HAL
open science

Dépendances à longue distance et genres textuels

Lolita Bérard

To cite this version:

Lolita Bérard. Dépendances à longue distance et genres textuels. 4ème Congrès Mondial de Linguistique Française, Jul 2014, Berlin, Allemagne. pp.2349-2365, 10.1051/shsconf/20140801336 . hal-01103585

HAL Id: hal-01103585

https://hal.science/hal-01103585

Submitted on 2 Feb 2015

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

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

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

Dépendances à longue distance et genres textuels

Bérard, Lolita

Université Paris 3 Sorbonne Nouvelle et LaTTiCe (UMR 8094)
lolita.berard@atilf.fr

1 Introduction

Les dépendances à longue distance (long distance dependency – LDD) sont des constructions qui mettent en jeu un interrogatif-relatif dépendant d'un verbe traditionnellement analysé comme enchâssé : « **qu'est-ce que** tu *veux* que je **fasse** » (Chomsky, 1977, Ross, 1967). La construction verbale placée entre les deux dépendants est nommée « pont » (terminologie d'Erteschik-Shir, 1973), et le verbe est nommé « verbe pont ».

Nous limitons notre étude aux énoncés qui appartiennent aux trois configurations syntaxiques suivantes : les interrogatives, configuration dans laquelle les dépendances à longue distance ont été le plus souvent étudiées, les relatives et les clivées. Dans tous les énoncés décrits, le verbe pont construit une subordonnée complétive, comme dans l'exemple (1) et non une infinitive, comme dans l'exemple (2).

(1) la symbolique de cette œuvre s'accorde parfaitement avec le lieu où son auteur a souhaité qu'elle soit implantée (CERF, POLI)

(2) Qu'est-ce que vous préférez euh fabriquer finalement euh des pots plats et objets utilitaires ou plutôt euh des objets euh plus figuratifs

Les exemples avec infinitive sont très nombreux dans les corpus, mais également très ambigus. Dès lors que l'élément antéposé n'est pas un complément sélectionné par le verbe enchâssé, il devient très difficile de savoir s'il y a une relation de dépendance avec le verbe à complétive ou avec le verbe enchâssé. Généralement, on considère que l'ajout a une portée large et qu'il modifie l'ensemble de l'énoncé. Dans l'exemple suivant, 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*).

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

Afin de constituer un corpus d'exemples de dépendance à longue distance cohérent et exhaustif, nous avons écarté dans cet article les occurrences en infinitive et soumis à la description uniquement les occurrences en complétive.

Nous écartons également les exemples de pseudo-dépendance à longue distance, dans lesquelles le pronom interrogatif-relatif réalisé à distance est repris dans la complétive ou l'infinitive. Le double marquage peut être réalisé sous forme pronominale ou lexicale. Ces constructions 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. (exemple 4).

(4) 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)

Néanmoins, on la trouve aussi 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 ? » (exemples 5, 6).

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

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

Article available at http://www.shs-conferences.org or http://dx.doi.org/10.1051/shsconf/20140801336

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

(6) 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)

Il semblerait que les pseudo-dépendances à longue distance en « dont » soient l'équivalent d'une dépendance à longue distance du sujet dont la forme est perçue comme fautive (exemple 7).

(7) 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).

Le fait qu'un pronom régi soit présent dans la complétive implique deux choses :

- d'une part, il n'y a pas réellement de dépendance à longue distance ; il n'y a qu'un contrôle sémantique entre le pronom interrogatif-relatif et le verbe.
- d'autre part, les contraintes sont forcément différentes des constructions sans reprise pronominale. D'après (Godard, 1988), la 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 générativistes (Godard, 1988 : 40-42). En revanche, la liste des verbes ponts des DPR est plus restreinte (Godard, 1988 : 71-73). Nous estimons également que les contraintes cognitives ou de processing doivent être différentes selon que le pronom est présent ou non, puisque la référence de l'élément antéposé est dans un cas éloignée et dans l'autre non.

C'est pourquoi nous avons écarté les constructions de pseudo-dépendance à longue distance de notre étude.

Les études sur corpus menées sur l'anglais et le néerlandais (Verhagen, 2006, Dąbrowska, 2008) ont révélé un emploi très particulier des constructions de dépendances à longue distance : le verbe pont et son sujet sont quasiment figés. La question est de savoir s'il existe en français un tel phénomène de figement ou de construction dominante et si par ailleurs la construction est sensible aux genres textuels. Pour répondre à la question de la variation en genres, nous nous appuierons sur un corpus de dix millions de mots, le CERF, constitué dans les années 2000 sous la direction de J. Véronis et divisé en dix tranches : neuf genres d'écrit et un d'oral (plutôt informel spontané – entretien/conversation). Chaque genre est représenté par un million de mots. Les neuf tranches d'écrit sont les suivantes :

- DIVS (Divers) : chansons, critiques cinéma, critiques littéraires, textes d'enfants, humour, journaux intimes, journaux de lycéens et d'étudiants, philosophie, religion, textes éducatifs
- FORM (Forum internet) : cinéma, culture, économie, média, monde, multimédia, politique, société, sports, témoignages
- INST (Institutionnels) : accords/traités internationaux, administration/service public, assemblée, parlement européen, Hansard, juridique/législatif, ministères, ONG, ONU
- LITA (Littérature Ancienne 1746-1923) : journaux/autobiographies, romans d'aventure/policiers/sciences fictions, contes, correspondances, littérature fantastique, nouvelles, poésies, romans, pièces de théâtre
- LITC (Littérature Contemporaine) : contes, littérature érotique, romans fantastique/d'horreur, nouvelles, nouvelles de science fiction, poésies, romans policiers, romans, pièces de théâtre
- POLI (Politique) : Chirac, Jospin, Laguiller, Mitterrand, PCF, PS, RPR, Syndicats, Verts
- PRAT (Vie Pratique et loisirs) : petites annonces, auto/moto/chasse/pêche, bourse/finance, cuisine, CV, magazines féminins, horoscopes, loisirs/sorties, presse people, sports

