vrai
être normal	Il est normal qu'elle soit riche -> Elle est riche Il n'est pas normal qu'elle soit riche -> Elle est riche	Factif positif
sembler important	Il lui semble important qu'elle soit riche -> Elle est riche ? Il ne lui semble pas important qu'elle soit riche -> Elle est riche (?)	implicatif simple A négatif, B vrai
être vrai	Il est vrai qu'elle est riche -> Elle est riche Il n'est pas vrai qu'elle est/soit riche -> ~~*Elle est riche*~~	Implicatif positif

TABLEAU 32. Test de factivité et classement des verbes du corpus non étudiés par (Mørdrup, 1975)

On peut conclure de cette présentation que les verbes ponts présents dans les corpus peuvent être des assertifs, comme cela était prévu, mais qu'ils peuvent aussi appartenir aux trois grandes classes définies par (Mørdrup, 1975): les factifs, les semi-factifs et les implicatifs. Nos résultats favorisent donc l'hypothèse de (Erteschik-Shir, 1973). Parmi les factifs, seule la sous-classe des factifs positifs est exemplifiée ; parmi les implicatifs, on trouve des implicatifs positifs et des implicatifs simples. Ceci nous amène à poser la question suivante : l'extraction serait-elle impossible dans le cas où le verbe pont appartient à la classe des factifs négatifs ou des implicatifs négatifs ?

(Mørdrup, 1975) liste quelques verbes factifs négatifs et de nombreux verbes implicatifs négatifs. Faisant l'hypothèse que la taille du corpus joue un rôle dans la présence ou l'absence d'un verbe, nous avons tenté de trouver ces verbes dans des constructions de dépendance à distance sur le Web, en centrant nos recherches sur les verbes qui se construisaient avec une

complétive, soit « se figurer », « s'imaginer » et « rêver » pour les factifs négatifs, puis « éviter » et « empêcher » pour les implicatifs négatifs.

Concernant les factifs négatifs, plusieurs occurrences de dépendances à distance ont été relevées pour chacun des trois verbes. Voici une attestation de chaque : « rêver » (266), « se figurer » (267) et « s'imaginer » (268).

(266) *Autre "plus" que je **rêve** qu'ils proposent - mais je suis quasiment sûr que cela ne sera pas le cas - c'est de pouvoir télécharger la version PDF d'un bouquin que l'on a acheté en dur.*[73] *(Web, forum)*

(267) *Quant à l'hypothèse d'une choré sur une musique à demi-vitesse doublée ensuite au montage, c'est une éventualité, mais ça serait dommage et stupide de sa part vu les aptitudes que je **me figure** qu'elle a.*[74] *(Web, forum)*

(268) *Je veux dire par là que, à presque 31 ans, je commence à accepter que, quel que soit la perception que les autres ont de moi, je ne me sens pas du tout femme, je n'ai pas les manières que je **m'imagine** qu'une femme doit avoir. Pour moi une femme, c'est une Monica Belluci, une Carole Bouquet ou autre Catherine Deneuve. Rien à voir avec moi quoi !*[75] *(Web, forum)*

Concernant les implicatifs négatifs, les attestations sont moins nombreuses. Nous n'en avons relevé qu'une, qui est une traduction de Sénèque *Hercule furieux* par Jean-Marie Georgeot, effectuée entre 1978 et 2006. Il est possible que la langue source ait influencé la construction et le lexique de la traduction française (269) :

(269) *Celui qui, sur terre, celui qui sur mer pourchasse les crimes et, de son bras justicier, brise des sceptres cruels, est aujourd'hui esclave loin de nous, il subit ce qu'il **empêche** qu'on fasse.*[76] *(Web, institutionnel)*

On constate donc qu'aucune des catégories de verbes liés à la présupposition n'est absente en verbe pont. Les factifs, semi-factifs et implicatifs peuvent tout à fait apparaître dans les constructions de dépendance à distance, au même titre que les assertifs.

[73] http://www.polaris.me/1259/polaris-en-pdf/, consulté le 17/09/10
[74] http://www.koreus.com/modules/news/article5735.html, consulté le 16/06/12
[75] http://www.carolinedaily.com/forum/comment-se-sentir-femme-t13334.html, consulté le 15/06/12
[76] http://digital.fides.org.pl/Content/106/CD_1/TOME17/2_Rome_Textes-paiens/08_Seneque.pdf, consulté le 15/06/12

3.3.1.2 Verbes ponts et verbes recteurs faibles

Les verbes recteurs faibles ont pour vocation d'amoindrir la valeur de vérité de l'énoncé, en le modalisant par un point de vue (Blanche-Benveniste & Willems, 2007). Les verbes ponts dans les constructions de dépendance à distance auraient la même fonction d'après (Verhagen, 2006). Les verbes recteurs faibles sont donc de bons candidats de verbes ponts.

(Blanche-Benveniste & Willems, 2007) ont commencé à lister les verbes faibles du français. Elles relèvent : « je crois », « je pense », « je trouve », « j'imagine », « je/on dirais/t », « j'espère », « il me semble », « je présume », « je suppose », « c'est clair », « il paraît », « j'ai l'impression », « il est certain », « c'est vrai » et « c'est sûr. » Selon elles, le verbe est recteur faible particulièrement dans la forme de la première personne.

Parmi ces seize formes, huit sont attestées en verbes ponts dans les corpus. Voici un exemple de chacun de ces verbes :

*(270) eh bien je cr- je pense je pense oui c'est une hypothèse enfin qui me paraît maintenant comme ça au bout quand même d'un certain nombre d'années de travail dans ce domaine c'est une hypothèse que **je crois** qu'on peut avancer (CRFP, PUB-STR-1)*

*(271) je suis dans une période de ma vie où **je pense** que jamais dans ma vie j'ai vu aussi peu loin ce qui pouvait m'arriver (CFPP, Duchemin_Ceret_11e)*

*(272) mais là je préfère rester enfermée chez moi plutôt que d'aller dans le Champ de mars où y a **je trouve** qu'il y a trop de monde (CFPP, Lebond_Zanotti_7e)*[77]

*(273) j'ai habité quand même pendant quatre ans dans le huitième donc euh c'est un quartier que j'ai un peu découvert que **je dirais** qui est pas v- enfin j'habitais à côté de la place Bauveau donc je dirais pas que c'est un quartier euh c'est un quartier bien sûr mais c'est pas un endroit où les où les gens je veux dire c'est des galeries on est à côté du faubourg saint Honoré de l'Elysée euh ça m'a pas paru un quartier euh (CFPP, Audin_7e)*

*(274) l'idée c'est de faire un film quoi [...] pas de tomber forcément dans un cinéma qui soit euh - qui soit euh une sorte de mur de lamentation(s) de de je sais pas du du scénario de de la pensée capitaliste ou des choses comme ça mais peut-être plus de voir euh comment ils peuvent faire résistance vis-à-vis de ben d'un endroit où **il me semble** que hum que finalement euh que finalement il y a il y a plein il y a il y a plein de croisements en fait d'idées de d'automatisation de de domination(s) de c'est un peu c'est le Japon (Nancy, Japon)*

[77] On peut voir dans cette occurrence un abandon de construction (*où il y a //)* suivi d'une reconstruction (*je trouve qu'il y a...)* mais nous l'avons comptée comme dépendance à distance, la prosodie étant réalisée en une unité.

(275) *Bien sûr, le ciné (du moins, celui dont **je présume** que tu parles), c'est facile, on t'apporte tout sur un plateau, à renfort d'effets spéciaux made in hollywood. Génial ! (CERF, FORM)*

(276) *Une révolution où l'armée ne renoncera pas entièrement à son pouvoir, et où **il est clair** que Musharraf ne compte pas se retirer de son poste à la tête du pays. (CERF, PRES)*

(277) *oui c'est c'est une période où **c'est vrai** qu'on se prend on se prend pas mal la tête pour des choses qu'en valent peut-être pas la peine (CFPP, Teixeira_11e)*

Il reste donc huit formes non attestées en pont dans le corpus de travail. Mais là encore, une recherche sur le Web montre que ces verbes peuvent bien être pont :

(278) *Je pense qu'il attendait de moi que j'apporte mon soutien à Pete. Il a pensé que nous nous entendrions bien parce qu'on est nés le même jour ou un truc dans le genre… et qu'on est tous les deux des créatures assez désordonnées. C'est pour ça que **j'imagine** qu'il me voulait comme soutien. Je pense que ça a marché.[78] (Web, oral transcrit)*

(279) *On a bien compris que c'est un 37 mais on DIRAIT pas. Et puis le problème n'est pas combien **on dirait** qu'il fait sur la photo, c'est combien il dit faire. [tour de bras en cm][79] (Web, forum)*

(280) *"Qu'est-ce que je vais leur dire ?" Je n'oublierai pas de sitôt sa réponse : "Peu m'importe ce que vous leur direz. Vous en savez aujourd'hui assez sur ce plan. De plus ils oublieront plus de la moitié de ce que vous aurez dit. Mais ce que **j'espère** qu'ils n'oublieront pas, c'est qui était là devant eux pendant une heure."[80] (Web, oral inventé)*

(281) *Donc si Mitterrand parle de pêché c'est bien qu'il reconnait qu'il n'est pas allé à Bangkok pour coucher avec des hommes (ce qu'on peut faire gratuitement à Paris où **je suppose** qu'il y a beaucoup de clubs de rencontres gratuites)[81] (Web, forum)*

(282) *Ce n'est pas moi qui vais plaindre les banquiers. L'une de celles épinglées par l'AUTORITE DE LA CONCURRENCE, celui à qui **il parait** qu'on peut parler, qui refusait de payer mes honoraires et a dû le faire après avoir été rendu ridicule par la taxation du bâtonnier, m'a oublié.[82] (Web, blog)*

(283) *Peut-être que je me trompe, et que c'est possible, mais le plan proposé correspond exactement à ce que **j'ai l'impression** qu'il ne faut pas faire[83](Web, forum)*

[78] http://www.soul-kitchen.fr/6195-interview-de-graham-coxon, consulté le 30/09/10
[79] http://www.musculaction.com/forum/photos-de-l-auteur-t3434-162.html, consulté le 01/10/10
[80] http://www.nouvellescles.com/article.php3?id_article=136, consulté le 30/09/10
[81] http://blog.lefigaro.fr/rioufol/2009/10/bloc-notes-cest-au-secours-de.html, consulté le 30/09/10
[82] http://avocats.fr/space/bernard.kuchukian/content/dossier-navista---les-amendes-du-conseil-de-la-concurrence_65A04352-FFEA-4980-80BD-A2909C3C274F/Web-print, consulté le 10/09/10
[83] http://fr.wikipedia.org/wiki/Discussion:Perception_de_l%27antis%C3%A9mitisme/Suppression, consulté le 10/09/10

(284) *en tout cas ce qu'**il est certain** qu'il ne faut jamais faire lorsqu'on tombe en arriere c'est de mettre les mains tendus pour ce rattraper c'est le meilleur moyen de se faire sauter les épaules!!!! Au contraire comme il a été dit plus haut dans le topic faut mettre les pieds fesses et toujours retenir sa tete pour pas qu'elle cogne!!![84] (Web, forum)*

(285) *bah ce qui **est sur** qu'il faut faire si ça n'a pas déjà été fait c'est d'aller se présenter à la gendarmerie de Dijon pour porter plainte.[85] (Web, forum)*

Tous les verbes recteurs faibles peuvent donc bien être verbes ponts. Malgré tout, comme nous l'avons vu, certains n'apparaissent pas dans les corpus. En outre, les huit verbes recteurs faibles qui apparaissent ne font pas forcément partie des verbes ponts plus fréquents, et ce même sans restreindre le relevé aux formes de première personne du présent. Le tableau suivant présente le rang et la fréquence des verbes recteurs faibles parmi les verbes ponts (TABLEAU 33) :

Verbe	penser	dire	croire	trouver	sembler	Etre vrai	présumer	Etre clair
Rang	3	4	7	8	9	13	14	23
Nb d'occ	25	16	5	5	3	2	2	1

TABLEAU 33. Fréquence et rang des verbes recteurs faibles parmi les verbes ponts

Si les verbes faibles étaient vraiment propices à la construction, les huit verbes auraient tous été regroupés dans les rangs 1-8. Or ici, les deux premiers rangs ne sont pas attribués à des verbes recteurs faibles. La locution verbale « être clair » est classée au 23ème rang (sur 34). L'ensemble des verbes faibles ne réunit que 59 occurrences de dépendance à distance sur 229, soit 25,8% des occurrences.

Les verbes recteurs faibles sont tous candidats à la place de verbe pont. Néanmoins ils ne dominent pas dans le classement des verbes ponts les plus fréquents et ils n'apparaissent que dans un quart des occurrences. Mais ce résultat peut s'expliquer en partie parce que ces verbes ayant la fonction de moduler une assertion peuvent être exclus de constructions où la complétive n'est pas assertée (en particulier les extractions et certaines relatives). La complexité des données conduit donc à éviter d'aborder le problème avec des thèses a priori

[84] http://archive.parkour.net/lofiversion/index.php/t13459.html, consulté le 30/09/10

[85] http://ovalie.forumactif.com/la-place-publique-espace-visiteurs-guests-f49/arnaque-en-ligne-t6219.htm, consulté le 30/09/10. Il est probable que l'orthographe de « qui » soit une erreur dûe d'une part à la prononciation courante de 'il' en [i] et d'autre part à l'existence de la formule « ce qui est sûr ». L'énoncé serait bien une dépendance à distance « ce qu'il est sûr qu'il faut faire » et non une succession de relative « ce qui est sûr, qu'il faut faire ».

et à procéder de manière purement descriptive à travers deux questions : Quels sont donc les verbes qui apparaissent ? Quelles relations ont les verbes ponts par rapport aux verbes en général et plus précisément par rapport aux verbes à complétive ?

3.3.1.3 Verbes ponts et verbes à complétive

Afin de découvrir la relation entre les verbes ponts et les verbes à complétive, nous avons recherché une liste la plus exhaustive possible des verbes à complétive français. Nous nous sommes tournés vers le Trésor de la Langue Française informatisé (TLFi), dictionnaire en ligne de l'ATILF[86], qui permet de sélectionner des entrées en fonction de constructions particulières. Le TLFi compte 8959 verbes français, dont 1857 sont une sous-entrée d'une autre forme (« gymnastiquer », dans « gymnastique »). Le dictionnaire fournit donc 7102 entrées de formes verbales. Parmi ces formes, 229 se construisent avec « que ». Mais certaines sont notées comme étant des constructions vieillies (« acertainer que », « se passer que »), rares (« tenir que », « regarder que ») ou littéraires (« mander que »). D'autres ne comportent pas d'annotations spécifiques mais ne semblent pas très courantes avec une complétive (« téléphoner que », « agréer que »). Il est donc probable que de nombreux verbes parmi les 229 indexés ne soient pas attestés. Mais il est également probable que quelques verbes à complétive ne soient pas répertoriés dans le dictionnaire, ou tout simplement que la construction en « que » ne soit pas explicitée dans l'entrée ou dans le champ 'construction' (« affirmer », « apprécier », « vouloir »). D'après (Gross, 1975, p. 215), il y aurait au moins 2800 verbes qui se construisent avec une complétive en « que », sur les 3000 qu'il traite.

Afin de connaître les verbes attestés dans les corpus, nous avons effectué une recherche sur les six corpus suivants : CFPP, C-ORAL-ROM, CRFP, CTFP, PFC et CorpAix, étiquetés avec TreeTagger (entraîné sur l'oral). Les corpus comportent 3025 formes verbales, dans environ 350 000 occurrences. A l'écrit, huit tranches du corpus CERF (à l'exception de LITA et ORAL) étiquetées avec Cordial produisent 6362 formes verbales, dans environ 1 160 000 occurrences. Au total, 6489 verbes sont présents dans l'ensemble des corpus observés (écrits et oraux), et produisent environ 1 500 000 occurrences.

[86] Voir le site http://atilf.atilf.fr/tlf.htm

Nous avons ensuite recherché les verbes à complétive attestés dans les mêmes corpus. L'outil de recherche CooLoX renvoie 13 310 occurrences de complétives régies par 449 formes verbales pour l'oral, et 28 018 occurrences de complétives, régies par 696 formes verbales pour le CERF (sans la tranche LITA). La requête biaise en partie les résultats car elle impose que la conjonction « que » soit consécutive au verbe : « [POS~"^VER"] + restriction à droite [lemme="que" & POS="KON"] » pour l'oral, « [POS~"^V"] + restriction à droite [lemme="que" & POS="SUB"] » pour l'écrit. Au total, pour les 41 328 occurrences de complétives présentes dans les corpus, on trouve 870 formes verbales. Seuls 175 verbes à complétive sont présents à la fois dans le TLFi et dans les corpus. Cela laisse présupposer un certain nombre d'erreurs à la fois dans l'étiquetage des corpus et dans les indications des entrées du dictionnaire. La lecture des verbes à complétive définis par l'étiquetage montre en effet des erreurs dans l'étiquetage des verbes (« aberrer », « liser », mais également « donné », « plaignait », « servi ») et dans l'étiquetage de « que » comme conjonction ce qui assimile certaines formes verbales à des constructions à complétives alors que le « que » était certainement restrictif (« acheter », « balader », « congeler », etc.).

Afin de contourner ces erreurs, nous avons effectué une recherche supplémentaire, beaucoup plus précise : nous avons relevé toutes les formes placées entre un pronom personnel sujet (« je », « j' », « tu », « il », « elle », « on », « ça », « nous », « vous », « ils », « elles ») et la forme « que/qu' », dans les six corpus oraux et dans le corpus écrit. Au total, 195 verbes à complétive sont identifiés. Nous avons conscience des biais de la requête, qui impose la présence d'un sujet pronominal adjacent au verbe et la consécutivité de la complétive, et qui ne permet pas de relever les locutions verbales et les verbes qui ont obligatoirement un objet clitique (*ça m'ennuie que*). Néanmoins, la requête présente l'avantage de repérer uniquement les verbes qui apparaissent dans la construction en complétive la plus simple, et donc certainement la plus fréquente. Nous y ajoutons 49 locutions verbales. C'est donc sur cette liste de 244 verbes à complétive que nous nous sommes basés (voir Annexe 5).

Parmi les 244 verbes à complétive de référence, 34 apparaissent dans des constructions de dépendance à distance dans la totalité des corpus. La proportion est donc faible (13,9%). Une hypothèse serait que les 34 verbes qui apparaissent en pont sont spécifiques et qu'ils ont une particularité en commun qui les rend aptes à entrer dans une construction de dépendance à distance. Il faudrait alors chercher quelle est cette particularité que les autres verbes n'ont pas

et qui les empêche d'être ponts. L'autre hypothèse est que les autres verbes peuvent aussi être ponts mais qu'ils ne sont pas attestés dans cette position dans les corpus. Il semble tout à fait contre-intuitif que des verbes comme « aimer », « déclarer » et « suggérer » soient impossibles en verbe pont.

Nous avons recherché l'ensemble des 244 verbes à complétive sur le Web, afin de découvrir d'éventuelles attestations de ces verbes en construction de dépendance à distance. Nous sommes bien sûr conscients des problèmes que pose le Web : ce n'est pas un corpus, il n'est ni stable ni fermé et nous n'avons aucun renseignement sur les auteurs (langue maternelle, âge, particularités face au langage). N'ayant pas la possibilité de consulter l'ensemble des sources, il est également impossible de connaître la fréquence et la proportion des phénomènes relevés. Mais nous accordons peu d'intérêt aux paramètres sociolinguistiques (voir ci-dessous, §3.5.1.1) et il nous a semblé qu'étendre les corpus au Web était la meilleure façon de rester dans l'esprit de la linguistique de corpus. Nous appuyons nos conclusions sur des données attestées, écologiques et non sur notre intuition personnelle.

3.3.2 Verbes ponts attestés sur le Web

Après une recherche de 244 verbes à complétive sur Google, il s'avère que 161 de ces verbes sont attestés en pont sur le Web. En outre, 5 des verbes ne sont pas attestés en pont sur le Web mais l'étaient dans les corpus, ce qui fait un total de 166 verbes ponts (68% des verbes à complétive). Ce résultat soulève une question importante : pourquoi y a-t-il une telle différence entre le lexique des verbes ponts attesté dans les corpus et celui attesté sur le Web ? Le corpus est-il représentatif de la langue produite dans un contexte « naturel », non sollicité ? Ou bien la différence vient-elle seulement de la taille des données ?

La recherche des verbes à complétive en construction de dépendance à distance sur le Web était du type « (que|qui|où|dont} je V_présent qu », c'est à dire l'un des quatre interrogatifs-relatifs cités, suivi du pronom personnel « je », du verbe à la première personne du singulier au présent et de la forme « qu » pour les verbes personnels ; et du type « (que|qui|où|dont} il V_présent qu » pour les verbes impersonnels. Pour les verbes personnels, nous avons parfois varié la personne (« il », « tu »), le temps (« j'ai écrit ») ou l'orthographe du verbe

(« j'admet », « je permet » ; « je garanti »)[87]. Nous avons trié manuellement les réponses renvoyées par le moteur de recherche *Google*, jusqu'à la page 10 incluse, soit 100 résultats pour chaque interrogatif-relatif. Dans un deuxième temps, nous avons mis de côté les attestations provenant de sites étrangers, de locuteurs étrangers à la France (lorsque cela était visible), de traductions et les attestations non contemporaines.

Malgré la procédure relativement stable de la recherche sur le Web, il est difficile de calculer la proportion d'occurrences ou la proportion de réponses renvoyées par rapport au nombre de mots contenu sur le Web. En effet, le Web n'est pas une ressource close et son contenu est constamment modifié. Le nombre de mots varie parallèlement au changement de contenus disponibles. Néanmoins, une expérience sur les corpus montre que lorsque l'on augmente le nombre de mots d'une ressource, le nombre d'occurrences de constructions de dépendance à distance augmente également. L'accroissement du nombre de dépendance à distance par rapport au nombre de mots se vérifie à la fois sur un corpus très précis d'entretiens privés (TABLEAU 34) et sur un corpus général d'oral tout venant (TABLEAU 35). Pour les entretiens privés, l'augmentation du nombre d'occurrences de dépendances à distance est identique à celle du nombre de mots (*2,2). Pour l'oral tout venant, l'augmentation est un petit peu plus importante pour le nombre d'occurrences de dépendance à distance (*9,6) que pour le nombre de mots (*7,9).

	CFPP	CFPP + CRFP
nb mots	340 000	760 000
nb occurrences LDD	13	28

TABLEAU 34. Augmentation parallèle du nombre de mots et du nombre d'occurrences de dépendance à distance dans un corpus [oral, privé, entretiens]

	CFPP	oral
nb mots	340 000	2 700 000
nb occurrences LDD	13	125

TABLEAU 35. Augmentation parallèle du nombre de mots et du nombre d'occurrences de dépendance à distance dans un corpus général [oral tout venant]

Lorsque le nombre d'occurrences de constructions de dépendance à distance augmente, le lexique des verbes ponts s'étend. La corrélation entre nombre d'occurrences et étendue du

[87] La variation de l'orthographe est souvent présente dans les pages Web de type forum et blog. Elargir les requêtes augmentait les chances de repérer des occurrences de dépendances à distance

lexique peut être prouvée par le test suivant : lorsqu'on sélectionne aléatoirement n occurrences de dépendance à distance (10, 20, 30, etc.) parmi les 125 occurrences d'oral et ce à 100 reprises, la moyenne montre clairement un accroissement du lexique verbal. (FIGURE.19).

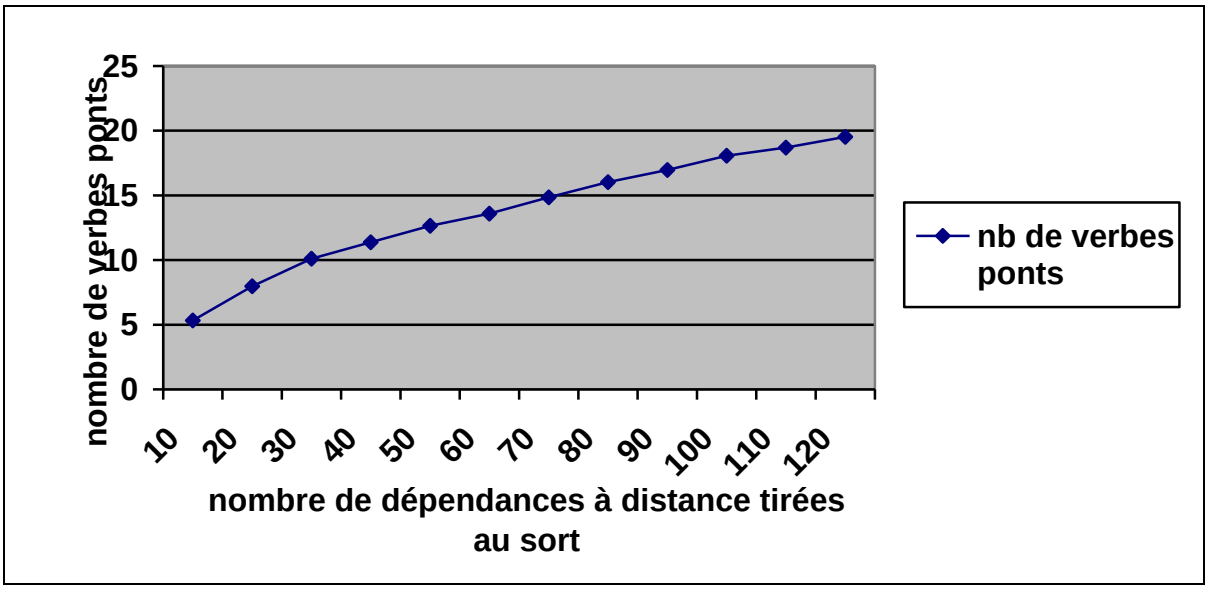

FIGURE.19 Corrélation entre le nombre d'occurrences de dépendance à distance et la richesse du lexique verbal

Ainsi, comme le Web est une ressource gigantesque, il doit y avoir beaucoup d'occurrences de dépendances à distance, ce qui explique qu'il y ait un lexique de verbes ponts beaucoup plus riche. Enfin, le corpus, majoritairement issu d'échanges provoqués, ne semble pas être non-représentatif de la langue habituelle des locuteurs puisque plus de 70% des 133 verbes qui sont ponts uniquement sur le Web produisent au maximum 5 occurrences de dépendances à distance dans les 400 premiers résultats fournis par Google. Pour comparaison, le verbe « vouloir », qui est très productif de dépendances à distance dans les corpus, fournit 96 occurrences de dépendances à distance dans les 400 premiers résultats fournis par Google. Le verbe qui est le plus productif de dépendance à distance sur le Web est « il arrive que », qui présente 36 occurrences (286). Ce verbe aurait certainement pu apparaître dans d'autres corpus.

(286) Tournées, concerts, sur la route de la Californie, Delia se sent une sorte de Janis Joplin avec qui il arrive qu'on la confonde. (Web, institutionnel)[88]

[88] http://www.amazon.fr/Retour-%C3%A0-Cayro-Dorothy-Allison/dp/2264029706, consulté le 03/08/11

Nous pouvons conclure qu'une grande majorité de verbes à complétive peut être employée en verbe pont. Si certains ne sont pas présents dans les corpus, c'est que leur emploi en pont est peu fréquent. Tous les verbes à complétive n'ont donc pas la même propension à apparaître dans une construction de dépendance à distance.

Nous allons chercher la raison qui fait qu'un verbe apparaît beaucoup, peu ou pas en verbe pont dans les corpus.

3.3.3 Sélection des verbes ponts parmi les verbes les plus fréquents ?

Nous allons examiner l'hypothèse d'une relation entre fréquence du verbe dans la langue et fréquence du verbe en position de pont. Nous supposons au départ que tous les verbes auraient la même proportion d'emploi en dépendance à distance (par exemple 0.5% de leurs emplois). Ainsi, les verbes les plus fréquemment utilisés dans la langue seraient également plus fréquents en verbe pont que les autres. Etant donné que les constructions de dépendance à distance sont rares, les verbes les moins employés seraient absents des corpus.

Les fréquences totales sont calculées sur les corpus oraux CRFP, PFC, CFPP, C-ORAL-ROM, CTFP et CorpAix uniformisés et étiquetés avec TreeTagger entraîné sur l'oral (voir §2.1.4) ; et sur huit tranches du corpus écrit CERF étiquetées avec Cordial.

- Rapport à la frequence globale : « Verbe »

Dans un premier temps, nous allons observer les fréquences de tous les verbes et voir dans quelle partie du classement apparaissent les verbes ponts.

Dans le graphique suivant (FIGURE.20), les verbes ont été ordonnés du plus fréquent au moins fréquent. Les rangs sont en abscisse et les fréquences en ordonnée (échelle logarithmique). Nous observons que les verbes attestés en ponts (en rouge) sont sélectionnés parmi les verbes les plus fréquents. Les verbes des premiers rangs qui ne sont pas ponts sont « avoir », « faire » et « pouvoir ». Le verbe attesté en pont le moins fréquent dans les corpus est « présumer ». Ce verbe fait tout de même partie de la première moitié du classement puisque son rang est de 1492 sur 6444, ou sur 5131 si l'on écarte les hapax.

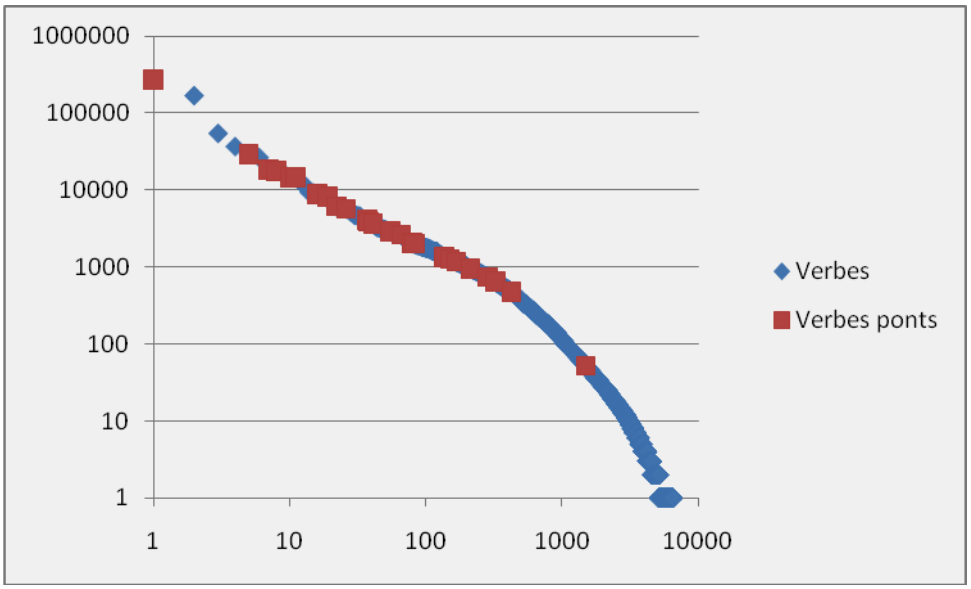

FIGURE.20 Représentation des verbes ponts parmi les verbes des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique)

La probabilité pour que les verbes attestés en pont soient répartis de cette façon aléatoirement est quasiment nulle (une chance sur $7*10^{26}$). En effet, nous observons parmi les 500 verbes les plus fréquents la présence de 24 verbes pouvant être ponts. Dans une distribution uniforme, il n'y en aurait eu que 2.

- Rapport à la frequence « Verbe+complétive »

Dans un deuxième temps, nous allons observer les fréquences des verbes immédiatement construits avec une complétive et voir dans quelle partie du classement apparaissent les verbes ponts (FIGURE.21). Ici encore, nous observons que les verbes qui ont des emplois en pont sont parmi les verbes à complétive les plus fréquents des corpus. Les verbes à complétive des six premiers rangs apparaissent dans les constructions de dépendance à distance. Le dernier verbe pont, « désirer », est classé au 189ème rang sur 871, ou sur 456 si l'on écarte les hapax.

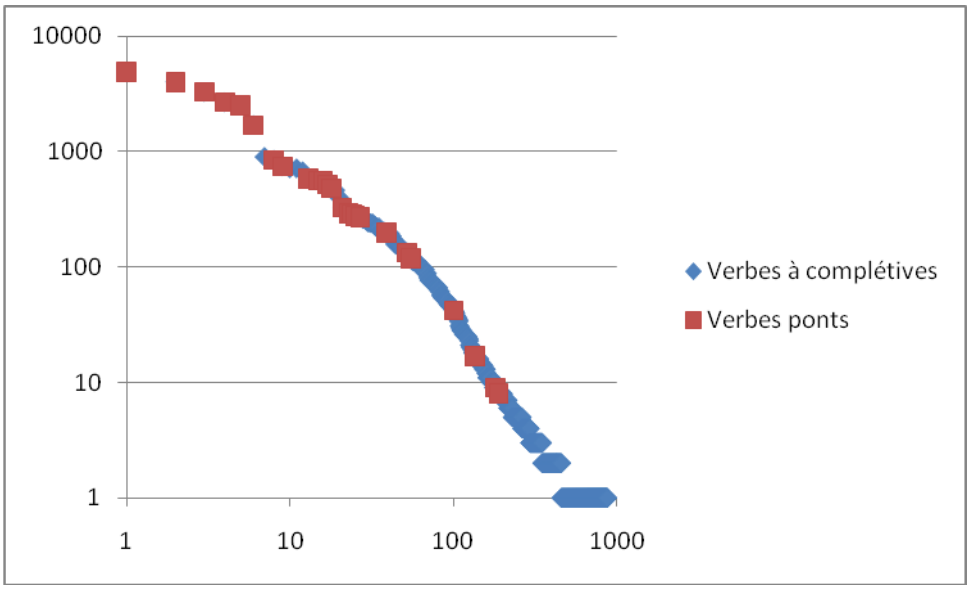

FIGURE.21 Représentation des verbes ponts parmi les verbes à complétive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique)

Comme précédemment, la probabilité pour que les verbes attestés en pont soient répartis de cette façon aléatoirement est quasiment nulle (une chance sur $1{,}6*10^{25}$). Nous observons parmi les 50 verbes à complétive les plus fréquents la présence de 19 verbes pouvant être ponts. Dans une distribution uniforme, il n'y en aurait eu que 2 ou 3.

Ainsi, malgré les erreurs d'étiquetage des verbes et du lemme « que » noté conjonction à la place d'adverbe, ce qui allonge la liste de verbes à complétive et biaise le rang des verbes, nous pouvons affirmer que les verbes attestés en pont font partie des verbes à complétive les plus fréquents.

Ce résultat peut être interprété comme une contrainte cognitive de processing. Ceux qui ont étudié ce type de contrainte avancent que la construction de dépendance à distance demande un effort à la fois au locuteur pour la produire et à l'interlocuteur pour l'interpréter. Cet effort ne vient pas tant du fait que le complément antéposé est éloigné de son verbe recteur que du fait qu'il ne doit pas être rattaché au verbe le plus proche. Le fait d'avoir en verbe pont un verbe très fréquent avec une complétive ouvre dès sa production une fenêtre V2 et laisse la possibilité d'une interprétation à distance.

S'il s'avère exact que les verbes ponts sont globalement sélectionnés parmi les verbes (ou les verbes à complétive) les plus fréquents, deux problèmes subsistent. Le premier est que

certains verbes à complétive très fréquents ne sont pas attestés en pont. Le second est que la fréquence d'un verbe à complétive n'implique pas la fréquence du verbe pont.

- Verbes non-attestés en pont = verbes peu fréquents ?

Une partie des verbes n'est pas attestée en verbe pont, ni dans les corpus, ni sur le Web en français contemporain de France (et à un instant T). Cela ne signifie pas qu'ils sont exclus de cette position. Certains verbes apparaissent sur le Web dans des sources francophones ou anciennes :

(287) *C'est pour renforcer cet élan que **j'annonce**, qu'à partir de l'automne prochain, la bourse d'équipement servie annuellement aux stagiaires passera de 300 DA à 2.000 DA.*[89] *(Web, institutionnel, Algérie)*

(288) *J'ai des difficultés avec la chronologie, je le sais. Je mélange les faits ou les colle les uns aux autres. Mais ce que **je raconte** qu'ils ont fait à Christine a été contrôlé et semble correspondre. Ma relation des faits est même plus précise que l'ancien dossier judiciaire.*[90] *(Web, blog littéraire, Belgique)*

(289) *Cette conscience du caractère épuisable des ressources naturelles, remarquable pour son époque, lui fait déclarer que l'homme n'ayant pas créé la matière naturelle, « c'est de cette matière que **je soutiens** qu'il ne peut y avoir que la possession et l'usage, sous la condition permanente du travail, abandonnant pour un moment [au travailleur] la propriété des choses qu'il a produites ».*[91] *(Web, institutionnel, 1840)*

(290) *lorsqu'un homme me loue son service d'une année pour une certaine somme, cette somme est dans l'intention des parties contractantes, le prix des services effectifs que **je compte** qu'il me rendra ; qu'ainsi je ne dois pas le prix entier, lorsque je n'ai pas reçu tous les services.*[92] *(Web, institutionnel, 1838)*

(291) *Pour montrer que, suivant Hippocrate, les tempéraments des saisons diffèrent les uns des autres par la chaleur, par le froid, par le sec et par l'humide, j'ai rassemblé plusieurs passages*

[89] Allocution du Président de la République d'Algérie lors de la rencontre nationale sur la Formation, Sidi Bel-Abbès, 4 mars 2009 : http://www.el-mouradia.dz/francais/Discours/2009/03/D040309.htm, consulté le 22/02/11

[90] http://old.radicalparty.org/belgium/x1_morg6.htm, consulté le 04/08/11

[91] Proudhon, 1840, *Qu'est-ce que la propriété ?* (chap. II, §4) dans : http://www.plusloin.org/refractions/refractions7/errinjaquet.htm, consulté le 01/03/11

[92] M. Toullier, 1838, *Le droit civil fraançais* (vol. 9) : http://books.google.fr/books?id=NYgPAAAAQAAJ&pg=RA1-PA265&lpg=RA1-PA265&dq=%22que+je+compte+qu%22&source=bl&ots=OHZUCO7BJ3&sig=ot51Xlh23D1kqJzimEkGblPwLUI&hl=fr&ei=8DwsTuXcCciEOoymhdsK&sa=X&oi=book_result&ct=result&resnum=4&ved=0CC8Q6AEwAw#v=onepage&q=%22que%20je%20compte%20qu%22&f=false, consulté le 24/07/11

*dans l'ouvrage où **j'établis** qu'il professe, sur les éléments, la même opinion dans le traité 'De la nature de l'homme' que dans ses autres écrits.*[93] *(Web, institutionnel, 1854)*

*(292) Je ne prétends point rappeller ici le nom de chaque manœuvre & son usage, pour désigner ensuite les nuances que **je juge** qu'il faudroit établir dans la quantité la plus avantageuse de les commettre.*[94] *(Web, institutionnel, 1781)*

*(293) Je répons que c'est précisément ce qu'ont fait la plupart de ceux qui se sont voulu mêler de policer les Nations ; ils ont cru, ou que l'homme étoit naturellement tel qu'ils l'ont trouvé à la naissance de leurs projets, ou qu'il devoit être ce que **je prouve** qu'il n'est point.* [95] *(Web, institutionnel, 1755)*

Ces verbes pourraient certainement apparaître également en français de France contemporain. Puisque les verbes attestés en pont dans le corpus sont majoritairement des verbes d'usage fréquent, on pourrait penser que les 77 verbes qui n'ont pas été attestés en construction de dépendance à distance, ni dans les corpus, ni sur le Web, sont des verbes très peu fréquents.

Mais il est difficile de définir un seuil au-dessus duquel la probabilité de trouver une occurrence pont pour un verbe soit forte et en-dessous duquel la probabilité de trouver une occurrence pont pour un verbe soit trop faible pour que l'on puisse dire s'il est pont ou non. Si, en moyenne, on compte pour un verbe 6720 occurrences non pont pour 1 occurrence pont, on remarque qu'un verbe apparaissant plus de 6721 fois a tout de même plus de chances de ne pas être attesté en pont dans le corpus (12/20 verbes dans notre cas) que d'être attesté en pont (8/20 verbes). Quant aux verbes à complétive, on compte 1 occurrence pont pour 182 occurrences non-pont en moyenne. Là encore, les verbes à complétive ayant une fréquence supérieure à 183 occurrences dans le corpus sont majoritairement non-attestés en pont (23/42).

Nous observons néanmoins qu'une partie des verbes à complétive non-attestés en pont est aussi fréquente que les verbes à complétive ponts (FIGURE.22). Si l'on écarte les verbes non-attestés en pont qui ont moins de 10 occurrences en construction de complétive dans le

[93] Traduction du grec ancien dans : Daremberg, 1854, *Oeuvres anatomiques, physiologiques et médicales de Galien* (tome 1) : http://remacle.org/bloodwolf/erudits/galien/trois.htm, consulté le 28/02/11

[94] Diderot & D'Alembert, 1781, Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers (vol.2), entrée « ALLONGER » :
http://books.google.fr/books?id=scM_AAAAcAAJ&pg=PA192&lpg=PA192&dq=%22que+je+juge+qu%22&source=bl&ots=o94Mpna6cR&sig=lo_Z8_EP7xYUMvAqMsJP859bl_o&hl=fr&ei=21aTTJuBA9S5jAf_tbDEBQ&sa=X&oi=book_result&ct=result&resnum=2&ved=0CBcQ6AEwATgK#v=onepage&q=%22que%20je%20juge%20qu%22&f=false, consulté le 16/09/10

[95] Morelly, 1755, Code de la Nature, ou le véritable esprit de ses Loix, de tout tems négligé ou méconnu.
http://www.taieb.net/auteurs/Morelly/codeii.html, consulté le 28/06/12

corpus, il reste 40 verbes, le dernier ayant le rang n°176. Parmi ces 40 verbes, huit ont une fréquence supérieure à 100 occurrences dans le corpus : « faire », « démontrer », « prouver », « annoncer », « signifier », « éviter », « découvrir », « persuader ».

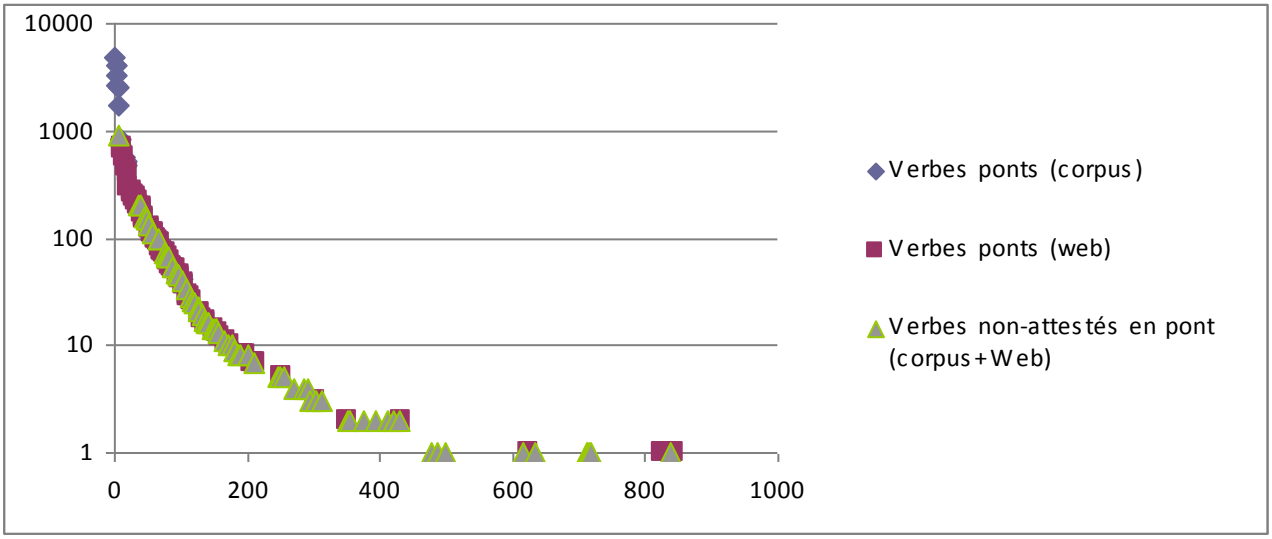

FIGURE.22 Représentation des verbes non-ponts parmi les verbes à complétive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique des fréquences)

Ces huit verbes sont donc relativement fréquents en tant que verbes à complétive. Il peut paraître étrange que l'on ait trouvé aucune occurrence de ces verbes en verbe pont, notamment sur le Web, qui offre un nombre très élevé d'énoncés. Un autre paramètre doit agir et expliquer pourquoi certains verbes à complétive sont attestés en pont et d'autres, pourtant fréquents, ne le sont pas.

- Prévision de la fréquence du verbe pont à partir de sa fréquence en absolu ou en verbe à complétive ?

Le classement comparatif des verbes, des verbes à complétive et des verbes ponts montre que la fréquence en construction pont n'est pas proportionnelle à la fréquence en construction avec complétive. Voyons par exemple les 23 verbes ponts de forme simple[96]. Les verbes sont ordonnés du plus fréquent au moins fréquent, par rapport au nombre d'occurrences des constructions en complétive.

[96] Nous écartons « être » qui regrouperait « être + ADJ » dans le total des verbes mais non dans celui des verbes à complétive et des verbes ponts.

	rang V	occ	rang Vcompl	occ	rang Vpont	occ
dire	1er	30029	1er	4899	4ème	16
penser	7ème	8187	2ème	3272	3ème	25
savoir	3ème	17882	3ème	2668	5ème	12
croire	8ème	6125	4ème	2515	7ème	5
falloir	5ème	14491	5ème	1695	2ème	37
trouver	6ème	8815	6ème	847	7ème	5
voir	2ème	18371	7ème	744	9ème	3
vouloir	4ème	14630	8ème	585	1er	88
rappeler	14ème	2634	9ème	561	14ème	1
comprendre	9ème	5695	10ème	558	14ème	1
estimer	18ème	1119	11ème	520	14ème	1
sembler	12ème	3624	12ème	480	9ème	3
souhaiter	15ème	2046	13ème	326	6ème	6
remarquer	19ème	951	14ème	293	14ème	1
supposer	20ème	736	15ème	288	14ème	1
imaginer	17ème	1359	16ème	277	14ème	1
attendre	11ème	3887	17ème	271	14ème	1
sentir	13ème	2898	18ème	197	9ème	3
prévoir	16ème	2006	19ème	118	12ème	2
entendre	10ème	4050	20ème	116	14ème	1
présumer	23ème	52	21ème	17	12ème	2
ressentir	21ème	653	22ème	9	14ème	1
désirer	22ème	481	23ème	8	14ème	1

TABLEAU 36. Rang et fréquence de 23 verbes dans trois situations : verbe, verbe+complétive, verbe pont

Les verbes « dire », « savoir » et « falloir » sont dans tous les cas dans le haut du classement (dans les cinq premiers). Le verbe « penser » possède un rang haut quasi-identique en construction à complétive et en pont.

Les verbes « remarquer », « supposer », « imaginer », « prévoir », « présumer », « ressentir » et « désirer » sont globalement dans le bas du classement, quelle que soit leur construction (dans les dix derniers). Les verbes « attendre » et « entendre » sont également dans le bas du classement pour les constructions en complétive et en pont.

Entre ces deux groupes, le classement des verbes varie. Le verbe « vouloir », par exemple, est au premier rang en verbe pont mais seulement au 8ème rang des verbes à complétive. De même, « souhaiter » est au 6ème rang en verbe pont mais seulement au 13ème rang des verbes à complétive. On ne peut donc pas prédire la fréquence d'un verbe pont à partir de son rang en verbe à complétive ou en verbe général. La proportion d'occurrences en pont par rapport au nombre d'occurrences total ou en complétive diffère considérablement d'un verbe à l'autre. Les verbes attestés en pont dans le corpus possèdent 1 occurrence pont pour au minimum 25 occurrences non-ponts (« présumer ») et au maximum 45 720 occurrences non-ponts (« être »). Pour un verbe à complétive, l'écart reste considérable : on compte 1 occurrence pont pour au minimum 6 occurrences non-ponts (« vouloir ») et au maximum 667 occurrences non-ponts (« être »).

Afin d'expliquer (1) pourquoi certains verbes fréquents ne sont pas attestés en pont et (2) pourquoi certains verbes sont très utilisés en pont et d'autres moins, nous allons examiner successivement les caractéristiques syntaxiques et sémantiques des verbes attestés et non-attestés en pont.

3.3.4 Sélection des verbes ponts parmi une classe syntaxique ?

Du point de vue des propriétés syntaxiques globales du lexème (et non de sa réalisation en pont), rien ne différencie les verbes ponts des verbes non-ponts. Dans les deux groupes se trouvent des verbes qui appartiennent à la quasi-totalité des tables définies par (Gross, 1975) pour les constructions « V+que » : les tables 6, 9, 10, 17 et 18.

Les verbes « juger » (non-pont) et « voir » (pont) sont dans la table 6 et acceptent la construction « N0 V qu P ».

(294) Il juge que Marie est belle

(295) Je vois que Paul travaille (Gross, 1975, p. 177)

Les verbes « démontrer » (non-pont) et « dire » (pont) sont dans la table 9 et acceptent la construction « N0 V qu P à N2 ».

(296) J'ai démontré à Pierre que Paul était venu (Gross, 1975, p. 191)

(297) Il dit à Pierre qu'il pourra le raccompagner

Les verbes « mettre » (non-pont) et « souligner » (pont Web) sont dans la table 10 et acceptent la construction « N0 V qu P prép N2 ».

(298) Lire ce rapport a mis sous ses yeux que tu avais fait cela (Gross, 1975, p. 194)

(299) J'ai souligné dans le texte qu'il avait fait cela. (Gross, 1975, p. 194)

Les deux dernières tables, les tables 17 et 18 comportent peu de verbes (respectivement 18 et 20 verbes). Certains verbes ponts appartiennent à la table 17 (« paraître », « falloir ») et acceptent la construction « il_impersonnel V (E+prép1 ce) qu P (E+prép2 N2) ». Les verbes non-ponts sont répartis dans les autres tables.

(300) Il paraît que Paul est arrivé (Gross, 1975, p. 212)

Aucun verbe pont n'appartient à la table 18, qui rassemble les verbes ayant pour construction « N0 V prép1 N1 prép2 N2 (E+pour) que P subj », c'est-à-dire des verbes qui ont deux compléments prépositionnels et une complétive au subjonctif, ou des verbes à deux compléments essentiels auxquels on a adjoint des compléments phrastiques en *pour*.

(301) Max a reçu de Luc comme mission qu'il examine les abords du fort

(302) Max a décrit un chauffeur à Léa, pour être sûr qu'elle le reconnaisse

(303) Ca a servi à Paul de prétexte pour qu'il ne vienne pas (Tolone, 2011, p. 99)

On pourrait supposer que les verbes de cette table sont exclus de la liste des verbes ponts, mais comme nous n'avons recherché que deux de ces verbes sur le Web : « poser » et « recevoir », nous ne pouvons tirer aucune conclusion. Même si effectivement tous les verbes de la table 18 étaient non-ponts, cela ne suffirait pas à différencier les ponts des non-ponts puisque des verbes non-ponts appartiennent à d'autres tables que la n°18.

Le critère syntaxique n'explique donc pas pourquoi certains verbes sont ponts de manière très fréquente ou non, et pourquoi d'autres ne le sont pas. Nous allons examiner si le critère sémantique peut fournir une explication.

3.3.5 Sélection des verbes ponts parmi une classe sémantique ?

Comparons les 15 verbes non-ponts les plus fréquemment attestés en verbes à complétive et les 15 verbes ponts également les plus fréquemment attestés en verbes à complétive[97] (TABLEAU 37).

	Non-pont	Pont
1	**(>500)** Faire	**(>1000)** Croire
2	**(>100)** Annoncer	Dire
3	Découvrir	Etre
4	Démontrer	Falloir
5	Eviter	Penser
6	Persuader	Savoir
7	Prouver	**(>500)** Comprendre
8	Signifier	Considérer
9	**(>50)** Compter	Espérer
10	Parler	Estimer
11	Raconter	Montrer
12	Réaffirmer	Rappeler
13	Révéler	Trouver
14	Soutenir	Vouloir
15	Stipuler	Voir

TABLEAU 37. Liste des quinze verbes à complétive les plus fréquents qui sont non-ponts et des quinze verbes à complétive les plus fréquents qui sont ponts

[97] Nous aurions pu comparer les verbes non-ponts avec les verbes les plus fréquents en construction de pont. Nous aurions également pu nous appuyer uniquement sur les corpus. Cela aurait changé peu de choses puisque les 8 des 9 verbes ponts les plus fréquents font partie de la liste des 15 verbes à complétive attestés en pont les plus fréquents. Pour rappel les 9 verbes ponts les plus fréquents sont : *vouloir, falloir, penser, dire, savoir, souhaiter, croire, trouver, voir.* Seul *souhaiter* est absent de la liste.

Du point de vue sémantique, certains verbes non-ponts semblent proches des verbes ponts : « annoncer » (non-pont) se rapproche de « dire » (pont_corpus) ou de « déclarer » (pont_Web), « découvrir » (non-pont) de « remarquer » (pont_corpus). Les dictionnaires qui fournissent des définitions courtes et généralement par synonymie, comme le dictionnaire en ligne de tv5[98], renvoient d'un verbe non-pont à verbe pont (TABLEAU 38).

Verbe non-pont	**Définition**	**Renvoi vers un verbe pont**
Annoncer	*faire savoir; faire connaitre d'avance; indiquer*	Indiquer
Découvrir	*trouver/révéler/dénoncer; voir/apercevoir*	Trouver, voir/apercevoir
Signifier	*être le signe de; exprimer/vouloir dire; déclarer positivement/notifier; notifier par voie de justice*	Vouloir dire ?, déclarer
Compter	*faire un calcul; mettre au nb de; mesurer; se proposer/projeter; estimer/croire/prévoir*	Estimer/croire/prévoir Proposer ?
Réaffirmer	*affirmer de nouveau*	Affirmer
Révéler	*faire savoir ce qui est inconnu, secret; indiquer/témoigner*	Indiquer

TABLEAU 38. Définition des verbes non-ponts (parmi les 15 verbes à complétive non-ponts les plus fréquents)

Inversement, la définition de certains verbes ponts comporte des synonymes non-ponts (TABLEAU 39).

Verbe pont	**Définition**	**Renvoi vers un verbe non-pont**
Croire	*tenir pour vrai; estimer/juger/penser;*	Juger
Considérer	*juger/estimer*	juger
Montrer	*faire voir/faire connaitre/faire paraitre/ démontrer*	démontrer
Trouver	*juger; éprouver*	juger

TABLEAU 39. Définition des verbes ponts (parmi les 15 verbes à complétive ponts les plus fréquents)

[98] http://dictionnaire.tv5.org/dictionnaire, consulté le 02/07/12

Il y a donc un mélange de termes ponts et non ponts dans les définitions des deux classes de verbes et de multiples renvois de l'une à l'autre.

Pour les autres verbes non-ponts, dans les définitions desquelles il n'y a pas de synonymie avec l'autre classe, il semble que la notion de 'description d'une action ou d'une activité' ou encore de 'relation objective entre les faits' soit généralement plus présente. On peut construire certains (« démontrer », « persuader », « soutenir », « prouver ») avec des adverbes comme « longuement », « délibérément », « volontairement » qui soulignent le caractère d'activité du procès. D'autres indiquent une relation de causalité entre deux faits (« faire », « éviter ») et décrivent donc des activités ou des relations entre activités. Or les dépendances à distance avec des verbes ponts très descriptifs semblent impossibles (304) : Les verbes tels que « entraîner » et « faire en sorte que » semblent appuyer cette hypothèse.

