



HAL
open science

Coordination et subordination en français oral

Dominique Labbé

► **To cite this version:**

Dominique Labbé. Coordination et subordination en français oral. Banks David. La coordination et la subordination dans le texte de spécialité, L'Harmattan, pp.161-182, 2007. halshs-00291040

HAL Id: halshs-00291040

<https://shs.hal.science/halshs-00291040>

Submitted on 27 Nov 2009

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.

Université de Bretagne occidentale
Faculté des Lettres et Sciences sociales
Equipe de Recherche en Linguistique Appliquée

Coordination/subordination dans le texte de spécialité

(Brest 14-15 novembre 2003)

Coordination et subordination en français oral

Dominique LABBE
CERAT-IEP - BP 48 - 38 040 Grenoble Cedex 9
dominique.labbe@iep.upmf-grenoble.fr

Résumé :

Comment analyser le français oral ? On présente les principales étapes à l'aide d'un corpus de plus de 300 entretiens comptant au total 2,3 millions de mots : règles de transcription, normalisation orthographique, balisage des textes, étiquetage des mots. Pour décrire les caractéristiques propres à l'oral, il faut connaître celles du français écrit. Nous donnons un exemple à l'aide d'un corpus de plus de 2000 textes comprenant plus de 7 millions de mots. Les différences entre l'oral et l'écrit sont considérables : vocabulaire, syntaxe, structure des phrases. L'examen de la coordination et de la subordination fait apparaître un système assez éloigné de celui que décrivent traditionnellement les grammaires du français. Des enquêtes d'usage et des corpus représentatifs sont indispensables pour une étude scientifique de la langue.

Mots clefs : Coordination - Subordination - Pronoms - Conjonctions - Français oral -
Corpus représentatifs.

Comment utilise-t-on la subordination et la coordination quand on s'exprime oralement en français ?

Pour répondre scientifiquement à cette question, il faudrait disposer d'un vaste échantillon représentatif des milliards de paroles échangées chaque jour par nos concitoyens. Un tel étalon existe pour la plupart des grandes langues contemporaines. Historiquement, le *British National Corpus* est le premier apparu au début des années 1990. Ce corpus est présenté dans le numéro 8-4 (1993) de *Literary & Linguistic Computing* et par le responsable des services informatiques de l'Université d'Oxford qui héberge ce corpus (Burnard, 1995). Sur la partie orale du BNC, voir particulièrement les articles de : Crowdy, 1993 et Nelson, 1997. Les derniers corpus représentatifs en date concernent le tchèque (Kucera, 2002) et l'écossais (Douglas, 2003).

Pour le français, il n'existe rien de tel. Sans prétendre combler une pareille lacune, nous allons présenter une petite expérience assez suggestive. Depuis une douzaine d'années, plusieurs sociologues nous ont confié des transcriptions d'entretiens pour en réaliser des analyses lexicométriques (détail des corpus en annexe). Le total représente près de 2,3 millions de mots recueillis auprès de plusieurs centaines de locuteurs d'une très grande diversité sociologique et géographique. Pourquoi ne pas confronter cette base de textes "oraux" au français écrit sur lequel nous disposons également d'une grande quantité de textes traités de la même manière ?

Naturellement, il s'agit d'une exploitation secondaire, opération courante en statistique mais assez délicate car la manière dont les données ont été collectées peut rejaillir sur les résultats. Par exemple, plus de la moitié du corpus oral est constitué d'entretiens réalisés par les sociologues d'EDF sur des sujets très divers — le chauffage, la climatisation, les appareils domestiques, l'isolation des logements anciens, les modes de vie, l'activité professionnelle, voire le budget des personnes en situation précaire (cf Labbé, 2002a) — mais tous abordent plus ou moins un sujet : l'électricité. Ce mot et l'adjectif "électrique" seront donc certainement sur-employés dans ces textes... De même, nous ne prétendons pas avoir un corpus représentatif de la manière dont nos contemporains écrivent effectivement le français. Dans notre corpus "français écrit", le XXe siècle ne pèse "que" 60% du total. Les textes littéraires occupent une proportion semblable... C'est sans doute trop (sauf si l'on partage le présupposé du "Trésor de la langue française" selon lequel la langue se résume à la production des "grands auteurs"). Il faudrait donc ajouter des textes techniques, de la correspondance et encore beaucoup de choses dont l'examen dépasse le cadre de ce propos...