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

- PRES (Presse) : presse d'Afrique-Subsaharienne, Courrier international, L'Humanité, Le Monde, Le Monde Diplomatique, Le Nouvel Observateur, Journal de Saône et Loire, presse satirique, Le Soir d'Algérie, La Tribune de Genève
- SCIE (Sciences et Techniques) : CNRS-éditions, CNRS-Infos, Hermes revues, Infoscience (web), manuels techniques, Pour la Science, La Recherche, Science et Avenir, sites web, thèses.

La division choisie dans ce corpus est contestable sur certains points, puisqu'elle mélange genre au sens de 'sujet abordé' et genre au sens de 'type de texte'. Cela provoque des recoupements entre les tranches : FORM.politique / POLI, FORM.sports / PRAT.sports, etc. Néanmoins, elle est globalement cohérente et présente des types de textes variés et assez nombreux pour être une base intéressante à une comparaison.

2 Description générale des constructions de dépendance à longue distance

Les précédentes études sur corpus effectuées sur d'autres langues semblent indiquer un certain figement du verbe pont dans les structures interrogatives. Verhagen (2006) remarque que peu de verbes sont verbes ponts en néerlandais : on compte seulement 4 verbes dans 50 occurrences d'oral et de presse. Parmi ces quatre verbes, l'un d'entre eux domine largement : « denken » (penser). Au delà du lexique du verbe pont, c'est l'ensemble de la construction qui semble figé : le verbe est très majoritairement employé avec le pronom personnel de deuxième personne, et conjugué en conséquence.

Dans les corpus oraux anglais, on retrouve presque la même construction : la seule différence est l'absence de la conjonction « that » (Verhagen, 2006, Dąbrowska, 2008). Cette construction majoritaire est très fréquente dans l'usage (67%), et plus encore lorsque les adultes s'adressent aux enfants (85%). C'est enfin quasiment la seule forme de CAD produite par les enfants (95%) (Dąbrowska et al., 2009).

D'après Dąbrowska, cette construction peut être qualifiée de "formule prototypique". En plus d'être très fréquente, il est très rare qu'elle subisse plus d'une modification. Généralement, un seul élément varie : soit l'auxiliaire (*did* dans l'exemple 8 ou *would)*, soit le sujet du verbe pont (*they* dans l'exemple 9), soit le verbe pont (*say* dans l'exemple 10), soit l'ajout de la conjonction *that* (exemple 11) ou d'un adverbe (*really* dans l'exemple 12).

(8) Wh- did you think …

(9) Wh- do they think …

(10) Wh- do you say …

(11) Wh- do you think that …

(12) Wh- do you really think …

Les exemples qui présentent plusieurs changements par rapport à la formule prototypique ne représentent que 2% à 4% des exemples relevés (exemples 13 à 15).

(13) Wh- did you say …

(14) Wh- does he think that …

(15) Wh- do you really say …

Les expériences réalisées par Dąbrowska (2008) montrent que si l'on modifie tous les éléments, l'énoncé est jugé agrammatical, comme si la formule n'était plus reconnue (exemple 16).

(16) ?? Wh- did he really say that …

Nous avons observé cette construction en français, dans le corpus écrit décrit plus haut (le CERF, 9M mots) et des corpus oraux (3M mots). Lorsqu'on limite les données aux structures interrogatives directes, comme l'ont fait Verhagen et Dąbrowska, on constate que le nombre de verbes qui apparaît dans les CAD

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

est restreint : 7 verbes pour 90 occurrences. Le verbe le plus fréquent n'est pas « penser » comme en anglais et en néerlandais mais « vouloir », qui se construit en français avec une complétive. La construction la plus fréquente dans les dépendances à longue distance des corpus français est composée du pronom interrogatif, du pronom personnel sujet à la deuxième personne, du verbe « vouloir » au présent et de la conjonction « que/qu' », c'est-à-dire « WH- tu veux qu- », « WH- vous voulez qu- » ou « WH- veux-tu qu- », « WH- voulez-vous qu- ». Ce modèle représente 72% des occurrences.

(17) qui voulez-vous que ce soit (CERF, DIVS)

Dans environ 14% des occurrences, la construction la plus fréquente subit une seule modification, du mode du verbe pont (exemple 18), du sujet du verbe pont (exemple 19) ou du lexique verbal (exemple 20).

(18) quand vous allez chez le coiffeur, vous espérez tous que ce sera un professionnel qui vous coupera les cheveux... Pourquoi voudriez-vous que ce soit des citoyens, la plupart du temps désinvestis politiquement, qui décident de l'avenir d'une nation ? ! ; -) (CERF, FORM)

(19) que veut-il que je fasse (Nancy)

(20) vers quand tu penses que ce sera (CERF, ORAL)

Mais elle subit plusieurs modifications dans presque autant d'occurrences (exemple 21 à 23).

(21) Mais que pensent-ils que nous fassions nous ici, en bas, pendant qu'eux passent leur temps devant les holoprojecteurs. (CERF, LITC) (lexique verbal + sujet)

(22) pourquoi tu as dit qu'il n'a pas eu une mort extraordinaire (Nancy) (lexique verbal + temps)

(23) qu'est-ce qu'il voulait que je fasse (CERF, LITC) (temps + sujet)

La modification la plus fréquente est celle du lexique verbal (21/25 occurrences) donc de l'élément qui constitue le coeur de la construction. On peut donc se demander s'il y a réellement une construction prototypique en français, surtout lorsqu'on compare les chiffres avec l'anglais : la construction la plus fréquente est moins dominante en français qu'en anglais (72% *vs* 82% en moyenne) et il y a beaucoup plus de modifications multiples (13% *vs* 2.6% en moyenne).