(304) *Le tremblement de terre a fait que les bateaux ont dû quitter le port*

**le port que le tremblement de terre a fait que les bateaux ont dû quitter*

Les verbes qui peuvent éventuellement apparaître dans certains contextes sont très difficiles à accepter lorsqu'ils régissent des compléments qui mettent en valeur leur sens descriptif (305).

(305) *Je vais démontre par l'absurde en trois étapes que la cigarette contient un produit toxique =>*

?? le produit toxique que je démontre par l'absurde en trois étapes que la cigarette contient

A l'inverse, les verbes ponts seraient plutôt l'expression d'une modalité, c'est à dire soit une attitude subjective du sujet à l'égard d'un procès : « voir », « savoir », « penser », « vouloir », « comprendre », « présumer » soit une évaluation de ses conditions de réalisation « falloir ». Plusieurs listes ou classements de verbes modaux existent. Nous en avons retenu deux, en les sélectionnant pour leur plus grande exhaustivité[99]. Le premier est fourni par un professeur de français de lycée[100]. Il comporte 39 verbes. Ces verbes sont répartis en deux groupes : les

[99] Notre objectif ici est simplement de vérifier que les verbes ponts sont des modaux. Nous établirons un classement plus précis des verbes ponts dans le chapitre suivant, notamment en nous appuyant sur le classement syntaxique de (Chu, 2008).

[100] M.Théry, Lycée Saint Rémi de Roubaix, classe de seconde, 2004. Ref. bibliog. : RIEGEL (M.) et al., *Grammaire méthodique du français*, PUF 1996; MAINGUENEAU (Dominique), *L'Analyse du discours*, Hachette, 1991, p.108,114; PAGES (A.) et PAGES-PINDON (J.), *Le français en seconde*, Nathan, 1992, et *Le français au lycée*, Nathan, 1984; CREPIN et al., *Français. Méthodes et techniques*, Nathan, 1996 ; Documents d'accompagnement du français au lycée, CNDP, 2001 ; Berthelier (M) et al., *Français-Méthodes, 2de-1re*, Hachette, 2004.

http://www.saintremi.com/fichesfrancais/5_outils_de_la_langue/5_2_grammaire/MODALISATIONS_tableau_sept_2004.doc, consulté le 04/07/12.

évaluatifs et les affectifs. Pour les évaluatifs, des sous-groupes précisent si la modalisation porte sur la vérité du propos (degré de vérité, degré de réalité ou de certitude), sur une appréciation de l'évènement (jugement de valeur) ou sur une volonté ou une nécessité qui entoure l'action. Le second classement retenu est celui de (Bally, 1942). Il comporte 23 verbes. Des précisions sont parfois ajoutées sur le type de modal : déclaratif, négatif, type 'vouloir'. Dans ces deux classements, la quasi-totalité des verbes modaux cités sont des verbes ponts (TABLEAU 40).

	Ponts	**Non-ponts**
Modaux de M. Théry	38	1
Modaux de Ch. Bally	20	3

TABLEAU 40. Répartition des modaux en ponts/non-ponts

Dans le premier classement, les modaux les plus nombreux sont ceux qui permettent au locuteur de prendre plus ou moins en charge les propos qu'il tient (degré de vérité, de certitude, de réalité). Les verbes modaux, et donc dans une large mesure les verbes ponts, mettent en évidence la distance, le degré de prise en charge, la subjectivité. Cette valeur est très visible dans les énoncés humoristiques. En voici deux exemples.

Le premier exemple (306) illustre la distance. On y définit la modestie, non seulement par le vocabulaire, mais également par la structure pont, volontairement exagérée par la récursivité.

(306) La véritable modestie consiste à ne jamais se prendre pour moins que ce qu'on estime qu'on croit qu'on vaut (CERF, DIVS).

Le second exemple (IMAGE 7) est également humoristique. Il illustre le fait que l'emploi des modaux et des ponts accentue l'importance du point de vue. Une affirmation est vraie uniquement pour une personne ou un groupe de personnes donné. Lorsque la dépendance à distance disparaît, c'est une situation objective qui est présentée.

IMAGE 7. Le verbe pont comme reflet de la subjectivité (Web, humour)[101]

L'hypothèse de (Erteschik-Shir, 1973) (voir p.57 et ss.) et de (Godard, 1980) (voir p.61 et ss.) pour expliquer la distinction entre verbes ponts et verbes non-ponts était que les verbes ponts seraient 'simples' et les verbes non-ponts 'complexes', la complexité étant définie par l'ajout d'un trait sémantique indépendant. Si l'on reprend les verbes du champ sémantique de DIRE, cités par (Godard, 1980), 22 font partie de la liste de référence des verbes à complétive : 4 sont étiquetés 'complexes' par (Godard, 1980) et 18 'simples'.

CLASSE	DIRE
Verbes complexes (=non-ponts ?)	(condition physique : force ou débit) Crier, murmurer (état d'esprit du locuteur) Pleurer (moyen de transmission) Ecrire
Verbes simples (=ponts ?)	(acte de parole) Avouer, confesser, jurer, promettre (volonté de faire savoir) Annoncer, apprendre, insinuer, mentionner, préciser, proclamer, révéler, souligner, suggérer (lien avec ce qui précède) Ajouter, confirmer, nier, rappeler, répéter

TABLEAU 41. Degré de complexité de certains verbes de DIRE d'après (Godard, 1980)

Les attestations semblent confirmer que les verbes complexes sont difficilement ponts : seul « écrire » est attesté, sur le Web et dans 2 occurrences (307)(308).

[101] http://www.demotivateur.fr/image/Leschats-22315, consulté 12/04/12

(307) *Alors, comment lire l'Evangile ? Tout d'abord, je dirais que la première "grille de lecture", la seule qui soit indispensable en permanence, est de recevoir ce texte pour ce qu'il est : le témoignage des apôtres. Ensuite, d'autres approches pourront se greffer sur celle-là, mais jamais, en aucun, cas prendre la place de la première ou l'exclure. Peut-être faut-il préciser ce que signifie cette expression "le témoignage des apôtres" (en effet, diverses écoles d'exégèse l'utilisent dans le but d'en contester la portée). [...] C'est donc en ce sens là que* ***j'écris*** *qu'il faut prendre le texte comme le témoignage des apôtres : des gens qui savent de quoi ils parlent, qui témoignent de ce qu'ils ont vu.*[102] *(Web, blog)*

(308) *Le hors-série en rupture de stock! J'ai reçu pas mal de mails concernant le hors-série que* ***j'ai écrit*** *qu'on ne trouve plus nulle-part. Je me suis donc renseignée auprès de mon éditeur et en effet il est en rupture de stock. Aucune réimpression n'est envisagée pour le moment et c'est bien dommage vu la demande que j'ai chaque semaine!*[103] *(Web, blog)*

Le petit nombre de verbes complexes recherchés sur le Web ne permet toutefois pas d'affirmer que le paramètre de complexité sémantique est lié à la propriété non-pont ou pont peu fréquent des verbes. En outre, la classe des verbes simples ne confirme pas non plus le lien entre non-complexité et possibilité de construction pont. Sur les 18 verbes de la classe, 8 n'apparaissent pas du tout en pont, y compris sur le Web (45%). Ils sont répartis dans l'ensemble des sous-types des verbes simples de DIRE : acte de parole, volonté de faire savoir, lien avec ce qui précède.

CLASSE	**DIRE**	
	NON-PONTS	**PONTS**
Verbes complexes	Crier, murmurer Pleurer	Ecrire
Verbes simples	- Confesser, jurer - Annoncer, insinuer, mentionner, proclamer, révéler - Nier	- Avouer, promettre - Apprendre, préciser, souligner, suggérer - Ajouter, confirmer, rappeler, répéter

TABLEAU 42. Mise en regard du degré de complexité de certains verbes de DIRE et de leur statut pont/non-pont

Même parmi les verbes simples attestés en ponts, certains sont très peu fréquents : 1 occurrence pour « promettre » (309), « préciser » (310), « souligner » (311), « suggérer » (312), « ajouter » (313), « rappeler » (314), 2 occurrences pour « répéter » (315)(316).

[102] http://cigales-eloquentes.over-blog.com/article-lire-l-evangile-64549520.html, consulté le 22/02/11.
[103] http://mlabroderieblog.canalblog.com/archives/mon_hors_serie/index.html, consulté le 22/02/11.

(309) *Celle avec qui **je promets** qu'un jour on fera le tour du monde.*[104] *(Web, forum)*

(310) *Le parc du Musée Caillebotte à Yerres (où **je précise** qu'il n'y a que des reproductions pour ceux qui ne le sauraient pas) est un endroit charmant où je viens me promener de temps à autre.*[105] *(Web, blog)*

(311) *Maintenant on peut prendre le mot travail dans un sens très large et général, englobant alors aussi les machines, qu'il a fallu construire mais surtout inventer, et c'est ici que **je souligne** qu'il y a du travail en quelque sorte gratuit qui n'est jamais rétribué à sa juste valeur ni encore clairement pris en considération.*[106] *(Web, forum)*

(312) *Bonjour, Je suis un simple usager du train et je vous invite à prendre connaissance et à soutenir une nouvelle vision du train où **je suggère** qu'en plus d'être un moyen de transport à grande vitesse nous avons la possibilité de faire du train un moyen de SAVOIR À GRANDE VITESSE.*[107] *(Web, forum)*

(313) *Je pensais plutôt a des "relevés mensuels" dont **j'ajoute** qu'il est question dans d'autres parties du texte.*[108] *(Web, forum)*

(314) *Auparavant, les titularisations étaient prononcées dans le corps des adjoints d'enseignement sans concours, mais sur liste d'aptitude. Ce corps a été mis en voie d'extinction. La seule solution d'accueil pour la titularisation à laquelle **je rappelle** qu'il était déjà procédé précédemment - est l'entrée dans le corps des professeurs certifiés. (CERF, institutionnel)*

(315) *Ces études aboutiront à la conclusion que j'ai donnée dans l'Énigme de Jésus-Christ, que **je répète** qu'il ne faut pas se lasser de répéter, car elle est le fondement vrai de toute l'histoire, de toute la mythologie du christianisme*[109] *(Web, institutionnel)*

(316) *Mais ce qui importe, en tout cas, c'est qu'on ne « dénature » point la chaire, et, au contraire, qu'on en fonde plutôt, et au plus tôt, Hue autre â Paris, où **je répète** qu'il n'y en a point.*[110] *(Web, institutionnel)*

Il est donc difficile d'affirmer l'existence d'une contrainte sémantique portant sur les verbes ponts. Il ne semble pas que les verbes non-ponts soient très différents des verbes ponts. Les

[104] http://people-book.heavenforum.com/t388-summ-if-you-want-to-be-loved-be-lovable-honey, consulté le 01/03/11

[105] http://dandylan.canalblog.com/archives/2009/05/04/13602099.html, consulté le 04/08/11

[106] http://www.agoravox.fr/actualites/economie/article/pourquoi-le-travailler-plus-met-en-52697, consulté le 27/07/11

[107] http://www.liberation.fr/economie/0102608276-reaction-sur-la-sncf-perd-son-monopole-du-transport-de-voyageurs:sortBy-lastCommented:page-1:target-1478873#1478873, consulté le 01/03/11

[108] http://forum.wordreference.com/showthread.php?t=1491206, consulté le 22/02/11

[109] http://www.mediterranee-antique.info/Religion/Masse/T_2/M_205.htm, consulté le 01/03/11

[110] http://www.archive.org/stream/annalesinternat02unkngoog/annalesinternat02unkngoog_djvu.txt, consulté le 01/03/11

définitions des deux groupes montrent des rapports de synonymie. La complexité sémantique des verbes, censée les exclure de la construction pont, n'est pas non plus clairement prouvée. Les données sont trop peu nombreuses lorsque l'on effectue le partage en classes. En outre, la restriction des requêtes effectuées sur le Web (pronom sujet de première personne, verbe au présent, proximité des éléments constitutifs de la structure : « pronom + sujet + verbe + qu ») limite le nombre d'occurrences repérées. Néanmoins, et pour les mêmes raisons, l'absence de contrainte sémantique n'est pas non plus établie. Les verbes qui sont les plus difficiles en position pont non-ponts ont un contenu descriptif plus important, ou marquant le lien entre deux descriptions que les verbes ponts. Notre conclusion sur ce point rejoint celle de (Abeillé, 2007) :

> « L'on pourrait souhaiter déduire le caratère pont ou non pont des prédicats de leurs propriétés sémantique. Intuitivement plus ils sont simples sémantiquement, meilleure est l'extraction, mais, faute d'une sémantique lexicale véritablement opérationnelle, nous avons préféré noter le contraste lexicalement (… par un trait booléen « pont »). » (Abeillé, 2007, p. 246)

Nous verrons que cette caractéristique non descriptive tient autant au sémantisme du verbe qu'à l'ensemble de la construction dans laquelle il est employé.

3.3.6 Synthèse

Les occurrences relevées indiquent que 34 tournures verbales peuvent apparaître en position de pont mais le lexique des verbes ponts attestés dans les corpus n'est qu'une partie des verbes ponts possibles. Les recherches effectuées sur le Web montrent que de nombreux verbes à complétive peuvent être ponts, dans divers types de sources (blog, forum, sites institutionnels). La classe des verbes ponts ne recouvrant pas celle des verbes à complétive, on ne peut poser une simple règle syntaxique pour obtenir les constructions à ponts à partir des constructions complétives. Une bonne description suppose de pouvoir caractériser la classe des verbes possibles en pont soit en termes syntaxiques soit en termes sémantiques.

D'un point de vue syntaxique, les verbes ponts n'ont comme point commun que le fait d'être des verbes à complétive. Ils ne relèvent pas d'un type particulier de constructions complétives parmi celles distinguées par (Gross, 1975). Il n'y a donc pas de contraintes syntaxiques au

niveau des structures. Il faudra cependant vérifier s'il n'y en a pas au niveau de la réalisation des constructions de dépendance à distance.

Du point de vue sémantique, plusieurs hypothèses proposées dans les études précédentes pour caractériser les verbes ponts ne sont pas validées. Les verbes ponts en jeu ne sont pas uniquement sélectionnés parmi les verbes non factifs. Toutes les catégories liées à la présupposition fournissent des verbes attestés en ponts : factifs positifs et négatifs, semi-factifs, implicatifs positifs, négatifs et simples. Ils ne sont pas restreints aux verbes recteurs faibles, mais tous les verbes recteurs faibles sont présents en ponts[111].

De même, les verbes ponts ne sont pas uniquement sélectionnés parmi les verbes « sémantiquement simples », au sens de (Godard, 1980), c'est à dire n'ayant pas de traits sémantiques additifs indépendants par rapport à la classe de référence (« dire », « penser »…). Compte tenu du nombre de données et du nombre d'occurrences repérées, il est difficile d'affirmer s'il se dégage une tendance à la simplicité des verbes ponts, et inversement une tendance à la complexité des verbes non-ponts. D'un point de vue sémantico-pragmatique, il reste à explorer l'hypothèse que les verbes ponts semblent relever du domaine de la modalité (épistémique, évidentielle, évaluative ou affective) tandis que les non-ponts seraient plutôt descriptifs.

Enfin, en termes de contraintes cognitives, on retiendra que les verbes ponts sont des verbes parmi les plus fréquents des verbes à complétive. Comme l'a signalé (Godard, 1980), il est important que le verbe soit « naturel » avec une complétive. D'un point de vue du processing, cela facilite la possibilité d'interpréter le complément antéposé comme dépendant du verbe enchâssé.

Il est fort probable que les différentes contraintes se combinent et se compensent afin de donner la répartition pont *vs* non-pont. Si l'on reprend les 22 verbes précédemment étudiés dans le paragraphe sur la complexité sémantique (voir p.185), on constate que le nombre d'occurrences joue également un rôle dans la possibilité pour un verbe d'être pont. Le verbe

[111] Il faut néanmoins noter que la liste des verbes recteurs faibles du français n'est pas finalisée. L'affirmation que tous les verbes recteurs faibles sont ponts ne vaut donc que pour l'état actuel des recherches sur les verbes recteurs faibles.

complexe qui est attesté en verbe pont, « écrire », a un nombre d'occurrences supérieur à 50 occurrences ; les autres sont largement inférieurs à 50 occurrences (TABLEAU 43).

Nombre d'occurrences du verbe à complétive	Verbes complexes	Pont
50-100	Ecrire : 57	Oui
0-50	Crier : 13, Murmurer : 5 Pleurer : 2	Non

TABLEAU 43. Rapport entre nombre d'occurrences, complexité et position de pont des verbes complexes

Les verbes simples sont répartis en trois groupes en fonction de leur nombre d'occurrences : les verbes qui apparaissent plus de 160 fois appartiennent au lexique des verbes ponts ; les verbes qui apparaissent moins de 50 fois n'appartiennent pas au lexique des verbes ponts ; les verbes qui apparaissent entre 50 et 160 fois forment un groupe mixte, avec des verbes ponts et des verbes non-ponts (TABLEAU 44).

Nombre d'occurrences du verbe à complétive	Verbes simples	Pont
> 160	Rappeler : 561 Préciser : 260 Souligner : 243 Avouer : 221 Apprendre : 221 Ajouter : 173	Oui
50-160	Annoncer : 158 Confirmer : 113 Suggérer : 77 Révéler : 68 Promettre : 52 Jurer : 52 Répéter : 51	Non Oui Oui Non Oui Non Oui
0-50	Mentionner : 25 Proclamer : 16 Nier : 16 Insinuer : 9 Confesser : 3	Non

TABLEAU 44. Rapport entre nombre d'occurrences, complexité et position de pont des verbes simples

Le lexique des verbes ponts est donc construit de façon complexe, et tient compte de plusieurs paramètres, sémantiques, pragmatiques et cognitifs. Ces considérations nous amènent à déplacer notre attention du lexème pont à l'ensemble de la cosntruction pont.

3.4 Autre source de contrainte ? La construction du « pont »

L'ensemble du pont, situé entre le complément antéposé (qui a la forme d'un pronom) et le verbe qui le régit, est constitué au minimum du sujet du verbe pont, du verbe pont et de la conjonction « que/qu' », comme on peut le prévoir du fait que le verbe pont est un verbe à complétive fini[112]. Le dernier élément essentiel du pont est le sujet du verbe de la complétive.

(317) "L'autre rive", ce "grand voyage", dont on pensait qu'il ne reviendrait pas (CERF, PRES)

Ensuite, il serait théoriquement possible que des éléments s'ajoutent à cette séquence minimale : d'autres compléments du verbe pont ou du V2 (objet, ajout), des éléments non régis et des disfluences. Nous allons examiner successivement chacun des éléments constitutifs du pont dans les productions des locuteurs.

3.4.1 Le sujet du verbe pont

Le sujet du verbe pont est très largement réalisé sous forme pronominale. Un pronom personnel est attesté dans 220 occurrences, soit dans 95,2% des cas (318). Les autres sujets sont des groupes nominaux (6 occurrences) (319), des indéfinis (2 occurrences) (320) et des noms propres (1 occurrence) (321). Deux occurrences, qui mentionnent un verbe impersonnel, ne comportent pas de sujet (322).

*(318) et pourquoi **tu** penses qu'ils ont pas voulu (PFC, 13bsa1)*

[112] Nous estimons que les exemples du type « le livre qu'a dit qu'aime le professeur » sont rarissimes.

(319) *En conclusion, il semble que ce que **les femmes** veulent qu'on leur dise, c'est "L'IVG quand vous voudrez, comme vous voudrez, où vous voudrez, et sans compte à rendre à personne !" (CERF, FORM)*

(320) *Le conflit pour le salaire prolonge le rapport de marchandage dans lequel **chacun** sait qu'il a besoin de l'autre : chacun s'efforce de rouler l'autre suffisamment pour avoir fait une bonne affaire, mais pas trop sinon le maintien de la relation deviendrait incertain. (CERF, SCIE)*

(321) *Le contraste entre les deux aurait difficilement pu être plus grand, et mit en sourdine tout l'impact que **Foucault** pensait que son travail devait avoir en France. (CERF, DIVS)*

(322) *c'est peut-être ça aussi l'idéal c'est d'apprendre à parler sans se dire alors là dans quel temps **[0]** va falloir que je mette de cette phrase pour qu'il(s) comprenne(nt) ce que j'ai voulu raconter (CorpAix, 6CELINE)*

Les pronoms sujets font référence à toutes les personnes, mais de manière inégale. La personne la plus représentée dans les ponts est l'interlocuteur (40 « tu », 55 « vous »). Dans le (TABLEAU 45), les singuliers et les pluriels d'une même personne ont été regroupés.

	Sujets pronominaux	**% par rapport aux sujets pronominaux**
P1 (je, nous)	31	14%
P2 (tu, vous)	95	43%
P3 (il, elle, on, ils, elles)	43	19,5%
Il impersonnel	52	23,5%

TABLEAU 45. Répartition des sujets pronominaux du verbe pont, personnels et impersonnels

Cette répartition est totalement différente de celle relevée dans l'ensemble des constructions des corpus. Les statistiques présentées ci-dessous sont celles fournies par Cordial, sur les corpus oraux CRFP, CFPP, CTFP, C-Oral-Rom, CorpAix et PFC et sur les tranches DIVS, FORM, INST, LITC, POLI, PRES et SCIE du corpus écrit CERF[113]. Elles comportent donc certainement des erreurs mais permettent d'indiquer une tendance. Des moyennes ont été calculées pour l'oral et pour l'écrit. Nous présentons ici la moyenne des deux groupes.

[113] La tranche PRAT n'apparaît pas car sa taille était trop importante pour qu'elle puisse être étiquetée par Cordial.

Dans les corpus, les sujets pronominaux représentent en moyenne 47.5% des sujets. La deuxième personne est la personne la moins fréquente (11.2%). Cela confirme les propos de Verhagen (cf p.77) : la forme de la construction n'est pas due à la fréquence de chaque élément. Il semble donc y avoir une spécificité de la construction pont, qui favorise les sujets pronominaux (95.6%) et le pronom personnel de deuxième personne (43.4%). Ce chiffre confirme également que la construction est employée dans un contexte d'intersubjectivité, contexte forcément peu présent dans le corpus écrit, et également peu présent dans certains genres d'oral (situations professionnelles, discours publics, entretiens)[114].

3.4.2 Le verbe pont

Dans le pont, le verbe à complétive est en grande majorité au mode indicatif (219 occurrences, 94,8%) (323). Quelques conditionnels sont aussi attestés (12 occurrences) (324).

(323) Otto Kranz est loin de tout contrôler, contrairement à ce qu'il ***souhaite*** *que vous pensiez. (CERF, LITC)*

(324) Je ne suis pas sur Terre pour plaire ou pour tenter de faire de moi ce qu'un autre ***voudrait*** *que je sois (CERF, LITC)*

Quel que soit le mode, le temps dominant du verbe pont est le présent. Il représente 85,7% des occurrences (325).

(325) le conservateur du musée - a décidé que nous ferions ensemble un itinéraire choisi par lui pour présenter certaines collections remarquables - euh sur lesquelles il ***souhaite*** *que nous insistions (CRFP, PUB-ORL- 1)*

L'écart se creuse entre le présent (198 occurrences) et les autres temps. L'imparfait regroupe 21 occurrences (326), le passé composé 7 occurrences (327), le futur 3 occurrences (328).

(326) Nous nous sommes contentés d'exclure un certain nombre de textes pour ne garder que la littérature qui semblait convenir à l'usage dont nous ***pensions*** *qu'il serait fait du corpus. (CERF, SCIE)*

[114] Dans nos données, des différences sont également notables en fonction notamment de la configuration syntaxique (voir §4.3.1)

(327) *La symbolique de cette oeuvre s'accorde parfaitement avec le lieu où son auteur **a souhaité** qu'elle soit implantée : au-dessus du vieux village de Craonne, sur le petit plateau au débouché de ce qu'on appelle ici la laie de Californie. (CERF, POLI)*

(328) *Tu m'entraînes chaque jour un peu plus dans un monde fait de dépendances dont il **faudra** que je me débarrasse pour être avec "Lui" ! (CERF, LITC)*

Les temps composés sont également peu présents : on relève une occurrence pour le plus que parfait (329), une pour le futur antérieur (330).

(329) *est-ce que vraiment on est condamné chaque fois à avoir une élection avec un enthousiasme et derrière une désillusion et des gouvernants qui reculent par rapport à ce qu'ils **avaient dit** qu'ils feraient (CorpAix, BAYROU)*

(330) *quand on aura atteint cette perspective où euh ben on sera tous salariés de notre société entre guillemets idéale où chacun **aura pris conscience** que son travail c'est plus un travail subordonné (Nancy, retraites)*

Le verbe pont est quasi-systématiquement à la voix active (98,7%) (331). Trois occurrences seulement sont réalisées au passif (332).

(331) *Communiste ou non, il sera possible à chacun de prendre la parole, le stylo ou l'ordinateur pour faire connaître son avis et débattre de ce qu'il **attend** que le Parti communiste français devienne. (CERF, PRES)*

(332) *C'est pour en diminuer l'importance - et non pour se livrer à un quelconque "tour de passe-passe budgétaire" – qu'il est prévu que France Télécom versera une contribution exceptionnelle à l'État.(CERF, INST)*

Parfois, le verbe pont est modalisé. Deux types de modalisation sont possibles. Il arrive que le verbe soit précédé d'un modal (4 occurrences) : pouvoir, falloir, croire (333).

(333) *Membre des Jeunesses contre le racisme en Europe, cofondateur du Collectif anti-expulsion, Aziz perd un peu courage en janvier et décide de se barrer en Grande-Bretagne, où il **croit savoir** que "les contrôles de police sont moins nombreux ". (CERF, PRES)*

(Chu, 2008) a noté les enchaînements possibles de modaux et les limitations qu'ils présentent. Dans nos données, nous ne relevons aucun enchaînement de modaux. L'autre type de modalisation est la négation (3 occurrences), marquée par « ne – pas » ou simplement par « pas » (334).

(334) *tous les hommes préférant le bonheur au malheur, le plaisir à la souffrance, ne fais pas aux autres ce que tu **ne voudrais pas** qu'on te fasse à toi-même. (CERF, DIVS)*

(Sportiche, 2011) a également remarqué que la négation combinée à l'extraction de sujet rendait l'énoncé agrammatical en relative. En interrogative, la négation ne semble jamais possible. Les verbes ponts ont donc un comportement similaire aux parenthétiques face à la négation. Ces derniers sont toujours produits dans une forme positive (335).

(335) *il se trouve que le le grand l'un des gros éléments de ce projet c'est un tunnel de cinquante-quatre kilomètres de long qui passe euh un tunnel transalpin le plus long d'ailleurs qui **je crois** existe (CorpAix)*

Comme le montrent les chiffres, les modalisations sont rares. La négation, surtout, semble poser problème aux locuteurs, qui reformulent et reconstruisent leur énoncé pour l'éviter. Voici un exemple présenté en grille (336).

(336) *je me dis bah il y a certaines choses peut-être*
qu'il aurait pas fallu que on se on se
que j'aurais peut-être pas dû faire comme ça
qu'il aurait fallu que je fasse autrement (PFC, 91ael1)

Dans l'ensemble des constructions des corpus (mêmes corpus et même source statistique que précédemment, p192), les verbes sont également en majorité à l'indicatif et au présent. Mais les proportions sont un peu différentes. L'indicatif est moins important en construction pont, au profit du conditionnel, comme l'illustre le (TABLEAU 46).

	Construction pont	Toutes les constructions
Indicatif	94,8%	97,7%
Conditionnel	5,2%	2,3%

TABLEAU 46. Proportion du mode des verbes en construction pont et dans l'ensemble des constructions

Le présent est beaucoup plus présent en pont que dans l'ensemble du corpus, comme l'illustre le (TABLEAU 47).

	Construction pont	Toutes les constructions
Présent	85,7%	66,5%

TABLEAU 47. Proportion du temps présent des verbes en construction pont et dans l'ensemble des constructions

Nous pouvons conclure que les constructions de dépendance à distance impliquent une forte présence de l'attitude modale du locuteur ou de l'interlocuteur. Celui-ci fait part de ses sentiments (présent) ou met en doute ses propos (conditionnel).

3.4.3 Les objets et les ajouts

Dans les constructions de dépendance à distance en complétive, le verbe pont, par définition, a au moins un complément, qui est la complétive. Dans la majorité des cas, le verbe ne régit pas d'autres compléments (96,1% des occurrences).

Lorsque le verbe comporte un autre complément que la complétive dans le pont (7 occurrences), peu de variations se présentent. Dans toutes les occurrences, le complément est un objet, qui se réalise sous la forme d'un pronom clitique (337). Ce complément est lié à la struture modale du verbe : source de la modalité dans *il me semble* et prise à témoin de la modalité avec *dire* (interlocuteur).

> *(337) ils interprètent leur sentiment comme de la colère, de l'euphorie ou comme rien du tout, en fonction de ce qu'on* ***leur*** *a dit qu'ils ressentiraient. (CERF, PRES)*

Dans le pont, les ajouts sont inexistants. On peut noter qu'en dehors du pont, il est également rare qu'un ajout soit présent, à la suite de la complétive. Nous avons relevé une occurrence, que nous présentons ici (338).

> *(338) une voix profonde, grave et lente comme un final de Beethoven, ce genre de voix dont on imaginerait que la Terre elle-même fût dotée* ***si elle n'était de sexe féminin****, qu'elle extrairait d'un gosier profond comme la fosse des Philippines, articulant avec des lèvres de la taille du Grand Canyon. (CERF, LITC)*

Un adverbe peut venir modaliser le verbe, en se plaçant après lui : « exactement, bien, particulièrement, réellement » (339)(340). Il renforce la valeur épistémique du verbe, ou lui donne un caractère intensif.

> *(339) À moins qu'il y ait des raisons que je n'ai pas perçues et qu'il faudrait* ***réellement*** *qu'on m'explique, là, mais, dans la manière dont, moi, j'ai lu le projet de loi, j'ai en quelque sorte été heurté par cet article (CERF, INST)*

(340) *il avait le choix entre revenir dans un processus de raison et dans un processus de paix ou s'enfermer dans l'impasse suicidaire qui était la sienne mais où il sentait **bien** que progressivement, il était de plus en plus isolé. (CERF, POLI)*

On peut conclure qu'il n'y a pas d'éléments qui contribuent à donner une valeur descriptive à l'ensemble du pont : pas d'adverbes de manière *(*ce que j'ai gentiment demandé qu'on m'adresse*) ni de déterminations spatio-temporelles pouvant caractériser un événement indépendant.

3.4.4 Le sujet du verbe à complétive

Le sujet du verbe à complétive est généralement un pronom clitique. (180 occurrences, 77,9%)

(341) *A qui veux-tu que **je** pense ? (CERF, DIVS)*

Mais il peut également être produit sous la forme d'un groupe nominal (43 occurrences) (342) ou d'un nom propre (2 occurrences) (343). Dans 6 occurrences, le sujet n'est pas dans sa position canonique puisque c'est l'élément qui est antéposé (344).

(342) *toutes euh comportent des paramètres qui leur sont spécifiques et qu'il faut que **l'élève** domine (CorpAix, ECRITUPC)*

(343) *Une révolution où l'armée ne renoncera pas entièrement à son pouvoir, et où il est clair que **Musharraf** ne compte pas se retirer de son poste à la tête du pays.(CERF, PRES)*

(344) *c'est **moi** qu'il faut qui parle maintenant non (CorpAix, BENOT)*

Le tableau suivant présente la répartition des sujets pronominaux en fonction de la personne à laquelle il est fait référence (TABLEAU 48) :

	Sujets pronominaux	**% par rapport aux sujets pronominaux**
P1 (je, nous)	80	44,4%
P2 (tu, vous)	10	5,6%
P3 (il, elle, on, ils, elles)	59	32,8%
Il impersonnel	31	17,2%

TABLEAU 48. Répartition des sujets pronominaux du verbe à complétive, personnels et impersonnels

On constate que le sujet du verbe à complétive est assez différent de celui du verbe pont. Les pronoms clitiques sont dominants dans les deux cas mais la proportion est beaucoup moins forte dans les sujets des verbes à complétive, au profit des sujets nominaux. Parmi les sujets pronominaux, la personne mentionnée est différente, comme l'illustre la comparaison en pourcentage du (TABLEAU 49) :

	sujets du verbe à complétive	**sujet du verbe pont**
Sujet clitique	**77,9%**	**95,6%**
P1 (je, nous)	44,4%	14%
P2 (tu, vous)	5,6%	43%
P3 (il, elle, on, ils, elles)	32,8%	19,5%
Il impersonnel	17,2%	23,5%

TABLEAU 49. Comparaison des sujets pronominaux des verbes à complétive et des verbes ponts

Ces observations mettent une fois encore en évidence que le verbe pont et son sujet sont employés dans un but d'intersubjectivité, la deuxième personne n'étant plus du tout majoritaire dans la suite de l'énoncé.

3.4.5 Les éléments non-régis et les disfluences

Les éléments non régis, tels que les commentaires en « parce que », les associés ou les adverbes énonciatifs ou modaux (« franchement », « peut-être ») ainsi que les disfluences (hésitations, répétitions, amorces, etc.) peuvent normalement apparaître dans toutes les positions. Pourtant, leurs occurrences sont rares dans le pont (9 occurrences).

On trouve dans le corpus d'exemples un double marquage du sujet, présent dans une occurrence (345). La fonction syntaxique est réalisée à la fois par le pronom personnel « elle » et par le groupe nominal « la religion ».

*(345) il y plein de questions que **la religion elle** te dit que c'est interdit de de répondre quoi (Nancy, Maroc)*

D’autres éléments associés sont réalisés sous la forme d’un adverbe (4 occurrences). L’adverbe peut être placé après le verbe mais former une locution particulière, sans lien avec le sens d’origine de l’adverbe (347) ou avoir une portée plus large, comme c’est le cas avec « ensuite », dont la valeur est argumentative dans l’exemple (346). Il peut aussi se placer avant la construction verbale, à la frontière du pont, et modifier plutôt l’interrogatif-relatif : « pourquoi diable », « où là » (347)(349).

*(346) Quand vous allez chez le coiffeur, vous espérez tous que ce sera un professionnel qui vous coupera les cheveux... pourquoi voudriez vous **ensuite** que ce soit des citoyens, la plupart du temps desinvestis politiquement, qui décident de l’avenir d’une nation ? ! ; -) (CERF, FORM)*

*(347) il ne me resterait plus de temps pour aller à l'étranger, où il faut **bien** que j'aille quelquefois (CERF, POLI)*

*(348) Dans la suite du film, Clo-Clo mourra, victime de sa séduction, sous les coups d’un obsédé sexuel (pourquoi **diable** fallait-il qu’elle se mette du rouge à lèvres dans une rue déserte en pleine nuit ?) (CERF, DIVS)*

*(349) il y a des avocats d’associations d’enfants battus il y a des avocats d’associations de femmes battues où **là** je pense que ils ont plus l’impression d’aider (CorpAix, AVOCAT96)*

On relève une occurrence de particule discursive : « ben » (350). Son rôle est de renforcer le caractère assertif de la séquence.

*(350) c’est pour ça d’ailleurs qu’on se rend compte **ben** que le livret A de la Caisse d’Epargne qui avait été complètement euh laissé à l’abandon dans les années soixante-et-dix quatre-vingts revient très très fort (CorpAix, PRVOYAN)*

En aucun cas ces éléments non régis ne contribuent à accroître la valeur descriptive de l’ensemble pont.

On compte seulement 3 occurrences présentant des disfluences : 3 hésitations « euh » (351) et 1 répétition du sujet « tu tu » (352). La dernière occurrence citée cumule hésitation et répétition.

*(351) effectivement les les gens qui sont installés dans des structures où euh ils sont bien à l’aise et où **euh** il faut que ça tourne rond (CorpAix, APOSTROP)*

*(352) tu as été dans d’autres euh enfin d’autres euh théâtres spectacles où **euh tu tu** entendais que c’était un français un peu euh (PFC, 75ccm1)*

En termes purement microsyntaxiques, la construction du pont est essentiellement réduite à la formule : « sujet non lexical, verbe, que ». Les rares développements de la séquence relèvent soit de la formation de l'énoncé et donc du domaine macrosyntaxique soit d'une modification non référentielle intégrée au lexème verbal (bien, réellement). Le faible nombre d'insertions non régies (parenthétiques, phatiques, disfluences) pourrait indiquer que la production de l'enchaînement entre le premier élément du pont et le verbe de la complétive est réalisée sans difficulté. Les deux verbes (et leurs constructions) sont réunis, sans phénomènes complexes de formulation. Cette cohésion syntaxique et prosodique renforce l'idée de construction à forme fixe et limitée dans ses extensions.

3.4.6 De la description de la construction du pont à l'interprétation de contrainte de *processing*

La description que nous venons d'effectuer peut-être interprétée en termes de contraintes de *processing*, dans la lignée de (Hofmeister & Sag, 2010).

3.4.6.1 Récapitulatif

La description de l'ensemble du pont, c'est-à-dire des éléments compris entre le complément antéposé et la complétive montre que la construction s'appuie sur une séquence verbale simple, composée du sujet, du verbe pont et de l'élément « que ». Il y a très peu d'éléments supplémentaires : quelques objets clitiques et pas d'ajouts spécifiques au pont. On peut noter que par rapport à d'autres constructions microsyntaxiques peu d'éléments non régis et de disfluences sont insérés. Le pont est donc généralement composé uniquement des éléments obligatoires : le sujet et le verbe. La composition de ces deux séquences est également simple : le sujet est pronominal, le verbe est conjugué au présent de l'indicatif, à la voix active, et il n'est pas nié ni modalisé. La comparaison avec l'ensemble du corpus montre que la construction pont favorise particulièrement les sujets pronominaux, notamment de deuxième personne, et le temps présent.

3.4.6.2 Absence de poids syntaxique

Le verbe pont possède donc les caractéristiques d'un verbe : sujet, marque de temps, de mode et de personne. Mais son fonctionnement peut être assimilé à celui d'un verbe modal plutôt qu'à celui d'un verbe constructeur. Il ne régit que des compléments obligatoires et possède une interprétation modale qui accompagne l'action décrite par le verbe de la complétive. Lorsque l'on tente d'ajouter plusieurs compléments au verbe pont, afin d'en favoriser l'interprétation descriptive (353), le jugement de grammaticalité de l'énoncé diminue beaucoup (354).

(353) Je démontre par l'absurde en trois étapes que Pierre a une sœur.

(354) ?? la sœur que je démontre par l'absurde en trois étapes que Pierre a

Le verbe pont peut donc être sélectionné parmi n'importe quel verbe à complétive à partir du moment où la construction qui l'englobe est peu développée syntaxiquement et réduite sémantiquement à une interprétation modale.

Le fait que le verbe soit modal ne signifie pas qu'il n'a pas d'importance dans l'énoncé du point de vue informatif. Au contraire, il arrive que le verbe pont soit justement l'élément sur lequel le locuteur met l'accent. Il présente d'abord un fait objectif puis le contraste avec le fait modalisé par un point de vue (355). Le verbe de la complétive, dans ce cas, est identique au verbe qui précède le verbe pont et le verbe pont apporte l'information principale.

*(355) Rappelez-vous : toutes ces tendances psychiques enfouies ne sont pas la réalité ; elles **sont** ce qu'il désire que la réalité **soit**. (CERF, LITC)*

3.4.6.3 Interprétation : contraintes cognitives de distance topologique

La taille moyenne du pont est de 2,3 mots. Le pont se développe en deux mots dans 79% des cas (182 occurrences) (voir FIGURE.17). Comme nous l'avons déjà signalé (voir §2.2.3.2), ces résultats ne sont pas dus à un biais de la requête. On peut être tenté d'expliquer ces résultats en termes de contraintes cognitives de performance sur la « reconnaissance » de la structure. La complexité de traitement de la construction de dépendance à distance serait contre-balancée par la simplicité syntaxique et lexicale du pont. En effet, dans l'hypothèse où la construction de dépendance à distance est dérivée ou directement reliée à une construction

où le verbe de la complétive est enchâssé dans la construction pont, le rattachement entre le complément antéposé et son verbe recteur demanderait un effort pour deux raisons : l'éloignement hiérarchique et l'éloignement topologique. Si l'éloignement hiérarchique ne peut pas être réduit, l'éloignement topologique, par contre, dépend du locuteur. En limitant le nombre de mots à l'écrit ou le nombre de syllabes à l'oral (3 en moyenne), ce dernier faciliterait le processus de rattachement. La mémoire serait ainsi moins mise à l'épreuve et les actants impliqués dans l'échange pourraient concentrer leur attention sur les connexions syntaxiques.

3.4.6.4 Poids du pronom interrogatif-relatif

(Hofmeister & Sag, 2010) avaient remarqué que lorsque l'interrogatif-relatif était complexe et spécifié, l'acceptabilité des énoncés de dépendance à distance était meilleure, car le rattachement sémantique était plus évident. Mais les occurrences avec pronom complexe sont assez rares dans les dépendances à distance du corpus (3 occurrences en « quel- »).

(356) par quel bout est-ce qu'il faut que je le prenne (CorpAix, POSIE)

Nous pouvons étendre leur idée en considérant pour les relatives la forme du pronom et la forme de l'antécédent. En effet, le verbe régit le pronom mais ce dernier est chargé du poids lexical et sémantique de son antécédent. Il présente donc plus de poids qu'un pronom interrogatif. Cela pourrait expliquer pourquoi les relatives acceptent plus de verbes ponts que les interrogatives. Le poids du pronom combiné à l'absence de poids du verbe pont, faciliterait le décodage de l'énoncé et rendrait la construction plus libre en termes de lexique.

L'autre interprétation est qu'un pronom complexe est certes plus facilement interprétable en termes sémantiques, mais surtout, il a de forte chance d'être assez spécifié pour pouvoir entrer dans le cadre de sous-catégorisation d'un seul des verbes. Dans l'exemple suivant (357), le lien *exécuter-mode* est certes plus probable que celui de *souhaiter-mode* mais il est surtout impossible d'avoir « *vous souhaitez dans ce mode ».

*(357) Tout d'abord, vous devez définir les macros d'initialisation du TNC et les placer **dans le mode dans lequel** vous souhaitez qu'elles s'exécutent (CERF, SCIE)*

La connexion syntaxique pourrait venir appuyer le rattachement sémantique, facilité lui-même par l'absence de poids du verbe pont.

3.4.6.5 Autre contrainte cognitive : connexion syntaxique

Les connexions syntaxiques, justement, pourraient dans certains cas permettre de contourner la difficulté de l'éloignement hiérarchique. Dans 159 occurrences, soit 69,4%, le complément antéposé ne peut pas être un objet du verbe pont[115]. La production du verbe pont, non compatible avec la proforme antéposée, signale immédiatement la présence d'une dépendance à distance. L'énoncé sans dépendance à distance serait agrammatical (358)(359)(360) .

*(358) * avec la meilleure des volontés où vous voulez*

*(359) *A qui veux-tu*

L'interlocuteur est en attente d'un verbe recteur compatible : « où + se garer » (360), « à qui + penser » (361).

(360) avec la meilleure des volontés où vous voulez qu'ils se garent (CorpAix, GENVE)

(361) A qui veux-tu que je pense ? (CERF, DIVS)

Les énoncés ne présentent pas d'ambiguïté de rattachement à l'un ou à l'autre verbe, et donc pas de mauvaise connexion qui nécessiterait ensuite une réanalyse. Le processus cognitif serait donc allégé.

La fonction de l'élément extrait pourrait également faciliter plus ou moins la dépendance à distance. L'élément extrait devrait être plus facilement rattachable au second verbe s'il est sujet plutôt qu'objet. D'après les études courantes en neurolinguistique, la relativisation du sujet est plus facile à décoder parce que la position du relatif est proche, voire identique, à la position du sujet de la déclarative, comme l'explique (Kaan & Stowe, 2002) :

> « Consider the subject relative in 8.2a) and the object relative in (8.2b).
> (8.2) a The senator who attacked the reporter admitted the error.
> b The senator who the reporter attacked admitted the error
> Numerous studies have shown that object relatives (cf e.g. Holmes, 1973; Wanner and Maratsos, 1978; King and Just, 1991[116]). [...] When the relative

[115] Il faut tout de même signaler que dans 29 occurrences, la proforme antéposée pourrait être ajout du verbe pont, ce qui porte à 56,8% la proportion d'occurrences dont le rattachement est totalement non ambigu lors de la production du verbe pont.

[116] Holmes V. (1973) Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior 12*, 285-293.

> pronoun [...] is encountered, the prediction is made that (i) a finite verb is needed to make a complete relative clause, and (ii) a trace is needed, by means of which the wh-phrase can receive a thematic role and be semantically and syntactically integrated into the representation. In the subject relative in (8.2a), these predictions are satisfied immediately at the next word: attacked is a finite verb, and a trace can be postulated in subject position. Both storage demands and integration costs are relatively low. In (8.2b), on the other hand, the predictions are met only after another noun phrase, the reporter, has been processed. » (Kaan & Stowe, 2002, p. 259-260)

Paradoxalement, dans notre corpus de constructions de dépendance à distance, comme dans les observations de (Godard, 1980), l'élément antéposé est très rarement sujet du verbe recteur (6 occurrences), que l'on examine les structures relatives, interrogatives (362) ou clivées (363).

(362) L2 et toi qu'est-ce que tu en retires euh - de ça qu'est-ce que tu trouves qui est super (Nancy, Marais)

(363) L2 il y a trois échographies d'ailleurs vous dites que c'est ce qui est couramment admis c'est pas forcément couramment admis c'est ce qu'on pense qui devrait être fait (CorpAix, FTUSPC)

Mais là interviendrait pour le français un autre phénomène. Le fait qu'en surface, la position ordinaire du relatif soit occupée par la forme 'inattendue' « que » (et celle du complémenteur par la forme 'inattendue' « qui ») pourrait gêner les locuteurs[117] dans leur processus de reconnaissance et donc expliquer la faible fréquence d'un pont entre un sujet et son verbe. En outre, cette antéposition du sujet suppose que le locuteur ait prévu d'introduire un pont au moment où il construit sa relative. La relativisation de l'objet laisse une plus grande liberté ; le locuteur a jusqu'au bout la possibilité d'intégrer un ou plusieurs ponts. C'est peut-être pour cette raison que l'antéposition de l'objet (55% des occurrences) est préférée à celle du sujet[118].

L'antéposition de l'objet est aussi préférée à celle de l'ajout (36% des occurrencesce qui s'expliquerait dans les termes de cette approche par le fait que le rattachement de l'ajout, qui,

Wanner E. & Maratsos M. (1978) An atn approach in comprehension. In M. Halle, J. B. and Miller G. (eds), *Linguistic Theory and Psychological Reality*. Cambrige, MA: MIT Press.

King J. & Just M. (1991) Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language 30*, 580-602.

[117] Tous les linguistes n'ont d'ailleurs pas la même analyse de ces « que » et « qui ». Nous n'entrerons pas dans ce débat ici. Pour plus de détails, consulter (Hirschbuhler & Labelle, 1996; Moreau, 1971; Sportiche, 2011)

[118] Pour une interprétation plus structurale voir (Sportiche, 2011)

par définition, peut être régi par n'importe quel verbe, sollicite un plus grand coût d'interprétation.

3.4.6.6 Corrélation entre les deux contraintes cognitives précédentes : distance topologique et connexion syntaxique

Il est intéressant de constater que les occurrences dont le pont est le plus développé en nombre de syllabes sont toutes des occurrences non ambiguës syntaxiquement (TABLEAU 50).

Nombre de syllabes du pont	Absence d'ambiguïté syntaxique entre la proforme et le verbe pont	Nombre de mots du pont	Absence d'ambiguïté syntaxique entre la proforme et le verbe pont
13	1/1 occurrence	5	4/4 occurrences
7	3/3 occurrences	4	11/12 occurrences
6	10/10 occurrences	3	27/31 occurrences
5	8/10 occurrences		
4	14/18 occurrences		

TABLEAU 50. Corrélation entre longueur du pont et absence d'ambiguïté syntaxique entre le complément antéposé et le verbe pont

(364) En application du paragraphe 11 de la résolution 53/102 de l'Assemblée générale en date du 8 décembre 1998, la Commission indique ci-après, pour chaque sujet, les points ***sur lesquels il pourrait être particulièrement intéressant que*** *les gouvernements expriment leurs vues, soit à la Sixième Commission, soit par écrit, afin de la guider utilement dans la poursuite de ses travaux. (CERF, INST)*

Deux interprétations de cause à conséquence sont possibles : soit le fait qu'il n'y ait pas d'ambiguïté syntaxique permet d'augmenter la distance topologique entre le complément et son verbe recteur, soit, inversement, le fait qu'il y ait une distance topologique importante contraint la structure syntaxique à être non ambiguë. Plus simplement, on pourrait conclure que les processus cognitifs limiteraient la complexité à un seul paramètre : la distance ou bien le rattachement syntaxique. Mais ce qui est intéressant, c'est que cette interprétation en termes de performance rejoint les considérations structurelles : la forme syntaxique réduite de la séquence pont favorise à la fois l'analyse en ligne de la séquence et l'interprétation comme

modal du verbe en bloquant toutes les expansions qui pourraient lui donner une interpétation descriptive. Au bout du compte, la construction pont est une forme spécifique associée à une interprétation spécifique.

Dans la section précédente, nous avons décrit les contraintes internes à la langue, c'est-à-dire les tendances lexicales et catégorielles qui apparaissent dans les occurrences. Dans la section suivante, nous allons observer s'il existe également des contraintes externes. Les éléments présents dans la construction de dépendance à distance varient-ils en fonction du type de données (genre) ? Varient-ils en fonction des acteurs du discours (locuteurs/interlocuteurs) notamment selon que l'interlocuteur est un adulte ou un enfant ?

3.5 Variation : Les contextes de productions

Dans la première section, nous examinerons si un type de textes ou un type de discours favorise la production de dépendances à distance. Dans la deuxième section, nous comparerons les corpus adultes-adultes et adultes-enfants.

3.5.1 Contexte d'apparition des dépendances à distance

Nous avions relevés les occurrences de dépendance à distance dans deux types de données : des corpus écrits et des corpus oraux. La répartition des 229 occurrences dans ces corpus est la suivante : 104 occurrences appartiennent au corpus écrit ; 125 occurrences proviennent de sources orales (FIGURE.23). Le nombre d'occurrences est donc approximativement équivalent, alors que la taille des deux corpus était très différente (presque 3 millions de mots pour l'oral et presque 9 millions pour l'écrit).

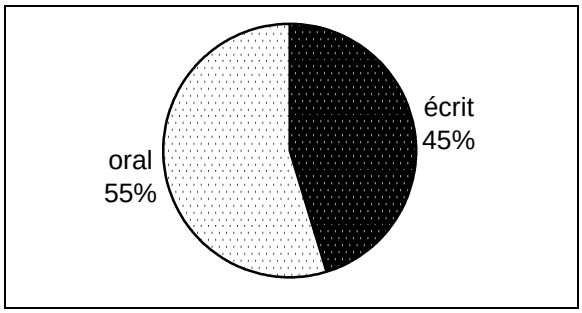

FIGURE.23 Répartition des occurrences de dépendance à distance relevées

A l'oral, la répartition par corpus est la suivante (FIGURE.24). Les corpus sont ordonnés en fonction de leur taille. Les occurrences apparaissent généralement dans les corpus dont la taille est supérieure à 55 000 mots (de CorpAix à CTFP). Néanmoins, le nombre d'occurrences n'est pas proportionnel à la taille du corpus : d'une part, les occurrences ne diminuent pas sur l'axe des abscisses ; d'autre part, le corpus Irit (7000 mots) comporte une occurrence de dépendance à distance.

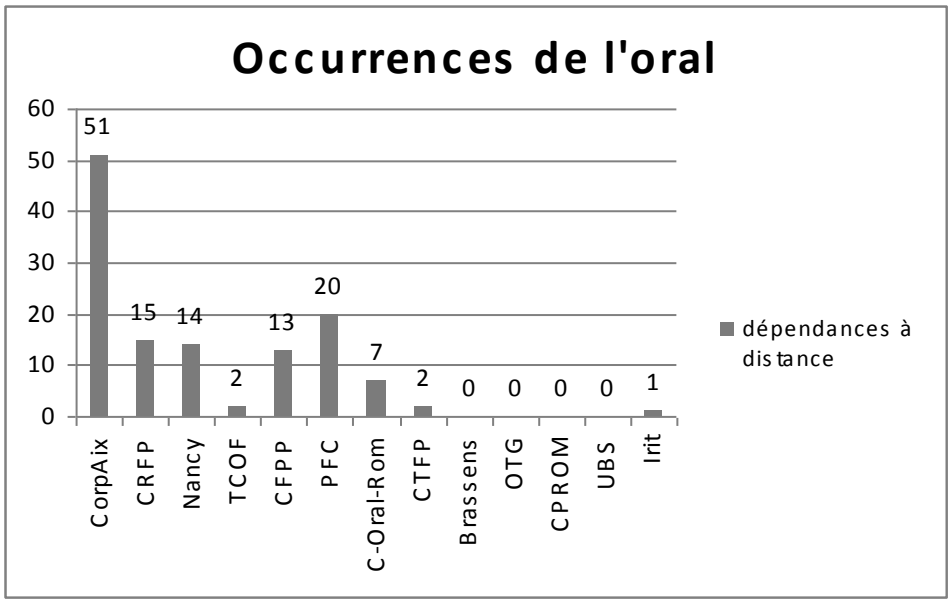

FIGURE.24 Répartition dans les corpus oraux des occurrences de dépendance à distance

A l'écrit, les tranches du CERF possèdent le même nombre de mots. Les tranches ORAL (CorpAix) et LITA (littérature/fiction ancienne) ayant été écartée, nous observons la répartition des occurrences dans les huit tranches restantes, qui sont, par ordre alphabétique : Divers (DIVS), Forum internet (FORM), écrits institutionnels (INST), Littérature/Fiction

contemporaine (LITC), Politique (POLI), Loisirs et vie pratique (PRAT), Presse d'informations (PRES), écrits scientifiques et techniques (SCIE). Comme pour les corpus oraux, la répartition des occurrences ne dépend pas du paramètre de la taille des corpus. Le nombre d'occurrences varie beaucoup en fonction des tranches (FIGURE.25).

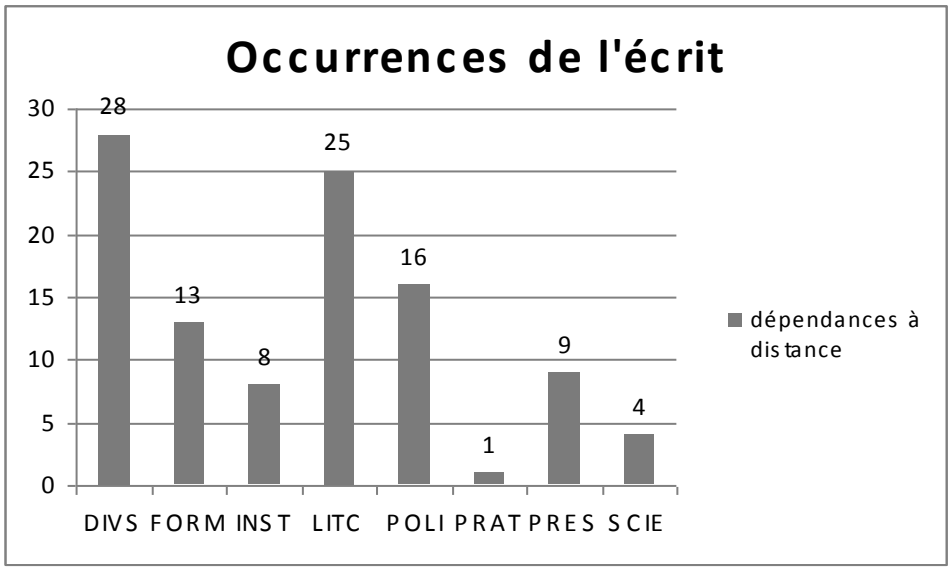

FIGURE.25 Répartition dans les corpus écrits des occurrences de dépendance à distance

L'observation de la répartition des occurrences de constructions de dépendance à distance tend à montrer que la production ne dépend pas du moyen de communication, ni du contenu thématique.