Avant de comparer la subordination et la coordination en français oral et en français écrit, nous allons présenter succinctement la manière dont ces deux corpus ont été constitués.

I. Saisie, balisage, étiquetage et traitement des corpus

La norme de saisie.

Il n'existe pas de véritable "code" de la transcription de l'oral à l'écrit. On a proposé des règles de saisie spécifiques visant notamment à restituer le rythme du propos (Blanche-Benveniste & Jeanjean, 1987). Bien que développée dans un laboratoire important et publiée il y a près de vingt ans, cette norme ne s'est pas imposée, sans doute parce que deux raisons fortes militent en sa défaveur :

- elle est complexe et très éloignée des conventions usuelles. Comment la faire accepter par les opérateurs qui sont chargés de la saisie de ces entretiens mais à qui l'on demande de continuer à utiliser des conventions profondément différentes pour le reste de leur travail ?

- à moins d'opérer une double saisie, on ne pourra comparer ces transcriptions avec les bases de données dont on dispose pour le français écrit comme nous voulons le faire.

Ne vaut-il pas mieux utiliser la norme standard quitte à accepter une certaine perte d'information ? Ce standard ne peut être que la norme "sténotypique", telle qu'elle est enseignée dans les écoles de secrétariat. Elle consiste, dans son principe, à faire entrer le "mieux possible"

la forme orale dans le lit de Procuste de l'écrit. Cela est particulièrement net pour la ponctuation où les résultats peuvent varier grandement selon l'opérateur. Si celui-ci privilégie la scansion au détriment de la syntaxe : la virgule représente une brève interruption; une interruption plus longue est transcrite grâce au point si elle est précédée d'une baisse de l'intonation et par un point d'exclamation ou d'interrogation lorsque l'intonation, ou la syntaxe le suggèrent. Si l'opérateur privilégie l'analyse syntaxique et sémantique, il ponctuera en fonction des périodes oratoires en ayant recours aux envahissants points de suspension quand la période ne se termine pas "logiquement" (du moins à ses yeux).

D'autres inconvénients paraissent inévitables. Signalons l'exemple très frappant du participe passé avec l'auxiliaire "avoir". Dans leurs propos spontanés, ou dans la vie quotidienne, la plus grande partie des locuteurs négligent l'accord (pour les verbes du 2e et du 3e groupes, le féminin doit s'entendre ; de même pour la liaison quand le 's' du pluriel est suivi par un mot commençant par une voyelle). Eh bien ! les secrétaires "corrigent" assez systématiquement, sans même en avoir conscience, malgré la consigne de transcrire le plus exactement possible ce qui est dit. Dans le cas d'une exploitation secondaire, on ne dispose généralement pas des bandes magnétiques, de telle sorte que sur ce point - crucial pour la réforme de l'orthographe -, ces corpus ne peuvent renseigner sur l'usage "réel" du français...

Enfin, sur ce premier point : une normalisation des graphies est indispensable, notamment pour les sigles, les abréviations, les noms propres, spécialement les patronymes et les toponymes étrangers (veiller à ce que Mao, Khrouchtchev ou Trotski n'aient chacun qu'une seule entrée ; décider que Florence et Firenze sont une seule et même ville...) Certes, cette tâche peut être partiellement effectuée par des automates mais les interventions manuelles sont nécessairement nombreuses et doivent suivre des règles bien précises...

Le balisage

Le texte ne suffit pas. Il faut ajouter quelques informations et normaliser ces informations. Par exemple, quand on consulte un corpus d'entretiens saisis par plusieurs opératrices différentes, on constate toujours que l'identification des questions et des réponses n'est pas stable. Par exemple, les questions sont parfois en italiques parfois précédés de "E : " (pour "enquêteurs"), quelquefois même sans aucun identifiant... Nous avons donc introduit des balises qui isolent les séquences du texte : en-têtes, remarques, questions et réponses (par exemple, le début du premier entretien de la base "négociation collective au Québec", reproduit dans le tableau I).

Grâce à ces balises, l'opérateur pourra isoler le texte des réponses — c'est ce que nous ferons dans la suite de cet exposé — ou celui des questions, s'il s'intéresse au style de la sociologie... etc. Cette opération ne supprime rien, au contraire, elle facilite le traitement de l'information recueillie.