Ces résultats ne concernent en outre que les données comportant une structure interrogative. Si l'on prend également en compte les structures interrogatives indirectes, les relatives et les clivées, l'existence d'une formule prototypique est très douteuse. Aucune construction ne se dégage vraiment (31% pour la plus fréquente) et le nombre d'occurrences dans lesquelles il y a plus d'une modification par rapport à cette construction est majoritaire (57%). Comme pour les interrogatives, c'est la variation du verbe qui est la plus fréquente : elle est présente dans 83% des modifications.

Bien sûr, chaque langue a son propre fonctionnement. On note notamment la différence de construction de « vouloir » et des verbes impersonnels en français et en anglais. Le verbe « vouloir », absent en anglais puisqu'il ne se construit pas avec une complétive, est très productif de dépendances à longue distance en français grâce à sa construction rhétorique. Les verbes impersonnels, qui peuvent se construire avec une complétive en français mais pas en anglais, augmentent le nombre de verbes ponts possibles et les modifications de la formule la plus fréquente en français.

(24) dans quel temps va falloir que je mette de cette phrase (CERF, ORAL)

En français, il est donc encore plus clair que les verbes ponts sont des modalisateurs, qui forment une chaîne verbale avec le verbe à complétive (Bérard, 2012). On trouve un continuum entre verbe modal + infinitif, dépendances à longue distance en infinitif et dépendances à longue distance en complétive mettant en jeu des verbes de type modaux au sens large. C'est ce principe de chaîne verbale qui rend si difficile le rattachement du relatif/interrogatif, notamment lorsqu'il est non sélectionné par le verbe, à son verbe recteur (V1 ou V2). Ainsi, plusieurs exemples que nous avons retenus peuvent faire débat. L'exemple 18, par exemple, reproduit ci-dessous, peut être interprété comme « pourquoi le voudriez-vous » ou bien reformulé en 18b auquel cas, « pourquoi » porte sur le V2 « être ».

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

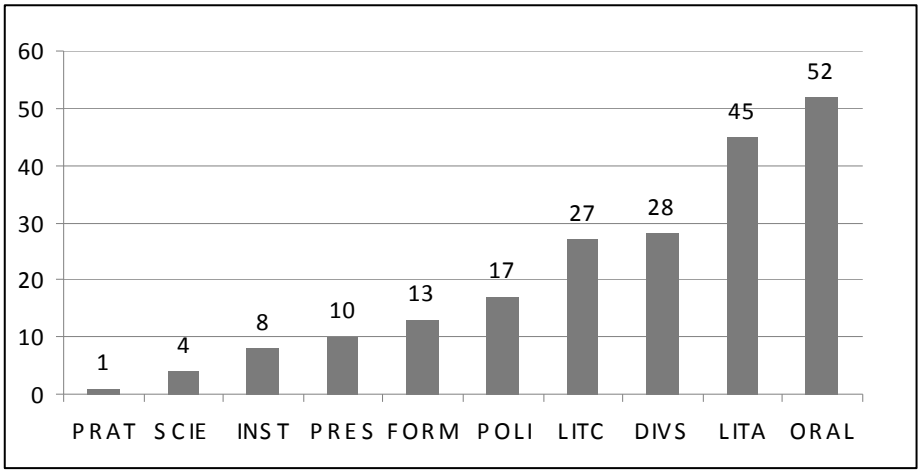
(18) quand vous allez chez le coiffeur, vous espérez tous que ce sera un professionnel qui vous coupera les cheveux... Pourquoi voudriez-vous que ce soit des citoyens, la plupart du temps désinvestis politiquement, qui décident de l'avenir d'une nation ? ! ; -) (CERF, FORM)

(18b) pourquoi ce serait des citoyens, la plupart du temps désinvestis politiquement, qui décideraient de l'avenir d'une nation ?

Dans les corpus français, plutôt qu'une formule prototypique, quatre verbes sont très représentés en verbe pont dans la totalité des énoncés mettant en jeu des dépendances à longue distance (interrogatives directes et indirectes, relatives, clivées). Il s'agit des verbes « vouloir », « falloir », « penser », et « dire ».

3 Variation en genre

La répartition des occurrences de dépendance à longue distance dans le corpus du CERF montre une grande disparité. Si l'on considère l'ensemble du corpus, on compte 205 occurrences de dépendance à longue distance : 88 relatives, 98 interrogatives, 15 clivées et 4 pseudo-clivées. Dans chaque tranche du CERF, le nombre d'occurrences varie entre 1 et 52.



Graphique 1. Répartition des occurrences de dépendances à longue distance dans le corpus CERF

Le nombre d'occurrences n'est pas le seul paramètre qui varie en fonction du genre de texte. Nous avons examiné différentes propriétés de la construction de dépendance à longue distance dans chaque genre : 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. Le dernier trait ne concerne que les relatives et les interrogatives mais celles-ci sont très majoritaires dans le corpus de LDD ; les clivées et pseudo-clivées ne sont représentées que par 19 occurrences. L'étude aboutit au résultat suivant : certaines tranches révèlent des comportements similaires. Trois groupes se distinguent, parmi lesquels deux sont opposés ; le troisième a un statut intermédiaire. Les paramètres prépondérants dans la classification sont d'abord la variété de verbe pont (nombre de verbes ponts différents par rapport au nombre d'occurrences de la construction) et ensuite la configuration syntaxique.