Nous avons cherché à classer l'ensemble des enregistrements des corpus afin d'identifier un éventuel critère dominant qui favorise la production de construction de dépendance à distance. Cette démarche amène une réflexion sur l'échantillonnage des corpus.

3.5.1.1 Echantillonnage

Echantillonner un corpus n'est pas une tâche facile. Que ce soit pour les corpus oraux ou écrits, il n'existe pas une norme qui permettrait de savoir comment regrouper les différentes productions de langue. Pourtant tout le monde s'accorde à dire qu'un corpus doit être varié et échantillonné, afin d'être représentatif de la langue. Pour décrire minutieusement un phénomène, il faut pouvoir l'observer dans de multiples contextes (Debaisieux, 2005).

> « Ces exemples illustrent bien la nécessité de prendre en compte « les genres » si l'on souhaite aboutir à une meilleure appréhension des faits de langue en dépassant notamment l'opposition souvent simpliste posée entre un écrit, dont on ne sait pas très bien par quoi il est représenté et un oral réduit à la seule modalité d'oral spontané de conversation. » (Debaisieux, 2005, p. 18)

Il est donc préconisé de « veiller à l'équilibre entre les enregistrements » dans les corpus oraux (Dister, Francard, Hambye, & Simon, 2009, p. 120).

Le corpus est donc paradoxal, puisqu'il doit d'une part être hétérogène et d'autre part homogène pour répondre à la définition de corpus. La définition de corpus implique un recueil de données authentiques reliées par un point commun (l'auteur, le thème, etc.).

> « Un corpus est tout simplement un ensemble de textes ou de documents, rassemblés pour être étudiés. [...] Il faut ensuite que ces derniers présentent une certaine cohérence entre eux pour former un corpus : comme il existe des corpus écrits autour de l'œuvre d'un auteur, les corpus oraux se forment autour d'une langue, d'un lieu géographique, d'une époque, d'un évènement, d'un phénomène oral, etc. » (Ministère de la Culture et de la Communication & Centre National de la Recherche Scientifique, s. d.)

Pour (Rastier, 2004), le terme 'corpus' s'oppose à celui de 'banque de données' justement parce que le premier possède un critère de rassemblement alors que pour le second, ce critère est absent :

> « Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications. » (Rastier, 2004, p. 1)

> « Ainsi une banque textuelle peut regrouper des textes numériques de statuts divers : aucun critère linguistique ne permet cependant leur totalisation, sauf l'hypothèse que la langue leur confèrerait une unité a priori » (Rastier, 2004, p. 2)

Pourtant, généralement, les corpus prônent leur variation comme une richesse. Cette variation est presque jugée indispensable, les corpus étant censés représenter le plus fidèlement possible l'utilisation du langage dans la vie réelle.

Il existe deux stratégies pour établir un échantillonnage. La première consiste à déduire les genres à partir de l'observation de la présence ou de l'absence de phénomènes (Biber, 1991). On établit donc une corrélation entre une constellation de traits linguistiques et les genres. La

difficulté de cette approche est de choisir les traits qui seraient corrélés sans avoir d'a priori. Il faudrait avoir décrit tous les traits pour pouvoir réellement déduire les genres. La seconde stratégie consiste à faire de l'échantillonnage a priori, d'après ce qu'on imagine qui suscite de la variété :

> « 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 » (Bilger, 2000b, p. 13).

On peut donc classer les données selon différents critères.

a) Critères d'échantillonnage

Le premier classement possible consiste à séparer les données en fonction de leur visée (Habert, 2000, p. 25) : distraire, informer, convaincre ; de leur point de vue : didactique, normatif, descriptif ; ou de leur type : descriptif, narratif, expositif, argumentatif, instructif. Le problème qui se pose est qu'il n'est pas rare que les différents visées, points de vus ou types se succèdent dans un même texte, dans un même discours.

Le deuxième classement possible est celui des thèmes ou domaines abordés par les données. Voici les groupes recensés par (Habert, 2000, p. 25) chez (Sinclair, 1996)[119] : religion, technique et technologie, droits, sport, belles lettres, politique, histoire, médecine, philosophie, économie, éducation, psychologie, sciences, sociologie, loisirs, civilisations, physique, biologie, mathématique, vie domestique, voyages, anthropologie, affaires militaires, média et communication, langage, littérature, architecture, mode et vêtement, informatique, agriculture, géographie, écologie et environnement, transport, chimie, finance. Comme nous le voyons dans ces quelques domaines, le recouvrement entre les catégories est difficile à éviter. Le domaine 'science' regroupe ceux de 'biologie', 'chimie', 'physique', etc. Le domaine 'religion' est proche du domaine 'philosophie'. Ce classement n'est donc pas transparent.

[119] Sinclair. (1996). EAGLES. Preliminary recommendations on Corpus Typology, *24*(1996), 16-19.

Le troisième classement possible, pour l'oral, est celui qui distingue les données selon des caractéristiques de la communication : le canal de transmission de l'information, la situation de parole ou la relation entre les participants. (Cresti & Scarano, 2000; Habert, 2000)

- le canal de transmission : face à face, téléphone, répondeur, radio, télévision. Ce critère a deux inconvénients. D'une part une catégorie 'face à face' réunirait des données très variées. D'autre part, ces catégories ne sont pas étanches. Il est tout à fait possible que deux personnes soient en communication en face à face (ou par téléphone) et que leur communication soit retransmise à la télévision ou à la radio.
- La situation de parole : on peut opposer la parole privée et la parole publique, en supposant que la parole publique est plus surveillée, plus proche de la norme (bien que les émissions de télé-réalité ou les émissions de radio libre, par exemple, puissent remettre en cause cette hypothèse). La parole provoquée par le chercheur (entretien) pourrait être opposée à la parole 'naturelle' (conversation familiale, échange commerciaux ; etc.). Cette distinction est intéressante mais présente deux difficultés : les données naturelles sont très diversifiées et difficiles à obtenir d'un point de vue juridique. En effet, il est délicat d'enregistrer des locuteurs sans avoir préalablement reçu leur accord (Baude et al., 2006). Dans ces conditions, les données peuvent perdre de leur naturel, et finalement se rapprocher des données provoquées. Enfin, les données pourraient être divisées en fonction de leur degré de préparation, du plus spontané au lu. Cette distinction est difficile à appliquer puisqu'on ne peut avoir accès à cette information que par l'intermédiaire du locuteur qui a une vision personnelle du degré de préparation.
- La relation entre les participants : un locuteur s'exprimerait différemment devant des familiers et des personnes qui lui sont inconnues, en fonction du degré de complicité qui l'unit à son interlocuteur. Il adapterait également sa façon d'écrire. Peu d'études l'ont démontré mais l'idée semble partagée par tous. Si les utilisateurs de SMS adaptent l'orthographe de leur texte en fonction du destinataire[120], il est probable que l'ensemble du langage soit adapter (vocabulaire, construction, etc.). Dans un contexte totalement différent, Lentin émet le conseil suivant :

  « il faudrait étudier tout particulièrement dans ses différentes réalisations la langue employée par l'adulte pour parler à l'enfant. Elle aussi sans doute

[120] Certains locuteurs l'admettent (information donnée par Marie-José Béguelin, lors des rencontres 'Macro-syntaxe' entre Nancy, Fribourg et Neuchâtel)

> présente des régularités qu'il serait indispensable de prendre en compte. » (Lentin, 1975, p. 16)

Plus précisément, c'est la position du locuteur par rapport à l'interlocuteur qui semble déterminante. Ainsi, un enfant qui explique une recette de cuisine à sa mère, ou qui s'adresse à un enfant plus jeune, pourra utiliser des constructions particulières.

> « La variation sociolinguistique est particulièrement liée au jeu des positionnements des interactants, ou plus précisément aux rapports de place qu'ils co-construisent au fil de l'interaction. » (Auzanneau, 2010, p. 27)

La difficulté qui se pose alors est de classer les échanges qui mettent en scène plusieurs participants qui ont des rapports complexes entre eux

Le quatrième classement possible est un classement basé sur des paramètres sociolinguistiques. Si ces critères sont admis comme facteurs de variation, peu d'études ont réellement démontré leur impact dans l'usage d'un phénomène. Voyons chacun des critères généralement indiqués dans les métadonnées des corpus : l'origine géographique, l'âge, le sexe, l'origine sociale, le niveau d'études du locuteur et le statut de l'interlocuteur.

- **L'origine géographique du locuteur :**

Les locuteurs, selon leur origine, attestent de certains traits de prononciation (Lyche, Detey, Durand, & Laks, 2010; Martinet, 1945). Les méridionaux produisent de façon quasi-systématique des voyelles fermées en syllabes ouvertes (« arrêt » [aʀe] ; « biologie » [bjoloʒi]) et des voyelles ouvertes en syllabes fermées ; les lorrains ont tendance à ouvrir les voyelles en fin de mot (« pied » [pjɛ] ; « vélo » [velɔ]). Le lexique, ainsi que les expressions, peuvent être spécifiques à une région : « chaler quelqu'un » en Provence (transporter quelqu'un sur son vélo ou en scooter), « des chaurées » en Lorraine (bouffées de chaleur). La syntaxe varie également selon les régions : changement de préposition, modification de la structure valencielle d'un verbe, etc.

Il faut nuancer toutefois l'importance de ce paramètre. Les variations concernent des points particuliers de la langue et la majorité se situe dans le domaine lexical. Les régionalismes sont aussi plus ou moins marqués en fonction des locuteurs. Enfin, les particularités régionales

évoluent rapidement et tendent à s'uniformiser dans notre société propice aux déplacements et à la communication à distance, et ce, malgré l'attachement des locuteurs aux particularités de leur langue.

- **L'âge du locuteur**

Outre la différence enfant-adulte, due au stade d'apprentissage de la langue (Lentin, 1971, 1975, 1988), peu d'études ont mis en évidence la variation de la langue en fonction de l'âge du locuteur. Le langage des jeunes a suscité de l'intérêt (Branca-Rosoff, 2007; Eckert, 1997; Lehka-Lemarchand, 2007; Podhorná-Polická, 2007; Trimaille & Billiez, 2007), mais il est souvent assimilé à la langue des banlieues. Les caractéristiques remarquées sont le déplacement de l'accent tonique, l'allongement de la syllabe pénultième (« Maaadame ») (Fagyal, 2003, p. 50) la verlanisation (« meuf, chelou, tèj, ouf, pécho, zyva »), la troncation (« dég » pour « dégueulasse », « zic » pour « musique »).

Des phénomènes tels que l'emploi des marqueurs pourraient varier selon l'âge des locuteurs. Ce n'est pas à proprement parler l'âge qui est un facteur (on ne modifie pas sa manière de parler en vieillissant) mais plutôt une différence due à l'évolution de la langue, comparable à l'évolution du français à travers les siècles. Les marqueurs se répartiraient en indice de génération moderne « quoi, enfin, bon » versus génération traditionnelle « c'est à dire ». D'autres marqueurs seraient employés par toutes les générations : « quand même, hein, si vous voulez » (Beeching, 2007).

- **Le sexe du locuteur :**

Dans les années 70, les études sur le genre s'intéressent au féminin dans la langue (pronom personnel) et à la féminisation des mots. Elles aboutissent généralement à une conclusion sur le statut de la femme dans la société et sur les clichés en vigueur concernant les hommes et les femmes (Aebischer, 1985[121] dans (Houdebine-Gravaud, 2003, p. 41)). Concernant réellement la langue, l'idée générale est que les femmes utilisent un langage plus proche de celui de la norme, et moins grossier. (Lakoff, 1975[122] dans (Bailly, 2007)).

Certains adjectifs affectifs seraient spécifiques aux femmes : « *charming* », « *lovely* », soit en français « mignon », « adorable ». L'emploi des adjectifs de couleur, de certaines formes

[121] Aebischer, V. (1985). *Les femmes et le langage* (Vol. 40). Paris: Presses universitaires de France.
[122] Lakoff, R. T. (1975). *Language and woman's place*. (M. Bucholtz, Éd.) (Vol. 56). Cambridge Univ Press.

modalisées (« peut-être », « sans doute ») ou de certaines formes interrogatives (« pouvez-vous », « ne croyez-vous pas ») seraient également caractéristiques des femmes. (Ritchie, 1974[123] dans (Houdebine-Gravaud, 2003, p. 40)).

Dans un contexte multilingue, à Alger, les femmes auraient tendance à utiliser l'alternance codique plus souvent et de manière plus longue que les hommes. (Ghanem Chentli, 2010)

Les études actuelles tendent à favoriser l'hypothèse d'un comportement langagier plutôt que d'un langage particulier aux femmes ou aux hommes. (Bailly, 2007; Eckert & McConnell-Ginet, 1999; Romaine, 2003).

Néanmoins, des études statistiques indiquent qu'en anglais, les hommes emploient les pronoms moins fréquemment que les femmes, tandis qu'ils emploient les déterminants « *a, an, the, no* » plus fréquemment que les femmes. (Koppel, Argamon, & Shimoni, 2002) dans (Juola, 2008, p. 301).

- **L'origine sociale du locuteur :**

L'origine sociale modeste, d'après les expériences de (Armstrong & Boughton, 1998) serait facilement identifiable, à travers un accent social.

> « Les tests perceptifs menés par les auteurs auprès des habitants de Rennes, à qui ils faisaient écouter les extraits d'enregistrements de Rennais et de Nancéens en leur demandant d'identifier leur appartenance sociale et régionale, montrent en effet que l'identification par les auditeurs de l'appartenance sociale du locuteur est de loin supérieure à leur identification de son appartenance régionale » (Lehka-Lemarchand, 2007, p. 12-13)

L'enquête de (Laks, 1980) indique que des particularités phonétiques, comme la chute du /r/ serait également une marque du statut social du locuteur. (Lehka-Lemarchand, 2007, p. 13)

L'origine sociale pourrait également être la cause de variations syntaxiques, telle que la réduction de la flexion dans les relatifs, ou le décumul des relatifs.(Guiraud, 1966, p. 41)

(365) Un pot qu'il y a quelque chose d'écrit dessus (Guiraud, 1966, p. 41)

Ce qui caractériserait le français bourgeois serait la sur-articulation ou la prononciation de toutes les liaisons, jusqu'à l'hypercorrection.

[123] Ritchie K. M. (1974) Male female language, New York, Metuchen, Scarecrow Press.

- **Le niveau d'étude du locuteur :**

Le niveau d'étude est très fréquemment demandé lors des enregistrements des locuteurs. Pourtant, encore une fois, peu d'études montrent que ce paramètre est réellement significatif dans la façon de parler. Néanmoins, il est possible que le niveau d'étude ait un impact sur la conscience qu'ont les locuteurs de leur langue et sur la volonté de s'approcher de la norme. (Beeching, 2007) note par exemple un emploi fréquent des marqueurs « c'est à dire, bon, enfin » par les diplômés, qui ont la volonté d'éviter le marqueur « quoi ».

De nombreux paramètres entrent donc en compte dans la parole d'un locuteur, à un instant T. Ces paramètres, bien sûr, n'ont pas une influence indépendante ; ils sont corrélés. Des chercheurs font l'hypothèse que dans les productions, le choix du lexique et des structures, analysable notamment à travers les n-grams, formerait le style d'un locuteur, qui pourrait être reconnaissable automatiquement (Juola, 2008). L'étude des mots et des structures permettrait de répondre aux questions suivantes :

> « was the document singly or multiply authored [47][124]? Was the author a native speaker of English? Of US or British English? A man or a woman [5, 89][125]? » (Juola, 2008, p. 239)

Ou encore combien y a-t-il d'auteurs dans ce groupe de textes et quels sont les textes qui ont été écrits par la même personne ? C'est ce qu'a expérimenté (Burrows, 1987, 2003)[126] :

> « Specifically, Burrows [3–5] demonstrated that a statistical analysis of common words in large samples of text could group texts by author. » (Juola, 2006, p. 3)

Mais (Brunet, 2004) montre par des études lexicométriques que le genre joue un rôle plus important que le style du locuteur, c'est-à-dire de l'écrivain à l'écrit. En comparant la proximité entre plusieurs œuvres de Lamartine, Musset et Hugo, la machine établit trois

[124] Graham, N., Hirst, G., & Marthi, B. (2005). Segmenting documents by stylistic character. Natural Language Engineering, 11(04), 397–415.

[125] Aragamon, S., & Levitan, S. (2005) Measuring the usefulness of function words for authorship attribution. Proceedings of ACH/ALLC. Consulté de http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.6935&rep=rep1&type=pdf

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. Literary and Linguistic Computing, 17(4), 401-412.

[126] Burrows, J. F. (1987). Word-patterns and story-shapes: The statistical analysis of narrative style. Literary and linguistic Computing, 2(2), 61–70.

Burrows, J. F. (2003). Questions of Authorship: Attribution and Beyond. Computers and the Humanities, 37(1), 5–32.

groupes. Ces trois groupes ne distinguent pas les auteurs mais le genre des textes : poétique, dramaturgique, narratif.

Les études, pour l'instant, n'ont pas vraiment mis en évidence que les paramètres sociolinguistiques étaient à l'origine de variation importante dans les structures linguistiques (Gadet, 2008). Les distinctions qui semblent être les plus importantes (Baude, 2006; Bilger, 2000a, 2000b; Bruxelles, Mondada, Traverso, & Simon, 2009) sont les suivantes :

- Echange provoqué par le chercheur / échange « naturel »
- Echange public / échange privé
- Echange formel / échange informel (cf lieux, relation avec l'interlocuteur)
- Echange préparé / échange spontané
- Echange spécifique / échange non spécifique (mode d'emploi, recette cuisine, technique, juridique…)
- Echange en face à face / échange à distance
- Echange avec un enfant / échange avec un adulte
- Echange en langue maternelle / échange en langue étrangère

Quels critères utiliser alors pour échantillonner les corpus écrits et oraux ? Il semble que l'un des objectifs soit de faire ressortir les constructions de grammaire première et celles apprises ultérieurement, en grammaire seconde (Blanche-Benveniste, 1990). Nous allons tenter de prendre une décision pour nos corpus oraux, en tenant compte des difficultés liées aux informations disponibles et en anticipant sur les problèmes posés par un système d'échantillonnage trop fin.

b) Echantillonnage du corpus écrit

Le corpus du CERF est pré-échantillonné en neuf genres d'écrit, le genre étant l'élément unificateur de la tranche. Dans cet échantillonnage, plusieurs critères sont malgré tout présents : le style (littérature), le thème principal du contenu (politique, divers, loisirs et vie pratique) et le support (forum). Certains genres sont déterminés par deux critères, celui du style et du support (presse), ou bien celui du contenu et du style (institutionnel, scientifique et

technique). Le fait que plusieurs critères aient contribué à établir les tranches du corpus génère quelques recouvrements. Ainsi, le sous-corpus FORM rassemble des textes au contenu scientifique, économique, politique, social, etc. qui peuvent tout à fait apparaître dans les tranches SCIE, PRES, POLI, DIVS et PRAT. Les textes de fiction ont été séparés des non-fictions mais les styles des sous-genres peuvent être identiques : DIVS_journal intime / LITT_journal autobiographique. Enfin, dans une même tranche, le contenu et le style peuvent être très différent : DIVS_critique littéraire / DIVS_texte religieux / DIVS_humour / DIVS_chansons ; LITC_contes / LITC_érotique / LITC_fantastique-horreur / LITC_poésie / LITC_policier / LITC_roman / LITC_théâtre.

Néanmoins, nous avons pu profiter de ce travail d'échantillonnage du corpus CERF, en gardant à l'esprit les réserves émises ci-dessus.

c) Echantillonnage du corpus oral

Chaque corpus oral dont nous disposons a été élaboré en suivant une organisation particulière, se basant sur des critères précis. Le plus souvent, les responsables ont essayé d'équilibrer les données dans chaque critère, dans un but comparatif. Le fait d'avoir réuni tous les corpus a détruit les équilibres qui existaient entre les sous-groupes.

Nous avons souhaité mettre en évidence la répartition des données selon les critères extralinguistiques, non pour recréer une organisation, mais pour pouvoir prendre en compte une éventuelle sur-représentativité d'un genre par rapport à un autre. C'est seulement grâce à cette représentation que nous saurons si un contexte favorise l'apparition des constructions de dépendance à distance, car la comparaison du nombre d'occurrences dans chaque genre pourra être relativisée par la taille du genre.

Nous nous sommes appuyés pour cette tâche d'échantillonnage sur les six corpus suivants : C-Oral-Rom, CRFP, PFC, CFPP, CTFP et CorpAix, soit environ 2 300 000 mots.

Hormis pour CorpAix, dont les métadonnées sont inaccessibles, les métadonnées présentes dans ces corpus d'origine étaient divisées en deux groupes : les informations qui concernaient l'enregistrement, et les informations qui concernaient les participants (majoritairement sociolinguistiques). Toutes les informations n'étaient bien sûr pas remplies dans l'ensemble des corpus. Voici le type d'informations fournies dans les corpus utilisés (TABLEAU 51).

Sujet	Informations	Corpus
Corpus, enregistrement	Date	C-Oral-Rom, CRFP, PFC, CFPP
	Lieu (quartier_CFPP)	C-Oral-Rom, CRFP, PFC, CFPP, CTFP
	Degré de préparation (formel/informel)	C-Oral-Rom, CRFP, CTFP
	Situation de parole (cours, débat, météo, prêche, émission, récit)	C-Oral-Rom, CRFP, CTFP
	Contexte (privé, public, professionnel)	C-Oral-Rom, CRFP, CTFP
	Canal (face à face, téléphone, média – radio/tv)	C-Oral-Rom, CRFP, CTFP
	Contenu thématique	C-Oral-Rom, CRFP, CTFP
	Qualité de l'enregistrement, durée, etc.	PFC, CFPP, CTFP
	Rubrique (témoignage, récit, enfants, métier-activité-explication)	CTFP
	Sexe	C-Oral-Rom, CRFP, PFC, CFPP, CTFP
Locuteur	Age / date de naissance	C-Oral-Rom, CRFP, PFC, CFPP, CTFP
	Lieu de naissance	PFC, CFPP
	Origine géographique	CTFP
	Niveau d'éducation - diplôme - âge de fin	 C-Oral-Rom, CRFP, CFPP, CTFP PFC
	Professions (passées et actuelle_CFPP)	C-Oral-Rom, CRFP, PFC, CFPP, CTFP
	Rôle (vendeur, guide, enseignant)	C-Oral-Rom, CRFP, CTFP
	Lieu d'habitation (domiciles successifs et actuel)	PFC
	Situation familiale, enfants	PFC, CFPP
	Relation avec les autres participants	PFC, CFPP
	Langues parlées, apprentissage	PFC
	Type de logement	PFC
	Intégration dans le voisinage	PFC
	Loisirs	PFC
	Parents et conjoint : date et lieu de naissance, profession, âge de fin d'études, langue parlée	PFC, CFPP

TABLEAU 51. Métadonnées fournies pour six des corpus oraux.

Les critères sociolinguistiques ne nous ont pas paru importants dans l'étude des constructions de dépendance à distance. En outre, n'ayant pas les informations pour CorpAix, il nous était difficile d'aboutir à une conclusion fiable sur un peu plus d'un million de mots. Les corpus dans lesquels peu ou pas d'occurrences de dépendances à distance apparaissaient ont

également été mis de côté. L'absence d'occurrences aurait pu avoir une signification mais la somme de travail manuel pour aboutir à la répartition nous a poussés à ne pas les prendre en compte.

Nous avons retenu cinq critères d'échantillonnage. Le premier critère est celui du contexte de la parole, privée ou public. Nous n'avons pas repris la distinction de contexte professionnel établie dans C-Oral-Rom, CRFP et CTFP. Dans ces corpus, l'étiquette professionnelle distingue les enregistrements qui se sont déroulés sur le lieu de travail de l'enquêté. Nous considérons que le lieu de l'enregistrement, même s'il est professionnel, n'interdit pas à l'échange de prendre une dimension privée (conversation entre collègues dans les temps de pause, ou même parfois en réunion), d'autant plus si l'enquêteur n'est pas un membre de l'entreprise. En revanche, nous définissons un critère spécifique/non spécifique, dans lequel le trait 'spécifique' pourrait par exemple être attribué aux explications sur la profession technique du locuteur.

Les deux critères suivants sont celui de situation formelle/informelle et de discours spontané/préparé. Ces deux critères, retenus mais distingués, permettent de décrire des situations comme celle citée par (Cresti & Scarano, 2000) :

> « Par conséquent, quand on fait allusion au français spontané, on s'attend toujours à ce que certaines caractéristiques sémiotiques, diaphasiques et linguistiques figurent dans les textes qui lui sont attribuables. On n'y classerait sûrement pas le discours inaugural de l'année universitaire, même quand le recteur est des plus démocratiques et parle sans ses notes. » (Cresti & Scarano, 2000, p. 341)

Le dernier critère retenu est celui du canal de communication, en face à face ou par téléphone. Ainsi que nous l'avons mentionné plus haut, la diffusion par radio ou télévision ne constitue pas un canal de communication. Sa fonction est plutôt de donner à la parole un statut public.

Les cinq critères d'échantillonnage peuvent s'organiser de manière à ce que chacun soit inclus dans un autre, jusqu'à former une seule hiérarchie (FIGURE.26) :

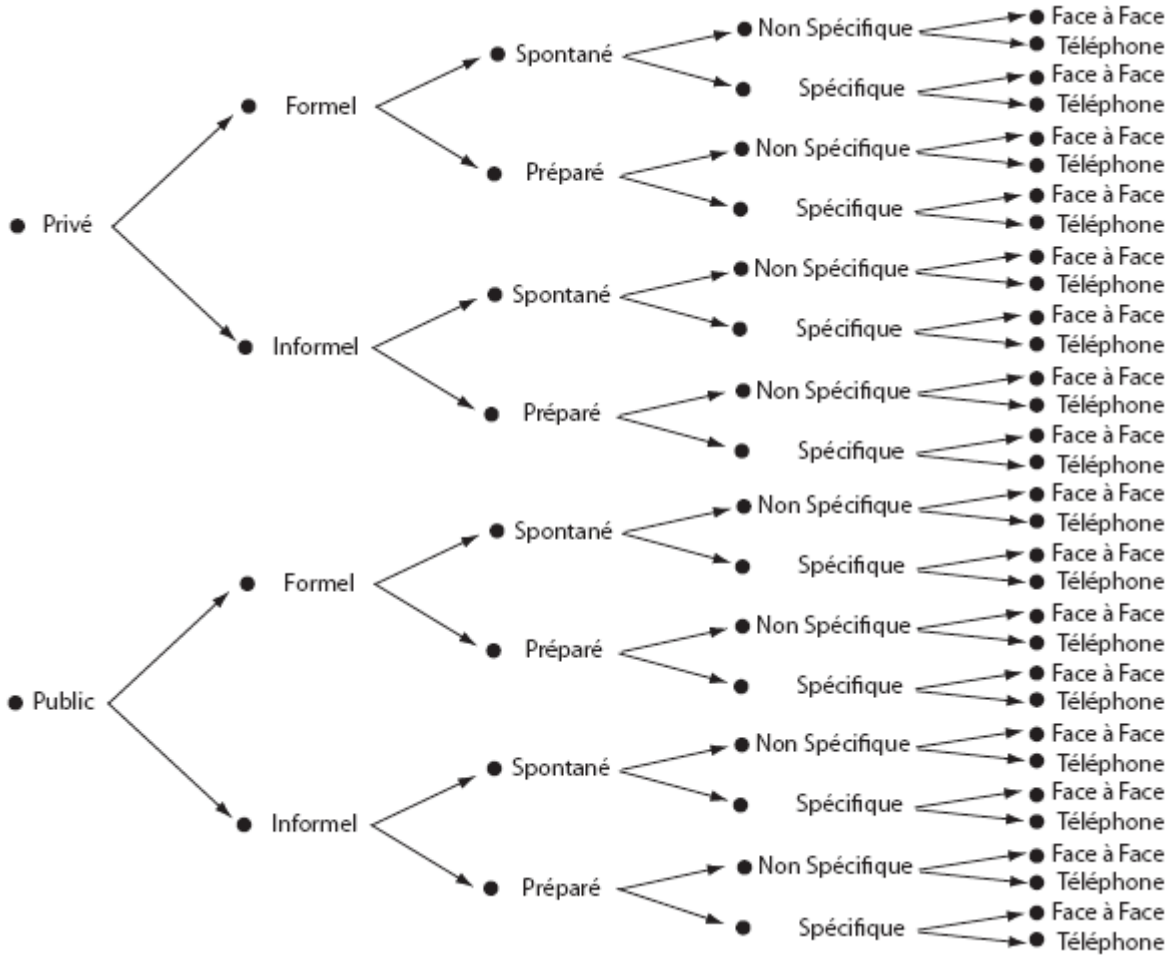

FIGURE.26 Hiérarchie des critères d'échantillonnage retenus

Cette hiérarchisation permet de classer des échanges très variés, de l'entretien téléphonique d'un locuteur qui souhaite postuler dans une entreprise (privé, formel, préparé, spécifique, téléphone) à la conversation d'un couple dont l'objectif est de donner une réponse, dans un jeu télévisé (public, informel, spontané, non spécifique, face à face).

Certaines situations sont tout de même ambiguës : dans un échange entre un journaliste et passant, le journaliste parle peu mais a préparé son interview, tandis que le passant parle plus et de façon spontanée.

Néanmoins, nous avons appliqué cette hiérarchie à l'ensemble des enregistrements des six corpus cités ci-dessus, en écoutant les fichiers audio ou en lisant les transcriptions (TABLEAU 52). Nous n'avons pas hésité à ne pas trancher en cas de doute. C'est pourquoi le nombre de fichiers total varie en fonction des critères. Lorsque nous nous sommes confrontés aux corpus, définir ce qui appartenait au discours spécifique ou non spécifique était assez

compliqué. Nous avons donc finalement laissé ce paramètre de côté. Nous estimons tout de même que les corpus sont majoritairement non spécifiques (environ 70%).

	Nb fichiers /748	% de fichiers	Nb mots /2 332 680	% de mots	Nb fichiers /748	% de fichiers	Nb mots /2 332 680	% de mots	
Privé	639	85,3%	1 977 845	84,8%	99	13,2%	331 362	14,2%	**Public**
Informel	590	78,7%	1 907 894	81,8%	127	17,0%	356 796	15,3%	**Formel**
Spontané	693	92,5%	2 180 488	93,5%	43	5,7%	117 910	5,1%	**Préparé**
Face à Face	695	92,7%	2 244 942	96,2%	39	5,2%	28 044	1,2%	**Téléphone**

TABLEAU 52. Répartition des six corpus oraux dans les catégories définies pour l'échantillonnage

L'échantillonnage n'est donc pas du tout équilibré entre les différents critères : entre 78% et 93% de fichiers sont étiquetés dans les catégories 'privé', 'informel', 'spontané' et 'face à face' ; 1% à 16% du corpus est classé dans les catégories 'public', 'formel', 'préparé' et 'téléphone'. Une représentation hiérarchisée est inutile dans ces conditions.

Maintenant que nous avons une vue globale des genres des corpus écrits et oraux, nous pouvons examiner dans quels genres les occurrences de dépendance à distance apparaissent et sous quelle forme.

3.5.1.2 Dépendance à distance et échantillonnage dans le corpus oral

Dans l'absolu, les constructions de dépendance à distance apparaissent beaucoup plus dans l'une des catégories suivantes : privé, informel, spontané et face à face. C'est ce qu'indiquent les proportions observées (TABLEAU 53) qui se lit comme suit : sur les 91 transcriptions qui comportent des constructions de dépendance à distance, 77 appartiennent à la catégorie 'privé', contre 12 dans la catégorie 'public'[127]. Les transcriptions qui comportent des dépendances à distance sont donc majoritairement des transcriptions d'échanges privés (85%).

[127] Dans les fichiers restants, il n'a pas été possible de trancher en faveur de l'une ou de l'autre catégorie.

	Nb transcriptions /91	% de transcriptions avec LDD		Nb transcriptions /91	
Privé	77	85%	13%	12	Public
Informel	74	81%	16%	15	Formel
Spontané	86	95%	5%	5	Préparé
Face à Face	90	99%	0%	0	Téléphone

TABLEAU 53. Répartition des transcriptions avec dépendance à distance dans les catégories définies pour l'échantillonnage

Un biais pourrait être introduit du fait de la différence de taille entre les deux catégories 'privé'/'public', 'informel'/'formel', 'spontané'/'préparé' et 'face à face'/'téléphone'.

Le tableau suivant présente la répartition des occurrences de dépendance à distance (TABLEAU 54). Sur les 108 occurrences de dépendance à distance présentes dans les six corpus oraux, 86% apparaissent dans des transcriptions d'échanges privés, 82% dans des transcriptions d'échanges informels, 95% dans des transcriptions d'échanges spontanés et 99% dans des transcriptions d'échanges produits en face à face.

	nb occ LDD /108	%LDD		nb occ LDD /108	
Privé	93	86%	12%	13	Public
Informel	89	82%	16%	17	Formel
Spontané	103	95%	5%	5	Préparé
Face à Face	107	99%	0%	0	Téléphone

TABLEAU 54. Répartition des occurrences de dépendance à distance dans les catégories définies pour l'échantillonnage

Les conclusions sont donc très semblables à celles fournies par les répartitions des fichiers. La très grande majorité des dépendances à distance est présente dans des productions privées, informelles, spontanées et réalisées en face à face (FIGURE.27).

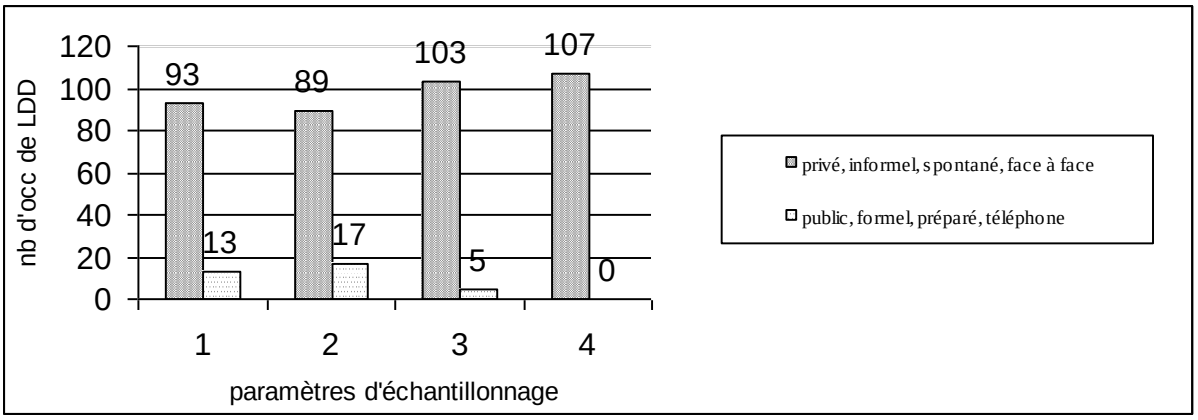

FIGURE.27 Répartition des occurrences de dépendances à distance dans les catégories définies pour l'échantillonnage

Cette répartition n'est pas forcément provoquée par les catégories de l'échantillonnage. Ce n'est pas parce que les occurrences sont plus nombreuses dans le discours privé que ce type de discours favorise la production de dépendances à distance. Comme nous l'avons vu (§3.5.1.1, TABLEAU 52) les catégories ne sont pas équilibrées dans l'ensemble des six corpus. Les échanges privés, par exemple, sont beaucoup plus représentés que les échanges publics. Il est donc plus probable que les occurrences de dépendances à distance soient repérées dans les types d'enregistrements les mieux représentés, tels que les échanges privés, plutôt que dans les types d'enregistrements les plus rares, tels que les échanges publics ou téléphoniques. En effet, si l'on calcule la proportion de fichiers avec dépendances à distance par rapport au nombre de fichiers de chaque catégorie, les résultats sont tout à fait différents (TABLEAU 55). Les proportions sont globalement similaires pour les deux catégories de 'privé'/'public', 'informel'/'formel' et 'spontané'/'préparé'. Seules les catégories 'face à face'/'téléphone' présentent une différence, puisque aucune occurrence de dépendance à distance n'a été relevée dans la catégorie de petite taille des échanges téléphoniques.

	% transcriptions avec dépendance à distance par rapport au nombre de transcriptions de chaque catégorie		
Privé	12,1%	12,1%	Public
Informel	12,5%	11,8%	Formel
Spontané	12,4%	11,6%	Préparé
Face à Face	12,9%	0%	Téléphone

TABLEAU 55. Proportion de fichiers avec dépendances à distance dans chaque catégorie définie pour l'échantillonnage, relativement au nombre de fichiers de chaque catégorie

Comme déjà indiqué, un calcul effectué sur le nombre de fichiers pourrait être biaisé par une longueur totalement différente des fichiers d'une catégorie par rapport à une autre. Néanmoins, le calcul basé sur le nombre d'occurrences de dépendance à distance proportionnellement au nombre de mots de chaque catégorie, ramené à un million de mots, montre un résultat identique (TABLEAU 56) : les proportions sont proches dans les deux pendants des trois premières catégories, et éloignées dans la dernière, à cause de l'absence d'occurrences dans les échanges téléphoniques.

	Rapport pour 1M mots		
Privé	47,0	39,2	Public
Informel	46,6	47,6	Formel
Spontané	47,2	42,4	Préparé
Face à Face	47,7	0	Téléphone

TABLEAU 56. Nombre d'occurrences de dépendance à distance pour un million de mots, dans chaque catégorie définie pour l'échantillonnage

Afin de savoir si les différences entre les catégories sont significatives malgré leur proximité, nous pouvons effectuer le test statistique du Khi2. Celui-ci s'applique uniquement si les catégories contiennent plus de cinq éléments. Il prend en compte à la fois le nombre de fichiers avec dépendance à distance et le nombre d'occurrences de dépendance à distance. L'hypothèse de base (H0) est qu'il n'y a pas d'influence des catégories sur la présence de dépendance à distance. Avec une marge d'erreur de 0,5%, le Khi2 théorique est de 0,455. Si le Khi2 observé est supérieur au Khi2 théorique, l'hypothèse H0 doit être rejetée. Dans notre cas, les Khi2 de chaque couple de catégorie ne sont pas supérieurs (TABLEAU 57). L'hypothèse H0 n'est pas rejetée ; il n'y a donc pas d'influence de la catégorie privée sur public, formel sur informel, spontané sur préparé ni inversement d'influence de la catégorie public sur privé, informel sur formel ou préparé sur spontané.

Critère	**Khi2 observé**
Privé / public	0,064
Formel / informel	0,023
Spontané / préparé	0,077

TABLEAU 57. Khi2 observé dans les catégories définies pour l'échantillonnage

Les résultats du test du Khi2 sont confirmés par ceux de la loi hypergéométrique[128]. Ce calcul prend en compte le nombre d'occurrences de dépendance à distance d'un groupe (par exemple privé et public), le nombre de mots d'un pendant de la catégorie (par exemple privé) et le nombre de mots du corpus total (privé et public). Il évalue à partir de ces données la probabilité d'obtenir n occurrences de dépendance à distance dans le pendant de la catégorie ciblée (dans notre exemple ci-dessus, le pendant 'privé'). La probabilité la plus forte indique le nombre d'occurrences le plus probable pour le pendant, si la répartition n'est influencée par aucun paramètre.

Pour chaque pendant de catégorie, le nombre d'occurrences observé est identique ou très proche du nombre d'occurrences le plus probable (TABLEAU 58). Ainsi, la différence entre les pendants n'est pas significative, puisque le nombre d'occurrences de dépendance à distance observé est pratiquement celui qu'on aurait eu le plus de chance d'observer en distribuant les dépendances à distance au hasard.

	Probabilité maximum de LDD	**Valeur observée**
Privé	91	93
Public	15	13
Informel	90	89
Formel	16	17
Spontané	103	103
Préparé	5	5
Face à Face	106	107
Téléphone	0	0

TABLEAU 58. Probabilité maximum (loi hypergéométrique) et valeur observée de dépendance à distance pour chacune des catégories définies pour l'échantillonnage

On peut donc conclure qu'à l'oral, les facteurs de contexte de discours (privé/public), de type de discours (formel/informel), du degré de préparation du discours (spontané/préparé) et du canal de communication (face à face/téléphone) n'ont pas d'influence sur les occurrences de dépendance à distance. Aucun de ces paramètres ne favorise la production de la construction.

La situation à l'écrit est différente. Bien sûr les paramètres observés afin de définir ce qui favorise la production de la construction sont très différents.

[128] Je remercie énormément Coralie Reutenauer de m'avoir indiqué et expliqué le principe de ce test, puis de l'avoir appliqué sur mes données.

3.5.1.3 Dépendance à distance et échantillonnage dans le corpus CERF[129]

La répartition des occurrences de dépendance à distance dans le corpus du CERF montre une grande disparité. Si l'on considère l'ensemble du corpus, hormis les sous-tranches de presse étrangère[130], on compte 205 occurrences de dépendance à distance. Dans chaque tranche du CERF, le nombre d'occurrences varie entre 1 et 52 (FIGURE.28).

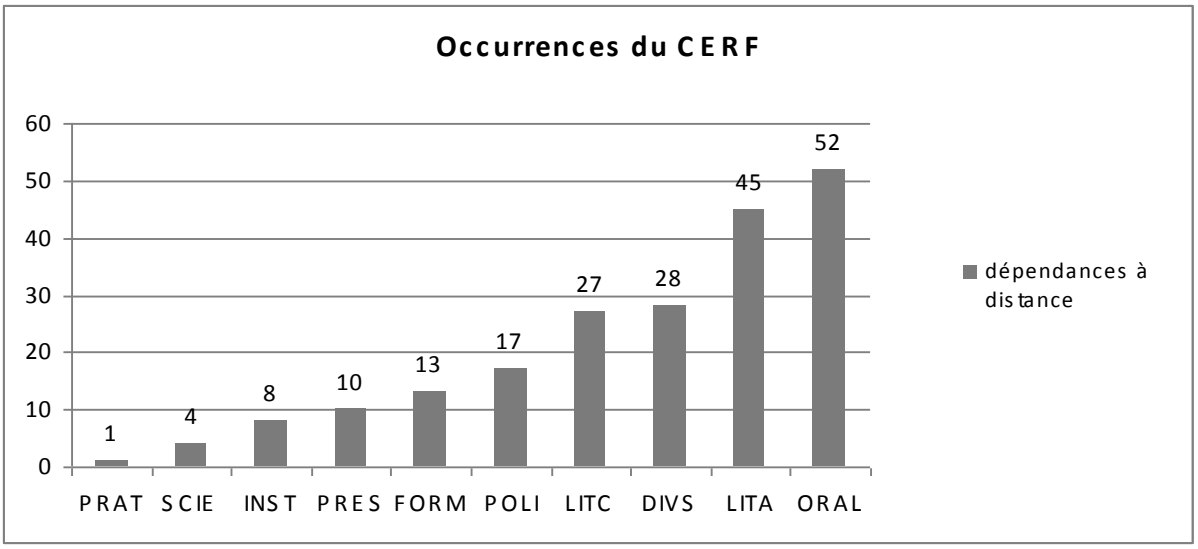

FIGURE.28 Répartition des occurrences de dépendances à distance dans le corpus CERF

Mais le nombre d'occurrences n'est pas le seul paramètre qui varie en fonction du genre de texte. Pour chaque genre, nous avons examiné différentes propriétés de la construction de dépendance à distance : le nombre d'occurrences, le nombre de verbes ponts différents dans ces occurrences, la configuration syntaxique la plus représentée, la personne à laquelle fait référence le sujet du verbe pont et le type d'interrogatif-relatif. Certaines tranches révèlent des comportements similaires et trois groupes se distinguent. Deux groupes sont opposés, et le troisième est intermédiaire.

[129] Les analyses présentées dans cette section ont fait l'objet d'une communication :
Bérard L. (2012) « Entre codification et liberté : variation des constructions de contrôle à distance en français contemporain. » Communication orale lors du colloque international de l'AFLS, *Le français à travers le temps : acquisition, changement, variation*, Newcastle, 1-3 juin

[130] En récupérant donc les occurrences écartées de Politique et de Littérature contemporaine, ainsi que la tranche Littérature Ancienne (LITA). Voir p.140

Le premier groupe réunit les tranches Loisirs et vie pratique (PRAT), Sciences et Techniques (SCIE), Institutions (INST) et Politique (POLI). Dans ce groupe (éléments caractéristiques en rouge), les occurrences de dépendance à distance sont peu nombreuses : entre 1 et 17 occurrences pour un million de mots. La richesse du lexique des verbes ponts, en revanche, est très grande, proportionnellement au nombre d'occurrences (80% en moyenne). Le verbe pont le plus fréquent est « penser ». La configuration syntaxique dominante est la relative. Le sujet du verbe pont de ce groupe est généralement à la troisième personne ou à la forme impersonnelle (P3b). Les pronoms interrogatifs-relatifs communs à ce groupe sont « lequel » (avec ses variations « laquelle », « lesquels », « lesquelles »), « dont » et « comment ». Voici le détail de ce groupe (TABLEAU 59) :

	PRAT	**SCIEN**	**INST**	**POLI**
Nombre LDD	1	4	8	17
Nombre verbes	1	4	7	10
Variété verbes	-	**100%**	**90%**	**60%**
Vpont le plus fréquent	penser	-	penser	penser
Configuration syntaxique	**interro indirecte**	**rel**	**rel (75%)**	**rel (60%)**
Personne sujet	P3	P2 (50%)	P3b (50%)	P3 (40%)
Type d'interrogatif-relatif	comment	l-quel- (3) dont	qu- (3) l-quel- (2) dont, comment	où (7) l-quel-, comment pourquoi, dont

TABLEAU 59. Comportement des dépendances à distance dans le groupe 1 du CERF

Le deuxième groupe réunit les tranches Littérature contemporaine (LITC), Divers (DIVS), Oral (ORAL) et Littérature ancienne (LITA). Ce groupe est à l'opposé du premier (éléments caractéristiques en bleu). Il présente cinq fois plus d'occurrences de dépendances à distance (même si ce nombre reste faible pour un million de mots) et presque trois fois moins de variété dans les verbes ponts, toujours proportionnellement au nombre d'occurrences de la construction (30%). Le verbe le plus fréquent est « vouloir ». La configuration dominante est

l'interrogative directe. Le sujet du verbe pont est un pronom personnel de la deuxième personne. Les pronoms interrogatifs-relatifs les plus fréquents du groupe sont « qu'est-ce que » et « que ». Les pronoms du premier groupe peuvent aussi apparaître mais de façon très occasionnelle (de deux à six occurrences dans chaque genre). On remarque que le genre LITA s'écarte légèrement du groupe (TABLEAU 60).

	LITC	**DIVS**	**ORAL**	**LITA**
Nb LDD	27	28	52	45
Nb verbes	9	8	12	18
Variété V	**33%**	**29%**	**23%**	**40%**
Vpont	vouloir	vouloir	vouloir	vouloir
Config stxq	**Interrogative directe (52%)**	**Interrogative directe (53,5%)**	**Interrogative directe (44%)**	Relative (55,5%)
Pers sujet	P2 (44%)	P2 (61%)	P2 (42%)	P2 (33%)
Type d'interro-relatif	qu'est-ce que (9), que (7), où (4) l-quel-, dont, comment (2) qui	qu'est-ce que, que (6) qui (5) pourquoi (3) où, dont (2) comment, combien, l-quel-	qu'est-ce que (15) où (13) que (6) quel- (3) qui, quand, pourquoi (2) l-quel-, combien	où (20), qu- (14) l-quel-, dont, qu'est-ce que, qui, combien, pourquoi

TABLEAU 60. Comportement des dépendances à distance dans le groupe 2 du CERF

Le troisième groupe réunit les tranches Presse (PRES) et Forum (FORM). Ce groupe est un intermédiaire entre les deux premiers. Comme pour le premier groupe, le nombre d'occurrences de dépendance à distance est assez faible et une grande variété de verbes est utilisée. En revanche, le verbe pont le plus fréquent est « vouloir », la configuration principale est l'interrogative directe, avec un verbe pont conjugué à la deuxième personne (même si la domination est faible), ce qui le rapproche du deuxième groupe. Les pronoms interrogatifs-relatifs des deux groupes précédents sont présents ; ils sont employés de façon quasi-équivalente. Le tableau suivant présente le détail de ces deux tranches : Forum et Presse (TABLEAU 61).

	FORM	PRES
Nb LDD	13	10
Nb verbes	8	6
Variété verbes	**60%**	**60%**
Vpont le plus fréquent	vouloir	vouloir
Config stxq	**interro dir (45%)**	**interro dir (50%)**
Personne sujet	P2 (45%)	P2 (50%)
Type d'interro-rel	qu- (3) pourquoi (2) dont, qu'est-ce que, où, comment	comment, qu- (3) où (2) qu'est-ce que, dont

TABLEAU 61. Comportement des dépendances à distance dans le groupe 3 du CERF

On constate donc que les trois groupes forment un continuum. Pour plus de facilité, nous les avons ordonnés et renommés en fonction de la richesse lexicale du verbe pont. Les chiffres fournis dans le (TABLEAU 62) sont le résultat de moyennes pour les quatre premiers paramètres et le résultat d'une addition pour les trois paramètres suivants.

	G1	G2	G3
	Prat, scien, inst, poli	**Forum, presse**	**Litt, divs, oral**
Nb LDD	7,5	11,5	38
Nb verbes	5,5	7	11,8
Variété V	85%	60%	30%
Configurat° syntaxique	Relative 78% (sans PRAT)	Interro directe 48%	Interro directe 50% (sans LITA)
V le +fréq	Penser (26% des occ)	Vouloir (43% des occ)	Vouloir (40% des occ)
Pers sujet	P3 / P3b : 53%	P2 : 43%	P2 : 43%
Type d'interro-rel	comment, l-quel-, dont : 56%	qu-, qu'est-ce que : 42% dont, comment : 32%	qu'est-ce que, qu- : 46%

TABLEAU 62. Récapitulatif du comportement des dépendances à distance dans les trois groupes du CERF

Nous présentons ci-dessous une attestation typique de chacun des groupes. Celle du G1 présente une relative en « lequel », et un sujet de verbe pont à la troisième personne (366).

(366) *Le conflit pour le salaire prolonge le rapport de marchandage* ***dans lequel chacun sait qu'il a besoin de l'autre*** *: chacun s'efforce de rouler l'autre suffisamment pour avoir fait une bonne affaire, mais pas trop sinon le maintien de la relation deviendrait incertain. (CERF, SCIE/HermesRev)*

Celle du G2 présente une relative en « dont », avec un verbe relativement recherché, qui reflète la richesse lexicale du groupe (367).

(367) *Bien sûr, le ciné (du moins, celui* ***dont je présume que tu parles****), c'est facile (CERF, FORM/Temoign)*

Dans ce groupe, peuvent également apparaître des interrogatives en « que », avec un verbe pont de forme « vouloir » et un sujet à la deuxième personne (368).

(368) *Ils nous accordent 35% des recettes fiscales alors que les charges grimpent constamment ;* ***que voulez-vous qu****'on fasse avec cela ?" (CERF, PRES / Monde Diplo)*

Enfin, l'attestation typique du troisième groupe présente une interrogative en « qu'est-ce que », avec un verbe pont de forme « vouloir » et un sujet à la deuxième personne (369).

(369) *qu'est-ce que vous voulez que je vous dise" (CERF, ORAL)*

Les regroupements observés grâce à ces propriétés peuvent sembler étranges. Pourquoi certaines tranches ont-elles un comportement similaire ? Qu'est-ce qui rapproche ou éloigne les genres observés ? Est-ce le contenu thématique, le degré d'élaboration, le caractère formel/informel du texte ou le degré de codification ?

La première hypothèse est que le paramètre qui est à l'origine du regroupement et du découpage des groupes soit le contenu thématique. Le contenu thématique est souvent pris comme critère d'échantillonnage, mais il est peu probable qu'il ait un impact sur la structure syntaxique des textes, et plus particulièrement sur celle des dépendances à distance. En outre, le contenu de chaque groupe est relativement hétérogène. Il y a certes des similitudes dans les contenus des tranches de chaque groupe, mais il y a également de grandes divergences. Dans le G1, les discours des hommes politiques ou des partis politiques (POLI) partagent certainement beaucoup de thématiques avec les textes du parlement, du sénat et de l'assemblée (INST) mais ils n'ont vraiment rien de commun avec les manuels techniques, les

revues scientifiques ou les thèses (SCIEN), qui elles aussi ne partagent pas leur contenu avec celui de la finance (PRAT). Dans le G2, les textes des forums ayant trait à l'économie, à l'actualité mondiale et à la politique (FORM) ont des contenus similaires à ceux de la presse (PRES) mais ils n'en ont pas avec les autres sous-parties de FORM qui concernent le sport, les médias ou les témoignages. Enfin, dans le G3, plusieurs sous-parties sont semblables : journal intime (DIVS) et journal autobiographique (LITT) ou éventuellement les dialogues entre les personnages au théâtre (LITT) et les conversations (ORAL). D'autres sont tout à fait différentes dans leur contenu : la critique littéraire, les textes religieux, les textes humoristiques, etc. (DIVS). Le contenu thématique n'est donc pas la cause de regroupement ou de séparation entre les tranches.

La seconde hypothèse est que l'élément unificateur des trois groupes soit le degré d'élaboration des textes. Mais ce critère ne semble pas pertinent. D'une part, la littérature, qui est forcément une production travaillée et élaborée, se trouve dans le même groupe que la tranche d'oral, majoritairement spontanée dans ce corpus. Il peut bien sûr y avoir des imitations de l'oral dans les romans ou le théâtre, mais cela n'est pas dominant dans la tranche littérature, qui comporte également de la poésie, des biographies, etc. D'autre part, les sujets lexicaux, qui sont une marque d'élaboration, sont présents dans tous les groupes. Ils se répartissent très inégalement en fonction des tranches. Dans chaque groupe, certaines tranches attestent de sujets lexicaux et d'autres pas (TABLEAU 63).

G1	PRAT	SCIEN	INST	POLI
Nb LDD	1	4	8	17
Sujet GN	0	0	0	4

G2	FORM	PRES
Nb LDD	13	10
Sujet GN	1	0

G3	LITC	DIVS	ORAL	LITA
Nb LDD	27	28	52	45
Sujet GN	1	1 NP	0	4

TABLEAU 63. Répartition des sujets lexicaux dans les groupes 1, 2 et 3 du CERF

La troisième hypothèse est que l'élément unificateur des groupes soit le caractère plus ou moins formel de la langue. Ce critère est courant pour distinguer les genres. L'inversion sujet-verbe est généralement considérée comme marque de langue formelle. Or certaines tranches présentent énormément d'inversions (entre 80% et 100% pour PRES, FORM, LITA et POLI) et d'autres relativement peu (entre 0% et 55% pour ORAL, LITC et DIVS). Si les tranches avec inversions appartiennent aux G1, G2 et G2, celles qui en comportent peu appartiennent toutes au G3. On peut y voir un indice de convergence.