Tableau I. Le balisage des textes

<Question>
Pouvez-vous me rappeler le nom de votre organisation ?
<Réponse>
Ville d'Anjou...
<Question>
Vos fonctions...
<inaudible : sonnerie du téléphone. Interruption>
<Réponse>
Je suis directrice des travaux publics.
Etc.

Normalisation des graphies, étiquetage des mots

Faut-il rappeler que la plupart des mots sont susceptibles d'avoir plusieurs graphies : majuscules ou minuscules, élisions, abréviations... Dans certains entretiens, "monsieur" est écrit en toutes lettres, ailleurs M (suivi ou non d'un point), dans d'autres encore : Mr... Les noms de lieux et de personnes sont transcrits "phonétiquement" si l'on n'a pas pris la précaution de les transmettre avant à l'opératrice. Les chiffres et les dates sont d'une infinie variété, parfois en lettres, parfois en chiffres, avec comme séparateur de millier : une virgule, un blanc ou rien du tout ; pour les décimales : un point ou une virgule, etc. Par exemple, 1990, '90 mais aussi, plus insidieusement : "I990" (sur le clavier, les I et O majuscules sont plus facilement accessibles que les chiffres...) Enfin, il faut compter avec les nombreuses fautes d'orthographe qui émaillent ce genre de document, même lorsqu'ils ont été saisis par des secrétaires expérimentées, ce qui était le cas pour la majorité des entretiens.

Dernières remarques de bon sens : des mots identiques peuvent s'écrire de plusieurs manières ("événement" et "évènement" ; "puis et peux", etc). Des milliers de mots français sont susceptibles d'avoir plusieurs orthographes différentes (la réforme "facultative" de l'orthographe a encore augmenté leur nombre). Et des mots différents peuvent s'écrire de la même manière ! Par exemple, "je suis" (suivre ou être ?), "l'est" (article + nom ou pronom + verbe ?) ou "prise (s)" : substantif féminin (prise de décision), adjectif "pris" (au féminin), verbe "prendre" au participe passé ou... verbe "priser". Dans tout texte français, ces "homographes" touchent plus du tiers des mots et ce sont des mots très fréquents. Par exemple, "que" peut être pronom ou conjonction comme nous le rappellerons par la suite.

Il faut donc normaliser les graphies (un mot, une seule orthographe) et ajouter à chacun des mots une étiquette qui indique sa catégorie grammaticale (on dit aussi "lemmatisation"). C'est en s'inspirant de cette idée simple que, il y a une quinzaine d'années, a été mis au point une chaîne de traitement du français contemporain d'abord dans le cadre de travaux sur le discours politique (Labbé, 1990) puis généralisée à l'ensemble des textes, oraux ou écrits, en français moderne.

La nomenclature des mots, apprise à l'ordinateur, est systématique (par exemple, en français, les substantifs se distinguent par le genre, donc tous les substantifs doivent se voir affecter le masculin ou le féminin), elle est exhaustive (tous les mots doivent y trouver leur place), elle exclut tout double compte, elle ne comporte pas de catégorie ad hoc, ou fourre-tout, etc. Elle a été élaborée à l'aide du *Dictionnaire général* de Hatzfeld et Darmesteiter — c'est la référence préconisée par C. Muller —, du *Larousse*, du *Littré*, du *Robert* et du *de Villers* ainsi que des principales grammaires (Blanche-Benveniste, Wagner & Pinchon, Arrivé & Al...). Par exemple, les conjugaisons d'un même verbe sont groupées sous son infinitif ou les pluriels du substantif sous le singulier ou encore les féminins et pluriels de l'adjectif sous le masculin singulier. Par exemple, "être v." regroupe toutes les formes conjuguées de ce verbe, tandis que "être n. m." ne se rencontre que sous le singulier et le pluriel.

Malgré ses faiblesses, cette nomenclature présente un avantage précieux : elle est bien connue des usagers du français et elle leur est devenue en quelque sorte "naturelle" au même titre que l'ordre alphabétique ou les PCS des sociologues. Le principe général consiste à regrouper les flexions d'un même mot sous une "forme vedette" ou "lemme" auquel est associée une catégorie grammaticale. Cette opération obéit à quelques règles simples dont la première, de bon sens, est de ne pas altérer le texte original mais d'ajouter à chaque mot une "étiquette" (tableau II).