3.1 Propriétés de la construction dans le Groupe 1

Le premier groupe réunit les tranches Loisirs et vie pratique (PRAT), Sciences et Techniques (SCIE), Institutions (INST) et Politique (POLI). Dans ce groupe, les occurrences de dépendance à longue distance

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

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 ». Le tableau suivant présente la synthèse des résultats pour ce groupe :

	PRAT	SCIEN	INST	POLI
Nombre de LDD	1	4	8	17
Nombre de verbes	1	4	7	10
Variété de verbes	-	**100%**	**90%**	**60%**
Vpont le plus fréquent	penser	-	penser	penser
Configuration syntaxique	interrogative indirecte	**relative**	**relative** (75%)	**relative** (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 (2) pourquoi, dont + 4 clivées

TABLEAU 1. Comportement des dépendances à longue distance dans le groupe 1 du CERF

Voici un exemple typique de ce groupe, formé d'une relative en « lequel » et d'un sujet de verbe pont à la troisième personne.

(25) 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)

3.2 Propriétés de la construction dans le Groupe 2

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. Il présente cinq fois plus d'occurrences de dépendances à longue 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 Littérature Ancienne (LITA) s'écarte légèrement du groupe.

	LITC	**DIVS**	**ORAL**	**LITA**
Nb LDD	27	28	52	45

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

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- + 1 clivée	qu'est-ce que (15) où (13) que (6) quel- (3) qui, quand, pourquoi (2) l-quel-, combien	où (20), qu- (14) qu'est-ce que, combien, qui (2) l-quel-, dont, pourquoi +1 clivée, 1 pseudo-clivée

TABLEAU 2. Comportement des dépendances à longue distance dans le groupe 2 du CERF

Voici une attestation typique de ce groupe, qui présente une interrogative en « qu'est-ce que » avec un verbe pont de forme « vouloir » et un sujet à la deuxième personne.

(26) qu'est-ce que vous voulez que je vous dise (CERF, ORAL)

3.3 Propriétés de la construction dans le Groupe 3

Le troisième groupe réunit les tranches Presse (PRES) et Forum (FORM). Ce groupe a une position intermédiaire entre les deux premiers. Comme pour le premier groupe, le nombre d'occurrences de dépendance à longue distance est assez faible et une grande variété de verbes est utilisée (première sous-partie du tableau). 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 (deuxième sous-partie du tableau). Les pronoms interrogatifs-relatifs des deux groupes précédents sont présents : « qu'est-ce que », « que » / « dont », « comment » ; ils sont employés de façon quasi-équivalente (troisième sous-partie du tableau).

	FORM	**PRES**
Nb LDD	13	10
Nb verbes	8	6
Variété verbes	**60%**	**60%**
Vpont le plus fréquent	vouloir	vouloir

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

Config syntaxique	**interro directe (45%)**	**interro directe (50%)**
Personne sujet	P2 (45%)	P2 (50%)
Type d'interro-relatif	qu- (3 occurrences) pourquoi (2 occurrences) dont, qu'est-ce que, où, comment (1 occ) + 2 clivées, 2 pseudo-clivées	comment, qu- (3 occ) où (2 occurrences) qu'est-ce que, dont (1)

TABLEAU 3. Comportement des dépendances à longue distance dans le groupe 3 du CERF

L'exemple (27) peut être considéré comme représentatif des constructions relevées dans ce dernier groupe. Il présente une relative en « dont », avec un verbe appartenant à un registre relativement recherché, qui reflète la richesse lexicale du groupe. Dans ce groupe, peuvent également apparaître des interrogatives en « que », avec un verbe pont de forme « vouloir » et un sujet à la deuxième personne (exemple 28).

(27) Bien sûr, le ciné (du moins, celui dont je présume que tu parles), c'est facile (CERF, FORM/Temoign)

(28) 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)

3.4 Synthèse

On constate que les trois groupes forment un continuum. Pour plus de facilité, nous avons ordonné et renommé ces groupes en fonction de la variété lexicale que présente le verbe pont. Les chiffres fournis dans le tableau suivant sont le résultat de moyennes pour les quatre premiers paramètres (nombre de constructions de dépendances à longue distance, nombre de verbes différents, variété lexicale du verbe pont et configuration syntaxique de la construction) et le résultat d'une addition pour les trois paramètres suivants (verbe pont le plus fréquent, sujet du verbe pont et type d'interrogatif-relatif).

	G1	**G2**	**G3**
	PRAT, SCIE, INST, POLI	**Forum, presse**	**Littératures, divers, oral**
Nb LDD	7,5	11,5	38
Nb verbes	5,5	7	11,8
Variété V	85%	60%	30%
Configuration syntaxique	Relative 78% (sans PRAT)	Interro directe 48%	Interro directe 50% (sans LITA)
Vpont le +fréq	Penser (26% des occ)	Vouloir (43% des occ)	Vouloir (40% des occ)
Pers sujet	P3 / P3b : 53%	P2 : 43%	P2 : 43%

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

Type d'interro-relatif	comment, l-quel-, dont : 56%	qu-, qu'est-ce que : 42%	qu'est-ce que, qu- : 46%
		dont, comment : 32%	

TABLEAU 4. Récapitulatif du comportement des dépendances à longue distance dans les trois groupes du CERF

Cette description a été réalisée en dehors de toute considération des variations générale éventuellement présente dans chaque genre. Mais si le verbe "penser" était plus fréquent dans les genres PRAT, SCIE, INST et POLI, il serait logique qu'il soit également plus présent dans les constructions de dépendances à longue distance relevées dans ces genres. Nous allons donc comparer dans chaque genre la variété des verbes à complétive, leur fréquence, le pronom personnel sujet le plus fréquent et pour finir, la combinaison "pronom personnel (P2 / P3) + verbe".