Le choix de vouvoyer ou de tutoyer est une autre marque de langue formelle/informelle. Seul LITA possède une majorité de tutoiement, ce qui peut paraître étonnant. Néanmoins, on constate une diminution de la proportion des « vous » au profit des « tu » dans les tranches qui correspond approximativement aux groupes constitués à partir du comportement des dépendances à distance (TABLEAU 64). Ce trait peut être considéré comme un deuxième indice de l'importance du critère formel/informel dans l'emploi des dépendances à distance.

Tranche	**Proportion**	**Groupe des dépendances à distance**
SCIEN	Vous	G1
PRES	Vous	G2
FORM	4 vous, 1 tu	
ORAL	15 vous, 7 tu	G3
DIVS	11 vous, 6 tu	
LITC	7 vous, 5 tu	
LITA	12 tu, 3 vous	

TABLEAU 64. Proportion de tutoiement et de vouvoiement dans les tranches du CERF

Enfin, le degré de formalité du type d'interrogatif-relatif, lui, correspond complètement au découpage des groupes. Le pronom « qu'est-ce que », informel, est réalisé dans le G3 tandis que les pronoms « dont » et « l-quel- », formels, sont réalisés dans le G1.

Il est donc fort probable que le critère de langue formelle intervienne dans le comportement des dépendances à distance, au moins en partie. En effet, le dernier critère observé semble déterminant.

Mais le principal critère unificateur semble être le quatrième critère proposé : celui de la formalisation, de la codification. La codification est le fait de « mettre des formes », c'est à dire de « donner [...] à un discours la forme qui est reconnue comme convenable, légitime, approuvée ». (Bourdieu, 1986, p. 43). Elle permet à la fois l'homogénéisation et la prévision. « On sait à quoi s'en tenir » (Bourdieu, 1986, p. 41). La codification implique une certaine « objectivation » (Bourdieu, 1986, p. 42). D'après le même auteur, « on y perd en charme... » (Bourdieu, 1986, p. 42). Avec un texte codifié, la communication est « minimale » (Bourdieu, 1986, p. 42).

Comme l'a démontré (Verhagen, 2006) (§1.3.5.2), la fonction la plus importante des dépendances à distance est l'expression de l'intersubjectivité. Elles permettent au locuteur d'exprimer ses sentiments, son point de vue, ou de demander son avis à son interlocuteur. Il devrait donc y avoir peu de dépendances à distance dans les textes codifiés. La codification est opposée au « jeu », à l' « improvisation », à l' « invention » (Bourdieu, 1986, p. 42-43). Il devrait donc y avoir beaucoup de dépendances à distance dans les fictions, l'humour, la chanson, etc. Les indications semblent correspondre aux groupes définis. Le G1, qui réunit des textes très codifiés (manuels techniques, thèses, discours politiques, etc.) présente peu d'attestations de dépendances à distance, alors que le G3, dont les textes sont peu codifiés (littérature, oral, divers), en présentent beaucoup plus. Enfin, le G2 est intermédiaire (forum, presse).

Cette conclusion est confirmée par quelques auteurs, pour qui la littérature est peu codifiée, tout comme l'oral. Il n'est donc pas étonnant que ces deux tranches puissent être classées dans le même groupe.

> « Le champ littéraire se caractérise par un très faible degré de codification »
> ((Bourdieu, 1992, p. 370-371) dans (Bois, 2009))

> « Plus la codification est forte, mieux le texte échappera, dans le cours du temps, au changement ; affaiblie ou inexistante, elle donne libre cours aux variations, qui ont souvent été (abusivement) considérées comme caractéristiques de l'oralité. » (Zumthor, 2008, p. 183)

La richesse lexicale du verbe pont dans les tranches du G1 pourrait sembler remettre en cause l'explication fournie par la codification. Si le texte est codifié, il ne devrait peut-être pas varier en construction et en lexique. Mais il est possible que l'effet soit inverse. Lorsque les textes sont très codifiés, la richesse lexicale peut être un espace de liberté pour le scripteur qui rend le texte moins monotone au lecteur. Parallèlement, l'absence de codification peut être vue comme une trop grande liberté pour les locuteurs, qui se 'rassurent' en formalisant les structures qu'ils choisissent. En outre, l'oral spontané, tranche dans laquelle apparaît le plus d'attestations, doit tenir compte du paramètre de temps de production, qui limite la recherche lexicale.

Pour résumer, la construction de dépendance à distance ne s'emploie pas comme un patron pouvant subir des variations minimales en français. Il existe plutôt des déclinaisons de la construction en fonction du degré de codification et de langue formelle/informelle du genre textuel. Parmi les 10 tranches prédéfinies du CERF, trois groupes forment un continuum et attestent d'un emploi dominant de la construction :

- dans le groupe de textes très codifiés, la construction majoritaire est une relative en « dont/l-quel- P3 V que », avec une grande variété de verbes ponts ;
- dans le groupe de textes peu codifiés, la construction majoritaire est une interrogative directe en « que/qu'est-ce que P2 V que », avec peu de variété des verbes ponts.
- dans le groupe de textes moyennement codifiés, les constructions des deux autres groupes se côtoient.

Il serait intéressant d'observer d'autres phénomènes sensibles à la codification afin de croiser les résultats et de vérifier que les groupes de tranches sont identiques.

Les dépendances à distance varient donc en nombre et en forme selon le type de textes dans lequel elles sont employées. Nous allons examiner si la construction est également sensible au type d'interlocuteur (adulte ou enfant).

3.5.2 Adultes-enfants

D'après (Dąbrowska et al., 2009), il y aurait cinq fois plus d'occurrences de dépendance à distance lorsque les adultes s'adressent aux enfants (corpus Manchester) que lorsqu'ils s'adressent aux adultes (corpus BNC-oral) (TABLEAU 65). Un enfant entendrait ainsi une construction de dépendance à distance environ une fois toutes les soixante-dix minutes, soit aussi souvent que les termes « *bus* », « *old* » et « *able* ». Parmi ces dépendances à distance, la formule prototypique « WH do you think S » représenterait 85% des occurrences et serait donc largement diffusée (voir §1.3.5.1). Nous avons voulu vérifier si le système était comparable en français.

Corpus	Nombre de mots	Nombre de dépendance à distance (adultes)
Manchester	1 000 000	225
BNC (oral)	1 000 000	42

TABLEAU 65. Nombre de mots et d'occurrences de dépendances à distance dans les corpus anglais : comparaison d'un corpus adulte et d'un corpus adultes-enfants

Dans le corpus adultes-enfants de TCOF (175 000 mots), constitué d'enregistrements de conversations et de narration, nous n'avons relevé que deux occurrences de dépendance à distance, toutes les deux dans la partie « conversation » des données. Ces occurrences présentent en verbe pont les deux verbes les plus fréquents de l'ensemble des occurrences de dépendance à distance dans les corpus adultes-adultes : « vouloir » et « falloir ». Les adultes qui les prononcent s'adressent à des enfants de 4 ans, 4 mois (370) et 5 ans, 7 mois (371).

(370) je marque /du,0/ maquillage après qu'est-ce qu' il faut que j'écrive ah oui les perles j'avais oublié les perles (TCOF, T_Nesrine1_Sed)

(371) parle dedans vas-y hein qu'est-ce que tu veux que je dise ah euh il fait beau aujourd'hui hein arrête (TCOF, T_Manon1_Lag)

Le faible nombre d'occurrences de dépendance à distance n'est pas dû à une quelconque particularité du corpus puisque nous obtenons des résultats relativement similaires à ceux relevés pour d'autres corpus de conversation entre adultes et enfants disponibles sur la plateforme Childes[131] : Paris[132] (Leroy & Mathiot, 2009; Morgenstern, sous presse;

[131] Voir le site http://childes.psy.cmu.edu/

Morgenstern & Parisse, 2007; Morgenstern & Sekali, 2009), Hammelrath[133], Lyon[134] (Demuth & Tremblay, 2008) (TABLEAU 66). En recherchant uniquement les deux constructions les plus fréquentes, du type « qu'est-ce que tu veux / il faut qu », nous avons repéré environ 11 occurrences de dépendance à distance tous les millions de mots. Le corpus de TCOF en comptait 18,9 pour un million de mots de conversation. La différence avec le corpus Manchester, qui en comptait 225, est énorme.

Corpus français de conversation	**Nb mots**	**Nb LDD en « vouloir » et « falloir » (adultes)**	**Nb occ pour 1M mots**
Lyon_childes	1 203 889	33	**27,4**
Paris_childes	595 005	4	**6,7**
Hammelrath_childes	163 758	0	**0**
MOYENNE			**11,4**

TABLEAU 66. Nombre de mots et d'occurrences de dépendances à distance dans trois corpus adultes-enfants français

Le corpus anglais ne semble pas être spécifique non plus. Les résultats auraient même pu être beaucoup plus nombreux. En recherchant uniquement la formule prototypique *do you think* dans d'autres corpus anglais de Childes (TABLEAU 67), nous repérons en moyenne 405 occurrences tous les millions de mots (corpus Abe (Kuczaj & Stan, 1977), Belfast (Henry, 1995; Wilson & Henry, 1998), Fletcher (Fletcher & Garman, 1988), Manchester (Theakston, Lieven, Pine, & Rowland, 2001)), auxquelles il faudrait ajouter les occurrences qui présentent des variations dans le lexique verbal, le temps, la personne, etc. pour connaître le nombre réel de constructions de dépendance à distance.

[132] Paris [2005-2008] : enregistrement de 3 enfants âgés de 7 mois à 3 ans, 5 mois, à Paris, chez eux. Les enregistrements ont eu lieu une fois par mois.

[133] Hammelrath : échantillon de parole de 258 enfants (125 filles et 133 garçons), âgés de 3 ans, 6 mois à 4 ans, 1 mois dans huit écoles (Nord et Pas-de-Calais). Un professionnel demande aux enfants de décrire trois images. 116 des enfants présentent du retard dans l'acquisition du langage. Le second échantillon est composé de 107 filles et de 121 garçons de 3 ans, 6 mois à 6 ans, 9 mois.

[134] Lyon [2002-2005] : enregistrement de 4 enfants (2 garçons et 2 filles), de 1 à 3 ans, à la maison dans leurs interactions avec leur mère. Les enregistrements sont effectués toutes les deux semaines et durent une heure. (185h de parole au total)

Corpus anglais de conversation	Nb mots	Nb LDD en « do you think » (adultes)	Nb d'occurrences pour 1M mots
Abe	364 445	121	332
Belfast	693 070	154	222
Fletcher	292 592	193	660
Moyenne			**405**

TABLEAU 67. Nombre de mots et d'occurrences de dépendances à distance dans trois corpus adultes-enfants anglais

La différence viendrait donc soit de la langue, soit de l'absence d'une formule prototypique en français.

En français, les adultes ne semblent pas employer plus de constructions de dépendance à distance lorsqu'ils s'adressent aux enfants plutôt qu'aux adultes. Ce résultat apparaît même si l'on compare uniquement les structures interrogatives directes (TABLEAU 68).

Corpus français	Nb mots	Nb LDD (interrogatives)	Nb occ pour 1M mots
Adultes	2 730 000	51	**19**
TCOF_enfants	175 000	2	**11**

TABLEAU 68. Nombre de mots et d'occurrences de dépendances à distance dans les corpus français : comparaison d'un corpus adultes et d'un corpus adultes-enfants

Si l'on réunit uniquement les occurrences en « qu'est-ce que tu veux qu » et « il faut qu » dans les corpus d'interactions entre adultes, on constate que proportionnellement au nombre de mots, la première construction est moins importante que dans le corpus adultes-enfants. La seconde est à peu près équivalente (TABLEAU 69).

Corpus français	Nb mots	« qu'est-ce que tu veux qu »	Nb occ pour 1M mots	« faut qu »	Nb occ pour 1M mots
Total enfants	2 037 652	21	**10,3**	18	**8,8**
Adultes_oral	2 730 000	13	**4,8**	26	**9,5**

TABLEAU 69. Comparaison de deux constructions ponts dans un corpus adultes et un corpus adultes-enfants français

Il est intéressant de noter que la forme la plus fréquente des interrogatives prend une interprétation différente dans le discours adultes-adultes et adultes-enfants. Dans le corpus

adultes-enfants, 19 occurrences de « qu'est-ce que tu veux qu » sur 21 expriment la volonté (90%) (372) ; dans le corpus adultes-adultes, 11 occurrences sur 13 sont des questions rhétoriques (85%) (479). Ainsi, si la forme est identique, l'emploi ne l'est pas du tout.

(372) qu'est-ce que tu veux que je dessine (LYON, Marie)

(373) Bien sûr. A chacun son rôle. Lui, il doit nous faire mourir, et nous, nous devons aller enterrer notre frère. C'est comme ça que ça été distribué. Qu'est-ce que tu veux que nous y fassions ? (CERF, LITC)

Si l'on regarde plus en détail les occurrences relevées dans les corpus français, on note que la répartition entre « vouloir » et « falloir » est équivalente. On remarque également une grande disparité dans le nombre d'occurrences, en fonction des différents corpus (TABLEAU 70).

Corpus enfants	**Nb mots**	**Nb occ de *qu'est-ce que tu veux qu***	**Nb occ pour 1M mots**	**Nb occ de *faut qu***	**Nb occ pour 1M mots**
TCOF_conversation	105 647	1	**9,5**	1	**9,5**
Lyon_childes	1 203 889	18	**14,9**	15	**12,5**
Paris_childes	595 005	2	**3,4**	2	**3,4**
Hammelrath_childes	163 758	0	**0**	0	**0**

TABLEAU 70. Comparaison de deux constructions ponts dans quatre corpus adultes-enfants français

Le corpus de Lyon comporte beaucoup plus d'occurrences de dépendance à distance que les autres corpus. Ces résultats pourraient s'expliquer par exemple par le contexte de l'enregistrement, l'âge des enfants ou l'identité des adultes. Ce corpus enregistre les enfants chez eux, avec leurs parents. En cela, il est identique au corpus de Paris. Or le corpus de Paris ne possède pas autant d'occurrences. L'âge des enfants pourrait être un facteur explicatif. Si l'on compare trois des corpus (Lyon, Paris et TCOF), on peut voir que les parents utilisent les dépendances à distance en s'adressant aux enfants de 0 à 3 ans. Les occurrences diminuent lorsque les enfants ont plus de trois ans (TABLEAU 71). Dans TCOF, les occurrences relevées s'adressent à des enfants un petit peu plus âgés (4-6 ans), tranche d'âge qui n'existe pas dans les corpus de Lyon et de Paris.

	Lyon	Paris	TCOF
0-2 ans	14	1	
2-3 ans	15	3	0
3-4 ans	4	0	0
4-5 ans			1
5-6 ans			1

TABLEAU 71. Répartition des occurrences de dépendance à distance selon l'âge de l'interlocuteur enfant, dans trois corpus français

Par ailleurs, les adultes enregistrés dans TCOF ne sont pas les parents des enfants. Il serait intéressant de vérifier s'il existe des différences dans la manière de s'adresser à l'enfant selon le statut de l'adulte (parents ou non). On pourrait également faire l'hypothèse d'une corrélation entre le statut de l'adulte par rapport à l'enfant et l'âge de l'enfant, sur des corpus plus importants. Il est pour l'instant difficile de tirer une quelconque conclusion sur si peu d'occurrences.

Une autre hypothèse concernant la particularité du corpus de Lyon quant au nombre d'occurrences de dépendance à distance est que les emplois soient idiolectaux. La comparaison des attestations de « qu'est-ce que tu veux qu » dans chaque sous-corpus de Lyon fait apparaître une sur-représentation de la construction chez la maman de Marie[135] (TABLEAU 72).

LYON	nb d'occurrences	Nb de locuteurs
***Anaïs**	2	2
***Marie**	7	1
***Nathan**	2	1
***Theotime**	4	1
Médiane	**3**	

TABLEAU 72. Répartition des occurrences de dépendance à distance selon les familles dans un corpus adultes-enfants français : Lyon

[135] Il n'y a en revanche pas de sur-représentation de dépendance à distance en *falloir* dans les sous-corpus de Lyon.

La même remarque peut être faite pour le corpus Manchester : le nombre d'occurrences de « *do you think* » en construction de dépendance à distance varie énormément selon les mères (FIGURE.29).

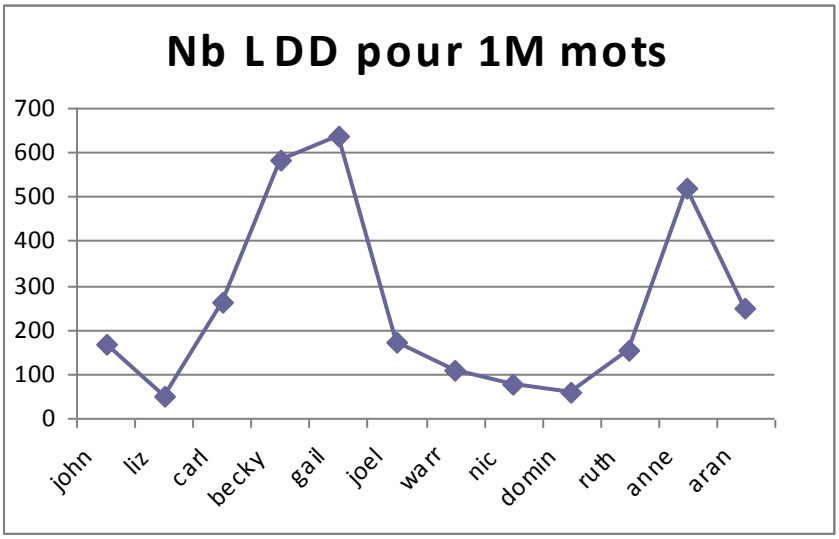

FIGURE.29 Répartition des occurrences de dépendance à distance selon les familles dans un corpus adultes-enfants anglais : Manchester (pour un million de mots)

La prépondérance d'une formule prototypique, qui est encore plus marquée dans les corpus adultes-enfants que dans les corpus adultes-adultes est une conclusion qui semble avérée en langue anglaise, bien que le facteur idiomatique, peu étudié, puisse également avoir un rôle. En français, si une construction domine dans les interrogatives, il n'est pas certain que l'on puisse prétendre qu'il s'agit d'une formule prototypique. Cette construction dominante dans les interactions adultes-adultes est très peu employée par les adultes qui s'adressent aux enfants, à l'exception de quelques rares personnes isolées. Cette construction en « vouloir » est certainement marquée, car son emploi est très différent en fonction des locuteurs et des types d'interaction. Elle possède, nous l'avons vu, un sens rhétorique entre adultes qui est totalement absent dans les interactions adultes-enfants. La seconde construction la plus fréquente, avec « falloir » en verbe pont, est utilisée de façon beaucoup plus uniforme : sa fréquence d'apparition est identique pour tous les locuteurs, et dans tous les types d'interaction.

3.6 Conclusion

En français, les constructions de dépendance à distance en complétive ne se réduisent pas à l'extension d'une formule prototypique en structure interrogative, contrairement à ce qu'il semble se passer en anglais et en néerlandais.

La construction doit répondre à certaines contraintes syntaxiques, qui semblent elles aussi moins importantes qu'en anglais. L'extraction à partir d'un sujet est a priori interdite (et non attestée), ainsi que l'extraction d'une partie d'un ajout, mais l'extraction de relative est fréquente en « dont » et l'extraction de complétives régies par un nom est possible avec certains noms non descriptifs associés à certains verbes supports.

Concernant les verbes ponts, aucune contrainte lexicale ne semble s'appliquer. Presque tous les verbes à complétive pourraient être ponts. La principale condition pour qu'un énoncé soit acceptable serait qu'il ne cumule pas trop d'éléments qui pourraient constituer une difficulté, soit pour l'encodeur, soit pour le décodeur. La production est donc facilitée si le verbe pont est fréquent, sémantiquement simple ou faible (modalisant), s'il est syntaxiquement peu développé, et si le processing entre l'élément antéposé et son verbe recteur est non ambigu. Le choix du verbe pont est encore moins contraint lorsque le reste de l'énoncé est déjà connu par les interlocuteurs, par exemple dans les cas de reprises.

Enfin, les dépendances à distance en complétive sont conditionnées par des paramètres externes. L'emploi dominant varie en fonction du degré de codification et du contexte formel/informel de la langue. L'emploi se différencie également en fonction du statut de l'interlocuteur (adulte *vs* enfant). La fréquence d'utilisation de cette construction semble être influencée par le locuteur et son idiolecte. Même dans les forums Internet, il n'est pas rare de remarquer que plusieurs emplois proviennent d'un même locuteur.

Les résultats précédents nous orientent vers une interprétation du phénomène en termes classiques de contraintes de performance sur des structures de compétence mettant essentiellement en jeu des enchâssements de complétives. L'enchâssement étant couteux, le pont est réduit afin que la production et la compréhension de l'énoncé soient plus aisées.

Mais ces résultats pourraient aussi être interprétés en terme de pure structure : la forme prototypique du pont (construction verbale réduite à « Sujet V que ») est en fait associée systématiquement à une interprétation : une modalisation au sens large de V2. On pourrait

donc considérer que cette séquence modale forme avec le constructeur une unité particulière irréductible à l'enchâssement, en reprenant la piste ouverte par TAG et le modèle sens texte modifié.

Dans le chapitre suivant, nous allons tenter de trancher entre ces deux analyses, notamment en comparant les dépendances à distance à verbe fini avec les dépendances à disance à l'infinitif.

CHAPITRE 4. VERS UNE INTERPRETATION DES RESULTATS : ENCHASSEMENT OU COMPLEXE VERBAL ?

L'analyse des dépendances à distance à verbe fini, et plus particulièrement l'examen de la construction du pont a démontré que la CV1 était réalisée de manière contrainte et spécifique : un sujet pronominal et un verbe. L'ensemble est interprété comme une modalité portant sur le V2. Dans ce chapitre, nous tentons de choisir l'interprétation théorique la plus probable : y a-t-il un enchâssement de la CV2 dans la CV1 (ce qui justifierait la dépendance à distance) ou un complexe verbal (ce qui écarterait la dépendance à distance) ?

Afin de répondre à cette question, nous examinerons dans la section suivante le fonctionnement des dépendances à distance dans un autre type de séquence : le cas où l'élément qu- dépend d'un verbe à l'infinitif, avec toujours la présence d'un verbe pont V1.

(374) c'est exactement ce que je rêvais d'entendre (CFPP, deBordes_7e)

La description de ce type de séquence nous permettra notamment de préciser la relation syntaxique entre le V1 et le V2.

4.1 Préalable : Comparaison du pont dans les séquences avec verbe fini et avec verbe infinitif

Nous avons décrit en détail les dépendances à distance dans les structures en complétive au chapitre précédent. Mais l'effet de dépendance à distance à un verbe peut aussi s'observer dans d'autres séquences. Nous allons maintenant décrire la construction dans les structures en infinitive. Il ne s'agira pas d'une étude exhaustive comme dans le premier cas, mais d'une sélection de ce type de données que nous jugeons suffisante pour éclairer notre problématique centrale. C'est la raison pour laquelle cette description figure dans ce chapitre consacré à l'interprétation des résultats

Dans un premier temps, nous justifierons l'intérêt de décrire ces données, qui ont été très peu étudiées dans cette perspective auparavant. Ensuite, nous décrirons le lexique et la syntaxe du verbe pont, comme nous l'avons fait pour les complétives. Cette démarche nous permettra de comparer les deux structures. Enfin, nous tenterons d'établir un classement des verbes ponts apparaissant dans les structures infinitives, en fonction de leur degré de verbalité (modaux, semi-modaux, constructeurs). Leur comportement syntaxique démontrera que que le lien entre le verbe pont et l'infinitif est particulièrement fort et justifie l'établissement d'une unité linguistique spécifique irréductible à un enchâssement : la chaîne verbale.

4.1.1 Contexte : pourquoi étudier les infinitives ?

De nombreux travaux ont étudié les contraintes qui pèsent sur la construction de dépendance à distance mais la majorité de ces travaux portent uniquement sur les complétives :

> « Les travaux à partir desquels a été peu à peu élaboré ce modèle comme ceux qui se sont ensuite fixé d'en tester la validité, ont presque toujours

> porté sur des phrases dont le verbe avait une forme finie. » (Huot, 1981, p. 21)

Deux raisons peuvent expliquer ce choix. La première est que ce sont les génératives qui ont étudié ce sujet les premiers. Or dans leur modèle, les constructions en complétive et en infinitive sont identiques formellement. Ayant la même structure profonde, les constructions en infinitive (interrogatives et relatives) ne sont pas censées remettre en cause la théorie des barrières :

> « Il était cependant admis, puisque les règles régissant le noeud COMP ne sont pas explicitement et régulièrement liées à la présence du Temps, que les constructions dont le modèle était censé rendre compte pouvaient aussi bien avoir un verbe non-fini. » (Huot, 1981, p. 21)

La deuxième raison est que les constructions en infinitive ne sont pas très fréquentes en anglais. Mais en français, comme le signale Huot, elles le sont beaucoup plus :

> « le problème des constructions infinitives semble beaucoup plus complexe en français qu'en anglais, ne serait-ce qu'à cause de la très grande variété de tours dans lesquels figure un infinitif. » (Huot, 1981, p. 22)

Quelle que soit la fréquence des deux constructions, une raison indépendante nous pousse à décrire les constructions en infinitive. Tous les chercheurs s'accordent sur le fait que le jugement d'acceptabilité est distinct pour les complétives et les infinitives introduites par un mot interrogatif. Ce fait a été noté par (Ross, 1967) :

> « (2.23a)
> he told me about a book which I can't figure out whether to buy or not
> he told me about a book which I can't figure out how to read
> he told me about a book which I can't figure out where to obtain
> he told me about a book which I can't figure out what to do about
>
> (2.23b)
> he told me about a book which I can't figure out why he read
> ? he told me about a book which I can't figure out whether I should read
> ?? he told me about a book which I can't figure out when I should read.
>
> For some reason that is obscure to me, I find sentences like those in (2.23a), where the embedded question consists of a wh-word followed by an infinitive, by and large more acceptable than corresponding sentences, like those in (2.23b), where the wh-word is followed by a clause with a finite verb. » (Ross, 1967, p. 27)

Hofmeister & Sag formulent une remarque similaire. Pour eux si l'acceptabilité est différente, c'est parce que d'autres paramètres, d'ordre sémantique et pragmatique, entrent en compte.

> « In general, these semantic and pragmatic accounts operate from the starting point that not all island effects are equally strong, and that syntactic constraints cannot or should not explain these differences. For instance, extraction out of a tenseless wh-clause is considered to be mildly degraded or marginally acceptable, compared to extraction out of a tensed wh-clause: This is a topic which John wondered whether to talk about >= This is a topic which John wondered whether she talked about. [...] » (Hofmeister & Sag, 2010, p. 20-21)

Toutes ces remarques, qui notent que l'acceptabilité de l'infinitive est plus forte que celle de la complétive, portent sur des constructions introduites par un interrogatif. Pourtant, ces formes semblent peu fréquentes. Nous n'en avons trouvé qu'un exemple durant les recherches des dépendances à distance en infinitive[136].

(375) on va les évacuer pour les remplacer par des professionnels qu'on ne sait pas où trouver (CRFP, PUB-POI-1)

Nul ne compare l'acceptabilité des complétives et des infinitives sans interrogatif, qui sont pourtant très nombreuses. Il semble aller de soi que ces constructions soient les mêmes. Il paraît néanmoins important de vérifier si le comportement des infinitives sans interrogatif est identique à celui des complétives.

Les questions qui se posent sont donc les suivantes : y a-t-il des contraintes dans l'usage des constructions de dépendance à distance en infinitive (sans interrogatif) ? Sont-elles moins fortes que celle des complétives ? Ou bien y a-t-il des contraintes présentes en complétive qui sont absentes en infinitive ?

Afin de répondre à ces questions, nous avons observé les données relevées dans les corpus. Comme nous l'avons indiqué (voir §2.4.2) nous avons conservé uniquement les occurrences d'antépositions des objets de l'infinitif dans cinq corpus oraux, c'est-à-dire 1234 occurrences.

[136] Les recherches des dépendances à distance en complétive spécifiaient le subordonnant « que » et nous n'avons donc pas pu trouver ce type d'exemples. Néanmoins, quelques requêtes en « savoir », « se demander », « dire », « chercher » et « comprendre », avec des interrogatives indirectes introduites par « où », « quand », « comment », « combien » et « pourquoi » n'ont donné aucun résultat, ni dans les corpus oraux, ni dans les corpus écrits.

4.1.2 Lexique et construction des verbes ponts

Dans la première section, nous examinerons le lexique des verbes ponts. Nous chercherons à savoir s'il est lié à la fréquence des verbes, s'il se réunit sous une classe sémantique particulière et s'il est restreint ou non. Dans la seconde section, nous examinerons l'ensemble de la construction pont.

4.1.2.1 Lexique des verbes ponts

Dans les 1234 occurrences de dépendance à distance en infinitive retenues, 76 lexèmes verbaux sont présents en position de pont. La liste suivante dresse l'ensemble des verbes attestés :

adorer
aimer
aller
apprécier
arrêter de
arriver à
avoir des difficultés à
avoir du plaisir à
avoir la chance de
avoir l'impression de
avoir à
avoir du mal à
avoir envie de
avoir le droit de
avoir l'habitude de
avoir tendance à
chercher à
commencer à
compter
conseiller de
continuer à
croire
décider de
demander de
désirer
devoir

dire de
entendre
envisager de
espérer
essayer de
être
être amené à
être capable de
être censé
être chargé de
être en droit de
être en train de
être forcé de
être incapable de
être obligé de
être (pas) prêt de
être susceptible de
être tenté de
faire
il est difficile de
il est inutile de
il est possible de
il est pratique de
il faut
il vaut mieux
il x'arrive de

il x'intéresse de
il y a lieu de
oser
oublier de
partir
penser
pouvoir
préférer
projeter de
répugner à
réussir à
revenir
rêver de
risquer de
s'attendre à
savoir
sembler
s'empresser de
souhaiter
suer à
tenir à
venir
venir de
vouloir

Nous avons conscience que tous les « verbes » cités ci-dessus ne répondent pas à l'ensemble des critères habituels de verbe ou locution verbale.

Le verbe « aller » est parfois considéré comme un auxiliaire, tant son fonctionnement est particulier. Certaines locutions, au contraire, ne sont pas totalement figées. Dans l'absolu, des adjectifs pourraient s'insérer et modifier les noms (*avoir la bonne impression* (376), *avoir la mauvaise habitude* (377), *avoir de grandes difficultés* (378), *avoir des difficultés importantes*).

(376) Je sors de chez Toyota Beaune et j'ai la bonne impression de m'être fait avoir... [137] *(Web)*

(377) Je ne promets jamais rien, parce que j'ai la mauvaise habitude de tenir mes promesses.[138] *(Web)*

(378) J'ai de grandes difficultés à prendre des décisions.[139] *(Web)*

Néanmoins, les adjectifs qui peuvent s'insérer dans ces structures sont peu nombreux. Dans « avoir des difficultés à » le nom peut varier en nombre : « avoir de la difficulté à » (379). La construction serait donc plutôt libre.

(379) J'ai de la difficulté à me concentrer.[140] *(Web)*

Pour vérifier le statut de la construction, on peut avoir recours au test de pronominalisation. Si la pronominalisation des noms ou des adjectifs rend l'énoncé agrammatical, cela prouve que les noms et les adjectifs n'ont pas les propriétés d'un complément régi par le verbe (380)(381).

*(380) J'ai l'impression qu'il va pleuvoir -> *je l'ai ; j'en ai l'impression*

*(381) J'ai la chance d'aller à Istanbul la semaine prochaine -> *je l'ai ; j'en ai la chance*

Mais plusieurs problèmes se posent. Certains verbes n'acceptent pas la pronominalisation du nom ou de l'adjectif, mais il n'est pas non plus possible de pronominaliser uniquement la partie infinitive (382)(383).

*(382) J'ai du mal à l'accepter -> *j'en ai, *j'en ai du mal*

*(383) Je ne suis pas près de faire cette concession -> *je ne le suis pas, *je n'en suis pas près*

Le second problème est qu'il n'est pas toujours simple de juger de la proportionnalité des énoncés lexicaux et pronominaux. La proportionnalité entre « avoir des difficultés/du plaisir »

[137] prius-touring-club.com/vbf/archive/index.php/t-3763.html, consulté le 20/07/12
[138] www.maphilo.net/citations.php?cit=4964, consulté le 20/07/12
[139] www.echelles-psychiatrie.com/echelle-beck.php, consulté le 20/07/12
[140] www.relationshumaines.ch/12.concentration.html, consulté le 20/07/12

et « en avoir » semble possible (384)(385), mais la pronominalisation semble donner trop d'importance au nom, alors que l'information se trouve dans la construction verbale à l'infinitif. En outre, il serait très difficile d'avoir un double marquage du type « j'en ai, des difficultés à lire cette affiche » ou bien « j'en ai, du plaisir à cuisiner pour mes amis », hormis dans des contextes particuliers (contrastif, pragmatique).

*(384) J'ai des difficultés à lire cette affiche -> ? j'en ai, *j'en ai des difficultés, *j'ai des difficultés à ça*

*(385) J'ai du plaisir à cuisiner pour mes amis -> ? j'en ai, *j'en ai du plaisir, ?j'ai du plaisir à ça*

Toutes les 'locutions' ne sont donc pas du même type, mais nous les conserverons toutes pour cette étude. Nous décrirons donc l'ensemble des occurrences des 76 lexèmes.

Nous verrons d'abord si la fréquence des verbes construits avec un infinitif a un impact sur la sélection des verbes ponts en structure infinitive. Ensuite, nous examinerons rapidement la sémantique lexicale des verbes ponts. Enfin, nous comparerons la richesse lexicale et la construction des verbes ponts en infinitive et en complétive.

La première question qui se pose est de savoir si les verbes pont en infinitive sont sélectionnés parmi les verbes les plus fréquents. Nous avons relevés les verbes suivis d'une infinitive dans les cinq corpus oraux étiquetés avec Cordial : CRFP, CFPP, C-Oral-Rom, CTFP et PFC. Les recherches ont autorisé au maximum un mot entre le verbe fini et l'infinitif, ce qui n'a pas permis de relever les locutions. La liste des verbes qui se construisent avec un infinitif comprend 182 verbes[141]. Ils sont utilisés avec un infinitif dans 14 465 occurrences dans les corpus.

Comme pour les complétives, les verbes ponts suivis d'une infinitive sont majoritairement sélectionnés parmi les verbes qui se construisent très fréquemment avec une infinitive (FIGURE.30).

[141] Le TLFi relève 340 verbes qui se construisent avec un infinitif.

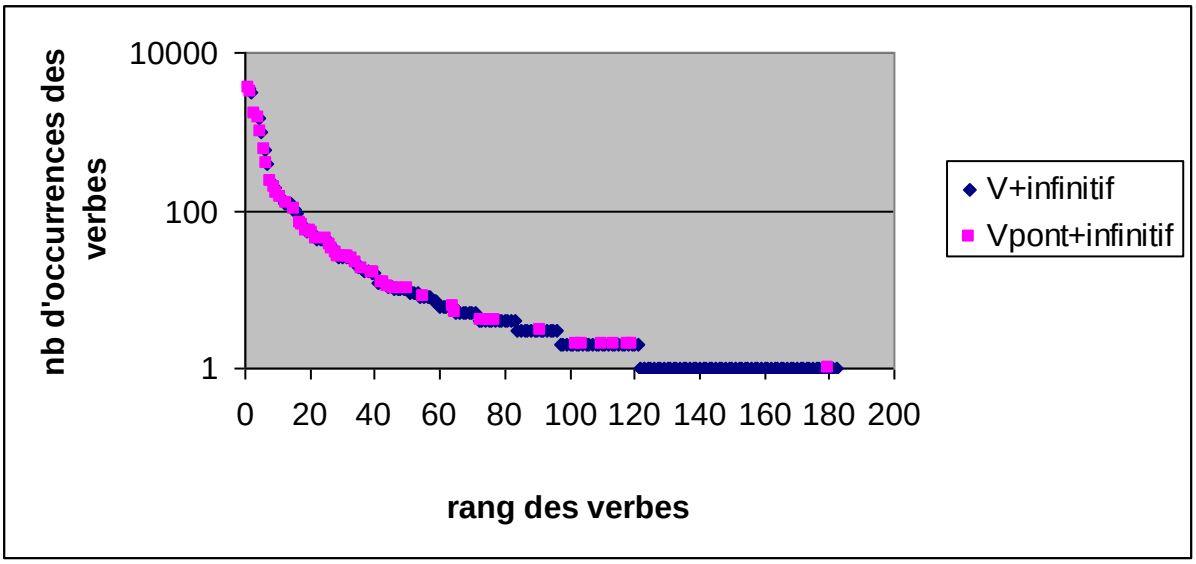

FIGURE.30 Représentation des verbes ponts parmi les verbes à infinitive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique du nombre d'occurrences)

Les 11 verbes les plus fréquents (plus de 140 occurrences) sont des verbes ponts. Sur les 35 verbes à infinitive les plus fréquents, 29 sont ponts ; sur les 50 verbes à infinitive les plus fréquents, 38 sont ponts.

Néanmoins, comme pour les complétives, la fréquence des verbes suivis d'une infinitive n'est pas équivalente à la fréquence des verbes en pont ; la fréquence du premier n'implique pas la fréquence du second. Voici pour illustration les 10 verbes les plus fréquents en Verbe+infinitif, et leur rang en verbe pont (TABLEAU 73).

	Rang Verbe + infinitif	Rang Verbe pont + infinitif
Aller	1	3
Pouvoir	2	1
Vouloir	3	2
Falloir	4	4
Devoir	5	5
Faire	6	11
Venir (0/de)	7	8
Aimer	8	6
Commencer	9	16
Essayer	10	9

TABLEAU 73. Comparaison des rangs de 10 verbes en Verbe+infinitif et en Verbe pont+infinitif

Toutefois, comme pour les complétives, les tendances générales montrent que les verbes du groupe le plus fréquent (plus de 20 occurrences) sont quasiment tous attestés en pont, tandis que les verbes du groupe le moins fréquent (une seule occurrence) ne le sont pas, excepté « suer à ».

D'un point de vue sémantique, les verbes ponts suivis d'infinitives sont clairement des verbes non descriptifs, ce qui confirme l'hypothèse formulée dans l'analyse des verbes ponts en complétive. Les plus fréquents sont « pouvoir » (328 occurrences), « vouloir » (227 occurrences), « falloir » (109 occurrences), « devoir » (63 occurrences) et le modal-auxiliaire « aller » (216 occurrences)[142]. Ils rassemblent à eux seuls 943 occurrences, soit 69% des occurrences.

*(386) c'est pas ce que tu **veux** faire toi hein (PFC, 12aps1)*

*(387) et à ton avis que **devraient** faire les adultes pour arrêter la pollution (CTFP, l'Alaska)*

*(388) je sais pas du tout quel âge elle **peut** avoir (PFC, 69asp1)*

*(389) et puis il y avait les CP à qui il **fallait** apprendre à lire (C-ORAL-ROM, ffammn15)*

Les autres verbes ont une valeur aspectuelle (« être en train de », « commencer à », « continuer à », « arrêter de », « venir de », « risquer de »), une valeur affective (« oser », « espérer », « préférer », « aimer », « adorer », « apprécier »), une valeur désidérative (« avoir envie de », « chercher à », « compter », « désirer », « envisager de », « être tenté de », « projeter de », « répugner à », « rêver de », « s'empresser de », « souhaiter », « tenir à »), une valeur évaluative (« avoir la chance de », « avoir du plaisir à », « il est difficile de », « il est inutile de », « il est pratique de »), une valeur évidentielle (« entendre ») ou une valeur épistémique (« croire », « penser », « sembler », « savoir »).

Le cas de « faire » est intéressant. Il est en effet non attesté en verbe pont avec une complétive et il occupe ici un rang honorable dans la fréquence des verbes admettant un infinitif. Mais il est clair que les deux types d'emplois sont très distincts : le verbe « faire » avec complétive indique une relation de cause à effet entre deux évènements, tandis que le verbe « faire » avec infinitif est un auxiliaire causatif qui restructure la valence d'un verbe. On ne peut donc comparer les deux verbes. Le cas de l'auxiliaire causatif exclut toute forme d'enchâssement.

[142] Il devrait y avoir une distinction entre les différents emplois de « pouvoir » et « devoir ». Nous n'avons pas opéré la distinction car les emplois étaient nombreux et ambigus (notamment avec le conditionnel)

Si l'on compare le lexique des verbes ponts en complétive et le lexique des verbes ponts en infinitive, on constate que ce dernier est plus riche, proportionnellement au nombre de verbes utilisés avec chacune des deux structures dans les corpus (TABLEAU 74).

	Verbe + infinitives/complétives	Verbe pont + infinitives/complétives	%
Infinitives	182	76	41,8
Complétives	195	34	17,4

TABLEAU 74. Comparaison de la richesse du lexique des verbes ponts dans les structures à complétives et à infinitives

Les deux structures ont un point commun qui concerne l'usage de leur lexique de verbe pont : les quatre verbes les plus fréquents en complétive et les quatre verbes les plus fréquents en infinitive atteignent approximativement la même proportion d'occurrences. Les verbes « vouloir », « falloir », « penser » et « dire » en pont avec une complétive rassemblent 72,2% des occurrences. Les verbes « pouvoir », « vouloir », « aller » et « falloir » en pont avec une infinitive rassemblent 71,2% des occurrences.

Cette répartition n'est pas celle attendue par la loi de Zipf, selon laquelle la fréquence d'un mot est proportionnelle à son rang. Selon cette loi, si le verbe le plus fréquent (donc de rang 1) couvre 500 occurrences, le deuxième verbe le plus fréquent (de rang 2) devrait posséder '500/2' occurrences, soit 250 occurrences, le troisième '500/3' occurrences, soit 167 occurrences, etc. Nous considérons que la répartition de Zipf donne une fréquence théorique des verbes, en partant de la première valeur réelle : la fréquence attestée du verbe de premier rang. La comparaison de la fréquence théorique et de la fréquence attestée des verbes fait ressortir une productivité démesurée des verbes les plus fréquents, et ce, pour les deux structures (FIGURE.31, FIGURE.32).

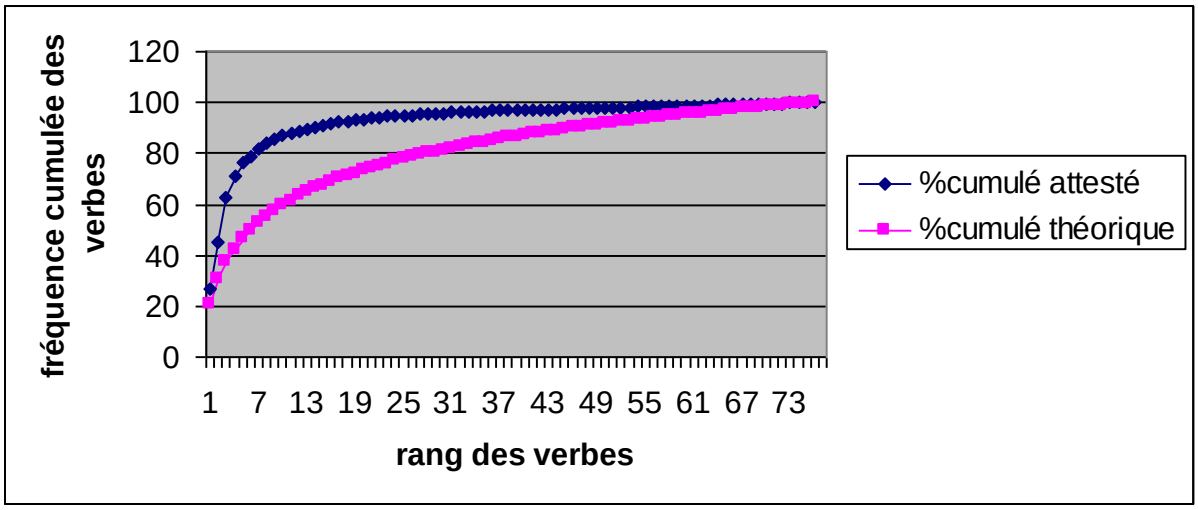

FIGURE.31 Comparaison des fréquences théoriques et attestées des verbes ponts dans les dépendances à distance en infinitive

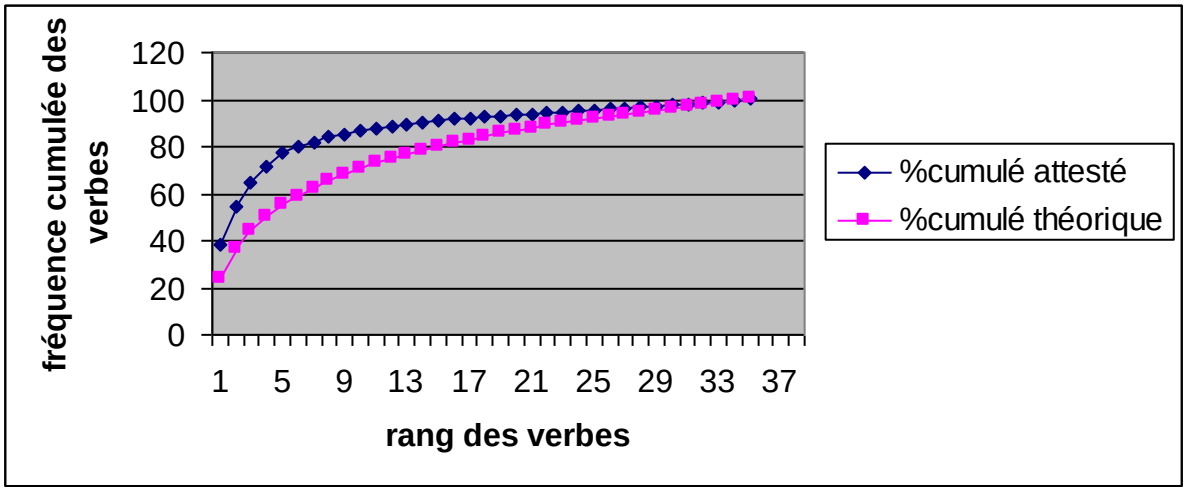

FIGURE.32 Comparaison des fréquences théoriques et attestées des verbes ponts dans les dépendances à distance en complétive

Dans les deux structures, ce sont les quatre premiers verbes qui s'éloignent particulièrement de leur valeur attendue. En outre, il est étonnant de voir que ces verbes fréquents en infinitive sont également très fréquent en complétive, du moins ceux qui acceptent dans leur construction à la fois les complétives et les infinitives. Il y a donc deux verbes, « vouloir » et « falloir », qui sont extrêmement employés en verbe pont[143]. Le lexique des verbes ponts en complétive et en infinitive présente donc de grandes similitudes à la fois dans sa répartition

[143] Il faut noter toutefois que « vouloir » est majoritairement employé avec une valeur rhétorique dans les occurrences en complétive, valeur qui est impossible en infinitive.

(fréquence plus élevée que celle de Zipf ; productivité intense des quatre verbes les plus fréquents) et dans son lexique (deux verbes identiques parmi les quatre plus fréquents de chaque structures).

Le lexique verbal ne semble donc pas constituer un point de différence entre les dépendances à distance en infinitive (sans pronom interrogatif) et ceux en complétive.

4.1.2.2 Construction du verbe pont

En structure infinitive, l'ensemble du pont est constitué au minimum du verbe pont et de son sujet (sujet qui n'est parfois pas sélectionné par le verbe pont).

Le sujet du verbe pont est majoritairement un sujet pronominal (93,4%) (390).

*(390) c' est ce qu'**ils** arrêtaient pas de nous dire (PFC, 44aar2)*

La personne référencée par le pronom sujet est d'abord la troisième personne (44,8%) puis la première (28%), puis la deuxième (14,7%). Les impersonnels représentent 12,5% des sujets. Enfin, une occurrence ne comporte pas de sujet exprimé. Dans 36 occurrences, le sujet est post-posé au verbe pont (391).

*(391) le chanteur par contre a créé euh de toutes pièces les paroles à partir des ambiances que pouvait lui inspirer **la musique**. (CRFP, PRI-PSO-4)*

Le verbe pont dans les constructions infinitives est conjugué au présent (74,6%) (392). Le deuxième temps le plus attesté est l'imparfait (17%). Généralement, le mode du verbe pont est l'indicatif (91,8%) ; quelques occurrences sont au conditionnel (94 occurrences – 7,6%) et au subjonctif (7 occurrences). Le verbe pont est quasiment toujours à la voix active, excepté dans 10 occurrences.

*(392) il y a une forme qui est mortelle qu'il **faut** soigner dans les quarante-huit heures (CRFP, PRI-NIC-1)*

Il est assez rare que le verbe pont soit nié (74 occurrences – 6%) (393). La deuxième partie de la négation est parfois modifiée par un adverbe (« pas du tout », « pas trop », « surtout pas »), comme cela est également possible entre un auxiliaire et un verbe.

(393) *il jouait notamment avec Claude Bolling des oeuvres qu'il **n'oserait pas** jouer ici (CRFP, PUB-AIX-1)*

Dans une occurrence, le verbe pont est affecté d'une restriction (394).

(394) *R.O.M. c' est les Read Only Memory c' est / les , des / mémoires qu' on **ne peut que** lire (CTFP, l'informaticien)*

Entre l'interrogatif-relatif et le verbe à l'infinitif (éventuellement précédé d'une préposition), des éléments peuvent être insérés. Dans 336 occurrences, soit 27% des occurrences, un ou plusieurs éléments s'ajoutent aux éléments obligatoires du pont : le sujet et le verbe. Ce chiffre concerne la distance topologique. Pour cibler la distance hiérarchique, il convient d'écarter plusieurs éléments :

- les marques de négation,
- les objets, ajouts et modifieurs du V2, qui peuvent se trouver entre les deux verbes,
- les disfluences (amorces, répétitions et pauses remplies « euh »), qui sont des éléments de performance, et non de langue,
- les épenthétiques, qui ne sont pas des compléments de V1, ni des non-régis (*les erreurs que **l'**on peut commettre*).

Il faudrait également écarter les éléments non régis, qui sont extérieurs à la construction. Nous souhaitons tout de même les examiner, comme indices de la difficulté de production de la construction. La frontière entre le pont et la CV2 n'étant pas marquée, il est difficile de savoir si un adverbe fait partie de l'un ou de l'autre. Nous tenterons de définir ici uniquement les éléments qui se rattachent au verbe pont. Ce tri aboutit à conserver 56 occurrences comportant 58 insertions.

Le verbe pont régit 32 des 58 insertions. Parmi elles, se trouvent 9 clitiques objets (395) et 23 adverbes (396). Les adverbes sont des modifieurs dont la portée est locale. Ils ne sont globalement pas des adverbes référentiels (TABLEAU 75).

(395) *moi ce que je **te** conseille de faire demain, c'est d'appeler le secrétariat (C-ORAL-ROM, ftelpv12)*

(396) *c'est effectivement le type de revendication qu'il faut **absolument** avoir (C-ORAL-ROM, fnatpd02)*

Classe d'adverbe	Lexique
comparatifs/superlatifs	le moins, plus, autant, plutôt
quantifieurs	tellement, beaucoup
manière	modestement
aspectuel	encore, toujours
intensifieurs	bien, réellement, absolument

TABLEAU 75. Classes d'adverbes régis dans les dépendances à distance en infinitive

Les 26 éléments restants ne sont pas régis par le verbe pont ni par le V2. On peut noter la présence de trois occurrences qui présentent une coordination de deux verbes ponts (comme on peut trouver des coordinations d'auxiliaires). Bien que cela ne soit pas à proprement parler un phénomène hors de toute rection, le poids du pont est tout de même augmenté (397)(398).

(397) selon ce que la personne euh souhaite ou peut mettre (CRFP, PRO-BEL-1)

*(398) c'est ce que **vous essayez ou vous avez essayé** de faire avec votre ami (CRFP, PRI-AMI-2)*

Dans une attestation, deux constructions verbales sont juxtaposées (399) et ont en facteur commun leur sujet : pour le premier, la dépendance du pronom relatif est courte mais pour le second, elle est longue. Le pont est alors fragmenté et l'auditeur doit non seulement garder en mémoire le relatif (qui est plus éloigné du fait de la juxtaposition) mais également opérer des liens entre les deux constructions verbales. La présence de « hein » semble aider à la compréhension : il marque la fin de la première construction ; le deuxième verbe, conjugué à la même personne et au même temps, est alors automatiquement interprété comme étant en liste par rapport au premier.

*(399) tu as quand même dans toutes les grandes villes des petits quartiers où les gens **se retrouvent hein ont tendance** à se regrouper (CFPP, Pujade_13e)*

Dans les éléments insérés non régis proprement dits, on trouve 2 incises, 3 pronoms toniques et 17 adverbes. Ils sont hors construction micro et donc n'alourdissent pas la syntaxe du pont. Les incises sont toujours positionnées entre l'interrogatif-relatif et le sujet du verbe pont ; elles ne créent pas une séparation entre les deux verbes.

*(400) ce que je qualifie enfin ici de monde profane c'est bien sûr la Roche- Mauprat hein pour des raisons sur lesquelles **je crois** il est inutile de de revenir (C-ORAL-ROM, fnatco03)*

Il en est de même pour les pronoms toniques, excepté dans une occurrence.

*(401) ils nous ont demandé où **nous** on voulait aller sachant qu'il y avait tant de places pour tel endroit et tant de places pour tel endroit (CRFP, PRI-NCY-1)*

Les adverbes, en revanche, sont très majoritairement placés entre les deux verbes. Néanmoins, ces adverbes pourraient tout à fait être présents entre un auxiliaire et un participe passé. Ils ne sont donc pas un obstacle au lien qui unit les deux verbes, d'autant qu'ils relèvent de la composante macrosyntaxique.

*(402) c'est une journée nationale et donc beaucoup de Français auront entendu parler de cette campagne que nous voulons **bien sûr** répercuter sur Montpellier et sa région (CRFP, PUB-MON-1)*

Les adverbes non régis sont beaucoup plus fréquents que les autres types d'éléments non régis. Ils appartiennent à diverses classes (TABLEAU 76).

Classe d'adverbe	Lexique
Argumentatif	ensuite, finalement, d'ailleurs, en fait, quand même, à la base, justement, encore, bien
Epistémique (valeur de vérité)	éventuellement, bien sûr, peut-être
Paradigmatisant	juste

TABLEAU 76. Adverbes non-régis attestés en pont dans les dépendances à distance en infinitive

Parallèlement à ces 335 énoncés dans lesquels on note une insertion qui augmente la distance topologique entre l'interrogatif-relatif et le verbe à l'infinitif, il existe 49 énoncés dans lesquels un élément est postposé alors qu'il aurait pu s'insérer entre les deux dépendants. On peut considérer qu'il y a là un indice de cohésion de la construction.

Dans 36 occurrences, le sujet du premier verbe est produit après le second verbe (403).

*(403) discours que m'avait demandé de prononcer **mon proviseur de Montluçon** à la veille de mon départ (CRFP, PRI-CLE-2)*

Dans d'autres occurrences, c'est l'adverbe modifiant le V2 qui est postposé, alors qu'il aurait tout à fait pu être placé entre les deux verbes, ou ailleurs dans le pont (404) (405) (406) (407).