Tableau II. L'étiquetage des textes

je, je, pronom	suis, être, verbe	directrice, directrice, nom fem.
Je	suis	directrice

Dans les étiquettes, en première position, on trouve la graphie normalisée : réduction des majuscules des mots communs, des graphies multiples, des abréviations, contrôle de l'orthographe des noms propres et de certains mots communs ; en seconde position : l'entrée de dictionnaire et, en troisième, la catégorie grammaticale. L'exemple ci-dessus comporte trois mots et deux homographies : "suis" (être ou suivre) et "directrice" (substantif "directrice" ou adjectif "directeur"). L'homographie est une situation très courante. En moyenne, dans un texte en français, plus du tiers des mots peuvent ainsi être rattachés à plusieurs entrées de dictionnaire. Remarquons que la lemmatisation doit être exhaustive et sans ambiguïté (chaque mot du texte se voit attribuer un lemme et un seul). Elle doit également être réversible, c'est-à-dire qu'on peut retrouver le texte original à partir du fichier des lemmes. Enfin elle ne doit pas comporter d'erreur...

Les bénéfices de ces opérations sont multiples. Par rapport aux traitements sur les graphies "brutes", la normalisation et la lemmatisation redonnent une existence aux verbes (en rassemblant leurs multiples flexions sous une étiquette commune). On peut enfin retrouver certains mots rares comme le point cardinal "est", les substantifs "être"; "avoir", "avons"... dont les occurrences sont habituellement noyées dans l'immensité des formes verbales homographes. On peut distinguer le "si" conjonction de l'adverbe homographe, les emplois pronominaux ou conjonctifs de "que", etc. Au-delà de ces avantages, la normalisation et la lemmatisation rendent possibles de nombreuses opérations statistiques dont cette étude donnera quelques exemples. En premier lieu, on peut enfin comparer toutes sortes de corpus entre eux. Par exemple, le français oral et le français écrit...

II. Principales caractéristiques du français oral

La comparaison des deux corpus fait apparaître des différences considérables à tous les niveaux : vocabulaire, catégories grammaticales ou syntaxe. Voyons d'abord ces différences d'ensemble avant de nous concentrer sur la coordination et la subordination.

Le vocabulaire de l'oral

Nous ne revenons pas sur le test statistique qui permet de déterminer quels sont les mots ou les catégories dont les fréquences d'emploi diffèrent significativement (du point de vue statistique) à l'écrit et à l'oral (voir à ce sujet : Labbé et Labbé, 1994).

Les listes des vocabulaires spécifiques sont très longues. Dans l'annexe II, seuls les pronoms, déterminants et conjonctions sont reproduits intégralement. Dans la lecture de ces listes, il faut naturellement démêler l'anecdotique du "structural". Pour l'anecdote : certains substantifs spécifiques manifestent simplement le poids relatif trop élevé de certains thèmes privilégiés dans les enquêtes EDF (chauffage, électricité, électrique) ou dans l'enquête québécoise (*négociation raisonnée* et, dans les pronoms : "nous autre", "vous autre", "eux autres"). En revanche, la présence dans le vocabulaire sous-employé des principaux mots de la politique est particulièrement significatif puisque, dans le corpus oral, figure plusieurs textes portant sur ce thème et notamment l'enquête de S. Pionchon sur la relation des Français à la politique. La caractéristique frappante de la plupart des entretiens, menés par S. Pionchon, c'est que les enquêtés s'évadent rapidement du thème et parlent d'autre chose, notamment de leur vie quotidienne.

Quelles sont les principales caractéristiques du vocabulaire de l'oral par rapport à celui de l'écrit ?

- un excédent de verbes usuels : les densités de "être" et "avoir" sont multipliées par 2 et spécialement des verbes "faire" (x 2,1), "dire" (x 2,7), "aller" (x 2,1) et "falloir" (x 2,4). Ce sont par excellence les outils de base de la communication orale ;

- les substantifs "chose", "gens", "problème" et "fait"... Pour le groupe nominal, une autre loi curieuse mérite d'être soulignée : dans la plupart des textes écrits, l'adjectif "grand" vient en tête ; dans la conversation, "petit" prend sa place !