Dans un premier temps, nous avons utilisé l'outil *Cordial Analyser*, réputé comme étant l'un des meilleurs taggueurs morphosyntaxiques du marché. Cordial effectue aussi une analyse syntaxique des phrases et fournit des statistiques sur les textes. Nous pouvons retenir des corpus étiquetés avec Cordial les informations suivantes:

- Dans chaque genre, le pronom sujet le plus fréquent est celui de la 3ème personne et le moins fréquent est celui de 2ème personne (sauf dans PRAT où le pronom le moins fréquent est celui de 1ère personne).
- De façon parallèle, dans chaque genre, les verbes sont conjugués majoritairement à la 3ème personne (entre 75% et 95% des verbes conjugués).

Le changement de personne sujet du verbe pont dans les dépendances à longue distance, entre les groupes 1 et 2-3, est donc important à noter.

- Les occurrences verbales sont plus nombreuses dans LITC, DIVS, Oral, LITA et FORM : 14 à 18% des formes sont des verbes. Dans les autres tranches, les verbes représentent entre 10 et 14% des formes.
- Les subordonnées sont plus présentes dans les tranches du G3 et dans FORM (>19000 occurrences par tranche) ; elles sont moins nombreuses dans les tranches du G1 et dans PRES (<15000 occurrences par tranche).

Le nombre de verbes et de subordonnées, plus important dans les tranches du G3, est congruent avec le nombre plus grand de dépendances à longue distance.

- Parmi les huit verbes "vouloir, penser, dire, falloir, savoir, croire, espérer et voir", le verbe le plus fréquent dans chaque tranche est :

	G1				G2		G3			
	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
V	savoir	dire	dire	falloir	dire	dire	dire	dire	dire	dire

TABLEAU 5. Verbe le plus fréquent (parmi 8 verbes) dans l'ensemble des constructions, par genre

Contrairement à ce qu'on aurait pu attendre à la suite de l'analyse des dépendances à longue distance, le verbe le plus fréquent des tranches du G1 n'est pas "penser" et celui des tranches du G2 et du G3 n'est pas "vouloir". Nous affinerons ci-dessous l'analyse en ne conservant que les occurrences des verbes avec complétives.

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

En outre, il est certain que des erreurs sont présentes dans l'étiquetage morphosyntaxique et syntaxique, notamment à l'oral pour lequel il n'y a pas de segmentation en énoncés et pour lequel l'outil n'est pas entraîné ni conçu. Les "quoi" ponctuant les fins d'énoncés à l'oral sont par exemple étiquetés conjonction de subordination.

Afin d'éviter les étiquetages erronés (notamment de "que"), nous avons utilisé dans un deuxième temps le concordancier *Antconc*, qui prend en entrée du texte brut. Nous avons recherché directement dans le texte les formes de pronom personnel sujet ("je, tu, il" etc.), suivies d'un mot et de que/qu'. Cette requête provoque bien sûr du silence (verbes pronominaux, locutions verbales, sujets lexicaux, sujets pronominaux et complétives un peu éloignée du verbe, inversion sujet-verbe) mais elle présente peu de bruit. Ayant vérifié manuellement les énoncés, nous sommes absolument sûre d'avoir pour chaque tranche un corpus d'exemples comparable de verbes à complétives. Voici ce que nous avons observé.

- Le nombre de verbes à complétive est légèrement plus grand dans les tranches des G2 et G3 (65 à 90 verbes *vs* 50 à 80 verbes pour le G1), ce qui diffère des constructions de dépendances à longue distance.
- Le nombre d'occurrences des verbes à complétive est plus important dans le G3 et dans FORM (>900 occurrences par tranche *vs* <800 occurrences dans le G1), ce qui pourrait expliquer qu'il y ait également plus d'occurrences de dépendances à longue distance dans ces tranches.
- Le verbe à complétive le plus fréquent de chaque tranche est différent du verbe pont le plus fréquent, hormis en INST

	G1				G2		G3			
	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
Vcompl	savoir	savoir	**penser**	falloir	penser	savoir	savoir	savoir	penser	savoir
Vcompl LDD	penser	/	penser	penser	vouloir	vouloir	vouloir	vouloir	vouloir	vouloir

TABLEAU 6. Verbe à complétive le plus fréquent dans les LDD et dans l'ensemble des constructions, par genre

- Les sujets pronominaux des verbes à complétives les plus fréquents sont différents de ceux des verbes ponts dans les dépendances à longue distance, et ce, dans toutes les tranches de corpus. Précisons que P1 regroupe "je, j', nous", P2 "tu, vous", P3 "il, elle, on, ils, elles" et P3B "il impersonnel, ça".

	G1				G2		G3			
	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
pp sujet Vcompl	P1	P3	P1	P1	P1	P3	P1	P1	P1	P1
pp sujet LDD	P3	P2	P3B	P3	P2	P2	P2	P2	P2	P2

TABLEAU 7. Personne sujet la plus fréquente dans les LDD et dans l'ensemble des constructions, par genre

- Dans l'ensemble des constructions de verbes à complétive, le pronom sujet "tu" est employé d'abord avec le verbe "penser" (rang 1) puis avec "croire" (rang 2) et "savoir" (rang 3). Le pronom sujet "vous" est employé avec "savoir" (rang 1) puis avec "vouloir" (rang 2). Il y a donc dans le corpus

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

global peu d'attraction entre P2 et vouloir, contrairement aux constructions de dépendance à longue distance.