*(404) rien ne justifie la violence et c'est cela d'abord qu'il faut condamner **clairement** (CRFP, PUB-BAY-1)*

*(405) quels sont les les grands noms que vous aimeriez faire connaître **un peu** (CRFP, PRI-GAP-2)*

*(406) et puis aussi le jeune maquettiste de la maison qui a fait un travail remarquable et que je tiens à à remercier **tout particulièrement** (CRFP, PUB-LIL-1)*

*(407) je pense que c'est un vin qu'il faut boire euh **tout de suite** (CRFP, PUB-DIJ-1)*

Enfin, il arrive même que la deuxième partie de la négation soit omise, ce qui rend la distance topologique entre les deux verbes inexistante (408).

(408) domicile que je n'ose qualifier de conjugal (C-ORAL-ROM, fnatla02)

Le pont dans les dépendances à distance en infinitive est donc peu développé par des éléments régis ou par des objets référentiels. En majorité, il n'est composé que du sujet (pronominal, à la troisième personne) et du verbe (au présent de l'indicatif, à la voix active et à l'affirmatif). Lorsqu'un élément s'intègre au pont, il s'agit soit d'un adverbe faisant « corps » avec le verbe (mitigatif, *aimer bien*), soit d'un non régi qui constitue une insertion au niveau de l'énoncé, comme on peut en trouver entre un auxiliaire et un verbe (*il a heureusement trouvé la solution*).

L'ensemble des traits du pont présente une tendance plus massive pour les structures en complétive que pour les structures en infinitive, comme le montre le tableau suivant (TABLEAU 77).

Traits	Tendance majoritaire	Complétives	Infinitives
Sujet	Pronom	95,6%	93,4%
	Personne	P2 : 43,4% il impers : 23,3%	P3 : 44,8% P1 : 28,0%
Verbe	Indicatif	94,7%	91,8%
	Présent	85,6%	74,6%
	Actif	98,7%	99,2%
Constitution du pont	Sujet + verbe uniquement (+non-régis)	95,6%	91,6%

TABLEAU 77. Comparaison de la construction du pont dans les structures à complétives et à infinitives

On remarque une exception pour la voix du verbe et la constitution du pont, où les chiffres sont équivalents. Les personnes des pronoms sujets, en revanche sont tout à fait différentes, ce qui s'explique par le fait que la valeur interactive se perd avec l'infinitif, comme le montre

l'exemple suivant, totalement accepté en complétive (409), mais difficilement acceptable en infinitive (410).

(409) A qui crois-tu qu'il l'a donné ?

(410) ??à qui crois-tu le donner ?

Les dépendances à distance en infinitive paraissent donc être légèrement plus variées que les dépendances à distance en complétive. Leurs verbes ponts sont d'un lexique et d'une morphologie plus riches. En revanche, ils admettent moins d'insertions. Mais nous avons jusqu'ici considéré toutes les occurrences d'infinitives, sans prendre en compte la relation syntaxique entre le verbe pont et le V2 qui est plus variée que dans le cas des complétives. Il reste donc à examiner ce point. Deux questions se posent : quels sont les verbes les plus employés en infinitive ? Quel est leur degré de verbalité ?

4.1.3 Lien V1-V2 - Classement syntaxique des verbes ponts

La relation qui unit le verbe pont au verbe recteur de l'élément antéposé n'est pas unique et parfois peu aisée à définir. Dans les structures en infinitive, divers types de verbes ponts sont présents. Ils possèdent des statuts particuliers. En nous appuyant sur la classification syntaxique de (Chu, 2008), on peut déterminer quatre types de verbes en verbes ponts : des modaux, des semi-modaux, des factitifs et des constructeurs.

Les verbes modaux sont caractérisés par le fait d'être utilisés avec un verbe à l'infinitif qui ne peut être ni pronominalisé (*en, *y, *le, *ça), ni interrogé. Les modaux sont employés avec les quatre types d'impersonnels présentés par (Chu, 2008, p. 48) (411) (412) (413) (414).

(411) Il pleut / il fait chaud => il commence à pleuvoir

(412) Il souffle un vent terrible => il commence à souffler un vent terrible

(413) C'est difficile de faire ça => il commence à être difficile de faire ça

(414) Il y a beaucoup de monde => il commence à y avoir beaucoup de monde

Dans notre corpus d'exemples, 13 verbes appartiennent à cette classe. Ils sont présents dans 658 occurrences. Il s'agit de la classe la plus productive de dépendance à distance (53,3% des occurrences). On trouve dans cette catégorie les verbes « aller », « pouvoir », « devoir », « venir de », « être en train de », « commencer », etc.[144] Ce sont les verbes qui correspondent aux verbes à montée des génératives.

Les verbes semi-modaux s'emploient également avec des infinitifs qui ne peuvent pas être facilement pronominalisés par des clitiques. En revanche, ils acceptent parfois les interrogations. Les semi-modaux ne s'emploient avec aucun des quatre types de verbes impersonnels.

Parmi les verbes ponts, nous avons relevé 44 verbes semi-modaux, présents dans 478 occurrences. Il s'agit par exemple de « vouloir », « avoir à », « essayer de », « arriver à », « savoir », « entendre », « oser », etc. On peut noter dans cette classe la présence de « être » produit sous la forme « j'ai été » + infinitif, dans un emploi semblable à « je suis allé » (415).

(415) celle de des copains que j'ai été aider (PFC, 54bfl1)

Dans les deux premières classes, la pronominalisation étant impossible ou restreinte, il ne semble pas qu'on ait affaire à une structure Verbe_constructeur + Verbe_construit, en tout cas qui constituerait une structure d'enchâssement. Dans la classe suivante, le lien entre les deux verbes est également particulier : Le V1 ajoute un élément de valence mais à l'ensemble des deux verbes.

Les factitifs « faire », « se faire », « laisser », « se voir », etc. ont un fonctionnement spécifique. Ils ne s'emploient jamais avec des verbes impersonnels. Le verbe à l'infinitif ne peut pas être pronominalisé, à l'instar des modaux et des semi-modaux. Les factitifs ne sont donc pas des verbes constructeurs. Néanmoins, et c'est là leur particularité, ils ajoutent un élément de valence au groupe verbal. Un verbe comme « tomber » possède un seul élément de valence : le sujet. Associé à un factitif, la construction possède deux éléments valenciels, soit dans l'exemple (416), « il » et « le vase ». Lorsque tous les compléments sont pronominalisés, les clitiques sont réunis avant le factitif et rien ne s'insère entre ce dernier et l'infinitif (417).

[144] Tous les verbes cités ici sont considérés comme modaux typiques par (Chu, 2008, p. 41)

(416) Le vase tombe => il a fait tomber le vase (Chu, 2008, p. 29)

(417) Il l'a fait tomber

La valence sujet du verbe peut également disparaître lorsque le factitif intervient (418).

(418) Ma soeur a écrit une lettre => j'ai fait écrire une lettre (Chu, 2008, p. 29)

Le verbe « faire » est le seul factitif attesté en verbe pont. Il réunit 10 occurrences.

La dernière classe déterminée par Chu est celle des constructeurs. Ces verbes ne s'emploient jamais avec des verbes impersonnels. Ils sont constructeurs : ils peuvent pronominaliser le verbe à l'infinitif, qui fait donc partie de leur rection (419).

(419) Il aime sortir avec son père => il aime ça / ça, il aime (Chu, 2008, p. 38)

On relève 19 verbes constructeurs parmi les verbes ponts des structures en infinitive. Ils sont présents dans 88 occurrences.

Le tableau suivant indique comment se répartissent les verbes ponts dans les quatre classes[145] (TABLEAU 78).

	modal	**semi-modal**	**factitif**	**constructeur**
Nb occ	658	478	10	88
Nb verbes	13	44	1	19
Pronominalisation	-	-	-	+
Interrogation	-	+/-	-	+
Infinitifs impersonnels	+	-	-	-

TABLEAU 78. Répartition des verbes ponts attestés en structure infinitive dans les quatre classes de verbes définies par (Chu, 2008)

[145] Nous avons légèrement adapté la classification de (Chu, 2008). Celui tient compte d'un paramètre supplémentaire que nous avons écarté. Selon lui, les modaux et les semi-modaux ne devraient pas s'employer avec des complétives. Nous considérons au contraire que la réalisation de la CV2 est indépendante du type de V1.

Et voici les verbes associés à chaque classe :

	modal	**semi-modal**	**factitif**	**constructeur**
Verbes	Pouvoir=328 Aller=216 Devoir=63 Venir de=20 Etre en train de=18 Commencer à=5 Continuer à=2 Avoir tendance à, être censé, être pas prêt de, être susceptible de, risquer de, sembler = 1	Vouloir=227 Il faut=109 Avoir à =31 Essayer de=13 Arriver à =11 Savoir, entendre=9 Oser=7 Compter=6 Souhaiter=5 Etre obligé de, réussir à, venir =4 Avoir du mal à, chercher à, il est difficile de =3 Il x'intéresse de, croire, il est inutile de, il vaut mieux=2 arrêter de, avoir la chance de, avoir le droit de, avoir l'impression de, avoir des difficultés à, avoir du plaisir à, désirer, être en droit de, être incapable de, être tenté de, être, être capable de, être forcé de, il est possible de, il est pratique de, il x'arrive de, il y a lieu de, partir, penser, revenir, s'empresser de, suer à, tenir à = 1	Faire	Aimer=33 Avoir envie=29 Décider de=5 Avoir l'habitude=3 Conseiller de, oublier de, préférer=2 Adorer, apprécier, demander de, dire de, envisager de, espérer, être amené à, être chargé de, projeter de, répugner à, rêver de, s'attendre à =1

Tous les verbes n'ont donc pas les mêmes propriétés. Comme l'ont indiqué (Blanche-Benveniste & Martin, 2010), il existe différents degrés de verbalité. Les verbes se distinguent des autres catégories grammaticales par un ensemble de traits : la conjugaison (temps, aspect, mode, personne), la valence (sélection des compléments par des contraintes lexicales et sémantiques), la rection (dépendance syntaxique des compléments, pour lesquels il existe une équivalence lexique/proforme), l'accord avec le sujet, la portée éventuelle d'une modalité (négation, la restriction ou l'interrogation) et le fait de rassembler les clitiques autour d'eux. Mais chaque sous-classe verbale ne répond pas à l'ensemble de ces traits, comme l'illustre le (TABLEAU 79).

	conjugaison	**valence**	**accord**	**modalités**	**clitiques**	**Exemple**
Verbes pleins	+	+	+	+	+	*Je n'y vais pas*
Verbes auxiliaires	+	[participe passé]	+	+	+	*Je le lui ai conseillé*
Verbes modaux	+	[infinitif]	+	+	[infinitif]	*Il a beau le lui dire*
Verbes supports	+	[nom]	+	+	+	*Il en a besoin, de ça*
« voici, voilà »	-	+	-	[négation seulement]	+	*Le voilà qui vient*

TABLEAU 79. Classement des verbes d'après différents critères morphosyntaxiques et syntaxiques – d'après (Blanche-Benveniste & Martin, 2010, p. 122)

Pour une grande majorité de verbes ponts, la pronominalisation de l'infinitif est impossible (92,9%). La construction ne peut pas être décrite en terme de V1_recteur + V2_régi. Il n'y a donc pour ces cas là qu'une illusion de contrôle à distance. Il paraît plus approprié de parler de dépendance locale à l'ensemble des deux verbes, que ce soit pour les modaux, qui n'ont pas de valence (420), pour les factitifs qui forment un bloc avec le V2 et unissent leur valence à celle du V2 (421) et pour les semi-modaux, dont la construction est plus proche de celle des modaux que de celle des constructeurs (422).

(420) c'est une époque où le raisin risque de pourrir (CRFP, PRI-MON-1)

(421) c'est des logiciels qu'elle a fait développer pour elle (CRFP, PRO-PSE-1)

(422) par où vaut mieux passer (CFPP, Rosier_12e)

C'est l'interprétation que propose (Blanche-Benveniste & Martin, 2010) en posant une unité 'chaîne verbale' :

> « Plus étoffés, les syntagmes verbaux intègrent des séries de formes verbales rassemblées ici sous le terme de « verbes modaux » (souvent désignées comme périphrases verbales ou semi-auxiliaires) : verbes exprimant le passé et ou le futur, *je vais* + V, *je viens de* + V ; verbes *devoir* et *pouvoir* dans leur sens de probabilité ; verbes indiquant l'apparence, *sembler, paraître, avoir l'air de* ; verbes indiquant des phases aspectuelles, initiale, durative ou terminale, *se mettre à, commencer à, risquer de, continuer à, être en train de, finir* de ; verbe causatif, *faire* + V. Leur caractéristique commune est de se faire suivre d'un verbe à l'infinitif, qui n'est pas pour autant leur complément, en constituant ainsi des « chaînes » qui peuvent compter jusqu'à quatre ou cinq formes verbales successives. Le premier verbe de la chaîne est le seul qui soit conjugué et la dernier, à l'infinitif, est le verbe recteur » (Blanche-Benveniste & Martin, 2010, p. 112)

Cette analyse implique que la chaîne formée par l'ensemble des deux (ou trois) verbes est un constructeur complexe comme l'on est amené à le poser dans le cas des expressions « il a l'intention de » : *il en a l'intention*.

La question qui se pose alors est de savoir s'il est légitime d'étendre cette analyse au cas où le verbe pont est un constructeur qui permet la pronominalisation du verbe à l'infinitif. N'y a-t-il pas dans ce cas enchâssement donc dépendance à distance ? Si c'est le cas, les deux verbes doivent avoir leur propre construction comportant objets et ajouts.

Dans le cas des constructeurs, la relation entre le verbe pont et le verbe à l'infinitif semble très forte. La construction du verbe pont est très contrainte. Sur les 88 attestations relevées, on

constate que 74 occurrences ne présentent aucun élément qui s'ajoute aux éléments requis dans le pont (84%).[146]

(423) c'est exactement ce que je rêvais d'entendre (CFPP, deBordes_7e)

Dans 14 occurrences, le verbe pont régit d'autres éléments que le sujet et l'infinitive. Dans 4 occurrences, il s'agit d'un objet clitique précédant le verbe pont (« conseiller », « dire » et « demander ») qui, comme dans le cas des complétives, est directement lié à la valeur modale du verbe. Dans les 10 occurrences restantes, il s'agit d'adverbes qui se trouvent généralement entre les deux verbes : *j'aime bien*, *j'aime beaucoup et* qui n'ont pas de valeur référentielle. On peut également noter que les sujets des verbes ponts constructeurs sont presque toujours pronominaux (97%). Ce pourcentage est plus élevé que la moyenne des sujets des verbes ponts en infinitive (93,4%).

La distance topologique entre l'interrogatif-relatif et l'infinitif est également peu développée. On compte, en plus des éléments régis du V1, trois éléments régis par le V2 (deux clitiques et un ajout), onze adverbes régis par le V2, six marques de négation et six marques de disfluences.

Autrement dit, le verbe constructeur n'a pas du tout de rection développée. On constate qu'il y a généralement aussi un rapprochement topologique des deux verbes. Ces deux observations nous amènent à proposer deux hypothèses : soit la construction est un enchâssement mais qui comporte de nombreuses contraintes de réalisation que l'on pourrait mettre au compte de la performance (facilité de « processing »), soit la construction « verbe_constructeur+infinitive » est en fait une extension de chaîne verbale, comme le propose (Blanche-Benveniste & Martin, 2010) :

> « A ces chaînes peuvent venir s'intégrer des verbes à contrôle, très différents des précédents parce qu'ils ont leur propre valence *décider de+ V, l'amener à + V, lui interdire de + V* » (Blanche-Benveniste & Martin, 2010, p. 112)

L'idée serait que, du point de vue de l'élément antéposé, toutes ces chaînes se comportent comme des constructeurs complexes de sorte que toutes les dépendances de l'énoncé restent locales, la différence résidant seulement dans les relations entre verbes à l'intérieur de la chaîne.

[146] Ne sont pas considérés comme insertions : la préposition entre les deux verbes, le pronom clitique objet du V2, la négation (même lorsqu'un adverbe vient la renforcer : *pas du tout, pas trop*).

Des facteurs sémantiques tendent à confirmer cette analyse de chaîne verbale. En effet, les énoncés ne présentent pas deux évènements distincts mais un évènement ou situation et une évaluation modale de cet évènement (424). Même si la pronominalisation est possible, les verbes ont un emploi très différent de celui d'un constructeur descriptif qui porte le poids informatif de l'énoncé (425).

(424) J'aime ça moi, flâner sur les bords de la rivière -> c'est sur les bords de la rivière que j'aime flâner

(425) Je le lui demande souvent avec insistance, de te téléphoner -> ?c'est à toi que je lui demande souvent avec insistance de téléphoner

4.1.4 Conclusion

Dans cette section, nous avons vu que le lexique des verbes ponts se comporte de façon relativement similaire dans les structures en infinitive et en complétive. Les verbes les plus fréquents sont extrêmement productifs. Deux verbes, « vouloir » et « falloir », qui acceptent les deux structures, se retrouvent parmi les quatre verbes les plus fréquents de chaque structure (« vouloir », « falloir », « penser », « dire » pour les complétives, « pouvoir », « vouloir », « aller », « falloir » pour les infinitives). Quant à la construction du pont, elle est également similaire : elle varie peu dans les deux structures. La majorité des occurrences ne présente qu'un sujet et un verbe dans le pont.

Les dépendances à distance en infinitive font intervenir différentes classes de verbes ponts (modaux, semi-modaux, factitifs, constructeurs). La majorité d'entre elles comportent des verbes sans valence ou à valence réduite ; ils ne permettent pas la pronominalisation du V2. Les modaux, semi-modaux et factitifs peuvent donc être analysés comme formant des 'chaînes verbales' avec le V2 et donc n'entraînant aucune dépendance à distance. Les verbes ponts constructeurs acceptent la pronominalisation de l'infinitif, ce qui est un indice de rection. Néanmoins, le pont est identique à ceux des classes citées précédemment : il ne comporte qu'un sujet et un verbe. Aucun indice déterminant de valence n'est relevé. La chaîne verbale pourrait donc s'étendre aux constructeurs.

La chaîne verbale formant une unité syntaxique, le rattachement entre l'élément antéposé et le verbe infinitif serait structurellement plus aisé. La combinaison de ces facteurs structuraux et la facilité à reconnaître en performance ces structures (faible distance topologique constatée

entre l'élément antéposé et le V2) pourrait expliquer pourquoi les séquences complexes avec infinitives en wh- sont mieux acceptées que celles mettant en jeu les complétives en wh-.

Mais ce résultat pose alors un problème. Quelle est la relation précise entre la chaîne verbale avec infinitif et la construction pont étudié dans le chapitre précédent ? S'agit-il de la même unité avec simplement un contenu différent ? S'agit-il au contraire de deux sous-types distincts de 'prédicats complexes' ? Outre les différences précédemment mentionnées un nouveau phénomène pourrait amener à distinguer les deux unités.

Théoriquement, il est possible pour un locuteur de choisir entre la construction en complétive et en infinitive, la seule contrainte étant que l'agent du V1 et du V2 soit identique. Il est donc intéressant de chercher à comprendre pourquoi le choix de la complétive, théoriquement plus difficile d'accès, serait effectué.

Dans le corpus d'exemples de dépendance à distance, seules 17 occurrences réalisées en complétive ont le même agent pour le verbe pont et le verbe recteur (17/229). Mais d'autres éléments favorisent la structure en complétive. Il n'y a pas de possibilité pour la structure en infinitive de différencier les temps des deux verbes, alors que cela est possible pour les complétives. Nous avons relevé 5 occurrences correspondant à ce cas de figure. (426).

(426) c'est là où on construit vraiment des enfin des amitiés qu'on sait qu'on gardera après (CFPP, Teixeira_11e)

Dans 3 occurrences, le second verbe est nié, ce qui serait plus difficile avec la forme infinitive (427). Pour comparaison, il n'y a aucune occurrence en infinitive où le V2 est explicitement nié. Lorsque la négation se fait uniquement par « pas », entre les deux verbes, il est plus difficile de déterminer quel verbe est nié. La première interprétation consiste plutôt à rattacher la négation au premier verbe (428).

(427) c'est pour ça que je dis que j'aime pas les gens à Paris (CFPP, Belamy_Hermano_KB)

(428) ça je pense que c'est quelque chose que je peux pas oublier (CRFP, PRI-AMI-2)

Dans 2 occurrences, on relève un phénomène de récursivité dans le pont (429). Mettre une série d'infinitive rendrait l'énoncé lourd et difficile à comprendre (430). Il est donc naturel que les complétives aient été choisies.

(429) "Monsieur le Président, Mesdames et messieurs les députés, Monsieur le député, La réunion informelle des chefs d'Etat et de gouvernement qui s'est tenue vendredi dernier à Petersberg, et

dont je pense qu'il est normal que je rende compte à la représentation nationale, a été positive à plus d'un titre. (CERF, POLI)

(430) ??réunion dont je pense être normal de rendre compte.

De même, lorsque le second verbe est modalisé (5 occurrences), la complétive est préférée. Elle permet d'éviter la succession d'un verbe fini et de deux infinitifs. (Chu, 2008, p. 117) a remarqué que l'ordre de combinaison des verbes modaux était important. Une contrainte similaire doit exister pour l'ensemble des verbes. Dans l'exemple (431), l'ordre des trois verbes ne semble pas autorisé à l'infinitif (432).

(431) Là où je sais que je peux avoir une réponse, je veux bien savoir cette réponse.(CERF, DIVS)

(432) ??Là où je sais pouvoir avoir une réponse

Dans le corpus d'exemples en infinitive, les chaînes de trois verbes sont produites en majorité lorsque le deuxième verbe est un factitif ou le modal « aller » (26 occurrences). Dans les autres cas (6 occurrences), le premier verbe est un verbe sans valence (« pouvoir », « commencer ») ou un verbe très fréquent avec des infinitifs (« vouloir », « falloir »).

L'habitude d'utiliser un verbe avec un infinitif ou non semble aussi avoir de l'importance dans le choix complétive/infinitive. Trois occurrences en complétive présentent en effet des verbes peu fréquents avec une infinitive : « sentir » (433), « ressentir » (434) et « remarquer » (435).

(433) il avait le choix entre revenir dans un processus de raison et dans un processus de paix ou s'enfermer dans l'espace suicidaire qui était la sienne mais où il sentait bien que progressivement, il était de plus en plus isolé (CERF, POLI)

(434) on n'aurait pas pu euh participer à cette évolution dont on ressentait qu'on avait vraiment besoin (PFC, 81arc1)

(435) il y a des périodes où vous remarquez que vous vendez plus ce produit (CorpAix, AZ)

Enfin, il arrive que le verbe pont dans un énoncé à complétive ait une interprétation différente de celle qu'il aurait en infinitive. Dans l'exemple (436), le verbe « savoir » a le sens de 'connaissance' ; avec une infinitive, il prendrait le sens de 'capacité' (437).

(436) c'est génial quoi mais faire un truc que tu aimes quoi et que tu sais que tu maîtrises (PFC12afl1)

(437) faire un truc que tu sais maîtriser

Il en est de même dans les exemples cités précédemment (426) et (429). Les verbes « savoir » et « penser » ont un sens différent avec une complétive et avec une infinitive. « Penser », par

exemple, a la valeur de « j'ai l'intention de » lorsqu'il est suivi d'un infinitif (438) ; il prend une valeur épistémique lorsqu'il est suivi d'une complétive : « à mon avis » (439).

(438) Je pense aller à Nancy

(439) Je pense que j'irai à Nancy

L'équivalence sémantique des deux constructions n'est donc pas toujours attestée.

Formuler une complétive ou une infinitive ne constitue donc pas un choix totalement libre, même lorsque l'agent est identique pour les deux verbes. L'hypothèse qui consiste à poser que la différence d'acceptabilité entre infinitive et complétive est due au type de structure et de relation entre le V1 et le V2 est donc plutôt renforcée par la complémentarité des deux constructions.

La prise en compte de l'analyse des constructions en infinitif nous amène à deux conclusions. L'hypothèse que la relation entre deux verbes ne se réduit pas à un simple enchâssement est appuyée par la variété des relations « V1-V2_inf ». Il n'est donc pas *ad hoc* de poser que la relation « verbe pont conjugué - verbe constructeur » décrite dans le chapitre III constitue un objet syntaxique spécifique, en tout cas distinct de l'enchâssement de deux constructions verbales. La comparaison des deux constructions montre aussi que la chaîne de deux verbes finis est sans doute distincte de celle que constitue une chaîne verbale à l'infinitif (sens différent des verbes ponts, information supplémentaire dans le cas des complétives). Dans la section suivante, nous allons donc approfondir la question de la modélisation des constructions de dépendance à distance à verbe fini.

4.2 Une construction spécifique du français : le complexe verbal fini

Deux types d'analyses sont possibles pour les dépendances à distance à verbe fini : une extension de la notion de chaîne verbale avec de fait une dépendance locale ou une analyse par enchâssement avec dépendance à distance. Dans les structures avec enchâssements, deux

modèles sont possibles. Soit on considère (générativistes) qu'il y a en structure une relation de dépendance non locale (à distance) entre l'élément extrait et une catégorie vide en position canonique qui est donc localement régie par le V2 enchâssé dans le V1. Soit on considère que la relation de dépendance non locale s'exprime entre un élément antéposé et le verbe recteur enchâssé dont il dépend par le biais d'une réalisation non canonique du complément (*slash*). Dans les deux cas, l'écart entre les possibilités théoriques posées par le modèle et les faits attestés est mis au compte de contraintes de diverses natures qui excluent certaines configurations (contraintes syntaxiques : impossibilité d'antéposer une partie d'un sujet ou une partie d'un ajout ; contraintes de performance, qui réduisent la liberté de construction).

Comme le notent (Culicover & Jackendoff, 2005, p. 232-234) (voir §1.3.4), cette approche revient à partir des structures générales offertes théoriquement par le modèle et à les limiter pour les rendre compatibles avec les faits. Ils proposent de procéder en sens inverse et donc de partir des structures attestées et de ne modéliser qu'elles. C'est la démarche qui est la plus naturellement compatible avec notre description. Nos résultats nous paraissent de ce point de vue particulièrement faciles à interpréter à partir des propositions de modèles comme TAG ou de l'analyse par arbres à bulle (voir §1.1.1 et §1.1.2). Ces deux modèles permettent en effet de poser au départ une structure de dépendance locale entre un gouverneur et un dépendant et de donner une liste fermée des possibilités d'enrichir le gouverneur, que ce soit par des adjonctions (TAG) ou par des procédés plus variés (bulles). Evidemment, on pourrait penser qu'il s'agit de déplacer la complexité ou les contraintes des enchâssements à la constitution du gouverneur, mais (Blanche-Benveniste, Non publié) avait bien vu un premier avantage de cette solution dans la citation que nous rappelons ici :

> « Quelle que soit l'analyse interne qu'on choisit pour ces séquences, il paraît très utile d'avoir un terme qui permette de désigner l'ensemble, au moins provisoirement, par une saisie externe » (Blanche-Benveniste, Non publié, p. 4)

On peut en effet réutiliser les mêmes règles qu'avec un verbe simple pour décrire la globalité de la structure. Il faut ensuite entrer dans la 'chaîne' ou le 'complexe verbal', mais nos résultats montrent que son analyse interne est finalement simple. Notre solution penche donc pour une extension de l'unité « chaîne verbale », pour les raisons que nous allons développer ci-dessous.

4.2.1 Une construction complexe comme unité syntaxique

Dans les emplois avec effet de dépendance à distance (avec V2 fini et infinitif), nous avons observé une tendance très forte du verbe pont à constituer une séquence minimale : « sujet + verbe + que ». De plus, le sujet est quasiment toujours un pronom clitique ; le verbe est un verbe fréquent, sémantiquement simple, et à interprétation modale ; il est conjugué à la forme la plus courante (présent de l'indicatif, voix active) ; les rares objets sont également des pronoms clitiques ; il y a très peu d'expansions (modal, adverbe, négation) ni même d'injections macrosyntaxiques (éléments non régis ou même disfluences). L'ensemble est réalisé très rapidement, en quelques syllabes.

Le coeur des dépendances à distance possède donc une forme très particulière. Si le phénomène des dépendances à distance était seulement une extraction d'un élément, y aurait-il une telle régularité dans le pont ? L'environnement cognitif (mémoire, encodage, décodage, processing, etc.) peut-il à lui seul influencer à ce point les locuteurs pour qu'ils limitent systématiquement les possibilités d'expansion qu'offre dans d'autres contextes la construction verbale finie ? Les particularités formelles du pont nous orientent plutôt vers une origine structurale.

L'analyse des structures en infinitive a montré de façon plus évidente que le lien entre les deux verbes, V1 (verbe pont) et V2, ne pouvait pas reposer sur un enchâssement. Un continuum peut être établi, qui irait de « aller », verbe intermédiaire entre les auxiliaires et les modaux, les modaux traditionnels (« pouvoir », « devoir »), les verbes syntaxiquement transparents (« sembler », « commencer », etc.), les causatifs (« faire », « laisser ») et les constructeurs à valeur modale. Le V2 sélectionne les compléments lorsque le V1 est un auxiliaire, un modal et un verbe syntaxiquement transparent ; le V1 ajoute un complément lorsque le V2 est un causatif. Dans ces contextes-là, la modélisation par enchâssement occulte que les deux verbes ne sont pas deux prédicats ou constructeurs distincts pourvus de toute la palette d'expansions possibles. La situation est plus fidèlement décrite si l'on pose que les deux verbes forment une construction complexe. Dans le cas où le verbe pont est un constructeur à valeur modale, les deux verbes forment toujours une chaîne verbale. Chaque verbe peut avoir une valence mais la valence du premier n'est jamais développée au point de

lui conférer une valeur descriptive. Les compléments sont globalement disposés autour de la chaîne. Cette dissymétrie de développement de la valence se traduit par une dissymétrie sémantique entre les deux verbes, le premier n'étant qu'un modalisateur du second.

Notre hypothèse est que la notion de chaîne verbale peut être étendue aux cas des « verbes ponts + verbes finis ». En effet, ces verbes ponts se comportent de la même manière : il n'ont pas de valence développée et sont des modalisateurs du V2. Par ailleurs, le faible nombre d'insertions macrosyntaxiques indique que la construction est aussi unie que celle en infinitif. Pour les locuteurs, les deux verbes ne feraient qu'un. Il y aurait donc bien lieu de parler de 'chaîne verbale'. Dans ce cas, l'élément extrait ne serait pas régi à distance, puisqu'il dépendrait des deux verbes, réunis en 'constructeur complexe'.

Nous remettons donc en cause l'idée qu'un recteur soit obligatoirement constitué d'une catégorie unique pour un complément donné, et plus généralement qu'un constructeur complexe ne puisse comporter qu'un seul verbe fini. Nous considérons que, de même que les chaînes « auxiliaire + verbe » et « modal + verbe infinitif » régissent le sujet et les compléments, les chaînes 'étendues', ou complexes verbaux, « verbe pont + que +verbe fini », « verbe faible + que + verbe fini » régissent l'ensemble des compléments. Cette multi-rection est le pendant des multi-régis. Nous savons qu'un recteur ne régit pas forcément une seule catégorie. Il existe des rections complexes (*je trouve ce discours intéressant*). Le complexe verbal serait donc un type particulier de prédicat complexe dont d'autres formes sont : verbe+nom (*avoir l'intention de*), verbe+adjectif (*être capable de*), verbe+verbe (*vouloir faire*). Dans ce cas, il n'y a plus lieu de parler de 'dépendance à distance'. La distance disparaît avec la formation du constructeur complexe, qui gouverne localement l'ensemble des compléments.

Les séquences avec effet de dépendance à distance manifestent donc un constructeur complexe ou une chaîne verbale de forme particulière, avec une construction syntaxique identifiable (peu de négatives, peu d'injections possibles, pas de parenthétiques, peu de disfluences). Sa valeur sémantique est également constante : elle modalise la description d'une action ou d'une situation. L'action est posée et accompagnée explicitement d'un point de vue ou d'un degré de certitude qui invite à prendre des précautions sur la valeur de vérité de l'énoncé ou apporte une évaluation subjective.

Compte tenu de cette association constante entre particularités syntaxiques et sémantiques fortement liées, nous pouvons poser l'hypothèse que les effets de dépendance à distance peuvent être ramenés à l'existence d'une 'construction' du français, au sens indiqué par (Lebaud, Paulin, & Ploog, 2006) :

> « Cette théorie de Construction Grammar a pour source de nombreux travaux tels Fillmore (1985[147], 1988[148]), Fillmore & Kay (1993[149]), Goldberg (1995[150]). Ce que nous allons présenter maintenant est la grammaire de constructions telle qu'elle a été formulée par Goldberg (1995). Dans cette théorie, le lexique prend pour unité de base des constructions, c'est à dire des correspondances entre une forme et un sens, qui existent indépendamment des éléments qui les composent.
> "C is a construction iff defC is a form-meaning pair <Fi, Si > such that some aspect of Fi or some aspect of Si is not strictly predictable from C's component parts or from other previously etablished constructions" (Goldberg 95, p.4) » (Lebaud et al., 2006, p. 94)

Si l'on pose qu'il s'agit d'une 'construction' qui allie une forme précise à un sens de 'commentaire modal', que l'on pourrait appeler « complexe verbal », deux conséquences en découlent. En premier lieu, on pourrait mieux comprendre la difficulté à cerner la valeur lexicale des verbes qui entrent dans la construction. Les verbes à valeur modale seraient tout à fait adaptés et fréquents dans la construction. Mais la construction elle-même pourrait également influencer l'interprétation en favorisant l'interprétation modale d'un verbe polysémique. Des effets de 'coercition' pourraient ainsi se manifester avec des verbes dont la valeur initiale paraît peu modale (les verbes dits 'complexes', tels que « hurler »), notamment dans certains contextes favorables, par exemple des contextes de reprise (440). Cela donnerait lieu à des continuums d'acceptabilité.

(440) Qu'est-ce que tu lui as hurlé qu'il devait faire ? tu devrais être plus cool avec lui.

Les seuls verbes qui résisteraient à une telle coercition seraient des verbes impossibles à attirer sur le terrain modal de par leur caractère intrinsèquement objectif (« faire que », « vouloir dire que »). En second lieu, s'il s'agit d'une construction, il est probable qu'elle soit

[147] Fillmore 85, "Syntactic intrusion and the notion of grammatical construction » *Berkeley Linguistic Society*, 11, p73-86.
[148] Fillmore Charles J. 88, "the mecanisms of 'Construction Grammar' », *Berkeley Linguistic Society*, 14, p35-55.
[149] Fillmore & Kay Paul 93, "Construction Grammar Coursebook », manuscrit, University of California at Berkley, Department of linguistics.
[150] Goldberg Adele E 95, Constructions. A Construction Gramma Approach to Argument Structure, Chicago : University of Chicago Press.

employée dans d'autres environnements syntaxiques, ce qui sera abordé dans le paragraphe suivant.

4.2.2 Ré-emploi dans le *sluicing*

Le *sluicing* est un autre phénomène remarqué par les générativistes (Ross, 1969). Il apparaît dans les contextes d'interrogatives indirectes. Il se présente le plus souvent sous la forme d'un interrogatif (403).

(441) Je viendrai mais je ne sais pas quand

Il est généralement interprété comme un type d'ellipse. L'approche générativiste considère qu'il y a un mouvement puis une suppression des éléments ellipsés. Mais l'hypothèse du mouvement pose problème car certains exemples montrent que le *sluicing* ne respecte pas les contraintes de mouvement du wh-. (Merchant, 2006) présente un exemple de ce type mettant en jeu les constructions de dépendance à distance. Si l'on reconstruit l'interrogative indirecte complète, l'énoncé viole la contrainte du nom complexe (442), de l'extraction du sujet (443), de l'extraction de l'objet (444), etc.

(442) They want to hire someone who speaks a Balkan language, but I don't remember which. (Merchant, 2006, p. 274)

**I don't remember which (Balkan language) they want to hire someone [who speaks_]. (Merchant, 2006, p. 275)*

(443) A biography of one of the Marx brothers is going to be published this year. Guess which_!

**Guess which (Marx brother) [a biography of _] is going to be published this year. (Merchant, 2006, p. 275)*

(444) Ben will be mad if Abby talks to one of the teachers, but she couldn't remember which.

**Ben will be mad if Abby talks to one of the teachers, but she couldn't remember which (of the teachers) Ben will be mad [if she talks to _]. (Merchant, 2006, p. 275)*

Une hypothèse alternative considère qu'il y a copie des éléments ellipsés, qui disparaissent ensuite. Dans ce cas, les énoncés sont agrammaticaux dans leur forme complète mais l'effacement d'une partie des éléments leur rend leur grammaticalité. Cette solution est peu satisfaisante. On peut tout aussi bien considérer dans ce cas qu'il y a un mouvement qui

devrait aboutir à une agrammaticalité, l'ellipse permettant à l'énoncé de retrouver sa grammaticalité. Mais comment justifier qu'un effacement rende grammatical un énoncé agrammatical ?

La troisième hypothèse s'appuie sur le caractère flou de la reconstruction syntaxique à l'œuvre dans l'ellipse et propose de reconstruire un mouvement court, accepté dans sa forme non ellipsée (445). Mais la forme reconstruite est très arbitraire et cette solution ne résout que les ellipses de constructions verbales.

(445) They want to hire someone who speaks a Balkan language, but I don't remember which <he should speak _ >. (Merchant, 2006, p. 277)

Une analyse totalement différente a été proposée par d'autres linguistes (Culicover & Jackendoff, 2005). Ces derniers ne considèrent pas le *sluicing* comme un type d'ellipse. Ils refusent l'idée d'une structure syntaxique non-prononcée, absente ou disparue et posent que l'interrogatif est le complément du verbe ; il est l'unique noeud de la complétive. Etant la tête d'un énoncé autonome, c'est le contexte qui permet une interprétation sémantique de ce fragment comme une proposition interrogative Le *sluicing* ne serait donc pas une construction dérivée de la construction complète, ce qui explique que la reconstruction de celle-ci puisse être agrammaticale (442)(443)(444). L'absence d'ellipse serait également prouvée par des exemples dans lesquels le lien de dépendance entre l'interrogatif et le verbe présupposé est absent (446). La reconstruction est là-encore agrammaticale (447).

(446) Pierre a mangé mais je ne sais pas quoi

*(447) *Pierre a mangé mais je ne sais pas quoi il a mangé*

Par l'observation des combinaisons possibles entre interrogatifs et préposition dans les *sluicing* anglais (du type « *who with* », « *what from* », « *?when before* »), (Culicover & Jackendoff, 2005) arrivent à la conclusion que la construction est une liste de locutions idiomatiques :

> « This pattern suggests that the acceptable case of sluice-stranding, like irregular past tenses, are learned one by one, to some extent with the help of family resemblance to other case. There is therefore not a general rule for sluice-stranding, but rather a collection of related idiomatic lexical entries. Such a solution of course precludes deriving sluice-stranding, from full sentences by general processes of wh-movement and deletion, as is assumed in much of the literature. » (Culicover & Jackendoff, 2005, p. 171-172)

En français, lorsque le *sluicing* est constitué du verbe et de l'interrogatif, il est généralement détaché (apposé par juxtaposition ou coordination), ce qui le rend relativement autonome. Il est également mobile et peut être placé en début, en milieu ou en fin d'énoncé. On peut en conclure que le *sluicing* a un statut parenthétique (Rigaud, 2010, p. 360).

Cette construction, qui est donc totalement différente de celles étudiées précédemment (relatives, interrogatives et clivées), peut également faire intervenir les chaînes « VqueV » (448) ou « V-Vinf » (449).

(448) Je peux vous assurer que je ne lâcherai pas et si on me dit que j'ai tort, ***il faudra qu'on me dise*** *pourquoi ! [PRAT.Feminin] ((Rigaud, 2010), Annexes)*

(449) Je viens de traverser une ville où tout le monde courait... Je ne ***peux pas vous dire*** *laquelle... [DIVS.Humour] ((Rigaud, 2010), Annexes)*

Or, les verbes présents dans ces chaînes ont été identifiés comme verbes ponts. Ils partagent également une rection syntaxique minimale. On peut donc considérer qu'il y a un ré-emploi du complexe « verbe à valeur modale + verbe » dans un autre contexte syntaxique que celui que nous avons étudié. On pourrait estimer ici que l'ensemble joue le rôle de modifieur sémantique du pronom qui serait déplaçable et qui serait la tête de la construction (450)(451).

(450) Laquelle, je peux pas vous dire

(451) Pourquoi, il faudra qu'on me dise

La chaîne verbale forme donc une construction à part entière, employée dans plusieurs environnements syntaxiques. Il faudrait formaliser les énoncés d'une manière qui mette en évidence ce complexe verbal et qui traite la question de la structure interne du complexe.

4.2.3 Représentation par le nucléus verbal

Dans un premier temps, les deux modélisations qui nous avaient semblé les plus intéressantes étaient celle de TAG (Joshi, 1987; Joshi et al., 1975) et celle du nucléus verbal (Kahane, 1997; 2000) qui adaptent les réseaux de dépendance traditionnels (voir §1.1).

TAG présente l'avantage d'indiquer d'une certaine manière que le verbe principal est le V2 puisqu'il est tête d'un arbre initial. Le pont s'insère à l'énoncé dans un deuxième temps, par adjonction, comme arbre auxiliaire. Cette représentation est malheureusement loin d'être parfaite. Si l'on présente le V1 comme une adjonction Phrase sur un nœud Phrase au reste de l'énoncé, comment expliquer par exemple que le V1 contraint le mode du V2 ? En effet, lorsque le V1 appartient au lexique « vouloir », « falloir », « valoir mieux », « souhaiter », être intéressant », le V2 est au subjonctif. Les deux verbes doivent donc être pris comme un ensemble et il faut une stipulation particulière pour amener le subjonctif. En outre, la description d'énoncés authentiques nous a permis de conclure au statut hautement restreint de la construction pont. Si cette construction est catégorisée comme phrase dans les séquences comme *à qui veux-tu que je m'adresse* pour être adjointe à la phrase initiale, il faudra toute une série de contraintes pour empêcher cettte phrase de recevoir les expansions ordinairement possibles (autres compléments ajouts...). Le formalisme n'exprime pas naturellement le caractère spécifique de la construction pont.

En revanche, c'est exactement ce que propose la représentation de (Kahane, 1997; 2000). Une rection peut être assumée par un seul verbe ou par un complexe verbal, également nommé « nucléus verbal » (Kahane, 2000). L'analyse des dépendances à distance est alors largement simplifiée, comme l'explique (Kahane, 2008) :

> « Il est possible dans une relative de remplacer le verbe principal de la relative par n'importe quelle chaîne de verbes :
>
> *Voilà la personne à laquelle Félix* / *parle.*
> / *veut parler.*
> / *devrait penser à parler.*
> / *aurait aimé que Luce puisse accepter de parler.*
>
> [...] On peut alors formuler la contrainte sur l'extraction de manière très simple : seul peut être relativisé un dépendant immédiat du verbe principal, sachant que le verbe principal peut être un verbe simple, mais aussi un nucléus verbal » (Kahane, 2008, p. 15 - en ligne).

L'important est que ce qui est inséré soit un verbe ou une chaîne de verbes et non une phrase ce qui permettrait de limiter a priori les expansions de la structure.

Comme l'avait proposé (Blanche-Benveniste, Non publié), la syntaxe de l'élément « extrait » ne tient pas compte de l'organisation interne du nucléus verbal. Cependant la modélisation,

pour être complète doit rendre compte de l'organisation interne du nucléus. Une solution serait de considérer qu'il s'agit là d'une 'structure plate' sans hiérarchie : les verbes seraient simplement juxtaposés dans le nucléus à la manière d'un nom composé ou de certaines prépositions complexes : « en raison de », « au regard de ». L'autre solution serait de définir une hiérarchie interne, de la même manière que pour « de sorte que ». Si la séquence est analysée comme une locution conjonctive en syntaxe externe, une syntaxe interne « préposition + nom » est nécessaire pour expliquer par exemple l'insertion possible d'un adjectif (*de telle sorte que*). Pour (Kahane, 2000), à l'intérieur du nucléus verbal, la relation entre les deux verbes reste hiérarchique, comme le montre le sens des flèches dans le réseau ci-dessous.

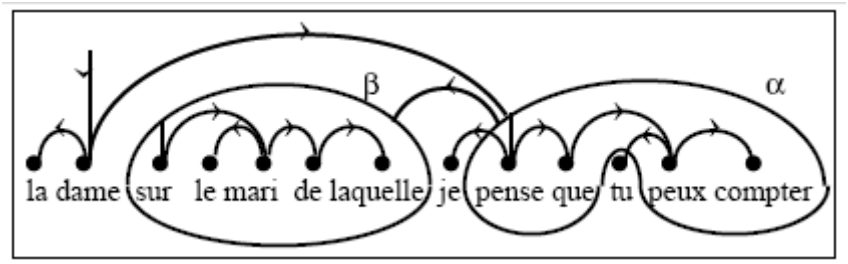

FIGURE.33 Arbre à bulle d'une dépendance à distance en complétive (Kahane, 2000, p. 267)

La question reste ouverte et liée à la description d'autres constructeurs complexes que ceux à base verbale. Mais la modélisation qui nous semble à préserver est celle qui fait des chaînes verbales et des complexes verbaux des constructions spécifiques irréductibles à des phrases non canoniques.

Nous avons défini le phénomène des dépendances à distance comme étant globalement une 'construction' : un constructeur complexe avec expansion modale. Mais n'y aurait-il pas en fait plusieurs 'constructions', de forme et de sens plus précis, en fonction de la configuration syntaxique hôte : interrogative, relative et clivée ? C'est ce que nous allons examiner dans la section suivante.

4.3 La chaîne verbale : unité fonctionnant dans les trois configurations syntaxiques

Dans les analyses précédentes, nous avons observé la totalité des occurrences comportant des dépendances à distance en distinguant uniquement les structures en complétive des structures en infinitive. Nous sommes partis de l'idée des génératives qui analysaient le phénomène des dépendances à distance comme un phénomène général d'extraction et nous sommes arrivés à la conclusion qu'il s'agissait plutôt d'un phénomène spécifique de constitution de constructeur complexe. Dans ce cheminement, nous avons toujours considéré que les contraintes ou la composition du complexe verbal étaient a priori identiques pour toutes les configurations syntaxiques : les relatives, les interrogatives et les clivées.

Lors de la comparaison du comportement de la construction selon les genres représentés dans les tranches du corpus CERF, nous avons observé qu'un ensemble de traits était commun à certaines tranches, ce qui nous a permis de les regrouper (voir §3.5.1.3). Parmi ces traits, se trouvaient la configuration syntaxique majoritaire et les traits descrptifs du pont. Bien que la corrélation entre configurations syntaxiques et traits particuliers n'ait pas été prouvée, on peut y voir un indice montrant que les contraintes sont dépendantes des configurations.

Nous allons donc vérifier dans cette section s'il existe des différences de comportement entre les différentes configurations syntaxiques qui hébergent le phénomène : les relatives, les interrogatives et les clivées. Bien évidemment, pour distinguer plusieurs constructions de complexe verbal, il faudrait que ces contraintes ne découlent pas directement de la simple composition entre la construction hôte et le complexe.

4.3.1 Comparaison des configurations syntaxiques

Chaque configuration syntaxique implique des propriétés sémantiques et discursives particulières. Celles-ci influencent les contraintes qui pèsent sur la construction de

dépendance à distance. Le comportement de chaque configuration syntaxique est donc distinct dans le détail.

Les dépendances à distance en clivées et en pseudo-clivées ne sont pas suffisamment représentées dans les corpus : 22 occurrences avec verbe fini et 20 occurrences à l'infinitif pour les clivées, 4 occurrences avec verbe fini et 10 occurrences à l'infinitif pour les pseudo-clivées. La description des configurations syntaxiques portera donc sur les relatives et les interrogatives[151]. Nous décrirons de façon parallèle la structure en complétive et la structure en infinitive.

4.3.1.1 Relatives vs interrogatives

Les tendances générales des dépendances à distance (voir §3.4) ont masqué les différences de comportement de chaque configuration syntaxique. Or les dépendances à distance sont beaucoup plus contraintes dans les configurations interrogatives que dans les configurations relatives.

Cette différence est très visible pour les complétives. Les interrogatives s'éloignent moins souvent de la forme la plus fréquente. Les relatives sont plus variées à la fois dans le lexique et la morphologie du verbe pont[152], ainsi que dans les types de sujets (TABLEAU 80).

Traits	Tendance majoritaire	Relatives	Interrogatives
Nb occ		**95**	**108**
Sujet	Type = Pronominal	(88 occ) 93%	100%
	lexique	(30 occ) 32%	(8 occ) 7%
Verbe	Mode= Indicatif	(86 occ) 91%	(106 occ) 98%
	Temps = Présent	(69 occ) 73%	(103 occ) 95%
	Voix = Actif	(93 occ) 98%	100%

TABLEAU 80. Comparaison des ponts en relative et en interrogative, dans les constructions complétives

[151] Notre classification s'est appuyée sur le critère de la syntaxe externe (distribution de construction nominale pour les relatives, distribution de construction verbale pour les interrogatives indirectes), dans l'optique de faire apparaître une sélection de verbes ponts en fonction de la valeur pragmatique de l'énoncé.

Ex : j*e sais ce qu'il faut que je fasse vers qui je dois aller* (interrogative indirecte)

Ex : *c'est ce qu'on pense qui devrait être fait* (relative)

[152] Pour plus de clarté, nous continuerons à parler de « pont » et de « verbe pont », même si ce lexique n'est plus approprié dans une optique de constructeur complexe.

Dans les interrogatives, les possibilités du verbe pont sont plus limitées dans son lexique et sa morphologie, mais également dans sa construction : une seule occurrence présente un second complément du verbe pont. Ce complément est le récepteur de « dire », présent sous la forme de clitique objet. Les insertions macrosyntaxiques sont elles aussi plus rares, ce qui indique que le complexe est plus figé et mieux intégré (TABLEAU 81).

Traits	Tendance majoritaire	Relatives	Interrogatives
Constitution du pont	Sujet + verbe uniquement	(90 occ) 95%	(107 occ) 99%
	Objet Ajout	5 clitiques (1 ajout postposé)	1 clitique
Insertions macrosyntaxiques	Non-régi disfluence	7 adv, 1 double marquage 3 disfluences	2 adv

TABLEAU 81. Comparaison de la construction du verbe pont en relative et en interrogative, dans les constructions complétives

Dans les interrogatives, le verbe pont ne présente aucune variation de modalité. Il est toujours attesté dans une forme positive. Les relatives, en revanche, comportent parfois un verbe pont à la forme négative (3 occurrences) (452).

(452) tous les hommes préférant le bonheur au malheur, le plaisir à la souffrance, ne fais pas aux autres ce que tu ne voudrais pas qu'on te fasse à toi-même. (CERF, DIV)

Enfin, les types de verbes sont différents entre relatives et interrogatives : ces dernières ne produisent aucune occurrence avec un verbe pont factif, semi-factif ou implicatif alors que les relatives permettent les trois types de verbes liés à la présupposition.

Il est cependant possible de voir l'influence des propriétés pragmatiques des deux constructions sur une partie des différences au moins. Les interrogatives constituent une adresse à l'interlocuteur. Le pont a généralement pour rôle de convoquer l'opinion de ce-dernier. Il n'est donc pas surprenant que le sujet du verbe pont soit majoritairement à la deuxième personne dans les interrogatives (77%). Les relatives ne concernent pas l'interlocuteur aussi directement. Dans les constructions de dépendance à distance, leur pont met plutôt en relief l'opinion du locuteur (20%) ou d'une tierce personne (35%). La seconde personne est la moins représentée.

Les différences entre les relatives et les interrogatives sont moins marquées dans les constructions infinitives[153] (TABLEAU 82). On retrouve tout de même une plus grande variation dans la morphologie du verbe pont et dans les types de sujets en configuration de relative. La négation du verbe pont est présente uniquement dans les relatives. La construction du verbe pont est plus développée dans les relatives (objet clitique, adverbe). Comme dans les complétives, la deuxième personne est la personne la plus fréquente dans les sujets des verbes ponts en interrogative, et la personne la moins fréquente en relative.

Dans les constructions infinitives, en revanche, le lexique des verbes ponts en relative n'est pas plus varié qu'en interrogative. Le faible nombre d'occurrences d'interrogative peut expliquer ce résultat.

Si le trait de factivité ou d'assertion d'un verbe suivi de l'infinitif n'est a priori pas absent (Mørdrup, 1975), il est plus difficile à observer que dans le cas d'un verbe à complétive. En effet, le verbe infinitif peut souvent être interprété comme un procès non actualisé. Comment savoir alors s'il est asserté ou présupposé ? Nous nous abstiendrons donc de définir ce trait dans les constructions en infinitive.

Traits	**Tendance majoritaire**	**Relatives Inf_constr**	**Interrogatives Inf_constr**
Nb occurrences		73	11
Sujet	Type = Pronominal	(70) 96%	100%
	Personne	P1 (27) 37% P2 (15) 21% P3 (31) 42%	P1 (1) 9% P2 (10) 90%
Verbe	Lexique	(16) 22%	(4) 36%
	Mode = Indicatif	(51) 70%	(9) 82%
	Temps = Présent	(45) 63%	100%
	Voix = Actif	(71) 97%	100%
	Négation	(5) 7%	0
Constitution du pont	Sujet + verbe uniquement	(62) 85% 3 clitiques 9 adv régis, dans 8occ	(9) 82% 1 clitique 1adv régi

TABLEAU 82. Comparaison de la construction du verbe pont en relative et en interrogative, dans les constructions infinitives avec verbe pont constructeur

[153] Afin que la comparaison soit judicieuse, seules les occurrences avec un verbe pont constructeur sont observées.

Les relatives sont assez différentes des interrogatives dans les constructions de dépendance à distance. Mais une comparaison plus fine peut encore être effectuée : celle des interrogatives directes et indirectes.

4.3.1.2 Interrogatives directes vs interrogatives indirectes

Les interrogations directes (excepté les interrogations à valeur rhétorique) expriment des demandes explicites à un interlocuteur. Les interrogations indirectes, en revanche, n'attendent pas forcément de réponse (453), et donc ne mettent pas directement en cause l'interlocuteur. Elles sont à la croisée des interrogations et des assertions. Leur rapport à l'énonciation est donc différent de celui des interrogatives directes.

(453) Je me demande où j'ai mis mes clés

Cette différence se ressent dans l'emploi des constructions de dépendance à distance, notamment en construction complétive. Les interrogatives directes, qui sont plus fréquentes, sont légèrement plus contraintes (TABLEAU 83). Le lexique des verbes ponts est moins riche, la morphologie du verbe pont est plus souvent conforme au trait dominant (présent, indicatif, actif). De même, les sujets des verbes ponts sont dans les deux cas des pronoms qui réfèrent à la deuxième personne mais dans les interrogatives directes, leur proportion est beaucoup plus importante.

Dans les interrogatives directes, les verbes ponts ne comportent jamais d'autres compléments régis que la complétive. Les seules insertions dans le pont sont des adverbes non-régis et elles restent très rares. Dans les interrogatives indirectes, par contre, on relève une occurrence dans laquelle le verbe pont régit un pronom clitique objet (454).

*(454) Ce n'est déjà pas facile si je commence à me juger je ne suis pas sortie du bois. Alors vous comprenez pourquoi je **vous** dis que je ne dois pas trop anticipé {sic} (CERF, DIV)*

En revanche, dans les deux types d'interrogation, on constate le même phénomène d'évitement des verbes ponts factifs, semi-factifs et implicatifs et d'évitement de la négation du verbe pont.