- Dans chaque tranche, les pronoms sujets "tu" et "vous" sont employés majoritairement avec les verbes à complétive :

	G1				G2		G3			
	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
TU	x	x	x	x	dire	savoir	savoir, **vouloir**	croire	penser	savoir
VOUS	voir, estimer	**vouloir**	savoir	savoir	dire	comprendre	savoir	voir	savoir	savoir

TABLEAU 8. Verbe à complétive le plus fréquent employé avec « tu » et « vous » dans l'ensemble des constructions, par genre

Les seuls genres dans lesquels le P2 est lié au verbe "vouloir" sont LITC (ex aequo avec "savoir") et SCIE. Or dans ce dernier, "P2-vouloir" n'est pas représentatif des dépendances à longue distance.

P2 est moins fréquent dans les tranches du G1 que dans les tranches du G3 ("tu" y est même inexistant) mais il en va de même pour les pronoms personnels sujets P1, P3 et P3B.

La proportion de P2 dans chaque genre ne permet pas non plus d'expliquer la forme sujet dans les dépendances à longue distance du G2 et G3. On remarque qu'elle est inconstante dans les tranches du G1 (de 3% pour SCIE à 18.3% pour PRAT).

	G1				G2		G3			
P2 sujet des V complétive	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
% par rapport aux autres personnes	18,3	3,183	4,239	4,887	9,091	2,257	10,31	9,134	15,2	12,35

TABLEAU 9. Personne sujet la plus fréquente dans l'ensemble des constructions, par genre

- Enfin, si la proportion de P2-vouloir est légèrement plus élevée dans les tranches du G3 (12,6% des occurrences d'un genre en moyenne vs 10,9% dans le G1), la proportion de P3-penser est relativement constante dans toutes les tranches du corpus (entre 5% et 10% des occurrences d'un genre). La forme des dépendances à longue distance, particulière à chaque groupe de genre, ne peut donc pas être expliquée par des traits inhérents aux genres eux-mêmes.

Pour résumer, le nombre d'occurrences des dépendances à longue distance et la richesse lexicale des verbes ponts sont peut-être liés au genre dans lequel elles se trouvent. En revanche, la forme du pronom personnel sujet, la forme du verbe pont, et la combinaison "pronom sujet + V" est spécifique aux dépendances à longue distance :

	G1				G2		G3			
	PRAT	SCIE	INST	POLI	FORM	PRES	LITC	DIVS	ORAL	LITA
Vcompl	savoir	savoir	**penser**	falloir	**penser**	savoir	savoir	savoir	penser	savoir

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

Vcompl LDD	penser	/	penser	penser	vouloir	vouloir	vouloir	vouloir	vouloir	vouloir
pp sujet	P1	P3	P1	P1	P1	P3	P1	P1	P1	P1
pp sujet LDD	P3	P2	P3B	P3	P2	P2	P2	P2	P2	P2

TABLEAU 10. Verbe à complétive et sujet les plus fréquents dans les LDD et dans l'ensemble des constructions, par genre.

Dans la partie suivante, nous allons tenter d'expliquer ce qui rapproche des tranches PRAT, SCIE, INST et POLI d'une part, et les tranches LITC, DIVS, ORAL et LITA d'autre part et ce qui distingue ces deux groupes (le troisième étant intermédiaire). Nous pourrons alors assumer que le trait découvert forme des classes de textes dans lesquelles la forme particulière des dépendances à longue distance prend part à leur grammaire locale. Nous n'irons pas jusqu'à parler de diglossie (Massot & Rowlett, 2013).

4 Critère de groupement

Les regroupements constitués à partir des propriétés peuvent sembler étranges. On peut se demander pourquoi certaines tranches ont un comportement similaire. Qu'est-ce qui rapproche ou éloigne les genres observés ? Nous proposons quatre hypothèses pour répondre à cette question : le contenu thématique, le degré d'élaboration/de spontanéité, le caractère formel/informel du texte et le degré de codification.

4.1 Le contenu thématique

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 (Sinclair, 1996), mais il est peu probable qu'il ait un impact sur la structure syntaxique des textes, et plus particulièrement sur celle des dépendances à longue 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 Groupe 1, 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 rien en commun avec les manuels techniques, les revues scientifiques ou les thèses (SCIEN), qui à leur tour se distinguent du vocabulaire de la finance (PRAT). Dans le Groupe 2, les textes des forums ayant trait à l'économie, à l'actualité mondiale et à la politique (FORM) présentent des contenus similaires à ceux de la presse (PRES) mais diffèrent des autres sous-parties de FORM qui concernent le sport, les médias ou les témoignages. Enfin, dans le Groupe 3, plusieurs sous-parties possèdent des thématiques similaires : journal intime (DIVS) et journal autobiographique (LITA) ou éventuellement les dialogues entre les personnages au théâtre (LITA, LITC) et les conversations (ORAL), mais d'autres sont tout à fait différentes : la critique littéraire, les textes religieux, les textes humoristiques, etc. (DIVS).

Le contenu thématique n'explique donc pas les regroupements et les séparations entre les tranches.

4.2 Le degré d'élaboration des textes

La seconde hypothèse que nous allons envisager comme élément unificateur et séparateur des trois groupes est le degré d'élaboration ou de spontanéité des textes. 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

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

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 du discours, sont présents dans le corpus d'exemples de tous les groupes, en tant que sujet du verbe pont. Ils se répartissent très inégalement en fonction des tranches. Dans chaque groupe, certaines tranches attestent de sujets lexicaux alors que d'autres n'en présentent aucun. (Tableau 11).

G1	PRAT	SCIEN	INST	POLI
Nb LDD	1	4	8	17
Sujet lexical du verbe pont	0	0	0	4
Proportion par tranche				23.5%

G2	FORM	PRES
Nb LDD	13	10
Sujet lexical du verbe pont	1	0
Proportion par tranche	7.6%	

G3	LITC	DIVS	ORAL	LITA
Nb LDD	27	28	52	45
Sujet lexical du verbe pont	1	1 nom propre	0	4
Proportion par tranche	3.7%	3.5%		8.8%

TABLEAU 11. Répartition des sujets lexicaux des verbes ponts dans les groupes 1, 2 et 3 du corpus CERF

Dans l'ensemble du corpus, les sujets lexicaux sont présents dans toutes les tranches. La proportion de sujets pronominaux (*vs* lexicaux) est plus importante dans les tranches du G3 et dans FORM mais les sujets sont majoritairement lexicaux dans toutes les tranches, hormis LITC et ORAL. Ces conclusions s'appuient sur l'analyse syntaxique de Cordial, qui est à prendre avec précaution notamment pour l'oral où nous avons remarqué beaucoup d'erreurs. Nous ne pouvons donc pas affirmer que les groupes formés à partir de l'analyse des dépendances à longue distance concordent avec ceux ayant une élaboration plus ou moins complexe.

4.3 Le caractère formel / informel

La troisième hypothèse est que l'élément unificateur et séparateur 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

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

énormément d'inversions dans le corpus d'exemples des dépendances à longue distance (entre 80% et 100% pour PRES, FORM, LITA et POLI) et d'autres relativement peu (entre 0% et 55% pour ORAL, LITC et DIVS). Toutes les inversions se trouvent dans la configuration syntaxique d'interrogative directe. Si les tranches avec inversions appartiennent aux G1 et G2, celles qui en comportent peu appartiennent toutes au G3 (voir Tableau 12).

La convergence partielle des groupes d'inversion du sujet avec les groupes de dépendance à longue distance pourrait indiquer que ces derniers répondent au trait formel/informel.

Tranche	Proportion dans les interrogatives directes des LDD	Groupes formés à partir de l'analyse des LDD
PRAT	- (0/0)	G1
SCIE	- (0/0)	
INST	- (0/0)	
POLI	100% (2/2)	
PRES	80% (4/5)	G2
FORM	83% (5/6)	
LITC	35% (5/14)	G3
DIVS	53% (8/15)	
ORAL	0% (0/23)	
LITA	87% (14/16)	

TABLEAU 12. Répartition des inversions sujets-verbes ponts dans les LDD des groupes 1, 2 et 3 du corpus CERF

Les formes de vouvoiement et de tutoiement sont également une marque de langue formelle/informelle. Dans les constructions de dépendance à distance, 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 longue distance : 100% de vouvoiement dans les LDD du G1, 80% à 100% de vouvoiement dans les LDD du G2, moins de 70% de vouvoiement ou majorité de tutoiement dans les LDD du G3 (Tableau 13). Dans la totalité des tranches, si on concentre notre attention sur les constructions « pronom personnel + verbe + qu » (requête effectuée avec Antconc, voir supra), la tendance est similaire, bien que des différences notables existent dans les tranches du G3 : 100% de vouvoiement dans le G1, 70% à 90% de vouvoiement dans le G2, moins de 65% de vouvoiement ou majorité de tutoiement dans le G3.

Cette seconde convergence des groupes nous amène à prendre en compte l'importance du critère formel/informel dans l'emploi des dépendances à longue distance.

Tranche	Proportion dans les LDD	Proportion pour les verbes à dans la totalité des tranches	Groupes formés à partir de l'analyse des LDD
PRAT	/	Vous (41)	G1
SCIE	Vous (2)	Vous (12)	

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

INST	Vous (1)	Vous (34)	
POLI	Vous (2)	Vous (39)	
PRES	Vous (6)	9 vous (90%), 1 tu	G2
FORM	4 vous (80%), 1 tu	80 vous (70,1%), 34 tu	
LITC	7 vous (58,3%), 5 tu	53 vous, 65 tu (55%)	G3
DIVS	11 vous (64,7%), 6 tu	60 vous (61,8%), 37 tu	
ORAL	15 vous (68,1%), 7 tu	132 vous, 299 tu (69,3%)	
LITA	3 vous, 12 tu (80%)	73 vous (62,9%), 43 tu	

TABLEAU 13. Proportion de tutoiement et de vouvoiement dans les tranches du CERF

Enfin, le degré de formalité du type d'interrogatif-relatif, lui, correspond véritablement au découpage des groupes. Le pronom « qu'est-ce que », jugé informel, est réalisé dans les dépendances à longue distance du Groupe 3 tandis que les pronoms « dont » et « l-quel- », rattachés au registre formel, sont réalisés dans les dépendances à longue distance du Groupe 1. Dans la totalité du corpus, on retrouve la même tendance. Le pronom « qu'est-ce qu » est plus représenté dans les tranches du G3 et les pronoms « dont » et « l-quel- » sont majoritaires dans les tranches du G1.