Comme nous l'avons déjà remarqué (§3.4.6), plus la structure est contrainte, plus elle est ambiguë (ou inversement, plus la structure est ambiguë, plus elle est contrainte). Ce

phénomène pourrait être expliqué par le fait que l'ambiguïté peut être mieux résolue par l'auditeur compte tenu d'une certaine habitude et d'une attente d'une structure typique. La comparaison des deux types d'interrogative vérifie cette observation : le rattachement du pronom antéposé avec le V1 ou le V2 est ambigu dans plus de la moitié des occurrences dans les interrogatives directes. Il est beaucoup moins ambigu dans les interrogatives indirectes.

INTERROGATIVES (108)	**Directe**	**Indirecte (dans l'échantillon sélectionné)**
Nb occurrences	92	16
lexique	(7) 8%	(5) 31%
Verbe : Mode=Indicatif Temps= Présent Voix=Actif	 (91) 99% (88) 96% 100%	 (15) 94% (15) 94% 100%
Nb compléments	0 (+valoir mieux)	1 clitique objet (*je vous dis*)
Sujet : Categ personne	 Pronom / zéro (1) P2 (78), P3 (4), impers (9), zéro (1)	 Pronom (100%) P1 (1), P2 (5), P3 (3), impers (7)
Ambiguité process	(50) 54%	(5) 31%
Insertions macrosyntaxiques Non-régi disfluences	 2 adv associés (*diable, ensuite*) 0	 0 0

TABLEAU 83. Comparaison des ponts en interrogative directe et indirecte, dans les constructions complétives

La comparaison doit être interprétée avec prudence. En effet, le faible nombre d'occurrences d'interrogatives indirectes ne permet pas de certifier que les différences présentes dans les occurrences relevées sont représentatives de l'emploi général des constructions de dépendance à distance en interrogative.

Dans les constructions infinitives avec V1 constructeur, la répartition est de 10 interrogatives directes et d'une interrogative indirecte. Il n'y a donc pas lieu d'effectuer une comparaison.

L'observation de la richesse lexicale et de la construction du verbe pont semble indiquer que les interrogatives sont plus contraintes que les relatives, et que les interrogatives directes sont

légèrement plus contraintes que les interrogatives indirectes. Mais rien ne permet d'affirmer l'existence de plusieurs constructions car les restrictions supplémentaires observées pour les interrogatives semblent liées à la fonction pragmatique marquée de ces dernières. Nous allons voir dans la section suivante si le lexique des verbes ponts diffère en fonction de la construction hôte.

4.3.2 Cadre syntaxique et lexique des verbes ponts

Nous avons déjà noté (§4.3.1.1) que les verbes liés à la présupposition étaient attestés dans les relatives mais pas dans les interrogatives. Y a-t-il d'autres répartitions de ce genre ? Comment peut-on classer les verbes ? Quelle classe est favorisée dans quel contexte ?

Les verbes généralement employés dans les constructions de dépendance à distance ont une interprétation modale. Si la présence d'un complément ou d'un adverbe pousse à une interprétation en verbe descriptif, l'acceptabilité diminue beaucoup (455).

(455) ??Qui Pierre a dit récemment à Marie que tu avais rencontré

Les verbes ponts ne sont pas employés en tant que verbes constructeurs, avec une rection développée, mais comme opérateurs modaux. Nous avions établi une classification assez sommaire des valeurs modales (voir §3.3.5), le but étant alors de distinguer modaux et verbes descriptifs. Dans cette section, nous allons étudier plus en détail les diverses modalités en jeu : déontique, épistémique, évidentielle, évaluative, désidérative et affective.

Les déontiques décrivent

> « un état ou un fait du monde dont l'existence est présentée comme une obligation et non pas comme une vérité » (Chu, 2008, p. 156).

Ils sont représentés par « falloir », « être prévu », « amener à », « valoir mieux » et « charger de ». Cette modalité est présentée comme objective. On s'attend donc à la retrouver dans tous les contextes.

(456) qui il vaut mieux que je voie (CorpAix, BUSABCD)

Les évaluatifs permettent au locuteur de porter un jugement de valeur sur le contenu de l'énoncé : « être intéressant / normal », « sembler important ».

(457) la Commission indique ci-après, pour chaque sujet, les points sur lesquels il pourrait être particulièrement intéressant que les gouvernements expriment leurs vues (CERF, INST)

Les désidératifs montrent la volonté du locuteur : « vouloir », « désirer », « souhaiter », « avoir envie », « envisager », « rêver », « décider », « attendre », « demander ». L'action désirée n'est pas encore réalisée.

(458) pourquoi vous voulez qu'on rentre (PFC, 75xad1)

Les affectifs marquent les goûts et les sentiments du locuteur : « adorer », « aimer », « apprécier », « espérer », « répugner à », « préférer ». Ils sont totalement subjectifs.

(459) et vous qu'est-ce que vous préférez euh fabriquer finalement euh des pots plats et objets utilitaires ou plutôt euh des objets euh plus figuratifs (C-ORAL-ROM, ffamdl28)

Les épistémiques et les évidentiels indiquent respectivement :

> « l'attitude du locuteur vis-à-vis de la véracité du fait rapporté ou de l'évaluation de sa connaissance d'un fait ou d'un état dans le monde » (Chu, 2008, p. 156) et
> « l'engagement du locuteur à l'égard de la vérité de son propos. » (Chu, 2008, p. 156).

On simplifie souvent les définitions en 'degré de certitude du locuteur' et 'source de l'information'. La source peut être extérieure si l'information a été lue ou entendue, mais aussi intérieure et ressentie par le locuteur : « dire », « rappeler », « voir », « entendre », « sentir », « ressentir », « trouver », « imaginer », « remarquer », « se rendre compte », « prendre conscience », « avoir le sentiment », etc. Parmi les épistémiques se trouvent les verbes « croire », « penser », « estimer », « supposer », « être vrai / évident / clair », « présumer », « sembler » et « savoir ». Les deux modalités peuvent être confondues lorsque le locuteur n'est pas certain de ce qu'il dit parce qu'il n'est pas à la source de l'information. Il se désengage en citant sa source et donne par là un indice de la véracité de ses propos. Ce sont les modaux les plus employés dans les constructions de dépendance à distance.

(460) à partir de quand tu crois que c'est important (CorpAix, 13GURIN)

(461) *ça fait combien de temps que vous dites qu'il est euh l'oeuvre est ici (C-ORAL-ROM, fpubmn02)*

La répartition de ces valeurs dans les constructions hôtes est la suivante (TABLEAU 84) :

	Interrogatives		Relatives	
	Directes (92)	Indirectes (16)	Appositives (31)	Déterminatives (58)
Déontiques	X (10)	X (7)	X (6)	X (6)
Désidératifs	X (70)	X (4)	X (4)	X (16)
Epistémique	X (9)	X (2)	X (12)	X (21)
Evidentiels	X (3)	X (3)	X (13)	X (13)
Evaluatifs	X (web)		X (2)	

TABLEAU 84. Répartition des modaux dans les différentes configurations syntaxiques

Nous avons cité des exemples de toutes les catégories de modaux en interrogative hormis les évaluatifs. Ceux-ci ne sont pas présents dans les corpus mais on en trouve un exemple sur le Web.

(462) *L'éthique dit à l'homme ce qu'il est bon qu'il fasse et devienne, mais elle ne dit pas pourquoi il est bon qu'il y ait des hommes, et qu'il y ait un monde*[154]*. (Web, institutionnel)*

Globalement les diverses catégories de modaux sont attestées dans les deux constructions. On s'attendait en particulier à voir les déontiques figurer dans tous les types de constructions hôtes. Mais en prenant en compte les sous-types de constructions, nous avons remarqué deux contraintes de cohabitation entre modaux et structures syntaxiques.

154 http://books.google.fr/books?id=8WDmj3Qa5hMC&pg=PA256&lpg=PA256&dq=%22ce+qu%27il+est+bon+qu%22&source=bl&ots=X-Dxp7kz0V&sig=QPPwXaagtdVVf5TkD9v6QS2p5FA&hl=fr&ei=CTxfTuDxD4yn8QO0i7HGAw&sa=X&oi=book_result&ct=result&resnum=3&ved=0CCcQ6AEwAg#v=onepage&q=%22ce%20qu%27il%20est%20bon%20qu%22&f=false, consulté le 01/09/11

4.3.2.1 Verbes faibles et relatives déterminatives

Toutes les modalités, et plus encore les modalités épistémiques et évidentielles, affaiblissent la validité de l'assertion. Leur sens est généralement faible (Blanche-Benveniste & Willems, 2007; Urmson, 1952) (voir §1.4.1.2), c'est à dire qu'il indique une opinion (463) et non l'activité mentale de 'croire', 'penser', etc. (464).

(463) Je crois que Jean est en retard (opinion)

(464) Je crois que Dieu existe (croyance)

Les verbes faibles ayant pour fonction de restreindre la validation de l'assertion ou d'anticiper la validation de l'interlocuteur dans les interrogatives, ils ne devraient apparaître que dans les parties assertées ou interrogées des énoncés et non dans les parties présupposées. On les trouve fréquemment dans les interrogatives. Dans les clivées (Blanche-Benveniste, 2000), on observe régulièrement que les épistémiques sont dans la partie assertée : tous les verbes à lecture faible sont ponts dans les clivées non contrastives (465) où la partie assertée est introduite par « que », tandis que les verbes à lecture forte sont ponts dans les clivées contrastives (466).

(465) j'ai eu une relation privilégiée avec le pays à cette époque-là + et c'est c'est pour ça que je pense que j'ai autant /d', ø/ envie de retourner au Portugal (CorpAix, PORTUGA)

(466) C'est parce que la sécurité sociale est en fait une partie du salaire (une part socialisée, collective qui s'ajoute à la part individuelle) que nous pensons que ce sont les syndicats qui ont vocation à solliciter des suffrages de l'ensemble des assurés sociaux dépendant du régime général. (CERF, POLI)

Dans les relatives, les verbes faibles à valeur épistémique devraient se trouver dans les relatives appositives et prédicatives, mais pas dans les relatives déterminatives, dans lesquelles le contenu de la construction verbale est présupposé. Effectivement, certains types de relatives déterminatives interdisent la modalisation par des épistémiques faibles (467) alors qu'ils autorisent d'autres types de modaux (468), des épistémiques de sens fort (469) et d'autres modaux forts (470).

*(467) *Emporte la tasse que je crois que tu laves*

(468) Emporte la tasse qu'il faut que tu laves

(469) quels sont les aspects avec lesquels on sait qu'un enfant ne sera pas viable (CorpAix, FTUSPC)

(470) mais là il faut que tu sélectionnes l'ordre dans lequel tu veux qu'il te replace ça (CRFP, STR PRI 2)

Mais il existe des cas intéressants de relatives clairement déterminatives où l'on trouve des modalisations épistémiques par verbes recteurs faibles. Dans ces énoncés, l'antécédent est très peu référentiel (« ce ») (471). Par ailleurs, les effets de contrastes favorisent ces emplois.

(471) c'est pas forcément couramment admis c'est ce qu'on pense qui devrait être fait (CorpAix, FTUSPC)

4.3.2.2 « Vouloir » faible

Le verbe « vouloir » présente une particularité. Il peut prendre un sens faible ou fort. Le sens fort volitionnel (objectif) est possible dans toutes les structures. Le sens faible, au contraire, n'est possible que dans les questions rhétoriques. Il remplace une réponse négative et marque l'exaspération ou l'évidence.

(472) Mais que voulez-vous que j'y fasse (CERF, LITC)

(473) C'est toujours Alexandre Qui voulez-vous que ce soit (CERF, DIVS)

Dans les interrogatives directes, le sens faible du verbe « vouloir » est le plus fréquent. Il représente 84% des occurrences (59/70 occurrences). Parmi ces occurrences en « vouloir » faible, une expression est récurrente : *qu'est-ce que tu veux que* ou *qu'est-ce que vous voulez que* (35 occurrences). On peut penser que l'on a affaire dans ce cas à une construction idiomatique, et non plus à une dépendance à distance. Il arrive en effet que *qu'est-ce que vous voulez que* soit employé dans des tours où la reconstruction est possible mais difficile (474) (= *je le sais, comment ça fonctionne*) ou même avec des verbes qui n'attendent pas de complément (475).

(474) T'en as pour longtemps à me soûler avec tes salades, dis, hein ? Qu'est-ce que tu veux que je sache, moi, les atomes et les molécules comment qu'ça fonctionne ? (CERF, DIVS)

(475) qu'est-ce que vous voulez que j'argumente ? (CERF, FORM)

On pourrait très bien l'imaginer avec des verbes intransitifs (476)(477). Lorsque le locuteur emploie le sens faible du verbe, il n'interroge pas sur le vouloir mais affirme une impossibilité.

(476) Qu'est-ce que tu veux que je tombe ? C'est tout plat ! (exemple inventé)

(477) Qu'est-ce que tu veux que je dorme, là ? T'as vu le bruit qu'il y a ? (exemple inventé[155])

Ce tour ressemble à la construction « pourquoi voulez-vous que » (478) où le verbe « vouloir » signifie 'dire'. L'interrogation ne porte pas sur la cause de l'évènement ; il s'agit en fait d'une demande de justification de l'assertion. Dans cette construction, le temps du verbe « vouloir » est bloqué au présent, à l'imparfait ou au conditionnel. Avec un temps différent, l'interprétation est modifiée et la cause revient au premier plan.

(478) Pourquoi voulez-vous qu'il arrive en retard ?

Dans le tour en « qu'est-ce que tu veux que », le présent, l'imparfait et le conditionnel sont également les seuls temps qui permettent l'interprétation rhétorique. La seconde contrainte est que le sujet doit être un pronom clitique de deuxième personne (ou éventuellement de troisième personne).

(479) alors il dit ben il peut pas y mettre une braguette à mon chien elle lui fait qu'est-ce que vous voulez que j'y fasse si le chien il a /pi-, fait/ contre votre façade (CorpAix, PAVRAYST)

Dans les autres cas, l'interprétation volitionnelle est choisie.

(480) « Mets tes mains plus bas. » Qu'est-ce qu'il voulait que je fasse ? Mes mains rampèrent jusqu'à la ceinture.(CERF, LITC)

Il n'est pas possible dans ce cas là de considérer que l'élément antéposé est complément du V2 développé en complexe verbal, puisqu'il y a des cas où il ne peut entrer dans sa valence. Ces exemples demandent une autre analyse, celle de la formation d'interrogatifs complexes à partir de verbes (*à qui je veux / je peux*). Cette question ne relève donc pas de la constitution d'un constructeur complexe ; elle ne sera pas approfondie ici.

Une premier bilan se dégage de ces résultats : la majorité des modaux peuvent s'employer comme ponts dans toutes les constructions et dans toutes les structures syntaxiques. Les

[155] Je remercie mon cousin de s'être intéressé à mon travail et de m'avoir fourni cet exemple.

contraintes pragmatiques favorisent l'interprétation faible de certains verbes dans certains contextes (« vouloir » en interrogative) et la rendent très difficile dans d'autres, au point de l'interdire dans certains cas (relatives déterminatives). Nous pouvons donc conlure qu'il n'existe qu'une construction « complexe verbal », dont les différentes réalisations sont simplement le résultat de combinaisons avec les propriétés syntaxiques et discursives des constructions hôtes.

CONCLUSION

Notre étude s'est concentrée sur les constructions « QU- CV1que CV2 ». Nous nous sommes, limités aux relatives, aux interrogatives et aux clivées, c'est à dire aux structures dans lesquelles il existe un lien syntaxique explicite entre QU- et le reste de la construction. Notre objectif était de décrire les occurrences attestées de cette construction et d'observer d'une part si elles étaient conformes aux contraintes formulées par les études précédentes (à partir de données construites ou de données attestées mais non francophones) et d'autre part quel cadre théorique l'analysait de la façon la plus juste. Pour ce faire, nous nous sommes appuyés sur 229 occurrences de la construction, relevées dans environ trois millions de mots de corpus oraux (en grande partie harmonisés) et neuf millions de mots de corpus écrits.

Résultats obtenus

Nous avons rapidement écarté l'idée d'une construction prototypique quasiment figée, que ce soit dans les corpus adultes-adultes ou adultes-enfants. En effet, nous avons relevé une grande variété de verbes ponts (ou V1). L'extension des corpus au Web, méthode souvent contestée pour des raisons de fiabilité, a dans notre cas été bénéfique car la masse de données (de données écologiques[156], par surcroît) n'aurait jamais pu être atteinte par un corpus. Ces verbes ne correspondent pas particulièrement aux verbes assertifs ni aux verbes recteurs faibles, comme aurait pu le laisser supposer les études antérieures. Presque tous les verbes à complétives peuvent être ponts, particulièrement dans les textes codifiés. Les paramètres qui font que certains le sont plus régulièrement sont leur fréquence dans la langue et l'importance de leur sens modal. Ce sens modal est renforcé par l'absence de compléments et d'adverbes. Ainsi, les ponts sont généralement très courts et de structure syntaxique minimale (sujet pronominal-verbe). L'absence de poids syntaxique et sémantique du pont permettrait un meilleur rattachement entre le V2 et l'élément antéposé, notamment lorsque la connexion syntaxique est ambiguë (mais généralement, il ne l'est pas).

[156] Données non sollicitées par le linguiste

A plusieurs reprises, nous avons constaté que le verbe pont se comportait comme un verbe parenthétique (construction syntaxique limitée : uniquement des compléments clitiques difficulté de négation,) ou un auxiliaire (même type d'adverbes insérés entre les deux verbes). La récursivité des verbes ponts est extrêmement rare. Le V1 semble former une unité avec le V2 et y apporter une modalisation. La comparaison avec les constructions infinitives a fourni un éclairage intéressant. Les emplois de la construction avec un infinitif sont comparables à ceux de la construction avec verbe fini : le V1 est un verbe fréquent, modal et sa construction syntaxique est limitée. Or parmi ces verbes suivis d'un infinitif, on distingue un continuum de trois classes ayant des degrés de verbalité : les modaux, les semi-modaux et les constructeurs. Un locuteur, bien sûr, ne fait pas de différences entre ces verbes. Il construit dans tous les cas une chaîne de verbes. Par ailleurs, même dans le cas des constructeurs (avec un verbe fini ou infinitif), la pronominalisation est difficile (*?on l'a décidé, d'habiter là ; ?tu ne le voudrais pas, qu'on te fasse ça*).

L'interprétation théorique est donc guidée par deux arguments majeurs : l'un négatif envers la représentation en enchâssement, l'autre positif envers le complexe verbal. L'extrême rareté de la récursivité ne favorise pas l'idée d'un enchâssement de construction accompagnée d'une extraction qui franchirait des frontières de constructions verbales. Inversement, la proximité topologique et syntaxique des deux verbes (problème de pronominalisation de V2 par V1) appuie l'hypothèse du constructeur complexe. Cette construction de « complexe verbal » peut être modélisée dans le cadre de la grammaire de dépendance admettant les arbres à bulle proposée par (Kahane, 2000). En syntaxe externe, la chaîne verbale forme une bulle. L'élément antéposé n'est donc plus un dépendant à distance mais un dépendant du constructeur verbal. La syntaxe interne conserve peut-être une certaine hiérarchie entre les deux verbes mais nous n'avons pas creusé cette hypothèse, la syntaxe interne étant totalement transparente dans les constructions de dépendance à distance.

Le constructeur complexe, de façon attendue, réagit aux paramètres interactionnels et on relève une spécialisation du lexique du verbe modal en fonction de la configuration syntaxique. Il peut être employé dans d'autres contextes syntaxiques, tels que le *sluicing*. Enfin, il peut donner lieu à des constructions plus figées, en particulier la formule rhétorique *qu'est-ce que tu veux que j'y fasse ?*.

Questions en suspens

L'analyse lexicale des verbes ponts reste à approfondir. Nous avons conclu que seuls les verbes complètement descriptifs d'une relation entre les actions (causatifs) ne pouvaient pas être ponts. Pour les verbes polysémiques ou en partie modaux, l'emploi semble possible. Mais est-ce l'interprétation modale du lexique qui les rend acceptables ou bien la construction de constructeur complexe qui donne au verbe une interprétation modale ?

L'examen des occurrences de dépendance à distance montre que le pont comporte un sujet et un verbe. Certains interprètent cette réduction comme une contrainte de performance : le pont est réduit pour faciliter le rattachement syntaxique et la conservation en mémoire. Pour nous, il est nécessaire que le pont ne soit pas développé car l'ajout de complément lui donnerait un sens descriptif (et pas seulement un poids référentiel et topologique). C'est une combinaison « sujet-verbe » qui est intégrée, et non une phrase. Il serait intéressant de clarifier la répartition entre compétence et performance, peut-être par des expériences.

Perspectives et prolongements

Pour confirmer l'interprétation de constructeur complexe, il faudrait d'abord réunir plus d'occurrences de dépendance à distance, et pour cela, réunir plus de corpus. Une analyse précise ne peut se faire qu'aux prix de nombreux exemples, relevés dans les situations variées. La standardisation et l'étiquetage des corpus existants devront être poursuivis. Le projet ANR ORFEO permettra de réaliser cette tâche.

Ensuite, il faudrait bien sûr compléter l'étude par l'examen des constructions « X CV1 CV2 » qui ont été écartées : les relatives en « dont » et les antépositions sans marque syntaxique explicite (*à Pierre, je trouve que tu ressembles*). Il serait également pertinent d'analyser les autres constructions dans lesquelles des chaînes verbales apparaissent (*sluicing*, comparatives) afin de comparer les propriétés du V1. Enfin, une étude sur les constructions avec ellipse de V2 (*il lui a proposé d'aller au cinéma mais elle voulait pas*) pourrait apporter d'autres informations, la négation, par exemple, ne semblant pas être une contrainte aussi importante que dans notre construction.

La modalisation par bulle est la plus adéquate à notre description. Les autres modèles proposent des insertions de noeuds « phrases » qui provoquent des sur-générations d'énoncés.

D’abord parce qu’ils autorisent indéfiniment la récursivité, et ensuite parce qu’ils ne limitent pas l’insertion aux éléments « sujet + verbe ». Par la représentation en bulle, les deux verbes sont clairement liés, mais finalement, celle-ci permet également la récursivité et le développement rectionnel du V1. Il serait peut-être intéressant d’intégrer le sujet (éventuellement en spécifiant même sujet pronominal) à la bulle. Ainsi, le groupe serait complètement dans la bulle.

Il serait également intéressant de réfléchir à une implémentation de cette représentation en chunk verbal. L’existence d’un chunk du complexe verbal faciliterait certainement l’analyse syntaxique automatique. Néanmoins, il est d’abord nécessaire d’avoir une idée claire de la syntaxe interne du complexe, afin de pouvoir donner toutes les informations concernant la construction.

ANNEXES

1.Typologie des verbes liés à la présupposition et à l'implication (Mørdrup, 1975) 297

1.1 Factifs 297

1.1.1 1er groupe : + indicatif (*Pierre sait que Marie a cessé de battre ses enfants*) 297

1.1.2 2ème groupe : + subjonctif 297

1.1.3 Cas spéciaux 298

1.2 Factifs négatifs 298

1.3 Semi-Factifs 298

1.4 Implicatifs positifs 299

1.4.1 1er groupe : + infinitif (sujet coréférentiel) ou « à ce que » 299

1.4.2 2ème groupe : pas co-référentiel 299

1.5 Implicatifs négatifs 300

1.5.1 1er groupe 300

1.5.2 2ème groupe 300

1.6 Implicatifs simples 300

1.7 Verbes assertifs 301

2. Liste des étiquettes de Cordial (tirée de la documentation du logiciel) 303

3. Occurrences écartées 307

4. Occurrences de dépendances à distance 311

5. Liste des verbes à complétives prise comme référence (tirée des corpus) 327

1. Typologie des verbes liés à la présupposition et à l'implication (Mørdrup, 1975)

1.1 Factifs

1.1.1 1er groupe : + indicatif (*Pierre sait que Marie a cessé de battre ses enfants*)

Cacher	Reconnaître	Ignorer
Dissimuler	Oublier	Savoir
Révéler	Se rappeler	
Méconnaître	Se souvenir	

1.1.2 2ème groupe : + subjonctif

Complétive = OD (ou OI)		Complétive = sujet réel		Complétive = OI ou sujet réel Verbes réfléchis	
Abuser	Se jouer	Arranger	Intéresser	S'affecter	S'honorer
Accepter	Se marrer	Attendrir	Intriguer	S'affliger	S'impatienter
S'accommoder	Se moquer	Blesser	Oppresser	S'agacer	S'indigner
Admirer	Pester	Bouleverser	Outrer ?	S'amuser	S'inquiéter
Applaudir	Se plaindre	Chiffonner	Peiner	S'attrister	S'irriter
Apprécier	Regretter	Compliquer	Refléter	Se chagriner	Se justifier
(Dés)Approuver	Remercier	Coûter	Refroidir	Se choquer	S'offenser
Bénir	Ressentir	Déconcerter	Renverser	Se consoler	S'offusquer
Blâmer	Souffrir	Dépasser	Reposer	Se contenter	Se réjouir
Critiquer	Supporter	Déranger	Séduire	S'effaroucher	Se satisfaire
Se dépiter	Trahir	Désarçonner	Simplifier	S'effrayer	Se scandaliser
Déplorer		Désobliger	Soulager	S'émerveiller	Se targuer ?
Endurer		Désoler	Stupéfier	S'enorgueillir	Se venger
Enrager		Embêter	Surprendre	S'épouvanter	
Envier		Empoisonner	Toucher	S'étonner	
Exulter		Ennuyer	Tourmenter	S'excuser	
Se féliciter		Epater	Tracasser	Se fâcher	
Jouir		Gâter	Troubler	Se ficher ?	
Jubiler		Gêner	Vexer	Se foutre ?	
Se lamenter		Impressionner		Se glorifier	

Accepter, envier, oublier, se moquer, s'amuser : deviennent implicatifs (positif ou négatif) quand ils sont suivis d'un infinitif.

1.1.3 Cas spéciaux

Complétive indirecte (de ce que) = circonstanciel [provenance, cause]		Infinitif :	
S'augmenter de Inférer de Jaillir de Naître de Provenir de Venir de	Aimer de Jalouser de Mépriser de Profiter de En vouloir à qqn de	**Infinitif passé** Pâtir de Se rengorger de Se repentir de Punir qqn de Récompenser qqn de **Infinitif présent :** Se borner à Se restreindre à	**Infinitif présent :** Se casser à S'épuiser à S'éreinter à S'exterminer à Se tuer à S'user à Se dépêcher de S'empresser de/à Se hâter de Se presser de Tarder à

1.2 Factifs négatifs

Affecter de
Feindre de
Simuler de
Se figurer
S'imaginer
Rêver

1.3 Semi-Factifs

Apercevoir
Apprendre
Concevoir ?
Constater
Comprendre ?
Découvrir
Deviner
Se douter
Entendre (sauf s'il exprime la volonté)
Entrevoir
Eprouver
Indiquer ?
Informer ?
Marquer ?
Montrer ?
Observer
Réaliser
Remarquer
(Se rendre compte)
Retenir
Sentir
Voir

1.4 Implicatifs positifs

1.4.1 1er groupe : + infinitif (sujet coréférentiel) ou « à ce que »

Cplmt faux avant le début de l'action	Cplmt vrai avant le début de l'action	Cplmt faux avant le début de l'action + On a tenté d'accomplir l'action avant le moment de la réussite	Pas de présupposition temporelle, Simultanéité de l'action de la principale et de celle du cplmt
S'abaisser à S'abandonner à Accepter de S'aventurer à Commencer à Condescendre à Consentir à Daigner Descendre à S'enhardir à Finir par Se hasarder à	S'acharner à S'adonner à S'attarder à ? Continuer à S'entêter à S'obstiner à Persévérer à Persister à	Aboutir à Arriver à Atteindre à Parvenir à Réussir à	S'amuser à Se consacrer à Se délecter à Se divertir à Se plaire à

1.4.2 2ème groupe : pas co-référentiel

(Ré)Accoutumer qqn à
Acculer qqn à
Assujettir qqn à
Astreindre qqn à
Conduire qqn à
Contraindre qqn à
Convaincre qqn de
Décider qqn à
Déterminer qqn à
Entraîner qqn à
Faire faire qqch à qqn
Forcer qqn à
(Ré)Habituer qqn à
Laisser
Obliger
Persuader qqn de
Plier qqn à
Porter qqn à
Pousser qqn à
Provoquer qqn à
Réduire qqn à

1.5 Implicatifs négatifs

1.5.1 1er groupe

Cplmt vrai avant le début de l'action +verbes imperfectifs	+verbes perfectifs	+verbes à sens ambigus : perfectif ou imperfectif ?	
Achever de S'arrêter de Cesser de (Ne pas) Démordre de Finir de S'interrompre de	Echouer à	S'abstenir de Ajourner de Dédaigner de Différer de Se dispenser de Epargner de Eviter de / que S'exempter de Faillir (à) Ne pas laisser de	Manquer à/de Négliger de Omettre de Oublier de Se passer de Se priver de Refuser de Renoncer à Se reposer de Résister à

1.5.2 2ème groupe

Décourager qqn de
Dégoûter qqn de
Désaccoutumer qqn de
Déshabituer qqn de
Détourner qqn de
Dissuader qqn de
Distraire qqn de ?
Empêcher qqn de / que
Garder qqn de
Guérir qqn de
Préserver qqn de
Rebuter qqn de
Retenir qqn de

1.6 Implicatifs simples

+ -> vrai	+ -> faux	- -> faux	- -> vrai
Démontrer que Prouver que Faire que Impliquer que Voir Regarder Entendre Ecouter	[rien]	Pouvoir Avoir l'occasion de Avoir la possibilité de Etre en mesure de Etre à même de Etre en état de Trouver l'occasion de	Balancer à Craindre de Détester (de) Hésiter à Rechigner à Reculer à ? Renâcler à ? Répugner à Contester que ? Soupçonner qqn

1.7 Verbes assertifs

accorder
accréditer
accuser
admettre
adorer
affirmer
aider
aimer
ajouter
alléguer
ambitionner
annoncer
appeler
s'appliquer
apporter
appréhender
s'apprêter
arguer
arrêter que
aspirer
assigner
assurer
s'attacher
attendre
attester
augurer
autoriser
avertir
aviser
avouer[25]
brûler
certifier
changer
charger
chercher
choisir
citer
commander
comploter
comporter?
compter
concéder
conclure
concourir
condamner
confesser[25]
confier
confirmer
conjurer
conseiller
considérer
consigner
consister
conspirer
convenir
convier
conter
convoiter
crier
croire
décider
déclarer
déconseiller
décréter
déduire
défendre
défier
délibérer
demander
démentir
déposer
désespérer
désirer
destiner
devoir
se dévouer
dire
discipliner
disposer
divulguer
donner
s'écrier
écrire
s'efforcer
s'employer
encourager
engager
enjoindre
enseigner
entreprendre
envisager
équivaloir
s'escrimer
espérer
essayer
estimer
établir
s'évertuer
exceller
exciter
exclure
s'excuser
exercer
exhorter
exiger
expliquer
s'exposer
se flatter
former
gager
gagner
garantir
geindre
gémir
glisser
grogner
grommeler
gueuler
habiliter
haïr
hurler
imaginer
imposer
inaugurer
inciter
incliner
induire
inférer
s'ingénier
insinuer
inspirer
instruire
interdire
intimer
inviter
juger
jurer
se lasser
lire
méditer
menacer
mentionner
mériter
mettre
narrer
nécessiter
nier
noter
occuper
offrir
s'opposer
ordonner
organiser
paraître
parier
parler
pencher
penser
permettre
se piquer
porter
poser
postuler
pourvoir
prêcher
préciser
préconiser
prédire
prédisposer
préférer
préméditer
préparer
présager
prescrire
presser
présumer
présupposer
prétendre
se prêter
prétexter
prévenir
prévoir
prier
proclamer
projeter
promettre
proposer
protester
publier
raconter
rapporter
réclamer
redire
redouter
réfléchir
régler
remontrer
répandre
répéter
répondre
représenter
reprocher
requérir
réserver
résoudre
rétorquer
ricaner
riposter
risquer
ruminer
sembler
signaler
signifier
solliciter
sommer
songer
se soucier de
souhaiter
souligner
soupirer
soutenir
spécifier
stimuler
stipuler
suffire
suggérer
supplier
supposer
tâcher
télégraphier
téléphoner
témoigner
tendre
tenir
tenter
tolérer
trancher
travailler
trembler
trouver
valoir
se vanter
veiller
vérifier
viser
vouloir

25: ces verbes peuvent prendre le sens de *reconnaître*. Dans ce cas, ils sont factifs.
26. *signifier* peut prendre le sens d'un verbe-SI.

2. Liste des étiquettes de Cordial (tirée de la documentation du logiciel)

0. ADJMS Adjectif Masculin Singulier
1. ADJMP Adjectif Masculin Pluriel
2. ADJFS Adjectif Féminin Singulier
3. ADJFP Adjectif Féminin Pluriel
4. ADJMIN Adjectif Masculin Invariant en Nombre
5. ADJFIN Adjectif Féminin Invariant en Nombre
6. ADJSIG Adjectif Singulier Invariant en Genre
7. ADJPIG Adjectif Pluriel Invariant en Genre
8. ADJI Adjectif Invariant en Nombre et en Genre
9. DETDEM Adjectif Démonstratif
10. DETPOSS Adjectif Possessif
11. ADJNUM Adjectif Numérique Cardinal
12. ADJORD Adjectif Numérique Ordinal
13. ADV Adverbe
14. DETDMS Article Défini Masculin Singulier
15. DETDFS Article Défini Féminin Singulier
16. DETDPIG Article Défini Pluriel Invariant en Genre
17. DETIMS Article Indéfini Masculin Singulier
18. DETIFS Article Indéfini Féminin Singulier
19. AHMS Adjectif Masculin Singulier débutant par un "h" aspiré
20. COO Conjonction de Coordination
21. SUB Conjonction de Subordination
22. INT Interjection
23. PREP Préposition
24. NCMS Nom commun Masculin Singulier
25. NCMP Nom commun Masculin Pluriel
26. NCFS Nom commun Féminin Singulier
27. NCFP Nom commun Féminin Pluriel
28. NCHMS Nom commun Masculin Singulier débutant par un "h" aspiré
29. NCHFS Nom commun Féminin Singulier débutant par un "h" aspiré
30. NCHSIG Nom commun Singulier Invariant en Genre débutant par un "h" aspiré
31. NCMIN Nom commun Masculin Invariant en Nombre
32. NCFIN Nom commun Féminin Invariant en Nombre
33. NCSIG Nom commun Singulier Invariant en Genre
34. NCPIG Nom commun Pluriel Invariant en Genre
35. NCI Nom commun Invariant en Nombre et en Genre
36. PPER1S Pronom Personnel 1 e Personne du Singulier
37. PPER2S Pronom Personnel 2 e Personne du Singulier
38. PPER3S Pronom Personnel 3 e Personne du Singulier
39. PPER1P Pronom Personnel 1 e Personne du Pluriel
40. PPER2P Pronom Personnel 2 e Personne du Pluriel

41. PPER3P Pronom Personnel 3 e Personne du Pluriel
42. PDS Pronom Démonstratif Singulier
43. PDP Pronom Démonstratif Pluriel
44. AINT Adjectif Interrogatif
45. AIND Adjectif Indéfini
46. NCHMP Nom Masculin Pluriel débutant par un "h" aspiré
47. PRMS Pronom Relatif Masculin Singulier
48. PRFS Pronom Relatif Féminin Singulier
49. AHFS Adjectif Féminin Singulier débutant par un "h" aspiré
50. AHSIG Adjectif Invariant en Genre débutant par un "h" aspiré
51. PRI Pronom Relatif Invariant en Genre et en Nombre
52. PIMS Pronom Indéfini Masculin Singulier
53. PIFS Pronom Indéfini Féminin Singulier
54. PIMP Pronom Indéfini Masculin Pluriel
55. PIFP Pronom Indéfini Féminin Pluriel
56. PISIG Pronom Indéfini Singulier Invariant en Genre
57. PIPIG Pronom Indéfini Pluriel Invariant en Genre
58. PII Pronom Indéfini Invariant en Genre et en Nombre
59. PP Pronom Possessif
60. NCHMIN Nom Invariant en Nombre débutant par un "h" aspiré.

64. NPMS Nom Propre Masculin Singulier
65. NPMP Nom Propre Masculin Pluriel
66. NPFS Nom Propre Féminin Singulier
67. NPFP Nom Propre Féminin Pluriel
68. NPHMS Nom Propre Masculin Singulier débutant par un "h" aspiré
69. NPHFS Nom Propre Féminin Singulier débutant par un "h" aspiré
70. NPHSIG Nom Propre Singulier Invariant en Genre débutant par un "h" aspiré
71. NPMIN Nom Propre Masculin Invariant en Nombre
72. NPFIN Nom Propre Féminin Invariant en Nombre
73. NPSIG Nom Propre Singulier Invariant en Genre
74. NPPIG Nom Propre Pluriel Invariant en Genre
75. NPI Nom Propre Invariant en Nombre et en Genre

100. VINF Verbe à l'infinitif
101. VINDP1S Verbe indicatif présent 1ère personne du singulier
102. VINDP2S Verbe indicatif présent 2e personne du singulier
103. VINDP3S Verbe indicatif présent 3e personne du singulier
104. VINDP1P Verbe indicatif présent 1ère personne du pluriel
105. VINDP2P Verbe indicatif présent 2e personne du pluriel
106. VINDP3P Verbe indicatif présent 3e personne du pluriel
107. VIMPP1S Verbe indicatif imparfait 1ère personne du singulier
108. VIMPP2S Verbe indicatif imparfait 2e personne du singulier
109. VIMPP3S Verbe indicatif imparfait 3e personne du singulier
110. VIMPP1P Verbe indicatif imparfait 1ère personne du pluriel
111. VIMPP2P Verbe indicatif imparfait 2e personne du pluriel
112. VIMPP3P Verbe indicatif imparfait 3e personne du pluriel
113. VINDPS1S Verbe indicatif passé simple 1ère personne du singulier
114. VINDPS2S Verbe indicatif passé simple 2e personne du singulier
115. VINDPS3S Verbe indicatif passé simple 3e personne du singulier

116. VINDPS1P Verbe indicatif passé simple 1ère personne du pluriel
117. VINDPS2P Verbe indicatif passé simple 2e personne du pluriel
118. VINDPS3P Verbe indicatif passé simple 3e personne du pluriel
119. VINDF1S Verbe indicatif futur simple 1ère personne du singulier
120. VINDF2S Verbe indicatif futur simple 2e personne du singulier
121. VINDF3S Verbe indicatif futur simple 3e personne du singulier
122. VINDF1P Verbe indicatif futur simple 1ère personne du pluriel
123. VINDF2P Verbe indicatif futur simple 2e personne du pluriel
124. VINDF3P Verbe indicatif futur simple 3e personne du pluriel
125. VSUBP1S Verbe subjonctif présent 1ère personne du singulier
126. VSUBP2S Verbe subjonctif présent 2e personne du singulier
127. VSUBP3S Verbe subjonctif présent 3e personne du singulier
128. VSUBP1P Verbe subjonctif présent 1ère personne du pluriel
129. VSUBP2P Verbe subjonctif présent 2e personne du pluriel
130. VSUBP3P Verbe subjonctif présent 3e personne du pluriel
131. VSUBI1S Verbe subjonctif imparfait 1ère personne du singulier
132. VSUBI2S Verbe subjonctif imparfait 2e personne du singulier
133. VSUBI3S Verbe subjonctif imparfait 3e personne du singulier
134. VSUBI1P Verbe subjonctif imparfait 1ère personne du pluriel
135. VSUBI2P Verbe subjonctif imparfait 2e personne du pluriel
136. VSUBI3P Verbe subjonctif imparfait 3e personne du pluriel
137. VCONDP1S Verbe indicatif conditionnel 1ère personne du singulier
138. VCONDP2S Verbe indicatif conditionnel 2e personne du singulier
139. VCONDP3S Verbe indicatif conditionnel 3e personne du singulier
140. VCONDP1P Verbe indicatif conditionnel 1ère personne du pluriel
141. VCONDP2P Verbe indicatif conditionnel 2e personne du pluriel
142. VCONDP3P Verbe indicatif conditionnel 3e personne du pluriel
144. VIMPP2S Verbe impératif présent 2e personne du singulier
146. VIMPP1P Verbe impératif présent 1ère personne du pluriel
147. VIMPP2S Verbe impératif présent 2e personne du pluriel

149. VPARPRES Verbe participe présent
150. VPARPMS Verbe participe passé masculin singulier
151. VPARPFS Verbe participe passé féminin singulier
152. VPARPMP Verbe participe passé masculin pluriel
153. VPARPFP Verbe participe passé féminin pluriel

201. PCTFAIB Ponctuation faible
209. PCTFORTE Ponctuation forte

3. Occurrences écartées

Occurrences de littérature ancienne et de francophonie du CERF (écrit)

Tranche	Contexte gauche	pont	Contexte droit
LITA	Ah ! je l'avoue, plus d'une fois je fus tenté de me prendre pour un héros, et il m'est arrivé de gémir sur la désolante petitesse de ce globe	où Juvénal témoigne qu	Alexandre le Grand ne trouvait pas le moyen de respirer. Mon Dieu ! oui, j'eusse aimé à escalader le firmament, à négocier dans la Lune,
LITA	" René, " répondit-elle, " que veux-tu faire ? Laboure ton champ. - Il ne peut pas nous nourrir ; laissez-moi partir. - Eh bien, " dit la mère, " va donc	où Dieu veut que	tu ailles. " Elle embrassa l'enfant en sanglotant. Le soir même mon père quitta la ferme maternelle,
LITA	il a paru fort inquiet, m'a demandé où était cette lettre, et est descendu immédiatement à la cuisine	où il savait qu	'était votre pourpoint.
LITA	En arrivant à la hutte, je frappai selon mon habitude, et, ne recevant pas de réponse, je cherchai la clef	où je savais qu	'elle était cachée, j'ouvris la porte et j'entrai. Un beau feu flambait dans le foyer.
LITA	Et après, je repartirai pour les pays du Barrage,	où je pense que	vous ne tarderez point à paraître.
LITA	Aussi renonça-t-il à l'idée de promener le noble Anglais dans des hypogées déjà cent fois parcourus, et dédaigna-t-il de lui faire entreprendre des fouilles à des endroits	où il savait qu	on ne trouverait rien, pour en avoir extrait lui-même depuis longtemps et vendu fort cher ce qu'il y avait de curieux.
LITA	tandis que l'orage éclatant après une journée de vide, de tristesse, je rentrai le front bas, comme il convenait. Je revins chez nous un peu après l'heure	où je savais que	mon père avait coutume d'y être. Il " savait " donc. Je me promenai dans le jardin, attendant que mon père me fît venir.
LITA	DUCHOTEL (même jeu) : Je ne les ai pas quittés... (Calme.) Donc, tu en conviens, tu vas trouver madame Chardet, là	où tu sais qu	'habite madame Chardet et madame de Fontenac, là où tu sais qu'habite madame de Fontenac ?
LITA	DUCHOTEL (même jeu) : Je ne les ai pas quittés... (Calme.) Donc, tu en conviens, tu vas trouver madame Chardet, là où tu sais qu'habite madame Chardet et madame de Fontenac, là	où tu sais qu	'habite madame de Fontenac ?
LITA	Nonobstant, nous vous invitons par la présente à venir déjeuner (midi ou 1h) vendredi de cette semaine, jour de votre fête	où il est dit que	la Sainte Vierge s'est trompée.
LITA	les soirs où des étrangers, ou seulement M. Swann, étaient là, maman ne montait pas dans ma chambre. Je dînais avant tout le monde et je venais ensuite m'asseoir à table, jusqu'à huit heures	où il était convenu que	je devais monter ; ce baiser précieux et fragile que maman me confiait d'habitude dans mon lit au moment de m'endormir, il me fallait le transporter de la salle à manger dans ma chambre et le garder
LITA	Ils firent ce repas frugal remplis de joie par le souvenir de la bonne action qu'ils avaient faite le matin ; mais cette joie était troublée par l'inquiétude	où ils se doutaient bien que	leur longue absence de la maison jetterait leurs mères. Virginie revenait souvent sur cet objet ; cependant Paul, qui sentait ses forces rétablies, l'assura qu'ils ne tarderaient pas à tranquilliser

LITA	Le père Ubu ne se tient plus debout, et j'ai rendez-vous chez lui aujourd'hui à 10 heures, avec le docteur Saltas, pour tenter de le décider (il a déjà accepté d'ailleurs) à entrer à la Charité - d'	où il est extrêmement probable qu	il ne sortira pas vivant.
LITA	Mon Edouard, Voici encore que nous allons changer de domicile : nous quittons Morden Road samedi pour aller nous installer à Londres,	où il va être indispensable que	je sois pour mon travail et mes affaires. Il va falloir se remettre à l'oeuvre et à la vie utile,
LITA	qu'à glisser leurs bras sous son dos bombé, à le détacher de la gangue du lit, à se baisser avec leur fardeau, et ensuite uniquement à le laisser avec précaution opérer son rétablissement sur le sol,	où dès lors on pouvait espérer que	les petites pattes auraient enfin un sens. Mais, sans compter que les portes étaient fermées à clé, aurait-il vraiment fallu appeler à l'aide ?
LITA	Bref, au bout de quelques semaines, nous étions intimes et nous emmenions M. de Sivrac à notre terre de la Creuse,	où mon mari lui avait persuadé qu	'ils reviseraient ensemble ses manuscrits.
LITA	Les jours	où il semble que	tout le monde ait envie de mourir
LITA	parfois un vers (…) vient rouvrir Quelque plaie	où le feu désire qu	'on l'attise ;
LITA	Mais au moins ai-je toujours regardé le jour de Noël, quand il est revenu (…), comme un beau jour, un jour de bienveillance, de pardon, de charité, de plaisir, le seul, dans le long calendrier de l'année,	où je sache que	tous, hommes et femmes, semblent, par un consentement unanime, ouvrir librement les secrets de leurs coeurs et voir dans les gens au-dessous d'eux de vrais compagnons de voyage sur le chemin du tombe
LITA	-quel est donc, reprit l'ami de mon père, ce Jean-François les bas-bleus qui annonce les événements à cent lieues de distance, au moment	où il suppose qu	ils doivent s'accomplir ? Un somnambule, un convulsionnaire, un élève de Mesmer ou de Cagliostro ?
LITA	- Nous avons trop parlé, dit précipitamment Passepoil, et nous n'avons rien dit. Nevers et sa botte secrète valent de l'or, mes compagnons, voilà ce	qu'il faut que	vous sachiez. Avez-vous envie de faire d'un coup votre fortune ?
LITA	Ce n'est pas mal du tout ça, dit le soldat. Mais	qu'est-ce qu'il faudra que	je te donne à toi, la vieille ? je suppose que tu veux quelque chose.
LITA	Un jour qu'il y avait foire à la ville, le paysan décida d'y aller, et avant de partir il demanda à sa femme ce	qu'elle voulait qu	'il lui rapporte. " Un peu de viande, lui dit-elle, et quelques brioches ; enfin, tu sais bien ce qu'il faut pour la maison. "
LITA	Mais le Point, à son sens, celui	qu'il faut qu	on voie Et croie, est ceci dont elle la complimente : Le libre arbitre pèse, arguë et parlemente, Puis le pauvre-de-coeur décide et suit sa voie.
LITA	et qui n'employait point ses loisirs à jouer à la boule ou à la paume, comme bien d'autres, mais à faire des comptes, à lire le Bocage des six corporations, et à apprendre un peu d'espagnol,	qu'il était bon qu	un marchand sût parler, comme aujourd'hui l'anglais, a cause de la quantité de personnes de cette nation qui habitaient dans Paris.
LITA	Gérard de Nerval, Petrus Borel, Théophile Gautier, Alphonse Brot, Augustus Mac-Keat et Philothée O'Neddy. - Gérard de Nerval avait publié sous la Restauration des poésies nationales et napoléoniennes,	qu'il ne voulait pas qu	'on lût, déclarant tout le premier que c'était du poncif. Il ne montrait rien ou presque rien des excellentes choses qu'il préparait alors.
LITA	Bien de n'être pas dupe en ce monde d'une heure, Mais pour ne l'être pas durant l'éternité, Ce	qu'il faut à tout prix qui	règne et qui demeure, Ce n'est pas la méchanceté, c'est la bonté.
LITA	Qu'est-ce que vous allez lui dire ? demanda le chancelier à M. Hippolyte Patard. M. le secrétaire perpétuel, dont la tête se troublait un peu à la suite de toutes ces ridicules histoires, répondit vaguement :	Qu'est-ce que vous voulez que	je lui dise ? ... Je lui dirai : "Du courage, mon ami... "

LITA	Lecteur, c'est peut-être la haine	que tu veux que	j'invoque dans le commencement de cet ouvrage !
LITA	Et pour	qui voulez-vous qu'	à présent je la brode
LITA	où il n'y avait de lumière qu'à une seule fenêtre... cette fenêtre que précisément j'ai emportée dans ma mémoire et que j'ai là, toujours, sous le front ! ... La maison, dans	laquelle on ne pouvait pas dire que	cette lumière brillait, car elle était tamisée par un double rideau cramoisi dont elle traversait mystérieusement l'épaisseur, était une grande maison qui n'avait qu'un étage, - mais placé très haut.
LITA	C'est peu d'être homme ; nous sommes encore bien près du singe,	dont on dit que	nous procédons. Soit ; raison de plus pour nous éloigner de lui et pour être au moins à la hauteur du vrai relatif que notre race a été admise à comprendre ;
LITA	Mon cher ami, j'ai soif d'aimer sérieusement, de me donner et de posséder ;	pourquoi faut-il que	j'aie trouvé l'écueil sous la mer calme ? Je tremble d'être à Caen ;
LITA	" Dis-moi, Henry, demanda-t-il soudainement,	combien de chiens prétends-tu que	nous avons ?
LITA	-et la matière des astres et des mondes,	que penses-tu qu'	elle soit auprès de la matière subtile qui remplit l'espace ?
LITA	-	que veux-tu que	je te réponde ? ... il n'y a point d'expression possible pour comparer des corps si grossiers à un
LITA	MONSIEUR. - Mais, nom d'une pipe !	que veux-tu donc que	je te réponde ?
LITA	Marie Immaculée, amour essentiel, Logique de la foi cordiale et vivace, En vous aimant	qu'est-il bon que	je ne fasse, En vous aimant du seul amour, Porte du ciel ?
LITA	Sous la glycine et le cytise, Tous deux seuls,	que faut-il qu'	on dise ?
LITA	"Dis donc, Joseph,	combien penses-tu que	nous en avons vu passer depuis 1804 ?
LITA	et j'irais le voir, poursuit-il, tenant toujours dans sa main les deux écus. Eustache,	que penses-tu qu'	il arriverait si, selon le voeu du marchand, la justice était juste ?
LITA	(haut.) maintenant, dites-moi un peu,	que croyez-vous que	vous donne votre oncle ?
LITA	Sabre à finances, évidemment !	que veux-tu qu'	il me fasse, ce petit sagouin de quatorze ans ?
LITA	au point où nous voilà,	qui veux-tu qui	conspire ?
LITA	Eh bien, après,	que veux-tu que	j'y fasse ? ce n'est pas moi qui le leur ai dit. Cependant, Messieurs des Finances, préparons-nous
LITC	Rosalie :	Qu'est-ce que tu veux que	je te dise ?
LITC	Émile :	Qu'est-ce que tu veux que	je fasse, y'est deux fois plus gros que moi.
POLI	Max van den Berg, Pays-Bas : Nous devons expliquer	pourquoi nous pensons que	la politique étrangère, les politiques liées à l'immigration, à la justice doivent être harmonisées.
PRES	Et, le lièvre sortit fièrement de sa cachette : "Je vous l'avais dit ;	comment voulez-vous qu'	un candidat incapable de se retenir puisse valablement défendre vos intérêts ? "
PRES	compliqué de savoir si une plante est modifiée ou non. En France, on milite pour l'affichage. Mais	comment voulez-vous que	le commerçant qui vend des corn-flakes sache si le producteur a mis dans sa boîte du maïs modifié

4. Occurrences de dépendances à distance

		Contexte gauche	pont	Contexte droit
CERF	DIVS	Nous apprenons, chaque jour et chaque minute, que le Messie nous a précédés sur le chemin étroit	où Il veut que	ses disciples marchent. Il nous en donnera la force. Ainsi nous serons les fils de notre Père qui est dans les cieux, Lui qui fait lever son soleil sur les méchants et sur les bons, Lui qui fait pleu
CERF	DIVS	je ne me contente pas d'une première réponse, je dois aller plus loin. Je sais que je n'aurai jamais toutes les réponses à mes questions, je crois que ce n'est pas nécessaire non plus, mais là	où je sais que	je peux avoir une réponse, je veux bien savoir cette réponse. Aussi loin que je me souvienne j'ai toujours analysé, maintenant je le fais beaucoup mieux, je ne me pose plus de questions existentielle
CERF	DIVS	ssentie comme quelque chose dont on m'avait privée pendant longtemps... c'était comme si tout un nouveau monde s'ouvrait devant moi ! Comme s'il y avait toute une partie de moi-même que j'ignorais et	qu'il fallait que	j'apprenne à connaître ! Car, par opposition à la musique classique où tu joues tout ce qu'il y a d'écrit devant toi, dans le jazz il s'agit d'exprimer tout ce qui se trouve en toi.
CERF	DIVS		Combien vous croyez qu	il y a de gens qui travaillent a la sécurité sociale ?
CERF	DIVS	exprime notamment dans les "Cinq préceptes" de base, s'appuira donc sur un principe universel : tous les hommes préférant le bonheur au malheur, le plaisir à la souffrance, ne fais pas aux autres ce	que tu ne voudrais pas qu	on te fasse à toi-même.
CERF	DIVS	llait être de loin plus influente avait été publié en 1974, L'Archipel du goulag de Soljenitsyne. Le contraste entre les deux aurait difficilement pu être plus grand, et mit en sourdine tout l'impact	que Foucault pensait que	son travail devait avoir en France. Devant ce récit irrésistible de la torture physique et mentale, menée par un régime que beaucoup en France considéraient toujours comme l'avant-garde du progrès so
CERF	DIVS	! ! ...	Qu'est-ce que vous voulez que	j'aille faire à Sète ?
CERF	DIVS	d tirage ça hein ! Très bien pour allumer le feu donc ! Vous avez ici " Le Pape est mort " et alors en bas, une publicité malencontreuse : " Grande Braderie au marché Saint-Pierre. " Enfin,	qu'est-ce que vous voulez que	je vous dise ? Tiens, garde-le, parce que l'autre est pas mort et on peut encore se faire du blé avec celui-là...
CERF	DIVS		Qu'est-ce que vous voulez qu	'j'lui dise ?
CERF	DIVS	LE PÈRE - Fous-moi la paix !	Qu'est-ce que tu veux que	je fasse, moi, avec les atomes, les rnolécules... Pose-moi des questions qui n'aient pas trait aux atomes ni aux molécules et je te répondrai.
CERF	DIVS	Concernant ses disciples Jésus enseigne : "Tout ce	que vous voudriez que	les hommes fassent pour vous, faites-le de même pour eux : c'est là tout l'enseignement de la Loi et des Prophètes", et encore : "... Entrez par la porte étroite... gardez-vous des faux prophètes,
CERF	DIVS	Qui qu'à cassé le cocotier Qui qu'à caché les allumettes Qui qu'à pas fait sa gymnastique Qui qu'à pris la pipe à pépé C'est Alexandre C'est encore Alexandre C'est toujours Alexandre	Qui voulez-vous que	ce soit Qui qu'est tombé dans les égouts Qui qu'à fait tomber le facteur Qui qu'à mangé les cornichons Et qui c'est qu'à mordu Mirza C'est Alexandre C'est encore Alexandre C'est toujours Alexandre Qui voulez-vous q

CERF	DIVS	Qui qu'est tombé dans les égouts Qui qu'à fait tomber le facteur Qui qu'à mangé les cornichons Et qui c'est qu'à mordu Mirza C'est Alexandre C'est encore Alexandre C'est toujours Alexandre	Qui voulez-vous que	ce soit A - LE - XAN - DRE Ouah Qui qu'habite la rue Quinquempois Qui qu'à bu mon sirop d'orgeat Qui qui s'est fait un coupe-papier Avec l'épée d'Ivanhoé C'est Alexandre C'est encore Alexandre C'est
CERF	DIVS	AN - DRE Ouah Qui qu'habite la rue Quinquempois Qui qu'à bu mon sirop d'orgeat Qui qui s'est fait un coupe-papier Avec l'épée d'Ivanhoé C'est Alexandre C'est encore Alexandre C'est toujours Alexandre	Qui voulez-vous que	ce soit Ça ne peut être qu'Alexandre C'est toujours Alexandre C'est encore Alexandre C'est peut-être Alexandre Alexandre Alexandre Alexandre Ridicule Mais qu'est ce que ça veut dire Hein Assez !
CERF	DIVS	nsible résurgence peut s'expliquer ainsi précisément qu'il peut en appeler à un savoir autre qu'un savoir sur le cinéma, prenant appui sur des faits réels, ou en "référence à des arts antérieurs, sur	lesquels il est supposé qu'	une connaissance, ou au moins certaines habitudes de pensée, sont acquises. "La localisation d'un point d'ignorance suppose maintenant au préalable la reconstruction du contexte en termes de savoirs.
CERF	DIVS	Le texte (incluant l'appropriation qui en est faite par les acteurs) est soumis à trois contraintes : 1. - celle du bruit que l'artifice rend plus prégnant, bien qu'il s'agisse toujours du son direct, et	dont il faut que	la voix émerge (il y a un effet de lutte de la parole, et divers degrés d'effort pour s'arracher au bruit).
CERF	DIVS	Mais lorsque le croyant prend conscience d'Allah, de la raison pour laquelle Allah l'a créé et de la façon	dont Il veut que	nous nous servions de ce monde, lorsque le croyant prend réellement conscience de la profondeur de l'Islam et de sa raison d'être, alors, comme à ses premiers jours, passant d'un monde à l'autre, il
CERF	DIVS	C'est en cela	qu'on dit qu	'il use de "moyens habiles". Chacun interprètera (en fonction de ses propres "constructions mentales") ses actes et son enseignement à la mesure de son propre degré d'avancement sur la Voie, de son il
CERF	DIVS	Quand à la Saint Médard il tombe de la pluie, de la neige, de la grêle, des hallebardes et de la suie, on est tranquille pour quarante jours plus tard, parce que, à part tout ça,	qu'est-ce que vous voulez qu	'il tombe ? Oui, je sais, mais enfin c'est plutôt rare.
CERF	DIVS	LE PÈRE : T'en as pour longtemps à me soûler avec tes salades, dis, hein ?	Qu'est-ce que tu veux que	je sache, moi, les atomes et les molécules comment qu'ça fonctionne ?
CERF	DIVS	déjà entreprises, parce que j'envisageais cette possibilité plus tôt que prévu je me suis jugée. Ce n'est déjà pas facile si je commence à me juger je ne suis pas sortie du bois. Alors vous comprenez	pourquoi je vous dis que	je ne dois pas trop anticipé. Même si il va en répit régulièrement, je sais qu'il revient ici, alors qu'un placement il n'y a pas de retour, sauf si il vient pour une fin de semaine. Je crois que si
CERF	DIVS	J'aimerais savoir	pourquoi vous dites que	la musique guérit ?
CERF	DIVS	A quoi ça sert l'amour Si c'est un aller sans retour Y a plus que du vide à la place Mais	que veux-tu que	j'en fasse ? A quoi ça sert la vie Quand on meurt petit à petit S'il ne reste plus que l'absence A
CERF	DIVS	j'en fasse ? A quoi ça sert la vie Quand on meurt petit à petit S'il ne reste plus que l'absence A	qui veux-tu que	je pense ? Je jure si tu reviens Je saurai être heureux d'un sourire Je jure que j'apprendrai à écr
CERF	DIVS	our de plus est un jour de trop Est-ce que tu reviendras bientôt ? Il ne reste plus que l'absence A	qui veux-tu que	je pense ? Chaque jour de plus est un jour de trop Je plie déjà sous le fardeau Chaque jour de plus
CERF	DIVS	Dans la suite du film, Clo-Clo mourra, victime de sa séduction, sous les coups d'un obsédé sexuel (	pourquoi diable fallait-il qu'	elle se mette du rouge à lèvres dans une rue déserte en pleine nuit ?) non sans avoir transmis aupar
CERF	DIVS	de poing dans la gueule, des coups de pied dans les couilles et des coups de genou dans les fesses,	comment voulez-vous qu'	on interroge ? Des fois, ils ne parlent même pas notre langue.

CERF	DIVS	La véritable modestie consiste à ne jamais se prendre pour moins que ce	qu'on estime qu'on croit qu'	on vaut
CERF	FORM	en revisionnant les cassettes de la perte de leur belle apparence physique plutôt que des stupidités allignées pour la galerie qu'ils débitent à longueur de journée) mais bon c'est leur probléme. Là	où je trouve que	tu t'égares quaybir c'est quand tu vois une évolution de la langue dans le sabir des lofteurs. Moi j'y vois plutôt une régression, une sorte de degré zéro (comme dirait barthes pas fabien, l'autre) d
CERF	FORM	Dans la bourse tout est affaire de psychologie et le cours de l'action ne représente pas tant la valeur intrinséque des actifs et passifs de la société que le prix	qu'on pense que	les autres sont prêts à payer. Un peu comme l'immobilier d'ailleurs sinon comment expliquer que le même appartement dont le coût de construction est similaire se vende 4 fois plus cher à Paris qu'à G
CERF	FORM	te il n'en demeure pas moins en effet qu'elle est envahissante. L'Europe d'ailleurs hormis ses langues (et encore même pas sûre !) et son histoire ressemble de plus en plus aux States c'est pour celà	qu'il me semble que	c'est une cause juste de défendre la "diversité culturelle" d'un pays afin qu'il ne perde pas son identité !
CERF	FORM	Paris est une cité, on s'y entasse également, c'est pour cela	qu'il faut bien comprendre qu	en mai 68, le beans n'était pas tant une question de confort au sens qu'on lui connaît.
CERF	FORM	En conclusion, il semble que ce	que les femmes veulent qu	on leur dise, c'est "L'IVG quand vous voudrez, comme vous voudrez, où vous voudrez, et sans compte à rendre à personne ! "
CERF	FORM	es comédiens, toutes les mises en scène, c'est comme pour tout, sacro-saint cinéma compris), il faut sûrement avoir un imaginaire très peu développé. Quel dommage ! Bien sûr, le ciné (du moins, celui	dont je présume que	tu parles), c'est facile, on t'apporte tout sur un plateau, à renfort d'effets spéciaux made in hollywood. Génial !
CERF	FORM	st une question moi que je pose ! On débat à partir d'un discours de Messier qui a peut-être été interprété hâtivement sur une notion d'exception culturelle Française dont je cherche la définition...	qu'est ce que vous voulez que	j'argumente ?
CERF	FORM	émocraties ne sont pas si democratiques que cela et sans doute ne le seront jamais... Quand vous allez chez le coiffeur, vous espérez tous que ce sera un professionnel qui vous coupera les cheveux...	pourquoi voudriez vous ensuite que	ce soit des citoyens, la plupart du temps desinvestis politiquement, qui décident de l'avenir d'une nation ? ! ; -)
CERF	FORM	le "remet les pendules a l'heure". La, je trouve qu'il y a veritablement une vision de societe globale sous-tendant cette affirmation. Et je le trouve paradoxale. Tout d'abord, je n'ai jamais compris	pourquoi il faudrait que	l'Etat se serve (grassement, d'ailleurs) sur la transmission du patrimoine. C'est mon droit d'accumuler de la richesse non pour moi uniquement, mais pour la transmettre a la generation suivante. Cett
CERF	FORM	l'état hébreu a réussi à imposer une démilitarisation. Tout cela n'a pu être imposé au Liban, alors	que voulez-vous que	le peuple palestinien comprenne, n'a-t-il toutes les raisons du monde pour que les pierres de 2000
CERF	FORM	Les politiques sont sur une autre planète. FILLON hier à la télé disait : mais	où croyez-vous que	nous vivions, nous connaissons les pb que vous soulevez.
CERF	FORM	onde, les médisants parlaient de chance. Après l'Euro, les mêmes trouvent cette équipe banale. Mais	que faut-il qu'	elle fasse : gagner pendant 10 ans toutes les coupes ? Gagner tous ses matches 10-0 ?
CERF	FORM	Mais	comment veux-tu que	je puisse te suivre dans ton point de vue si tu ne me proposes pas un seul exemple concret de l'amé
CERF	INST	Ça, c'est des éléments, en tout cas,	qu'on pense que	le gouvernement n'a pas évalués pour déterminer effectivement les impacts ou les conséquences que cela pourrait avoir de transférer de facto et de façon unilatérale les employés sous la responsabilit

CERF	INST	ce en augmentant le nombre de percepteurs, mais de ne pas créer une classe nouvelle de sous-percepteurs à l'intérieur du système de justice. À moins qu'il y ait des raisons que je n'ai pas perçues et	qu'il faudrait réellement qu	on m'explique, là, mais, dans la manière dont, moi, j'ai lu le projet de loi, j'ai en quelque sorte été heurté par cet article, qui est l'article 42, et je voyais, au contraire, qu'il aurait été plus
CERF	INST	Il s'agit naturellement d'une charge nouvelle et lourde pour l'Etat. C'est pour en diminuer l'importance - et non pour se livrer à un quelconque " tour de passe-passe budgétaire " -	qu'il est prévu que	France Télécom versera une contribution exceptionnelle à l'État. Cette contribution ne compensera que partiellement la charge incombant désormais à l'État. Son montant sera fixé en 1996 en loi de fin
CERF	INST	nos consommateurs sans garantie que ce soient des mesures pertinentes, valables, qui donnent toute l'information justifiée et souhaitée par la population ? Ce sont là des questions, M. le Président,	qu'il semble très important qu	un gouvernement responsable doit se poser et surtout auxquelles nous devons avoir les réponses afin de prendre une décision tout à fait éclairée pour l'ensemble des personnes touchées par une décisio
CERF	INST	ns étaient prononcées dans le corps des adjoints d'enseignement sans concours, mais sur liste d'aptitude. Ce corps a été mis en voie d'extinction. La seule solution d'accueil pour la titularisation à	laquelle je rappelle qu'	il était déjà procédé précédemment - est l'entrée dans le corps des professeurs certifiés.
CERF	INST	ES POUR LA COMMISSION 27. En application du paragraphe 11 de la résolution 53/102 de l'Assemblée générale en date du 8 décembre 1998, la Commission indique ci-après, pour chaque sujet, les points sur	lesquels il pourrait être particulièrement intéressant que	les gouvernements expriment leurs vues, soit à la Sixième Commission, soit par écrit, afin de la guider utilement dans la poursuite de ses travaux.
CERF	INST	Ensuite ils m'écrivent : "Nous aimerions savoir	comment vous pensez que	l'actuel gouvernement israélien devrait réagir au fanatisme raciste et violent du Hamas et d'autres groupes intégristes islamiques palestiniens. ¯
CERF	INST	ef de l'État, c'est la reine de l'Angleterre, et du Canada, et du Royaume-Uni, représentée à Ottawa par le gouverneur général et ici par la lieutenante-gouverneure. Ce sont des impératifs lourds, ça,	dont nous voulons que	notre société se défasse. Nous voulons une république avec un système présidentiel digne de ce qu'est devenue notre nation et sa modernité. Je regrette, là, mais le député de Rivière-du-Loup, il nous
CERF	LITC	'ouverture que je voudrais si singulière de mes cuisses peut-être trop blanches ? Contrastes de couleurs. Le noir épilé pour toi en désir d'être sous ta main. Le rose presque rouge au bord du gouffre	où je voudrais que	tu te penches. OEil de reproduction fragile te regardant fixement, presque enfant, pénétré tant de fois avec force, rage, sans compassion. Essayaient-ils tous de me faire taire déjà ?
CERF	LITC	tout comme je l'avais imaginé. " Je me recouvris de mes mains, comme pour lui signifier : malgré tout je suis restée sous le spot, essayant de ne pas avoir l'air maladroite, regardant vers l'endroit	où je pensais qu	il se tenait, et incapable de le voir.
CERF	LITC	nctivement ralenti sa respiration et ne faisait aucun bruit. Près d'une minute passa ainsi, puis la jeune femme prit un couteau dans chaque main et se pencha rapidement en les avançant vers l'endroit	où elle présumait que	la bête était tombée.
CERF	LITC	tu me guidas là	où tu voulais que	j'entre en toi
CERF	LITC	rcing, les tatouages et le bondage. Et même l'hypnose, bien qu'on n'en sache pas plus à ce sujet que ce qu'on en avait trouvé dans un petit bouquin à quatre sous. Ce qu'on voulait faire à l'autre, ce	qu'on voulait que	l'autre nous fasse.
CERF	LITC	Rappelez-vous : toutes ces tendances psychiques enfouies ne sont pas la réalité ; elles sont ce	qu'il désire que	la réalité soit. J'ignore de quoi il s'agit. Peut-être sont elles une seule et même chose. En quelque sorte.

CERF	LITC	Mets tes mains plus bas. "	Qu'est-ce qu'il voulait que	je fasse ? Mes mains rampèrent jusqu'à la ceinture. " Plus bas. " Voulait-il que je me caresse ? Je n'étais pas prête à ça. Je ne le ferais pas. Pas s'il me regarde. C'était bien trop bizarre. " Plus
CERF	LITC	- Et je n'ai pas la moindre idée de ce	qu'il faut que	je fasse pour revenir dans ma chambre en dix-neuf cent quatre-vingt-seize !
CERF	LITC	CYNTHIA. - Otto Kranz est loin de tout contrôler, contrairement à ce	qu'il souhaite que	vous pensiez.
CERF	LITC	Lâchez-moi ! Vous n'avez pas d'yeux ! Mais	qu'est-ce qu'il faut que	je fasse pour que tu me lâches ? Tiens !
CERF	LITC	Je ne suis pas sur Terre pour plaire ou pour tenter de faire de moi ce	qu'un autre voudrait que	je sois
CERF	LITC	Et alors,	qu'est-ce-que vous voulez que	ça me fasse, Mercier ?
CERF	LITC	Bien sûr. A chacun son rôle. Lui, il doit nous faire mourir, et nous, nous devons aller enterrer notre frère. C'est comme ça que ça été distribué.	Qu'est-ce que tu veux que	nous y fassions ?
CERF	LITC		Qu'est-ce que vous voulez que	je vous dise ! Vous ne parlez pas de votre naissance... en général c'est rédhibitoire. Enfin, nous verrons bien... Et après...
CERF	LITC	Pensez-vous. Il nous vient des Chinois, des Hindous.	Qu'est-ce que vous voulez qu	'ils fassent d'un fauteuil second Empire ?
CERF	LITC	Et moi,	qu'est-ce que vous voulez que	j'en fasse ? savez-vous qui j'étais ? Bah ! ça n'a aucune importance. Après tout, je vivais toujours dans des meubles que je n'aimais pas et des situations fausses ; j'adorais ça. Une situation fauss
CERF	LITC	Imaginez l'excitation que peut procurer le fait de se trouver en tête à tête avec un étranger duquel il n'y aurait rien à craindre (ni pervers, ni séropo) et avec	lequel on sait qu'	on peut éventuellement aboutir au lit. Mais qui demeure mystérieux. Un danger, quoique inoffensif.
CERF	LITC	Il ne comprenait pas. Il en avait vu de toutes les couleurs, mais là, c'était le record. De ses mains crasseuses, il frotta ses yeux et son front derrière	lesquels il semblait que	toute la folie du monde jouait du marteau-piqueur.
CERF	LITC	t, illuminant une photo de son meilleur visage et la voix de Dieu -c'était la voix de Dieu, non ? - prononçait son nom, une voix profonde, grave et lente comme un final de Beethoven, ce genre de voix	dont on imaginerait que	la Terre elle-même fût dotée si elle n'était de sexe féminin, qu'elle extrairait d'un gosier profond comme la fosse des Philippines, articulant avec des lèvres de la taille du Grand Canyon. Pendant c
CERF	LITC	Tu m'entraînes chaque jour un peu plus dans un monde fait de dépendances	dont il faudra que	je me débarrasse pour être avec "Lui" !
CERF	LITC	- Et	qui veux-tu que	ce soit ? C'est bien à moi que tu t'apprêtes à jouer un vilain tour, non ? Alors si je ne me défend
CERF	LITC	Mais	que voulez-vous que	j'y fasse
CERF	LITC	- Ils nous somment ! La réaction de Borig fut explosive, Mais	que pensent-ils que	nous fassions, nous ici, en bas, pendant qu'eux passent leur temps devant les holoprojecteurs. Nous
CERF	LITC	s comparaisons inopportunes, les oublis destructeurs, les résumés trompeurs, alors, comprenez-vous,	comment voulez-vous qu'	on y voit clair après tout ça, même avec toute la bonne volonté, et la bonne volonté, on aurait auss
CERF	LITC	Je suis sèche. Je ne peux ni recevoir ni donner ;	comment voulez-vous que	je vous aide ? Une branche morte, le feu va s'y mettre. (Un temps ; elle regarde Estelle qui a la t

CERF	POLI	et si je devais comme je l'ai fait aujourd'hui avec grand plaisir, me préoccuper de leurs problèmes en allant quelquefois jusqu'aux détails, il ne me resterait plus de temps pour aller à l'étranger,	où il faut bien que	j'aille quelquefois... Je crois avoir remarqué d'ailleurs qu'on l'avait remarqué.
CERF	POLI	11 novembre 1918, dans ces hauts lieux de la mémoire combattante, nous venons ensemble d'inaugurer le monument réalisé par Haïm Kern. La symbolique de cette oeuvre s'accorde parfaitement avec le lieu	où son auteur a souhaité qu	elle soit implantée : au-dessus du vieux village de Craonne, sur le petit plateau au débouché de ce qu'on appelle ici la laie de Californie. Cette statue a été financée en partie par le Conseil génér
CERF	POLI	la Serbie a subi des dommages considérables, il avait le choix entre revenir dans un processus de raison et dans un processus de paix ou s'enfermer dans l'impasse suicidaire qui était la sienne mais	où il sentait bien que	progressivement, il était de plus en plus isolé. "
CERF	POLI	Français et Sud-Américains, marcher "la mano en la mano". Et bien c'était, comme tout ce qu'a toujours fait le Général de GAULLE, quelque chose de visionnaire car nous arrivons aujourd'hui à l'époque	où il est évident que	l'Europe et l'Amérique Latine doivent marcher la main dans la main.
CERF	POLI	en droit d'attendre du juge qu'il fasse justice. Il suffit de mesurer à cet égard ce que suscite comme amertume ou laisse comme frustration dans la conscience collective telle ou telle grande affaire	où le sentiment commun est que	la lumière n'a pas été faite et que la justice a été tenue en échec. En vérité, chaque jugement, parce qu'il participe à l'instauration d'un ordre de justice, sert, au-delà des justiciables qu'il vis
CERF	POLI	jours la conjonction du passé et de l'avenir bien entendu. Le présent même existe-t-il ? A peine l'a-t-on évoqué qu'il a disparu. Mais il est cependant des moments, et je crois que nous en vivons un,	où on a le sentiment qu	une porte s'ouvre sur l'avenir. Nous sommes en train de l'ouvrir nous-mêmes, nous sommes donc des ouvriers privilégiés de l'histoire.
CERF	POLI	qu'on le puisse, je le disais tout à l'heure, parmi les difficultés mais avec l'assurance d'accomplir son devoir, c'est ainsi que nous parviendrons à créer pour notre société un mode de communication	où les Français sauront qu	'ils ont encore une fois dans leur histoire de ces deux derniers siècles, défini l'exercice d'une liberté.
CERF	POLI	* C'est dans ce cadre de service public de proximité	que je pense qu	il faut réfléchir au rôle, au missions et aux moyens de la police et de la justice.
CERF	POLI	L'Éthique au nom de	laquelle nous estimons que	les critères moraux s'imposent aussi bien dans la sphère de l'entreprise que dans le monde politique ou syndical.
CERF	POLI	n de l'aide à apporter de toute urgence aux populations et aux victimes civiles, la mise en scène de l'aide humanitaire d'urgence ne doit pas occulter les projets de reconstruction de long terme dans	lesquels ils souhaitent que	l'Europe s'investissent massivement.
CERF	POLI	- "Monsieur le Président, Mesdames et messieurs les députés, Monsieur le député, La réunion informelle des chefs d'Etat et de gouvernement qui s'est tenue vendredi dernier à Petersberg, et	dont je pense qu'il est normal que	je rendre compte à la représentation nationale, a été positive à plus d'un titre. Elle a tout d'abord démontré que la France n'est pas isolée dans l'Union européenne lorsqu'elle déf
CERF	POLI	Là aussi, il me paraît souhaitable de conduire la réflexion, non seulement au plan national, mais encore au plan européen. C'est en effet dans le cadre européen	que le gouvernement souhaite que	s'harmonisent à l'avenir nos efforts de défense, l'analyse des besoins et la définition des réponses les plus appropriées aux défis du siècle prochain.
CERF	POLI	C'est ainsi	que je pense que	l'usage des libertés, et surtout l'usage qu'il en est fait dans le vocabulaire, finit par être considéré - par moi tout au moins et je ne suis pas le seul en France - comme un abus. On confond souven

CERF	POLI	n'est pas régulièrement validé par le suffrage universel ? C'est parce que la sécurité sociale est en fait une partie du salaire (une part socialisée, collective qui s'ajoute à la part individuelle)	que nous pensons que	ce sont les syndicats qui ont vocation à solliciter des suffrages de l'ensemble des assurés sociaux dépendant du régime général. Quelle légitimité pour le patronat surtout si, avec la moitié des sièg
CERF	POLI	es entreprises, d'une année sur l'autre, parfois même d'une saison sur l'autre. Dans ces conditions	comment voulez-vous qu'	ils prévoient l'activité économique tout au long d'une vie professionnelle ? Cela c'est un premier p
CERF	POLI	vivait rassemblée d'une génération à l'autre, quelques fois trois générations, parfois même quatre,	comment voulez-vous qu'	elle puisse se retrouver dans un espace aussi réduit ? La destruction ou une fraction de la destruct
CERF	PRAT	Indice de perspective d'avenir des consommateurs L'indice de confiance des consommateurs comprend un "sous-indice", lequel met l'accent sur les attentes des consommateurs afin de déterminer	comment ils pensent qu'	eux, et l'économie, réussiront pendant l'année qui vient et les années subséquentes. Comme c'était le cas auparavant, les consommateurs de l'Amérique du Nord sont généralement moins optimistes à propo
CERF	PRES	octobre 2002 et le Parlement dissous en juin 2001 reprendra ses fonctions. "C'est une révolution silencieuse", commente Naqvi. Une révolution où l'armée ne renoncera pas entièrement à son pouvoir, et	où il est clair que	Musharraf ne compte pas se retirer de son poste à la tête du pays. Parmi ses projets, la mise en place d'un Conseil national de sécurité, composé de militaires et de civils, pour contrôler l'exercice
CERF	PRES	t de lui son copain Rodolphe. Membre des Jeunesses contre le racisme en Europe, cofondateur du Collectif anti-expulsion, Aziz perd un peu courage en janvier et décide de se barrer en Grande-Bretagne,	où il croit savoir que	" les contrôles de police sont moins nombreux ".
CERF	PRES	Communiste ou non, il sera possible à chacun de prendre la parole, le stylo ou l'ordinateur pour faire connaître son avis et débattre de ce	qu'il attend que	le Parti communiste français devienne. Samedi et dimanche 14 et 15 septembre, au Parc paysager de La Courneuve. Sans interruption. De 16 heures à 18 heures, chacun pourra s'entretenir du 29e Congrès
CERF	PRES	fait prendre aux sujets d'une expérience un produit qui provoque une sécrétion d'adrénaline, ils interprètent leur sentiment comme de la colère, de l'euphorie ou comme rien du tout, en fonction de ce	qu'on leur a dit qu	ils ressentiraient.
CERF	PRES	votre mère a été une grande leçon de vie. Carlos. - C'était sublime ! Elle disait : " Je vous vois pleurer. Pleurez pas ! Ce n'est rien de grave. C'est juste une fin de vie qui se déroule.	Qu'est-ce que vous voulez que	je fasse de mon corps ? J'ai 80 ans. J'ai tout fait, tout donné, tout ce que je pouvais savoir, tout ce que j'ai appris. Laissez-moi tranquille. C'est ma deuxième naissance. " Elle est morte dans la
CERF	PRES	son bureau de la place Beauvau. Un visage bronzé mais amaigri. Des yeux qui brillent mais des traits encore creusés par l'épreuve. " Et puis il y a les mots. " L'autre rive ", ce " grand voyage ",	dont on pensait qu	il ne reviendrait pas. Une citation piochée chez Marc-Aurèle qui sonne comme une promesse de sérénité chez un homme qui fut surtout celui des certitudes. Une élocution moins sautillante qu'autrefois.
CERF	PRES	érale : "Ils nous accordent 35 % des recettes fiscales alors que les charges grimpent constamment ;	que voulez-vous qu'	on fasse avec cela ? Il nous faudrait au moins 50 % ! C'est pourtant ici, dans les provinces, que se
CERF	PRES	érale : "Ils nous accordent 35 % des recettes fiscales alors que les charges grimpent constamment ;	que voulez-vous qu'	on fasse avec cela ? Il nous faudrait au moins 50 % ! C'est pourtant ici, dans les provinces, que se
CERF	PRES	arti démocratique déclara : "Nous n'avons jamais reçu des autorités la liste électorale définitive,	comment voulez-vous que	nous contrôlions le vote ? " Quant au secrétaire local du Parti socialiste, il confia : "Nous ne po
CERF	SCIE	Time 60 1. Pour changer le chiffre, appuyez simplement sur le bouton à flèche jusqu'à ce que la durée pour	laquelle vous voulez que	le programme fonctionne apparaisse.
CERF	SCIE	. Conflits beaucoup plus destructeurs que les bras de fer destinés à tirer une part plus grande de la richesse collectivement créée. Le conflit pour le salaire prolonge le rapport de marchandage dans	lequel chacun sait qu'	il a besoin de l'autre : chacun s'efforce de rouler l'autre suffisamment pour avoir fait une bonne affaire, mais pas trop sinon le maintien de la relation deviendrait incertain. Le conflit pour le mai

CERF	SCIE	fférents modes pour un même TNC et il existe plusieurs étapes pour utiliser chaque jeu. Tout d'abord, vous devez définir les macros d'initialisation du TNC et les placer dans le mode dans	lequel vous souhaitez qu'	elles s'exécutent (et les désactiver à la fin de l'utilisation de ce mode). Les deux macros ci-dessous sont utilisées dans ce but :
CERF	SCIE	que sur le lexique ou quelques formes choisies pour échantillonnage. Nous nous sommes contentés d'exclure un certain nombre de textes pour ne garder que la littérature qui semblait convenir à l'usage	dont nous pensions qu	il serait fait du corpus.
CFPP	Audin_7e	spk2 [2491.438] : ah alors + j'ai habité quand même pendant quatre ans dans l'huitième + donc euh + c'est un quartier qu'j'ai un peu découvert	qu'je dirai qui	est pas v- enfin j'habitais à côté de la place Bauveau donc j'dirai pas qu'c'est un quartier + euh c'est un quartier bien sûr mais c'est pas un endroit où les + où les gens + j'veux dire c'est des ga
CFPP	Belamy_Hermano_KB	spk3 [3212.918] : mais voilà moi j'connais des maîtres qui ramassent donc euh c'est toujours le problème des c'est une question de civilité c'est pour ça	que je dis qu	'j'aime pas les gens à Paris parce qu'i(ls) ils sont pas respectueux ils pensent qu'à eux
CFPP	Liotard_Mallet_SO	spk4 spk2 [321.13] : mm ++ ouais mais:	qu'est-ce que vous voulez que	j'dise qu'est-ce ça fera quoi? +
CFPP	deBordes_7e	y a des quartiers qu' vous détestez dans Paris où vous n'iriez pour rien au monde ou simplement	où vous savez que	ça serait difficile de de vous intaller
CFPP	Duchemin_Ceret_11e	je je suis dans une période de ma vie	où je pense que	jamais dans ma vie j'ai vu aussi peu loin ce qui pouvait m
CFPP	Leblond_Zanotti_7e-1	spk3 [640.368] : mais là je préfère rester enfermée chez moi plutôt que + d'aller dans le Champ de mars	où y a je trouve qu	il y a trop de monde
CFPP	Matera_Mo_1	non dans tous les quartiers d'en haut il y a un plan	où on voit qu'	il y a beaucoup beaucoup d'endroits où il doit y avoir des
CFPP	Matera_Mo-1	au marché ils parlent leur(s) langue(s)	qu'est-ce que tu veux qu'	on /leur, 0/ parle
CFPP	Pelletier_IS	Jacqueline Pelletier la mondialisation	qu'est-ce que vous voulez que	je vous dise moi quand j'entends nos agr
C-Oral-Rom	ffamdl23	*CLE: et ma foi / ben #	qu'est-ce que tu veux que	je fasse d' autre ? # un peu passer le temps avec lui // et un peu à tricoter // # et je passe mon temps comme ça // #
C-Oral-Rom	ffamcv01	*CYR: <&euh il y a des différences> ou est-ce que c' est la même chose qui est [/] qui est <ressentie différemment / par les mêmes personnes ?> *VAL: / <mais	comment tu veux que	ça marche toi ?> *ALE: <tu &conc> [/] tu conçois hhh [/] mais &euh tu conçois ta vie de couple comment ? <enfin je veux dire> /
C-Oral-Rom	fmedsc02	*BCR: <voilà> à [/] à partir du moment où on est dans la zone	où on estime que	c' est là qu' il va falloir être très précis / # &euh on [/] on a nos électrodes qui sont connectées à des chaînes d' amplification
C-Oral-Rom	fnatpd02	# la charité / pour la recherche / pour tout un tas de choses hein ? # et donc de temps en temps / &euh il y a un scandale qui éclate / et on apprend que # ça allait pas très exactement / à ceux pour	qui on aurait voulu que	ça aille et pour ceux &euh # pour qui nous nous avions donné de l' argent // # donc ça pose tous ces problèmes-là // # maintenant # je le rappelle bien que c' était pas votre question / au Parlement

C-Oral-Rom	fpubdl03	mpris au niveau des familles // # il faut savoir les écouter // # il faut savoir répondre à leurs questions // # parce que même s' ils les posent pas directement / # il y a toujours des sous entendus	qu' il faut qu'	on arrive justement à récupérer // # donc la base numéro un c' est le dialogue // # une fois qu' on a fait ça / # je dirais que la partie est pratiquement gagnée // # donc &euh # il y a que cette sol
C-Oral-Rom	fpubmn02	*AME: &euh non [/] non [/] non // # &euh il [/] ça fait combien de temps	que vous dites qu	il est [/] &euh # l' oeuvre est ici ? #
corpaix	13Gurin	là mais bon comme quoi c'est pas c'est pas très très important quoi un accent circonflexe on peut le simplifier donc il y a il y a plein de petites choses comme ça quoi donc euh à partir à partir de	quand tu crois que	c'est important tu as l'impression que c'est important de modifier ou pas la langue euh que tu perds ta tu perds tes racines euh si tu modifies trop la langue quoi enfin changer un accent circonflexe
corpaix	28SPEED	fait tu te rends compte que + enfin je m- je me rends compte que je sais plus parler euh correctement le français + que j'ai du mal même quand je suis en face d 'un professeur ou de + de quelqu'un à	qui il faut que	je parle un français correct euh + et ben (j'ai, j'é-) j'ai vraiment du mal + je je sors toujours des mots mal à propos au milieu (qui ont, qui euh) qui (ont, euh) j'ai plus du tout l'habitude j- j'a
corpaix	32Verlan	omprend + ça se comprend + et euh en restant toujours dans l'histoire tu penses que l- les réformes justement à partir de l'abbé Grégoire à partir de ce qu'il a fait ont abouti à quelque chose on euh	quand tu penses que	cela a permis à la langue française de se développer par la suite + de bien se développer L2 j'avoue que je sais pas mais euh L1 ben enfin par rapport à l'enseignement si euh si je te dis euh je sais pas euh mille sept cent quatre-vingt quatorze mille hu
corpaix	34SOPH	L2 je pense que c'est utile pour le prof pour voir euh à	quel niveau il en est-ce qu'il faut qu	il leur fasse apprendre plus ou moins mais je pense pas que ce soit euh un moyen de de faire apprendre l'orthographe aux enfants
corpaix	6Celine	les règles mais au moins quand on les a apprises et ben on arrive à intégrer ça dans sa vie de tous les jours + c'est peut-être ça aussi l'idéal c'est d'apprendre à parler sans se dire alors là dans	quel temps va falloir que	je mette de cette phrase pour qu'il(s) comprenne(nt) ce que j'ai voulu raconter ça vient tout seul + non c'est vrai que les conjugaisons moi j'en ai pas spécialement des souvenirs euh + merveilleux m
corpaix	ALONSO	L2 et oui elle va m'entendre elle va m'entendre elle va me reconnaître elle me reconnaîtra - eh je sais pas moi	qu'est-ce que tu veux que	je lui dise à la cassette il faudrait que tu nous enregistres quand on travaille
corpaix	ALSACEA	L4 mais vous allez vous arrêter de la faire rire tous les deux + c'est pas possible + L3	qu'est-ce que vous voulez que	je dise + L4 ben normalement il(s) devrai(en)t déjà être couché(s) mais c'est + le +
corpaix	ANTOINE	L2 quand on quand on écrit une lettre + il suffit de savoir	où est-ce qu'il faut que	tu mettes (des, tes) SP + XXX où tu accordes où + voilà c'est tout pas besoin de savoir qu'il y a une con- une subordonnée conjonctive relative on s'en fout
corpaix	APOSTROP	mpris sur eux-mêmes des choses que sans Loïca peut-être ils auraient pas compris + bon et puis euh effectivement les les gens qui sont . installés dans des structures où euh ils sont bien à l'aise et	où euh il faut que	ça tourne rond et le syndicat pour certains ça peut être ça aussi ça peut être envisagé comme ça bon ben pour ceux- là euh une Loïca est inadmissible parce qu'elle a trop d'exigences L1 comment vos l
corpaix	AVOCAT96	cat ou les avocats qui qui qui font beaucoup d'aides judiciaires qui s'occupent beaucoup il y a il y a des avocats d'associations d'enfants battus + il y a des avocats d'assocations de femmes battues	où là je pense que	ils ont plus l'impression d'aider maintenant euh: l'avocat qui fait du droit des affaires il: défend
corpaix	AZ	L1 hum hum + il y a des périodes	où vous remarquez que	vous vendez plus ce produit que ce produit
corpaix	AZ	il y a des périodes	où c'est vrai qu'	il y a des produits qu'on vend pas et que coup sur coup on vous demande +

corpaix	BAYROU	de doute qui s'installe + est-ce que vraiment on est condamné + chaque fois à avoir une élection + avec un enthousiasme + et derrière une désillusion et des gouvernants qui reculent par rapport à ce	qu'ils avaient dit qu	ils feraient + eh bien je suis persuadé + euh que les Français attendent des gou- des gouvernants qu'ils aient cette double attitude + premièrement ayez le courage de nous dire ce qu'il faut faire +
corpaix	BEAUMETT	l y a des choses qui sont du simple au double + par exemple les les chaussures de basket quand on fait . une + bon + une relation avec les prix pratiqués par heu heu les Trois Suisses + la Redoute bon	où on sait que	c'est . un petit peu moins cher que les grandes surfaces + à l'intérieur c'est le double + et donc + à cet effet + ils interdisent que les familles rentrent des chaussures de basket + alors c'est . i
corpaix	BENOT	L2 tu restes à la ZOZO alors L1 et oui + ça ++	où tu veux que	j'aille hé L2 ouais + L1 tu sais on a beau regarder heu comme débouchés d'abord à Benot il y a rien ++
corpaix	BENOT	L1 eh eh c'est moi	qu'il faut qui	parle maintenant non
corpaix	BILLY	rle quoi + c'est évident + que euh + c'est + + comment dire + X on peut pas avoir le même résultat + ça veut pas dire que c'est pas aussi profond hein euh il y a + on a on a des + parfois des moments	où on sent que	c'est quand même fort puisque bon soit il rigole soit + il y a des moments c'est très chaleureux hein donc à la limite on (n') a même pas besoin de parler mais + c'est quand même chaleureux + mais il
corpaix	BRELPC	Belges + il y a une dame il y a pas longtemps qui m'a dit mais enfin monsieur Brel plus personne ne parle avec l'accent bruxellois pourquoi vous vous moquez comme ça tout le temps de nous +	qu'est-ce que tu veux que	je dise et et et moi ça ça me paraît charmant cet accent-là + seulement quand je le dis aux gens ils me croient pas + mais je le je trouve moi que c'est un truc charmant
corpaix	BUSABCD	L5 où c'est	qu'il faut qu	'elle(s) se présente(nt)
corpaix	BUSABCD	L4	qui il vaut mieux que	je voie
corpaix	CFA	oyeur le C.F.A. n'intervient pas dans la signature d'un contrat + à partir de ce moment-là l'employeur doit fournir au C.F.A. de son choix c'est-à-dire que c'est votre employeur qui choisit le C.F.A.	où il veut que	vous soyez c'est pas vous qui choisissez c'est l'employeur il choisit donc un C.F.A. et transmet un double de du contrat d'apprentissage + lorsque le C.F.A.: choisi reçoit le double vous êtes automat
corpaix	CHRABI	L2 même si on dit qu'on qu'on est . artisan mais même si on est intellectuel artisan même je préfère parce que tu écris et tout +	qu'est-ce qu'il faut qu	'on fasse pour lutter contre l'invasion -- euh une invasion qui est maintenant une épine dans le dos euh du peuple arabe -- et berbère dans notre région --
corpaix	CONVOCA	L2 et je n'ai pas de B.T.S. + donc c'est clair + ce	qu'il faut que	je fasse c'est préparer un B.T.S. et refaire une année de travail + et une fois que j'aurai tout ça je pourrai(s) tenter de d'y aller
corpaix	CONVOCA	leur avoir tant bourré le crâne et bien ils se sont aperçus que + + restée en rade parce que on ne lui avait rien proposé + donc maintenant c'est eux qui me proposent un stage où on va m'expliquer ce	qu'il faut que	je fasse vers qui je dois aller voilà + c'est tout à fait ça +
corpaix	CYRIL	est déstructurée par rapport n'a pas de structures par rapport à la langue académique en tout cas (elle, Ø) en a vachement moins + mais c'est justement l'intérêt de la langue de la rue c'est pour ça	que je pense qu	'elle est euh elle est dans un sens supérieure à l'autre + + je sais pas je peux partir ensuite sur euh les dangers justement de des lois Toubon etc. parce que finalement ce qu'elles cherchent ce qu'e
corpaix	Ecole	L4 on (n') a pas le droit de courir L2 ouais on joue on joue immobiles quoi L5 et	pourquoi vous pensez qu	'ils vous empêchent euh
corpaix	ECRITUPC	ture à d'autres fins cela dit tous les travaux donnent=à penser que euh toutes les situations d'écriture ne sont pas=équivalentes euh toutes euh comportent des paramètres qui leur sont spécifiques et	qu'il faut que	l'élève: domine + cependant il y a tout lieu de penser qu'effectivement euh l'élève qui construit un rapport euh distancié à l'écriture sera à même précisément de prendre en compte ces paramètres dif

corpaix	FLICIENM	uh + parler de mes propos de table + alors // + je lui ai dit ben Monsieur le Vicaire Général je savais pas du tout + que j'avais un voisin de table qui faisait toute cette littérature mais	qu'est-ce que vous voulez que	je fasse de ce paquet + alors il m'a dit + ah ben moi + moi ça m'est égal parce que + je n'en ferai rien je les ai pas lues + c'était assez gentil quand même + il connaissait d'ailleurs le + l'indivi
corpaix	FTUSPC	L1 mm madame Mulliez il faut rappeler que toute grossesse comporte un risque et quels sont euh les: les défauts qui font que l'enfant ne sera pas viable quels sont les aspects avec	lesquels on sait qu'	un enfant ne sera pas viable
corpaix	FTUSPC	L2 il y a trois échographies d'ailleurs vous dites que c'est ce qui est + couramment admis c'est pas forcément couramment admis c'est ce	qu'on pense qui	devrait être fait
corpaix	GENVE	L3 il faudrait quand même calmer un peu les gens ils euh ils peuvent passer ils peu- avec la meilleure des volontés	où vous voulez qu	ils se garent il y a pas de places bé ils se garent où ils peuvent eh +
corpaix	GUA151AB	L1 ben ça au départ vous savez c'était du temps	que je vous dit que	nous étions pas encore desur terre sûrement
corpaix	HISTOIRE	il recommence + il revient au bout de la troisième fois + il dit ça va comme ça deux fois c'est bon + Doume il le regarde il dit branle-toi il dit écoutez-moi j'en peux plus deux fois trois fois XXX	combien il faut que	j'en fasse + Doume il dit branle-toi ou je te tue bon ça va je X + je vais me branler ça va //// bougez plus XXX le mec il revient il en peut plus il est à quatre pattes par terre et tout + alors Dou
corpaix	LENJEU	ue chapitre euh et si elle me demandait de supprimer quelque chose je le faisais et je lui ai envoyé le chapitre beuverie euh elle m'a dit au téléphone tu es un peu dur j'ai dit écoute tu me diras ce	que tu veux que	je supprime je le supprimerai elle dit on en reparlera à New York car elle devait me rejoindre la semaine d'après + et puis je le téléphone a sonné à vide euh entre temps elle avait bu un litre de vo
corpaix	MAON	L2 voilà tenez regardez L1 et oui L2	qu'est-ce que vous voulez que	je vous dise hè L1 + et oui L2 voilà ce qu'il a fait le gars moi j'ai pris sur sur ce parce que
corpaix	MAON	qu'en étant là-bas qu'est-ce que vous voulez L2 ah qu'est-ce que vous voulez moi j'en ai un à aris hè de de fils /il, y/ et et mes deux petits fils /ils, y/ sont à aris hè L1 et ben oui L2	qu'est-ce que vous voulez que	j'y fasse moi on (n') y peut rien y faire L1 ben oui mais quand on a son mari euh qu'on est ensemble c'est pas pareil hè L2 hè + hè oui L1 moi quand j'avais L2 ++ bon écoutez vous me donnez une répon
corpaix	Merl3	e en l'air + mais ç- ça avait son importance parce que si on le couchait avec la queue dessous + l'échelle il était c'était en déséquilibre + alors qu'avec la queue en l'air la queue reposait + voilà	pourquoi on disait qu	'il fallait le poser avec la queue en l'air
corpaix	PAVRAYST	le chien naturellement il est venu faire contre sa façade pensez elle la colère qu'elle avait elle dit dites votre chien alors il dit ben il peut pas y mettre une braguette à mon chien elle lui fait	qu'est ce que vous voulez que	j'y fasse si le chien il a /pi-, fait/ contre votre façade mon Dieu elle était tout le temps là à surveiller sa et elle avait eu p- en haut alors mais de l'autre côté d'où j'habitais un magasin elle
corpaix	PORTUGA1	L2 spontanément -- ça en fait c'est c'est bizarre mais /ça, cela / dépend des jours + ça dépend il y a des jours	où je trouve que	+ les mots en français me viennent plus facilement mais mais je pense que je suis quand même plus-à l'aise dans le portugais
corpaix	PORTUGA3	au ortugal et ce ça m'a marquée parce que c'était la période où où où je où j'étais vraiment je sais pas c'est j'ai eu une relation privilégiée avec le pays à cette époque-là + et c'est c'est pour ça	que je pense que	j'ai autant /d', ø/ envie de retourner au Portugal

corpaix	POSIE	et demie + alors la grille d'analyse d'un texte puisque vous pouvez vous demander + quand vous avez lu un texte qu'est-ce-que je dois dire XX qu'est-ce-que je dois en penser qu- par	quel bout est-ce qu'il faut que	je le prenne pour parler vulgairement + et bien + toute cette grille c'est-à- dire le la struc- la structure apparente rythme et sonorité(s) vocabulaire syntaxe etc. tout sera donné dans le bulletin
corpaix	PRVOYAN	toutes les personnes que je rencontre surtout chez les professions libérales pour arriver donc à à avoir confiance dans un système de placement(s) plutôt qu' /en, à/ un autre c'est pour ça d'ailleurs	qu'on se rend compte ben que	le livret A de la Caisse d'Epargne qui avait été complètement euh laissé à l'abandon dans les années soixante-et-dix quatre-vingts revient très très fort parce qu'en fait là on sait au moins qu'on a
corpaix	RAPIN	L1	qu'est-ce que tu veux que	je t'explique +
corpaix	ROUREC-1	L1 quand il y en a je leur dis où faut	où il faut qu	'ils se garent
corpaix	SANDRA1	L1 parce que c'est un besoin qui doit être satisfait immédiatement + c'est pas comme la télé	où vous savez qu	'à huit heures et demi il y a [1] un film bon on vous y prépare à l'avance + la cassette non la cassette c'est tiens
corpaix	SAPHO	L2 ben évidemment vous me branchez sur des histoires constamment euh + sensuelles /ben, alors, Ø/	qu'est-ce que vous voulez que	je fasse moi je je vais être obligée de faire des choses abominablement euh + ah qu'est-ce qu'il(s) disai(en)t oui hi- hier j'ai regardé la télévision et il y avait un une citation très jolie
corpaix	SCA59AB	ichu machine(s) tout tout tout + on nous a emmené à Fréjus là-bas + et eh on a j'ai du rester neuf jours moi à Fréjus avec ma fille mon mari tout le monde et mon mari séparé de moi + alors	qu'est -ce que vous voulez que	je raconte de bien X
corpaix	SCA59AB	l y avait le grand autel /de, ø/ face comme ça + et tout le long il y avait tout des petits autels + qui XX tout en marbre tout joli comme tout il y a plus rien + on me l'a abîmé comme tout	qu'est-ce que vous voulez que	j'y dise vous dise + moi je voulais mon église + il y avait des chorales superbes
corpaix	SCA59AB	L3 eh oui tout ce que tu racontes L2	qu'est-ce que vous voulez que	je vous dise L1 oui
corpaix	SCA59AB	L2 et	qu'est-ce que vous voulez que	je vous dise les poissonnières faudra que je cherche j'ai une carte ancienne qu'il y a une poissonnière une toulousaine mais vous savez toulousaine marseillaise je vous la ferai voir mais c'est tout
corpaix	SCA59AB	s les jours il fallait que j'y trace le le chemin autrement il sortait pas il avait la voiture alors il me disait quand vous voulez aller à quelque endroit + je dites-le-moi je vous mènerai	qu'est-ce que vous voulez que	j'aille dire à un petit + qui serait été mon petit-fils menez-moi à tel endroit je disais rien + alors quand j'arrivais le soir à l'hôtel il me disait où vous êtes allé aujourd'hui alors je lui disai
corpaix	VINCENT	te ils sont montés sur sur leurs grands chevaux ils (m-, X) ils m'ont ils m'ont demandé comment je parlais euh qu'est-ce que c'était ces mots et que j'utilisais des mots compliqués euh alors c'est là	que je dis que	notre vocabulaire est très limité et ça je pense que euh un des facteurs principaux c'est que les jeunes d'aujourd'hui et ben ne lisent plus beaucoup quoi bon moi je suis pas un gros lecteur + je lis
CRFP	AMI PRI 1	L1 c'est une euh> c'est une - donc c'est pour ça	que je trouve que	c'est /simple, sain/ c'est une très bonne école quoi
CRFP	BEL PRI 1	L1 évidemment - surtout que je joue quand même avec des éléments de hasard aussi - hein je - ça m'arrive de faire quelque de faire une peinture	où je sais exactement ce qui	j'ai envie de faire - à la limite je la vois construite dans ma tête - mais euh malgré tout c'est souvent plus par impulsion et les choses qui arrivent le hasard font découvrir autre chose tiens tien

CRFP	BOR PRI 2	pris du travail à la DASS où je me j'étais auprès des handicapés - vous voyez je recherche toujours un petit peu ce qui - pas ce qui est facile enfin c'est pas c'est pas que c'est pas facile - mais -	où on sent que	les gens ont plus besoin de vous les familles ont besoin de vous - hein - parce que c'est vrai sur le secteur aussi on a besoin - mais j'avais l'impression que sur le secteur - d'abord il faut vous d
CRFP	LIM PRO 1	n sens le sens du spectacle la thématique tout le tralala - mais là - le comédien qui vient et qui est pas écouté par son public - c'est difficile pour lui c'est difficile - et ça c'est quelque chose	qu'il faut qu	'on apprenne c'est pas - il il le surmonte de toute manière et euh - et euh des fois c'est dangereux psychologiquement parce que euh parce que parce que s'il y a trois enfants qui ont parlé pendant to
CRFP	ORL PRO 1	c comme je vous l'ai dit - et d'autre part euh - le conservateur du musée - a décidé que nous ferions ensemble un itinéraire choisi par lui pour présenter certaines collections remarquables - euh sur	lesquelles il souhaite que	nous insistions - mais vous avez la possibilité - à la suite de cette visite de rester dans le musée - et de - de regarder de compléter par vous-mêmes - cette visite - avez-vous quelques questions -
CRFP	PER PRI 2	ça oblige l'artiste à se poser la question pourquoi - ils n'aiment pas - voire si à lui ça lui permet de - de - d'essayer de trouver - d'autres possibilités dans sa création - pour ne - c'est pour ça	que je dis qu	'il faut apprendre à le gérer - parce que c'est très bien que des personnes n'aiment pas et qu'ils en parlent - ça permet - euh de savoir - et de mettre à jour - que ce que l'on fait - n'est pas oblig
CRFP	PNO PRI 3	L1 voilà - après	qu'est-ce que tu veux que	je te dise bon alors on fait comme ça - sans arrêt jusqu'à ce que - à chaque fois on marque les points à chaque mène
CRFP	PNO PRI 3	L1 voilà - alors les boules ont des dessins différents - euh pour que chacun puisse reconnaître ses boules sur le terrain - voilà et puis	qu'est-ce que tu veux que	je te dise d'autre - donc à chaque mène on marque les points - et logiquement on va en combien en treize - on fait une une partie en treize points - et on peut
CRFP	PRI PRI 1	voilà bon ben tu as le dossier - donc euh	qu'est-ce qu'il faut que	je dise ben je sais pas alors je vais commencer par raconter un petit peu ma ben mon week-end
CRFP	PUB-TRO-1	comment le patron du textile va pouvoir augmenter les salaires si le prix lui est imposé vous comprenez ce que je cherche à dire hein si on lui dit ben ben tiens cette chemise là mon pote tu la fabriques pour dix balles	comment voulez-vous que	lui euh puisse répercuter puisse augmenter donner plus de salaire c'est pas possible d'accord
CRFP	ROU PRO 1	si c'est des questions qui concernent des formations c- bien souvent on a des étudiants qui viennent qui disent ben on aim- j'aimerais bien être animateur ou animatrice et puis	qu'est-ce qu'il faut que	je fasse quoi - voilà
CRFP	STR PRI 2	alors il m'est arrivé d'avoir euh - deux colonnes - et puis - je voulais les mettre l'une après l'autre alors au départ j'ai j'ai encadré le tout - mais là il faut que tu sélectionnes - l'ordre dans	lequel tu veux qu'	il te replace ça sur tout le texte
CRFP	STR PRO 1	L2 c'est pour ça	que je pense que	les les ça joue le fait qu'il y ait des niveaux bas que - et que peut-être ça pose aussi une question par rapport à l'accueil de ces jeunes dans à l'Education Nationale _____ Contexte = niveau bas + comportement diff => pb, si que comportement et bon ni
CRFP	STR PUB 1	tes - eh bien je cr- je pense je pense oui c'est une hypothèse enfin qui me paraît - maintenant comme ça au bout quand même d'un certain nombre d'années de travail dans ce domaine c'est une hypothèse	que je crois qu	on peut avancer - eh ben ces pertes qui sont autant de deuils - le deuil c'est la mort de l'objet hein c'est c'est c'est - le - pas seulement y renoncer hein comme diraient les moralistes - mais vrai
CTFP	la fleuriste	quand il y avait un client ben la patronne elle se dérangeait pas c'est moi	qu'il fallait qui	aille servir

CTFP	la fleuriste	L1 on on avait l'hôtel Plaza aussi Avenue Montaigne -	où il fallait qu'	on aille à chaque fois qu'il y avait des grandes réunions
IRIT	dialogue 1.8	en fait ça fait la croix avec la rue	où il faut que	tu ailles
Nancy	Cheillan	L3 oh la la les parents quand ils *** même parfois ils savent pas ce qu'ils veulent dire il y en a c'est vrai ceux qui sont compétents ça va mais ceux qui comprennent rien L2 hum hum L3	qu'est-ce que tu veux qu	'ils disent L2 oh oui L4 même si on les gronde les enfants même des jeunes hein eh ben les frères le père ils viennent /***,frapper de suite/ L3 oh ben oui ça parle *** L1 oui L1 à l'Estaque ***
Nancy	DIV assoc	à demander quelqu'un personne ne voudra aller là-haut	qui voulez-vous qui	grimpe comme-ça - mais qui est-ce moi j'ai dit on n'en
Nancy	DIV couture	t + là je pense qu'on a à peu près le le tout alors ce	qu'il faudrait peut-être qu'	on fasse parce que toi tu es débutant c'est qu'on < reg
Nancy	DIV seminaire	is je : je me mets à l'écoute de Dieu et de sa volonté	que veut-il que	je fasse bien souvent dans ma démarche vocationnelle e
Nancy	greve	us importante aux étudiants + pour l'instant on reste dans des chiffres dérisoires mais il y a quand même cette volonté premier aspect + deuxième élément tout ça s'intègre quand même dans un contexte	où nous avons vu que	euh aussi bien les frais d'inscription que les frais de sécurité sociale que le prix de ticket de restaurant universitaire augmentent bien plus vite que le montant des bourses qui sont aloués euh aux
Nancy	Japon	lamentation(s) de de je sais pas du du scénario de de la pensée capitaliste ou des choses comme ça mais peut-être plus de voir euh comment ils peuvent faire résistance vis-à-vis de - ben d'un endroit	où il me semble que	: hum que finalement euh - que finalement il y a il y a plein il y a il y a plein de croisements en fait d'idées de : d'automatisation de de domination(s) de : - c'est un peu c'est le Japon c'est c'e
Nancy	marais	L2 et toi qu'est-ce que tu en retires euh - de ça	qu'est-ce que tu trouves qui	est super euh -
Nancy	Marais	L2	pourquoi tu as dit qu	il n'a pas eu une mort extraordinaire L1 parce qu'il voulait mourir euh de façon euh ou euh quelque chose de de marquant en fait il est mort à l'hôpital comme n'importe quelle vieille personne quoi
Nancy	Maroc	L1	qu' est-ce que vous voulez que	je vous dise à propos du Maroc L2 alors quelles sont tes - tu es du Maroc Maroc ou
Nancy	Maroc	L1 je préfère ne pas y réfléchir puis ben pas y aller en courant pour l'instant je préfère en fait parce qu'il y plein de questions	que la religion elle te dit que	c'est interdit de de répondre quoi - c'est dit je peux pas répondre à ça tu te poses pas la question tu le prends comme ça soit tu le prends pas - et là j'arrive pas à comprendre tu vois tu as plein
Nancy	Retraites	propos des retraites depuis quatre-vingt treize ça fait quand même déjà une dizaine d'années + le statut de la fonction publique euh est caractérisé par le salaire à vie et c'est cette thématique là	qu'il faut que	nous soyons capables si en tout cas nous estimons que c'est une bonne chose de populariser auprès des salariés du privé la fonction publique c'est pas comme on le dit souvent l'emploi à vie hein les
Nancy	Retraites	as aller au-delà et puis euh et dire que la perspective à atteindre c'est celle-là et quand on aura atteint cette perspective où euh ben on sera tous salariés de notre société entre guillemets idéale	où chacun aura pris conscience que	son travail c'est plus un travail subordonné mais il a une rémunération euh bon bref euh je sais pas j'ai j'ai l'impression que c'est un peu ça que vous dites mais sans euh sans dire le mot euh très
Nancy	seminaire	rends compte davantage c'était Seigneur regarde qu'est-ce que je peux faire de beau pour toi et maintenant c'est plutôt Seigneur fais que je regarde ce que je pourrais faire de beau pour toi - et ce	que tu veux que	je fasse de beau pour toi - fais que je regarde ce que tu veux que je fasse et en fait /la, ma/ position est celle-ci main- je poursuis un discernement spirituel à l'heure actuelle j'ai un conseiller