Tranche	**Nombre d'occurrences dans les LDD**	**Nombre d'occurrences dans la totalité des tranches**	**Groupes formés à partir de l'analyse des LDD**
PRAT	/	3 qu'est-ce qu, 693 dont, 280 l-quel-	G1
SCIE	1 dont, 3 l-quel-	12 qu'est-ce qu,1114 dont, 724 l-quel-	
INST	1 dont, 2 l-quel-	13 qu'est-ce qu, 885 dont, 839 l-quel-	
POLI	1 dont, 2 l-quel-	26 qu'est-ce qu, 1118 dont, 529 l-quel-	
PRES	1 qu'est-ce qu, 1 dont	28 qu'est-ce qu, 1129 dont, 491 l-	G2
FORM	1 qu'est-ce qu, 1 dont	69 qu'est-ce qu, 823 dont, 351 l-quel-	
LITC	9 qu'est-ce qu, 2dont, 2 l-	250 qu'est-ce qu, 633 dont, 334 l-	G3
DIVS	6 qu'est-ce qu, 1dont, 2 l-quel-	218 qu'est-ce qu, 845 dont, 513 l-quel-	
ORAL	15 qu'est-ce qu, 1 l-quel-	627 qu'est-ce qu, 226 dont, 274 l-	
LITA	2 qu'est-ce qu, 1 dont, 1 l-quel-	137 qu'est-ce qu, 1359 dont, 540 l-quel-	

TABLEAU 14. Proportion de quelques interrogatifs-relatifs dans les tranches du CERF

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

Il est donc fort probable que le critère de langue formelle intervienne dans le comportement des dépendances à longue distance, au moins en partie.

4.4 Le degré de codification

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 : 43). Elle permet à la fois l'homogénéisation et la prévision. La codification implique une certaine « objectivation » (Bourdieu, 1986 : 42) qui a pour conséquence qu' « on y perd en charme... » (Bourdieu, 1986 : 42). Avec un texte codifié, la communication est « minimale » (Bourdieu, 1986 : 42).

Comme l'a démontré Verhagen (2006), la fonction la plus importante des dépendances à longue distance est l'expression de l'intersubjectivité.

> They belong to a set of prototypical patterns for explicitly signaling aspects of the intersubjective dimension of the construal configuration. (Verhagen 2006 : 340-341)

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 longue distance dans les textes codifiés. La codification est opposée au « jeu », à l' « improvisation », à l' « invention » (Bourdieu, 1986 : 42-43). Il devrait donc y avoir beaucoup de dépendances à longue distance dans les fictions, l'humour, la chanson, etc. Cette définition semble correspondre aux groupes que nous avons définis. Le Groupe 1, qui réunit des textes très codifiés (manuels techniques, thèses, discours politiques, etc.) présente peu d'attestations de dépendances à longue distance, alors que le Groupe 3, dont les textes sont peu codifiés (littérature, oral, divers), en présentent beaucoup plus. Enfin, le Groupe 2 est intermédiaire (forum, presse).

Cette analyse est confirmée par quelques auteurs, qui considèrent la littérature comme un genre 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 : 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 : 183)

La richesse lexicale du verbe pont que présentent les occurrences de dépendance à longue distance dans les tranches du Groupe 1 (85%) semble remettre en cause l'hypothèse de la codification, que nous privilégions. Si le texte est codifié, il ne devrait pas varier en construction ni 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 ainsi 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 d'employer. 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.

5 Conclusion

Les études sur corpus ont mis en évidence l'aspect formulaire du pont de la construction interrogative : « WH- do you think S » en anglais (Dąbrowska, 2008) et « WH- denk- P2 dat S » en néerlandais (Verhagen, 2006). En français, la construction de dépendance à longue distance ne s'emploie pas comme un patron pouvant subir des variations minimales. Il existe plutôt des déclinaisons de la construction en fonction du degré de codification et de langue formelle/informelle du genre textuel. Parmi les dix tranches

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)

SHS Web of Conferences 8 (2014)
DOI 10.1051/shsconf/20140801336
© aux auteurs, publié par EDP Sciences, 2014

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 aux faits de codification afin de croiser les résultats et de vérifier que les groupes de tranches que nous avons déterminés dans cette étude se maintiennent.

Références bibliographiques

Bérard, L. (2012). *Dépendances à distance en français contemporain. Etude sur corpus « c'est ce qu'on pense qui devrait être fait ».* Thèse de l'Université de Lorraine, Nancy.

Bois, G. (2009). *Les écrivains dominés du jeu littéraire. Définition de l'espace d'investissement et rapports aux enjeux littéraires.* Thèse de l'Université Lyon 2. http://recherche.univ-lyon2.fr/grs/index.php?page=39&notice=14, 05/03/12.

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.

Chomsky, N. (1977). On wh-movement. In *Formal Syntax*, ed. by P. Culicover, T. Wasow, & A. Akmajian. New York: Academic Press.

Dąbrowska, E. (2008). Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics*, *19 (3)*, 391-425.

Dąbrowska, E., Rowland, C. & Theakston, A. (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics, 20 (3)*, 571–597.

Erteschik-Shir, N. (1973). *On the nature of island constraints*. Thèse du Massachusetts Institute of Technology, Cambridge. http://dspace.mit.edu/handle/1721.1/12991.

Godard, D. (1988). *La syntaxe des relatives en français*. Paris : Editions du CNRS, Sciences du langage.

Ross, J. R., (1967). *Constraints on Variables in Syntax*. Thèse du Massachusetts Institute of Technology, Cambridge. http://www.eric.ed.gov/PDFS/ED016965.pdf.

Massot, B. & Rowlett, P. (2013). Le débat sur la diglossie en France : aspects scientifiques et politiques. *Journal of French Language Studies, 23(1)*, 1-16.

Sinclair,.J. (1996). Preliminary recommendations on Corpus Typology. Rapport technique. EAGLES (Expert Advisory Group on Language Engineering Standards).

Verhagen, A. (2006). On subjectivity and « long distance Wh-movement ». In *Subjectification: Various Paths to Subjectivity* ed. by A. Athanasiadou, C. Canakis, & B. Cornilie. Berlin/New York: Mouton de Gruyter.

Zumthor, P. (2008). Oralité. *Intermédialités : histoire et théorie des arts, des lettres et des techniques*, *12*, 169–202.

Article en accès libre placé sous licence Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0)