Nancy	seminaire	eux faire de beau pour toi et maintenant c'est plutôt Seigneur fais que je regarde ce que je pourrais faire de beau pour toi - et ce que tu veux que je fasse de beau pour toi - fais que je regarde ce	que tu veux que	je fasse et en fait /la, ma/ position est celle-ci main- je poursuis un discernement spirituel à l'heure actuelle j'ai un conseiller spirituel et euh je je continuerai à vivre dans l'Eglise sous quel
PFC	11agm1	E: Vers	quand tu penses que	ce sera?
PFC	12afl1	LS1: mais c'est euh non c'est génial quoi mais faire un truc que tu aimes quoi Et	que tu sais que	tu maîtrises. <FL1: Ah oui, oui.> Et là, ouais, ouais, là je m'éclate.
PFC	13bsa1	et	pourquoi tu penses qu'	ils ont pas voulu
PFC	21acp1	CP: En étant trop nombreux quoi. Enfin euh, (rires) non mais il faut qu'ils se <E:	Qu'est-ce que tu veux qu	'il fassent? Ah qu'ils se cassent ?> qu'ils se cassent quoi. Enfin vraiment quoi, et franchement et je sais pas pourquoi ils se, ils font autant chier pour les obtentions de visas et compagnie quoi, p
PFC	38agp1	GP: '	qu'est-ce que tu veux qu	'on fasse ?' ils nous, elle nous disait 'qu'est-ce que tu veux qu'on fasse ?'
PFC	38agp1	GP: 'qu'est-ce que tu veux qu'on fasse ?' ils nous, elle nous disait '	qu'est-ce que tu veux qu	'on fasse ?'
PFC	42aad1	M: Voilà oui, ah oui. <A: Et c'est pour ça	qu'il a fallu que	j'aille au, au, à l'informatique.>
PFC	75ccm1	tu as été dans d'autres euh enfin d'autres euh théâtres spectacles	où euh tu tu entendais que	c'était un français un peu euh
PFC	75ccr1	jusqu'à ce que votre enfant est l'âge	où où on vous dit que	vous vous n'avez plus droit à votre disponibilité
PFC	75clc1	n crapaud qui a eu la malchance de tomber dans une famille	où il fallait qu'	il s'investisse un peu
PFC	75csb1	A : Ben euh pff <VL : Il te fait pas confiance ?> <SB :	Qu'est-ce que tu veux qu	'elle y fasse là> c'est pas qui me fait pas confiance,
PFC	75csb1	A : Alors on lui file le can/ le bon candidat	qu'on veut qu	'il prenne et à côté on lui file des petits candidats de merde <VL : Oh non>
PFC	75xad1-75xrg1		pourquoi vous voulez qu'	on rentre vous allez fabriquer le petit frère
PFC	75xmm1	ise et la version originale japonaise la version française	où tu voyais qu'	il était très porté cul sur les meuf et tu voyais rien
PFC	81aar3	AX :	Qu'est-ce que vous voulez que	je fasse, <AR3 : Arrête, c'est fini.> écoutez, <E2 : Non mais, non mais on est, on est.> vous allez pas être bien reçus vous allez avoir du monde à (XX) dans dix minutes bon on s'arrête, <E2 : Non ma
PFC	81arc1	RC : J'allais te le dire, (pause) et	qu'est-ce que tu veux que	je dise moi je.
PFC	81arc1	et on n'aurait pas pu euh participer à cette évolution	dont on ressentait qu'	on avait vraiment besoin
PFC	91ael1	LE: je me dis bah il y a certaines choses peut être	qu'il y aurait pas fallu que	, on se on se que j'aurais peut être pas du faire comme ça qu'il aurait fallu que je fasse autrement
PFC	91amb1	E oui mais	comment ls veulent qu'	on qu'on consomme on a plus rien on a plus
PFC	81ajm1	connaît euh qui a peut-être peut-être pourra répondre à à	combien tu penses qu'	il y a de personnes à Lacaune qui parlent le pat

CFPP	Audin_7e	lus d'enfants en ce temps là je pense qu'effectivement c'étaient des gens croyants et c'étaient des gens qui avaient les moyens d'élever euh sept ou huit enfants et qui avaient des appartements dans	lequel il faut dire que	ces enfants s'entassaient un peu mais c'était à cette époque là c'était normal je veux dire qu'on envisageait pas que les enfants aient chacun leur chambre ça existait pas euh les enfants ils étaient
CFPP	daSilva_7e	parce que je je bouge beaucoup et euh c'est vrai que ça absolument rien à voir j'étais dans le dix-septième c'était un quartier qui est bon qui est très vivant très connu mais c'est un quartier dans	lequel euh je pense pas que	j'apprécierais autant de vivre que ici
CFPP	Teixeira_11e	c'est là où on construit vraiment des enfin des amitiés qu'on enfin	qu'on sait qu'	on gardera après ou qu'on fait des rencontres et
CRFP	PRI-PER-1	chaude le poil et un petit peu les filaments qu'il y a avec le poil tout partait bien gratté et voilà et après on ouvrait les cochons en deux et on découpait les cochons les euh les jambons euh les ventrèches les lards les costello les pieds les têtes to	où c'était prévu qu'	il devait aller on faisait des euh des confits de porc voyez
CFPP	Teixeira_11e	Julie Teixeira non mais c'est ça oui c'est c'est une période	où c'est vrai qu'	on se prend on se prend pas mal la tête pour des choses qu'en valent peut-être pas la peine mais c'est justement ça qui est qui est drôle et c'est là où on construit vraiment des enfin des amitiés qu'on enfin qu'on sait qu'on gardera après ou qu'on fait d
TCOF	T_Manon1_Lag	<|Adulte|>y parle dedans vas-y <|Adulte|>hein	qu'est-ce que <|Adulte|>tu veux que	je dise <|Adulte|>ah euh il fait beau aujourd'hui hein <|Adulte|>arrête
TCOF	T_Nesrine1_Sed	<|Adulte|>je marque du, 0/ maquillage <|Adulte|>après	qu'est-ce qu'<|Adulte|>il faut que	j'écrive <|Adulte|>ah oui les perles j'avais oublié les perles

5. Liste des verbes à complétives prise comme référence (tirée des corpus)

1. accepter
2. admettre
3. admirer
4. adorer
5. advenir (il)
6. affirmer
7. aimer
8. ajouter
9. amener
10. annoncer
11. anticiper
12. apercevoir (s'/0)
13. apparaître (il)
14. apprécier
15. appréhender
16. apprendre
17. argumenter
18. arriver (il)
19. asséner
20. assurer
21. attendre
22. attester
23. augurer
24. avancer
25. avouer
26. calculer
27. certifier
28. choquer
29. comprendre
30. compter
31. concéder
32. concevoir
33. conclure
34. confesser
35. confier
36. confirmer
37. connaître
38. considérer
39. constater
40. contester
41. convenir
42. craindre
43. crier
44. croire
45. décider
46. déclarer
47. découvrir
48. décréter
49. déduire
50. défier
51. demander (se/0)
52. demeurer (il)
53. démontrer
54. déplorer
55. désirer
56. détester
57. deviner
58. dire
59. douter (se/0)
60. écrire
61. embêter
62. empêcher
63. enchaîner
64. énerver
65. enlever
66. entendre
67. entrevoir
68. envisager
69. escompter
70. espérer
71. essayer
72. estimer
73. établir
74. étonner (s'/0)
75. être
76. étudier
77. éviter
78. exclure
79. exiger
80. expliquer
81. faire (ça)
82. faire gaffe
83. falloir (il)
84. gager
85. garantir
86. gêner
87. ignorer
88. il est avéré
89. il suit
90. imaginer (s'/0)
91. impliquer (ça)
92. importer (il)
93. imposer
94. indiquer
95. informer
96. insinuer
97. intégrer
98. interdire
99. juger
100. jurer
101. justifier
102. lire
103. maintenir
104. manifester
105. marronner
106. mentionner
107. mériter
108. mettre
109. minauder
110. montrer
111. murmurer
112. nier
113. noter
114. observer
115. ordonner
116. oser
117. oublier

118. paraître (il)
119. parier
120. parler
121. penser
122. percevoir
123. permettre
124. persuader
125. pleurer
126. poser
127. postuler
128. pouvoir (il se)
129. préciser
130. préconiser
131. préférer
132. pressentir
133. présumer
134. présupposer
135. prétendre
136. prévenir
137. prévoir
138. proclamer
139. promettre
140. proposer
141. prouver
142. raconter
143. rajouter
144. rappeler
145. réaffirmer
146. réaliser
147. recevoir
148. réclamer
149. reconnaître
150. redécouvrir
151. redouter
152. réfléchir
153. refuser
154. regretter
155. réitérer
156. remarquer
157. remercier
158. répéter
159. répondre
160. ressentir
161. ressortir (il)
162. rester (il)
163. résulter (il)
164. retenir
165. rétorquer
166. révéler
167. rêver
168. savoir
169. sembler (il)
170. sentir
171. seoir (il)
172. signaler
173. signifier (ça)
174. songer
175. souhaiter
176. souligner
177. soupçonner
178. soutenir
179. souvenir (se)
180. stipuler
181. subsister (il)
182. suffire (il)
183. suggérer
184. supplier
185. supposer
186. supputer
187. suspecter
188. tenir (il ne V qu'à X)
189. tolérer
190. trembler
191. trouver
192. valoir (il V mieux)
193. vérifier
194. voir
195. vouloir
196. avoir envie
197. avoir la certitude
198. avoir l'impression
199. avoir le sentiment
200. avoir peur
201. être certain
202. être convaincu
203. être d'avis
204. être heureux
205. être persuadé
206. être sûr
207. être surpris
208. prendre conscience
209. rendre compte
210. sembler important
211. il (m')est indifférent
212. il est bon
213. il est certain
214. il est clair
215. il est curieux
216. il est dommage
217. il est douteux
218. il est entendu
219. il est établi
220. il est étrange
221. il est évident
222. il est extraordinaire
223. il est impossible
224. il est improbable
225. il est incontestable
226. il est indispensable
227. il est intéressant
228. il est juste
229. il est manifeste
230. il est naturel
231. il est nécessaire
232. il est normal
233. il est possible
234. il est probable
235. il est prouvé
236. il est rare
237. il est souhaitable
238. il est sûr
239. il est temps
240. il est urgent
241. il est vrai
242. il est vraisemblable
243. il va de soi
244. il va sans dire

TABLE DES FIGURES

FIGURE.1 Arbre en constituants d'une phrase complexe (Fromkin, Rodman, & Hyams, 2010, p. 133) 26

FIGURE.2 Arbre en constituants d'une interrogative (transformée à partir d'une déclarative) (Fromkin et al., 2010, p. 156) 26

FIGURE.3 Formation d'un arbre de dépendances non bornées en TAG (B. Crabbé) 29

FIGURE.4 Représentation finale des dépendances non bornées en TAG (Abeillé, 1993, p. 248) 29

FIGURE.5 Représentation des dépendances non bornées en HPSG (Abeillé, 1993, p. 188) 30

FIGURE.6 Réseau de dépendance – ordre linéaire (Mertens, 2008, p. 89) 31

FIGURE.7 Réseau de dépendance – ordre structural (Mertens, 2008, p. 89) 32

FIGURE.8 Réseau de dépendance – Combinaison des ordres (Brunet-Manquat, 2000, p. 20) 32

FIGURE.9 Réseau de dépendance des exemples (17)(18) 33

FIGURE.10 Réseau de dépendance des 'dépendances à distance' (Dikovsky & Modina, 2000, p. 87) 33

FIGURE.11 Réseau de dépendance des relatives (Mel'čuk, 2003, p. 68) 34

FIGURE.12 Réseau de dépendance d'une phrase avec complétive et relative simple (Vergne, 1998, p. 8) 34

FIGURE.13 Réseau de dépendance d'une relative complexe 35

FIGURE.14 Arbre à bulle simplifié d'une dépendance à distance en infinitive 35

FIGURE.15 Arbre à bulle simplifié d'une dépendance à distance en complétive 36

FIGURE.16 Représentation de l'intersubjectivité de la troisième personne (Verhagen, 2006, p. 330) 79

FIGURE.17 Nombre de dépendances à distance repérées avec un empan de n mots entre le pronom et la complétive 135

FIGURE.18 Réseau de dépendance d'une relative complexe dans le French TreeBank : *homme que je pense que j'ai vu* 140

FIGURE.19 Corrélation entre le nombre d'occurrences de dépendance à distance et la richesse du lexique verbal 171

FIGURE.20 Représentation des verbes ponts parmi les verbes des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique) 173

FIGURE.21 Représentation des verbes ponts parmi les verbes à complétive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique)..... 174

FIGURE.22 Représentation des verbes non-ponts parmi les verbes à complétive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique des fréquences)..... 177

FIGURE.23 Répartition des occurrences de dépendance à distance relevées..... 207

FIGURE.24 Répartition dans les corpus oraux des occurrences de dépendance à distance 207

FIGURE.25 Répartition dans les corpus écrits des occurrences de dépendance à distance..... 208

FIGURE.26 Hiérarchie des critères d'échantillonnage retenus 220

FIGURE.27 Répartition des occurrences de dépendances à distance dans les catégories définies pour l'échantillonnage 223

FIGURE.28 Répartition des occurrences de dépendances à distance dans le corpus CERF..... 226

FIGURE.29 Répartition des occurrences de dépendance à distance selon les familles dans un corpus adultes-enfants anglais : Manchester (pour un million de mots) 240

FIGURE.30 Représentation des verbes ponts parmi les verbes à infinitive des corpus, ordonnés du plus fréquent au moins fréquent (échelle logarithmique du nombre d'occurrences) 250

FIGURE.31 Comparaison des fréquences théoriques et attestées des verbes ponts dans les dépendances à distance en infinitive..... 253

FIGURE.32 Comparaison des fréquences théoriques et attestées des verbes ponts dans les dépendances à distance en complétive 253

FIGURE.33 Arbre à bulle d'une dépendance à distance en complétive (Kahane, 2000, p. 267) 277

TABLE DES TABLEAUX

TABLEAU 1. Liste des principaux constituants et syntagmes utilisés dans une représentation arborescente 25

TABLEAU 2. Phénomènes complexes en grammaire de constituants et en grammaire de dépendances 43

TABLEAU 3. Combinaisons des structures syntaxiques et des fonctions de l'élément extrait dans les constructions de dépendance à distance 45

TABLEAU 4. Interprétation de l'adverbe et conséquences sur le verbe de la matrice, la subordonnée et la dominance sémantique 55

TABLEAU 5. Degré d'acceptabilité des verbes du champ sémantique de DIRE en verbe pont.. 59

TABLEAU 6. Classe des noms qui facilitent ou non l'extraction 63

TABLEAU 7. Résultat de l'expérience de (Ambridge & Goldberg, 2008) 66

TABLEAU 8. Observation de huit verbes allemands selon différents critères 68

TABLEAU 9. Proportion d'occurrences de dépendances à distance pour 1 million de mots dans un corpus adultes-enfants et dans un corpus adultes 77

TABLEAU 10. Récapitulatif des contraintes pouvant avoir cours dans les constructions de dépendance à distance 81

TABLEAU 11. Structure du corpus C-Oral-Rom, d'après (Moneglia, s. d., p. 12) 95

TABLEAU 12. Présentation globale des corpus sources 101

TABLEAU 13. Synthèse des conventions adoptées par chaque corpus 104

TABLEAU 14. Liste des fichiers supprimés dans chaque corpus 110

TABLEAU 15. Correspondance entre les corpus : fichiers au contenu identique 111

TABLEAU 16. Présentation globale des corpus oraux, après uniformisation d'une partie d'entre eux 114

TABLEAU 17. Comparaison du nombre de caractères de chaque corpus avant et après standardisation 115

TABLEAU 18. Requête des dépendances à distance en complétive 121

TABLEAU 19. Requête des dépendances à distance en infinitive 124

TABLEAU 20. Synthèse des requêtes effectuées dans un premier temps 125

TABLEAU 21. Comparaison des résultats fournis par le concordancier sur les corpus bruts et standardisés 131

TABLEAU 22. Répartition des résultats par interrogatif-relatif, avant et après standardisation 131

TABLEAU 23. Requête souple, permettant notamment de repérer dans l'empan des tokens séparés d'une apostrophe..... 132

TABLEAU 24. Résultats attendus pour la requête « qu' {2 mots} qu » sur l'exemple (196)..... 133

TABLEAU 25. Requêtes fournies à CooLoX et nombre de résultats obtenus 136

TABLEAU 26. Requêtes avec un verbe spécifique fournies à CooLoX et nombre de résultats obtenus 136

TABLEAU 27. Synthèse de la totalité des requêtes effectuées 138

TABLEAU 28. Répartition des occurrences d'interrogatives directes entre la forme la plus fréquente et les formes subissant une ou plusieurs modifications 151

TABLEAU 29. Comparaison des interrogatives directes en français et en anglais 151

TABLEAU 30. Typologie des verbes liés à la présupposition (Mørdrup, 1975)..... 159

TABLEAU 31. Verbes ponts du corpus classés par (Mørdrup, 1975) en fonction de leur trait de factivité 159

TABLEAU 32. Test de factivité et classement des verbes du corpus non étudiés par (Mørdrup, 1975) 162

TABLEAU 33. Fréquence et rang des verbes recteurs faibles parmi les verbes ponts 166

TABLEAU 34. Augmentation parallèle du nombre de mots et du nombre d'occurrences de dépendance à distance dans un corpus [oral, privé, entretiens]..... 170

TABLEAU 35. Augmentation parallèle du nombre de mots et du nombre d'occurrences de dépendance à distance dans un corpus général [oral tout venant] 170

TABLEAU 36. Rang et fréquence de 23 verbes dans trois situations : verbe, verbe+complétive, verbe pont 178

TABLEAU 37. Liste des quinze verbes à complétive les plus fréquents qui sont non-ponts et des quinze verbes à complétive les plus fréquents qui sont ponts 181

TABLEAU 38. Définition des verbes non-ponts (parmi les 15 verbes à complétive non-ponts les plus fréquents) 182

TABLEAU 39. Définition des verbes ponts (parmi les 15 verbes à complétive ponts les plus fréquents) 182

TABLEAU 40. Répartition des modaux en ponts/non-ponts 184

TABLEAU 41. Degré de complexité de certains verbes de DIRE d'après (Godard, 1980)..... 185

TABLEAU 42. Mise en regard du degré de complexité de certains verbes de DIRE et de leur statut pont/non-pont 186

TABLEAU 43. Rapport entre nombre d'occurrences, complexité et position de pont des verbes complexes 190

TABLEAU 44. Rapport entre nombre d'occurrences, complexité et position de pont des verbes simples 190

TABLEAU 45. Répartition des sujets pronominaux du verbe pont, personnels et impersonnels 192

TABLEAU 46. Proportion du mode des verbes en construction pont et dans l'ensemble des constructions 195

TABLEAU 47. Proportion du temps présent des verbes en construction pont et dans l'ensemble des constructions 195

TABLEAU 48. Répartition des sujets pronominaux du verbe à complétive, personnels et impersonnels 197

TABLEAU 49. Comparaison des sujets pronominaux des verbes à complétive et des verbes ponts 198

TABLEAU 50. Corrélation entre longueur du pont et absence d'ambiguïté syntaxique entre le complément antéposé et le verbe pont..... 205

TABLEAU 51. Métadonnées fournies pour six des corpus oraux. 218

TABLEAU 52. Répartition des six corpus oraux dans les catégories définies pour l'échantillonnage 221

TABLEAU 53. Répartition des transcriptions avec dépendance à distance dans les catégories définies pour l'échantillonnage..... 222

TABLEAU 54. Répartition des occurrences de dépendance à distance dans les catégories définies pour l'échantillonnage 222

TABLEAU 55. Proportion de fichiers avec dépendances à distance dans chaque catégorie définie pour l'échantillonnage, relativement au nombre de fichiers de chaque catégorie..... 223

TABLEAU 56. Nombre d'occurrences de dépendance à distance pour un million de mots, dans chaque catégorie définie pour l'échantillonnage 224

TABLEAU 57. Khi2 observé dans les catégories définies pour l'échantillonnage 224

TABLEAU 58. Probabilité maximum (loi hypergéométrique) et valeur observée de dépendance à distance pour chacune des catégories définies pour l'échantillonnage 225

TABLEAU 59. Comportement des dépendances à distance dans le groupe 1 du CERF 227

TABLEAU 60. Comportement des dépendances à distance dans le groupe 2 du CERF 228

TABLEAU 61. Comportement des dépendances à distance dans le groupe 3 du CERF 229

TABLEAU 62. Récapitulatif du comportement des dépendances à distance dans les trois groupes du CERF 229

TABLEAU 63. Répartition des sujets lexicaux dans les groupes 1, 2 et 3 du CERF 232

TABLEAU 64. Proportion de tutoiement et de vouvoiement dans les tranches du CERF 232

TABLEAU 65. Nombre de mots et d'occurrences de dépendances à distance dans les corpus anglais : comparaison d'un corpus adulte et d'un corpus adultes-enfants 235

TABLEAU 66. Nombre de mots et d'occurrences de dépendances à distance dans trois corpus adultes-enfants français 236

TABLEAU 67. Nombre de mots et d'occurrences de dépendances à distance dans trois corpus adultes-enfants anglais 237

TABLEAU 68. Nombre de mots et d'occurrences de dépendances à distance dans les corpus français : comparaison d'un corpus adultes et d'un corpus adultes-enfants 237

TABLEAU 69. Comparaison de deux constructions ponts dans un corpus adultes et un corpus adultes-enfants français 237

TABLEAU 70. Comparaison de deux constructions ponts dans quatre corpus adultes-enfants français 238

TABLEAU 71. Répartition des occurrences de dépendance à distance selon l'âge de l'interlocuteur enfant, dans trois corpus français 239

TABLEAU 72. Répartition des occurrences de dépendance à distance selon les familles dans un corpus adultes-enfants français : Lyon 239

TABLEAU 73. Comparaison des rangs de 10 verbes en Verbe+infinitif et en Verbe pont+infinitif 250

TABLEAU 74. Comparaison de la richesse du lexique des verbes ponts dans les structures à complétives et à infinitives 252

TABLEAU 75. Classes d'adverbes régis dans les dépendances à distance en infinitive 256

TABLEAU 76. Adverbes non-régis attestés en pont dans les dépendances à distance en infinitive 257

TABLEAU 77. Comparaison de la construction du pont dans les structures à complétives et à infinitives 258

TABLEAU 78. Répartition des verbes ponts attestés en structure infinitive dans les quatre classes de verbes définies par (Chu, 2008) 261

TABLEAU 79. Classement des verbes d'après différents critères morphosyntaxiques et syntaxiques – d'après (Blanche-Benveniste & Martin, 2010, p. 122) 262

TABLEAU 80. Comparaison des ponts en relative et en interrogative, dans les constructions complétives 279

TABLEAU 81. Comparaison de la construction du verbe pont en relative et en interrogative, dans les constructions complétives..... 280

TABLEAU 82. Comparaison de la construction du verbe pont en relative et en interrogative, dans les constructions infinitives avec verbe pont constructeur..... 281

TABLEAU 83. Comparaison des ponts en interrogative directe et indirecte, dans les constructions complétives 283

TABLEAU 84. Répartition des modaux dans les différentes configurations syntaxiques..... 286

TABLE DES IMAGES

IMAGE 1. Représentation prosodique de « le lendemain grande surprise » : Prénoyau – Noyau (Martin, 2006, p. 155) 40

IMAGE 2. Représentation prosodique de « à la caisse ils se pèsent » : Noyau – Postnoyau (Martin, 2006, p. 156) 40

IMAGE 3. Recherche des doublons dans les corpus 109

IMAGE 4. Doublons partiel dans les corpus 112

IMAGE 5. Solution adoptée pour les doublons partiels 112

IMAGE 6. Interface de recherche avancée d'AntConc 122

IMAGE 7. Le verbe pont comme reflet de la subjectivité (Web, humour) 185

BIBLIOGRAPHIE

Abeillé, A. (1993). *Les nouvelles syntaxes : Grammaires d'unification et analyse du français*. Paris: Armand Colin.

Abeillé, A. (2007). *Les grammaires d'unification*. Paris: Hermes science publications Lavoisier.

Abeillé, A., Clément, L., & Toussenel, F. (2003). Building a treebank for French. In A. Abeillé (Éd.), *Treebanks* (Springer., p. 165–187). Dordrecht: Kluwer.

Abeillé, A., Godard, D., & Sabio, F. (2008). Deux constructions à SN antéposé en français. *Actes du CMLF*.

Ambridge, B., & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, *19*(3), 357–389.

Antoine, J. Y., Letellier-Zarshenas, S., Nicolas, P., Schadle, I., & Caelen, J. (2002). Corpus OTG et ECOLE MASSY: vers la constitution d'une collection de corpus francophones de dialogue oral diffusés librement (Vol. 2, p. 319–324). Consulté à l'adresse http://www-clips.imag.fr/geod/User/jean.caelen/Publis_fichiers/CorpusDialogue.pdf

Apothéloz, D. (2003). Grammaticalisation ou différentiel de grammaticité? Le cas de la rection dite 'faible'. *Verbum*, *25*(3), 241-262.

Armstrong, N., & Boughton, Z. (1998). Identification and evaluation responses to a French accent: some results and issues of methodology. *Revue Parole*, *5-6*, 27–60.

Auchlin, A., Avanzi, M., Goldman, J.-P., & Simon, A. C. (2012). C-PROM corpus libre de parole multigenre. Consulté 13 janvier 2012, à l'adresse http://sites.google.com/site/corpusprom/

Audibert, L. (2001). *CooLoX [Computer Software]*. Université de Provence. Consulté à l'adresse http://www-lipn.univ-paris13.fr/~audibert/lox.php

Auzanneau, M. (2010, février 24). *La « langue des jeunes », Données et réflexions autour de l'usage de stagiaires de la formation continue en Ile de France*. MoDyCo. Consulté à l'adresse http://www.modyco.fr/fr/event/doc_download/476-24-02-10-auzanneaulanguedesjeunes.html

Bailly, S. (2007). Sexotypes langagiers et communication. In *Stéréotypage, stéréotypes: fonctionnements ordinaires et mises en scène: actes du colloque international de Montpellier, 21, 22 et 23 juin 2006, Université Montpellier III* (Vol. 2, p. 15). Montpellier. Consulté à l'adresse http://hal.archives-ouvertes.fr/docs/00/52/03/34/PDF/Bailly_2006_Sexotypes_langagiers_et_communication.pdf

Bally, C. (1942). Syntaxe de la modalité explicite. *Cahiers Ferdinand de Saussure*, (2), 3-13.

Baude, O. (2006). Corpus oraux: les bonnes pratiques d'une communauté scientifique. *Corpus, 7*, 22-73.

Baude, O., Blanche-Benveniste, C., Calas, M. F., Cappeau, P., Cordereix, P., Goury, L., & Mondada, L. (2006). *Corpus oraux, guide des bonnes pratiques 2006*. CNRS Éd.

Beeching, K. (2007). La co-variation des marqueurs discursifs bon, c'est-à-dire, enfin, hein, quand même, quoi et si vous voulez : une question d'identité ? *Langue française*, *154*(2), 78.

Bender, E. M., Sag, I. A., & Wasow, T. (2003). *Long distance dependencies*. Consulté à l'adresse http://hpsg.stanford.edu/book/slides/Ch14a.pdf

Benveniste, É. (1966). *Problèmes de linguistique générale. 1* (Vol. 1-1). Paris: Gallimard. Consulté à l'adresse http://www.sudoc.fr/070775656

Benzitoun, C. (2006, décembre 1). *Description morphosyntaxique du mot quand en français contemporain*. Université d'Aix-en-Provence.

Benzitoun, C., Fort, K., & Sagot, B. (2012). TCOF-POS : un corpus libre de français parlé annoté en morphosyntaxe. *Actes de JEP-TALN-RECITAL 2012*, *2*, 99-112.

Berrendonner, A. (1990). Pour une macro-syntaxe. *Travaux de linguistique*, (21), 25–36.

Biber, D. (1991). *Variation across speech and writing*. Cambridge Univ Press.

Biber, D., Conrad, S., & Reppen, R. (2002). *Corpus Linguistics: Investigating Language Structure and Use* (Vol. 4). Cambridge: Cambridge University Press.

Bilger, M. (Éd.). (2000a). *Corpus. Méthodologie et applications linguistiques, Actes du colloque international Questions de méthodes dans la linguistique sur corpus. 1998*. Paris: Honoré Champion et Presses Universitaires de Perpignan.

Bilger, M. (2000b). *Linguistique sur corpus: études et réflexions*. Presses universitaires de Perpignan.

Blache, P. (2001). Dépendances à distance dans les grammaires de propriétés : l'exemple des disloquées. *Actes de TALN*, *1*, 103-112.

Blanc, O., Constant, M., Dister, A., & Watrin, P. (2008). Corpus oraux et chunking. *Journées d'étude sur la parole (JEP)*. Consulté à l'adresse http://igm.univ-mlv.fr/~mconstan/papers/blanc_constant_dister_watrin_JEP08.pdf

Blanche-Benveniste, C. (Non publié). *Un statut de chaîne verbale, Mélanges offerts à G. Lazard*.

Blanche-Benveniste, C. (1989). Constructions verbales en incise et rection faible des verbes. *Recherches sur le français parlé*, (9), 53-73.

Blanche-Benveniste, C. (1990). Grammaire première et grammaire seconde: l'exemple de EN. *Recherches sur le français parlé, 10*, 51–74.

Blanche-Benveniste, C. (1998). Ponctuation et langue parlée. *Le discours psychanalytique*, 73–109.

Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français.* Ophrys.

Blanche-Benveniste, C., Bilger, M., Rouget, C., Van Den Eynde, K., Mertens, P., & Willems, D. (1990). *Le français parlé: études grammaticales*. CNRS éd.

Blanche-Benveniste, C., Deulofeu, J., Stéfanini, J., & Van den Eynde, K. (1987). *Pronom et syntaxe: l'approche pronominale et son application au français* (2ème édition augmentée.). Paris: SELAF.

Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé: transcription et édition*. Paris: Didier Erudition.

Blanche-Benveniste, C., & Martin, P. (2010). *Le Français : Usages de la langue parlée*. Peeters Leuven.

Blanche-Benveniste, C., Rouget, C., & Sabio, F. (2002). *Choix de textes de français parlé: 36 extraits*. Champion.

Blanche-Benveniste, C., & Willems, D. (2007). Un nouveau regard sur les verbes «faibles». *Bulletin de la Société de Linguistique de Paris*, *102*(1), 217–254.

Blanche-Benveniste, C., & Willems, D. (2009). Verbes' faibles' et verbes à valeur épistémique en français parlé: il me semble, il paraît, j'ai l'impression, on dirait, je dirais. *Proceedings of the International Congress of Romance Languages and Linguistics*.

Bois, G. (2009). *Les écrivains dominés du jeu littéraire. Définition de l'espace d'investissement et rapports aux enjeux littéraires* (Sociologie). Lyon 2. Consulté à l'adresse http://recherche.univ-lyon2.fr/grs/index.php?page=39&notice=14

Bourdieu, P. (1986). Habitus, code et codification. *Actes de la recherche en sciences sociales*, *64*(1), 40-44.

Bourdieu, P. (1992). *Les règles de l'art: genèse et structure du champ littéraire*. Paris: Seuil.

Boxell, O. (2012). Discourse-linking and long-distance syntactic dependency formation in real-time. *Proceedings of ConSOLE XIX*, 151-175.

Branca-Rosoff, S. (2007). Normes et genres de discours le cas des émissions de libre antenne sur les radios jeunes. *Langage et société (Maison des Sciences de l'Homme)*, *119*, 111-128, 207-208.

Branca-Rosoff, S., Fleury, S., Lefeuvre, F., & Pires, M. (2009a). Constitution et exploitation d'un corpus de français parlé parisien.

Branca-Rosoff, S., Fleury, S., Lefeuvre, F., & Pires, M. (2009b). Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000). Consulté 13 janvier 2012, à l'adresse http://cfpp2000.univ-paris3.fr/

Brunet, E. (2004). Où l'on mesure la distance entre les distances. *Texto ! [en ligne]*, *mars 2004*. Consulté à l'adresse http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html

Brunet-Manquat, F. (2000). Représentation adaptée pour l'analyse de dépendance. *Actes de Magistère*, 19-22.

Bruxelles, S., Mondada, L., Traverso, V., & Simon, A. C. (2009). *Grands corpus de français parlé : bilan historique et perspectives de recherches*.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711-733.

Bybee, J. (2007). Compte-rendu - Adele Goldberg, Constructions at work: The nature of generalization in language. Oxford: Oxford University Press, 2006. Pp. 280. *Journal of Child Language*, *34*(03), 692–697.

Canut, E., André, V., & Gaiffe, B. (2010). Mise à disposition de corpus oraux interactifs: le projet TCOF (Traitement des Corpus Oraux en Français). *Pratiques, 147-148*, 35-51.

Cappeau, P., & Seijido, M. (2005). Les corpus oraux en français. Inventaire. Consulté à l'adresse http://www.dglflf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 119). The MIT press.

Chomsky, N. (1977). On WH-movement. In P. Culicover, T. Wasow, & A. Akmajian (Éd.), *Formal syntax* (Academic Press., p. 71-132). New York: New York: Academic Press.

Chu, X. (2008). *Les verbes modaux du français*. Ophrys.

CLAPI-concordancier [Computer Software]. (s. d.). Consulté à l'adresse http://clapi.univ-lyon2.fr/req_nouveau.php?%20requete=CONCOR

Cresti, E., & Scarano, A. (2000). Sur la notion de parlé spontané. (M. Bilger, Éd.) *Corpus. Méthodologie et applications linguistiques.*, 340–349.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax* (1re éd.). Oxford University Press, USA.

Dąbrowska, E. (2008). Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics*, *19*(3), 391-425.

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*, *27*(1), 1-23.

Dąbrowska, E., Rowland, C., & Theakston, A. (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics*, *20*(3), 571–597.

Debaisieux, J.-M. (sous presse). *Apports linguistiques des analyses sur corpus. Subordination et insubordination en français*. Hermes.

Debaisieux, J.-M. (2005). Les corpus oraux: situation, exploitation linguistique, bilan et perspectives. *Scolia*, (19), 7-39.

DELIC. (2004). Présentation du Corpus de référence du français parlé. *Recherches sur le français parlé, 18*, 11-42.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of child language*, *35*(1), 99–127.

Denis, D., & Sancier-Chateau, A. (1994). *Grammaire du français* (Le livre de poche.). Paris: LGF.

Denis, P., & Sagot, B. (2009). Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art pos tagging with less human effort. In *Proceedings of PACLIC 2009*. Hong Kong, Chine.

Deulofeu, J. (2003). L'approche macrosyntaxique en syntaxe : un nouveau modèle de rasoir d'Occam contre les notions inutiles. *Scolia*, *16*, 77–95.

Deulofeu, J., & Blanche-Benveniste, C. (2006). C-Oral-Rom: French Corpus. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Éd.), *Spoken Language Corpus and Linguistic Informatics* (p. 181-198). Amsterdam, Netherlands: Benjamins.

Dikovsky, A., & Modina, L. (2000). Dependencies on the other side of the curtain. *Traitement Automatique des Langues (TAL)*, *41*(1), 79–111.

Dister, A. (2007). *De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelles orales VALIBEL*. Thèse de doctorat, Université de Louvain.

Dister, A., Francard, M., Hambye, P., & Simon, A. C. (2009). Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de banque de données textuelles orales VALIBEL (1989-2009). *Cahiers de l'Institut de linguistique de Louvain-CILL*, *33*(2), 113–129.

Durand, J., Laks, B., & Lyche, C. (2003). PFC : Un corpus numérisé pour la phonologie du français. In G. Williams, *Les linguistiques de corpus*. Rennes:Presses Universitaires de Rennes.

Durand, J., Laks, B., & Lyche, C. (2009). Le projet PFC: une source de données primaires structurées. In J. Durand, B. Laks, & C. Lyche (Éd.), *Phonologie, variation et accents du français* (p. 19-61). Paris: Hermes science publications.

Durand, J., & Lyche, C. (2003). Le projet 'Phonologie du Français Contemporain' (PFC) et sa méthodologie. In É. Delais-Roussarie & J. Durand (Éd.), *Corpus et variation en phonologie du français: méthodes et analyses* (p. 213-276). Toulouse: Presses Univ. du Mirail. Consulté à l'adresse http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=jdurand&subURL=LeProjetPFCetsaMethodologie.pdf

Durand, J., & Tarrier, J. M. (2006). «PFC, corpus et systèmes de transcription». *Cahiers de grammaire*, *30*, 139–158.

EAGLES. (1996). Preliminary recommendations on spoken texts. Consulté à l'adresse http://www.google.com/url?sa=t&rct=j&q=%22preliminary%20recommendations%20on%20spoken%20texts%22&source=web&cd=1&sqi=2&ved=0CCQQFjAA&url=http%3A%2F%2Fwww.ilc.cnr.it%2FEAGLES96%2Fpub%2Feagles%2Fcorpora%2Fspokentx.ps.gz&ei=sGWNTornJuam0QXzhbg6&usg=AFQjCNHfK03endlhZNvIO6zdad7aOHXngg&sig2=zNBhzoaJOQYC8ApTu7PNoA

Eckert, P. (1997). Age as a sociolinguistic variable. In F. Coulmas (Éd.), *The Handbook of Sociolinguistics* (p. 151-167). Oxford: Blackwell. Consulté à l'adresse https://www.msu.edu/course/lin/471/snapshot.afs/Articles-FS06/Eckert%201997.pdf

Eckert, P., & McConnell-Ginet, S. (1999). New generalizations and explanations in language and gender research. *Language in society*, *28*(2), 185–201.

Erteschik-Shir, N. (1973). *On the nature of island constraints.* (Thesis). Massachusetts Institute of Technology, Cambridge. Consulté à l'adresse http://dspace.mit.edu/handle/1721.1/12991

Eshkol, I., Tellier, I., Taalab, S., & Billot, S. (2010). Etiqueter un corpus oral par apprentissage automatique à l'aide de connaissances linguistiques. *Arxiv preprint arXiv:1003.5749*. Consulté à l'adresse http://arxiv.org/ftp/arxiv/papers/1003/1003.5749.pdf

F agyal, Z. (2003). La prosodie du français populaire des jeunes: traits héréditaires et novateurs. Le Français aujourdhui, (143 - n°spécial « Français de l'école et langues des élèves : quel statut, quelles pratiques ? »), 47-55.

Featherston, S. (2005). Bridge verbs and V2 verbs–the same thing in spades? *Zeitschrift für Sprachwissenschaft*, *23*(2), 181–209.

Fletcher, P., & Garman, M. (1988). Normal language development and language impairment: syntax and beyond. *Clinical linguistics and phonetics*, *2*(2), 97–113.

Ford, M., & Bresnan, J. (2010). Studying syntactic variation using convergent evidence from psycholinguistics and usage. *Research Methods in Language Variation and Change*. Consulté à l'adresse http://www.stanford.edu/%7Ebresnan/ford.bresnan.2010.doc

Fromkin, V., Rodman, R., & Hyams, N. (2010). *An introduction to language*. Wadsworth Pub Co.

Gachet, F. (2012). *Incise de dicours rapporté et autres parenthétiques: une étude grammaticale*. Université de Fribourg, Fribourg (Suisse).

Gachet, F., & Avanzi, M. (2009). Description prosodique des 'recteurs faibles en incise'. Consulté à l'adresse http://www2.unine.ch/files/content/sites/structuration_periodes/files/shared/new_am/49_.pdf

Gadet, F. (2008). Les corpus oraux et la diversité des productions langagières. *Verbum*, *XXX*(4), 261-273.

Gardelle, L. (2010). Dislocations à gauche et antépositions de compléments en anglais : formes, fonctionnalités et insertion en discours. *Discours*, (7).

Ghanem Chentli, C. (2010). *Analyse de discours : l'opposition masculin/féminin à travers une émission radiophonique d'Alger (chaine 3). Le cas de l'alternance codique*. Mémoire de Master, Université de Constantine, Alger. Consulté à l'adresse http://ns.umc.edu.dz/theses/francais/CHE1084.pdf

Godard, D. (1980). Les Relatives parenthétiques du français : la contrainte du syntagme nominal complexe (CNPC) et les « verbes-ponts ». *Linx*, (2), 35-88.

Godard, D. (1988). *La syntaxe des relatives en français*. Editions du CNRS, Sciences du langage. Paris, 237p.

Gross, M. (1975). *Méthodes en syntaxe régime des constructions complétives*. Paris: Hermann.

Guiraud, P. (1966). Le système du relatif en français populaire. *Langages*, *1*(3), 40-48.

Habert, B. (2000). Des corpus représentatifs : de quoi, pour quoi, comment. *Cahiers de l'Université de Perpignan*, *31*, 11–58.

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting*. New York: Oxford University Press, USA. Consulté à l'adresse http://books.google.com/books?hl=en&lr=&id=Q7Jcdt1ft_EC&oi=fnd&pg=PA3&dq=Belfast+English+and+Standard+English:+Dialect+variation+and+parameter+setting.&ots=7UREXgVYPJ&sig=weM1YWKyAyifS5rjMROPNPgUp1g

Hirschbuhler, P., & Labelle, M. (1996). Qui sujet conjonction ou pronom relatif ? In *Dépendance et intégration syntaxique, Subordination, coordination, connexion* (p. 67-76). Tubingen: M. Niemeyer.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366–415.

Houdebine-Gravaud, A.-M. (2003). Trente ans de recherche sur la différence sexuelle, ou Le langage des femmes et la sexuation dans la langue, les discours, les images. *Langage et société*, *4*(106), 33-61.

Huot, H. (1981). *Constructions infinitives du français : le subordonnant de*. Genève.

ICOR, G. (s. d.). Exploitation de la plateforme Corpus de Langue Parlée en Interaction (CLAPI) pour l'étude d'un phénomène interactionnel: l'usage des particules discursives dans les chevauchements de parole. *Cahiers de linguistique de Louvain*.

Joshi, A. K. (1987). An Introduction to Tree Adjoining Grammar. *The Mathematics of Language*, 87-114.

Joshi, A. K., Levy, L. S., & Takahashi, M. (1975). Tree Adjunct Grammars. *Journal of Computer and Systems Sciences*, *10*(1), 55-75.

Juola, P. (2006). Questioned Electronic Documents: Empirical Studies in Authorship Attribution. In Olivier & Shenoi (Éd.), *Research Advances in Digital Forensics II*. Heidelberg: Springer. Consulté à l'adresse http://www.mathcs.duq.edu/~juola/papers.d/forensics.pdf

Juola, P. (2008). *Authorship attribution* (Vol. 1). Delft: Now Publishers. Consulté à l'adresse http://www.mathcs.duq.edu/~juola/papers.d/fnt-aa.pdf

Kaan, E., & Stowe, L. (2002). Storage and Computation in Sentence Processing a Neuroimaging Perspective. In S. Nooteboom, F. Weerman, & F. Wijnen (Éd.), *Storage and computation in the language faculty* (p. 257–295). Kluwer Academic Publishers. Consulté à l'adresse http://www.let.rug.nl/neurolinguistics/pub/Kaan%20and%20Stowe%202002.pdf

Kahane, S. (1997). Bubble trees and syntactic representations. In Becker & Krieger (Éd.), *Proc. 5th Meeting of the Mathematics of Language (MOL5)* (p. 70-76). Saarbrücken: DFKI.

Kahane, S. (2000). Extractions dans une grammaire de dépendance lexicalisée à bulles. *Traitement Automatique des Langues*, *41*(1), 187–216.

Kahane, S. (2002). À propos de la position syntaxique des mots qu-. (P. Le Goffic, Éd.) *Verbum*, *24*(4), 399-433.

Kahane, S. (2008). Le rôle des structures et représentations dans l'évolution des théories syntaxiques. (G. Lecointre & J. Pain, Éd.) *Evolution : méthodologie, concepts, Les cahiers de l'Ecole Doctorale « Connaissance, Langage, Modélisation », Université Paris X, Nanterre*. Consulté à l'adresse http://www.kahane.fr/?u_act=download&dfile=DG-Darwin2008.pdf&

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, *17*(4), 401-412.

Kuczaj, I. I., & Stan, A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 589–600.

Labelle, F. (1990). Développements récents en analyse syntaxique par ordinateur. *Dialangue*, *1*, 20-22.

Lablanche, A. (2007). *L'infinitif complément d'un verbe, d'un adjectif, d'un nom : écriture d'un fragment de grammaire* (Thèse de doctorat). Université de Paris-Nanterre.

Laks, B. (1980). *Différenciation linguistique et différenciation sociale: quelques problemes de linguistique française*. Université de Paris VIII-Vincennes.

Laurent, D. (2008). *Cordial Analyser [Computer Software]*. Toulouse: Synapse Développement. Consulté à l'adresse http://www.synapse-fr.com/

Lebaud, D., Paulin, C., & Ploog, K. (2006). *Constructions verbales et production de sens*. Presses Univ. Franche-Comté.

Lehka-Lemarchand, I. (2007). *Accent de banlieue: approche phonétique et sociolinguistique de la prosodie des jeunes d'une banlieue rouennaise*. Université de Rouen, Rouen.

Consulté à l'adresse http://www.u-picardie.fr/LESCLaP/IMG/pdf/Lehka-Lemarchand__these_Vol_1_cle81a173.pdf

Lentin, L. (1975). Problématique de l'acquisition de la syntaxe chez le jeune enfant. *Langue française*, *27*(1), 14–23.

Leroy, M., & Mathiot, E. (2009). Pointing gestures and demonstrative words: Deixis between the ages of one and two. In A. Morgenstern, M. Johansson Falck, C. Lundmark, & M. Andrén (Éd.), *Studies in Language and Cognition* (p. 402-420). Cambridge: Cambridge Scholars Publishing.

Luzzati, D. (2009). Corpus d'hier et d'aujourd'hui : progrès quantitatifs ou progrès qualitatifs. *Cahiers de linguistique*, *32*(2), 97-111.

Lyche, C. (2005). Pour un renouvellement des données phonologiques : Le corpus PFC « phonologie du français contemporain » (French). *Lidil (Grenoble)*, (31), 119-138.

Lyche, C. (2007). Building a Large Corpus for Phonological Research: The PFC Project. In Y. Kawaguchi, T. Takagaki, N. Tomimori, & Y. Tsuruga (Éd.), *Corpus-Based Perspectives in Linguistics* (p. 91-113). Amsterdam, Netherlands: Benjamins.

Lyche, C., Detey, S., Durand, J., & Laks, B. (2010). *Les variétés du français parlé dans l'espace francophone : Ressources pour l'enseignement*. Editions Ophrys.

Martin, P. (2006). Intonation du français: parole spontanée et parole lue. *Estudios de Fonética experimental*, *15*, 134–162.

Martinet, A. (1945). *La prononciation du français contemporain; témoignages recueillis en 1941 dans un camp d'officiers prisonniers*. Paris: E. Droz.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. State University of New York Press.

Mel'čuk, I. (2003). Levels of dependency in linguistic description: Concepts and problems. (V. Agel, L. Eichinnger, H.-W. Eroms, P. Hellwig, H. J. Herringer, & H. Lobin, Éd.)

Dependency and valency. An International handbook of contemporary research, *1*, 188–229.

Merchant, J. (2006). Sluicing. In M. Everaert & H. Van Riemsdijk (Éd.), *The Syntax Companion* (Blackwell., p. 269-289). London. Consulté à l'adresse http://home.uchicago.edu/~merchant/pubs/SynCom.sluicing.pdf

Mertens, P. (2008). Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours. *Travaux de linguistique*, *56*(1), 97.

Ministère de la Culture et de la Communication, & Centre National de la Recherche Scientifique. (s. d.). Corpus de la parole : comment on parle en France aujourd'hui. Consulté 6 avril 2012, à l'adresse http://corpusdelaparole.in2p3.fr/spip.php?article28

Moneglia, M. (s. d.). Specifications on the C-ORAL-ROM Corpus. Consulté à l'adresse http://lablita.dit.unifi.it/app/coralrom/papers/Specifications-CORALROM.pdf

Mørdrup, O. (1975). Présuppositions, implications et verbes français. *Revue romane*, 125–155.

Moreau, M. L. (1971). L'homme que je crois qui est venu. Qui, que : relatifs et conjonctions. *Langue française*, *11*(1), 77–90.

Morgenstern, A. (sous presse). L'enfant dans la langue : de l'observation du naturaliste à l'analyse du linguiste.

Morgenstern, A., & Parisse, C. (2007). Codage et interprétation du langage spontané d'enfants de 1 à 3 ans. *Corpus*, (6), 55–78.

Morgenstern, A., & Sekali, M. (2009). What can child language tell us about prepositions? A contrastive corpus-based study of cognitive and social-pragmatic factors. In J. Zlatev, M. Johansson Falck, C. Lundmark, & M. Andrén (Éd.), *Studies in Language and Cognition* (p. 261–275). Cambridge: Cambridge Scholars Publishing.

Podhorná-Polická, A. (2007). *Peut-on parler d'un argot des jeunes ? Analyse lexicale des universaux argotiques du parler de jeunes en lycées professionnels en France (Paris, Yzeure) et en République tchèque (Brno)*. Université René Descartes - Paris & Université Masaryk - Brno. Consulté à l'adresse http://is.muni.cz/th/12093/ff_d/THESE_Alena_Podhorna-Policka_TOME_1.pdf

Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar* (CSLI, Stanford). Chicago: Chicago Press University.

Rastier, F. (2004). Enjeux épistémologiques de la linguistique de corpus. *Texto ! [en ligne]*, 1-16.

Rigaud, N. (2010). *L'art du vide : pour une approche descriptive des phénomènes elliptiques* (Thèse de doctorat). Aix-Marseille 1, Aix en Provence. Consulté à l'adresse http://sites.google.com/site/rigaudnat/these

Romaine, S. (2003). Variation in language and gender. In J. Holmes & M. Meyerhoff (Éd.), *The Handbook of Language and Gender* (p. 98-119). Oxford: Blackwell. Consulté à l'adresse http://people.cohums.ohio-state.edu/schwenter1/Romaine%202003.pdf

Romary, L., Salmon-Alt, S., & Francopoulo, G. (2004). Standards going concrete: from lmf to morphalou. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries, Coling 2004* (p. 22). Genève, Suisse.

Ross, J. R. (1967). *Constraints on variables in syntax.* (Thesis). Massachusetts Institute of Technology. Dept. of Modern Languages and Linguistics. Consulté à l'adresse http://www.eric.ed.gov/PDFS/ED016965.pdf

Ross, J. R. (1969). Guess Who. *Proceedings of the 5th Annual Meeting of Chicago Linguistics Society*, 252-286.

Sabel, J. (2002). A minimalist analysis of syntactic islands. *The linguistic review*, (19), 271-315.

Sagot, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In *7th international conference on Language Resources and Evaluation (LREC 2010)*. La Vallette, Malte. Consulté à l'adresse http://hal.inria.fr/inria-00521242_v1/

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of the International Conference on New Methods in Language Processing*, 44-49.

Schmid, H. (2009). *TreeTagger [Computer Software]*. Consulté à l'adresse http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

Sinclair. (1996). EAGLES. Preliminary recommendations on Corpus Typology, *24,* 16-19.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sportiche, D. (2011). French Relative Qui. *Linguistic Inquiry*, *42*(1), 83-124.

Stein, A., & Schmid, H. (1995). Etiquetage morphologique de textes français avec un arbre de décisions. *Traitement automatique des langues*, *36*(1-2), 23–35.

Tellier, I., Eshkol, I., Taalab, S., & Prost, J. P. (2010). POS-tagging for Oral Texts with CRF and Category Decomposition. *Research in Computer Science, special issue: Natural Language Processing and its Applications*, 79-90.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klinksieck. Consulté à l'adresse http://data.bnf.fr/12048639/lucien_tesniere_elements_de_syntaxe_structurale/

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*(01), 127–152.

Tolone, E. (2011). *Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français* (Thèse de doctorat). Université Paris Est, Paris. Consulté à l'adresse http://www.sudoc.fr/158124170

Trimaille, C., & Billiez, J. (2007). Pratiques langagières dejeunes urbains: peut-on parler de «parler»? *Les français en émergence*, *21*, 95.

Urmson, J. O. (1952). Parenthetical Verbs. *Mind*, *61*(244), 480-496.

Valli, A., & Véronis, J. (1999). Etiquetage grammatical des corpus de parole: problemes et perspectives. *Revue française de linguistique appliquée*, *4*(2), 113–133.

Van den Eynde, K., & Mertens, P. (2003). La valence: l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, *13*(1), 63–104.

Vergne, J. (1998). Entre arbre de dépendance et ordre linéaire, les deux processus de transformation: linéarisation, puis reconstruction de l'arbre. *Cahiers de grammaire*, *23*. Consulté à l'adresse https://users.info.unicaen.fr/~jvergne/Cahiers_de_Grammaire_Vergne.pdf

Verhagen, A. (2005). *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford University Press, USA.

Verhagen, A. (2006). On subjectivity and « long distance Wh-movement ». In A. Athanasiadou, C. Canakis, & B. Cornilie (Éd.), *Subjectification: Various Paths to Subjectivity* (p. 323-346). Berlin/New York: Mouton de Gruyter.

Véronis, J. (1998). *Contextes-concordancier [Computer Software]*. Consulté à l'adresse http://sites.univ-provence.fr/veronis/logiciels/Contextes/ContextesHelpFr/index.htm

Wilson, J., & Henry, A. (1998). Parameter setting within a socially realistic linguistics. *Language in Society*, *27*(01), 1–21.

Zumthor, P. (2008). Oralité. *Intermédialités : histoire et théorie des arts, des lettres et des techniques*, (12), 169-202.