- les pronoms "ce", "on" et "il". Les constructions "c'est" et "ce que" sont un trait caractéristique de l'oral de même que l'usage du "on" comme "nous" familier ;

- l'adverbe "pas" est 3,5 fois plus employé à l'oral qu'à l'écrit. Le système de la négation est d'ailleurs très révélateur (tableau III).

Tableau III. Les trois adverbes les plus fréquents à l'oral et à l'écrit

	Oral (%)	écrit (‰)	écart (oral/écrit)
pas	19,4	5,6	+245 %
ne	12,1	11,5	+6 %
plus	6,2	5,2	+19 %

L'adverbe "ne" est toujours suivi d'un verbe mais ce verbe peut être potentiellement suivi d'une multitude d'adverbes différents : "pas", "plus" et "point", naturellement, mais aussi des adverbes suivis de "que" (n'aimer vraiment que...) etc. En tout état de cause, si cette "règle" s'appliquait à l'oral, la fréquence du "ne" devrait toujours excéder le total des "pas" et des "plus"... En fait, "pas", à lui seul, dépasse "ne" de plus de 40%. On en tire une double conclusion évidente :

— "pas" a un spectre d'utilisation beaucoup plus large à l'oral qu'à l'écrit. Il n'est pas employé que dans le groupe verbal mais aussi couramment dans le groupe nominal ("un appareil pas au point", "une personne pas intéressante") ;

— la majorité des locuteurs "oublie" le premier terme de la négation dans le groupe verbal. Pour la majorité des Français, "j'en veux pas" est la forme canonique de l'oral et "je n'en veux pas" une manière un peu précieuse de s'exprimer. A ce propos, il faut rappeler que notre corpus ne donne pas la véritable mesure de ce phénomène car, comme nous l'avons souligné au début, une partie des secrétaires chargées de la transcription ont cru nécessaire de corriger ces phrases qui ne les avaient probablement pas choquées à l'audition mais qui, une fois transcrites, leur paraissaient incorrectes...

De nombreux auteurs ont déjà souligné ce phénomène, mais lequel aurait osé affirmer qu'il est aujourd'hui devenu la pratique dominante, et ceci dans toutes les couches de notre société ? S'agit-il encore d'une "faute" si elle n'est plus ressentie comme telle par la quasi-totalité des locuteurs d'une langue ?

Comme on le pressent, les corpus étiquetés représentatifs du français révéleraient bien d'autres phénomènes étonnants concernant la langue usuelle (et la profondeur de notre ignorance sur les usages réels !)

En attendant, on dispose là d'indicateurs simples et commodes pour reconnaître un oral spontané et démasquer le discours préparé à l'avance et appris par cœur...

La grammaire de l'oral

Nous allons nous concentrer sur les densités des catégories grammaticales et plus précisément sur la coordination et la subordination (tableau IV).

Les écarts sont massifs et l'on peut affirmer qu'ils sont statistiquement significatifs avec moins d'une chance sur mille de se tromper, sauf dans deux cas : les infinitifs et les conjonctions de coordination qui nous occupent justement.

Tableau IV.
Densités des catégories grammaticales dans le français écrit et dans le français oral.

Catégories	Français écrit (%)	Français oral (%)	Ecart (oral/écrit en %)
Verbes	14.9	19.4	+30
<i>Formes fléchies</i>	9.5	13.6	+43
<i>Participes passés</i>	1.9	2.7	+42
<i>Participes présents</i>	0.4	0.1	-73
<i>Infinitifs</i>	3.1	3.0	-3
Noms propres	2.6	0.7	-73
Noms communs	19.3	13.8	-29
Adjectifs	5.9	3.4	-43
<i>Adj. participe passé</i>	1.0	0.3	-69
Pronoms	11.2	19.4	+73
<i>Pronoms personnels</i>	6.6	10.4	+58
<i>Pronoms démonstratifs</i>	1.00	2.80	+180
<i>Pronoms possessifs</i>	0.04	0.01	-78
<i>Pronoms indéfinis</i>	0.43	0.55	+28
<i>Pronoms relatifs</i>	2.29	2.45	+7
Déterminants	18.7	12.9	-31
<i>Articles</i>	12.3	9.2	-25
<i>Nombres</i>	2.6	1.7	-35
<i>Possessifs</i>	2.1	0.7	-67
<i>Démonstratifs</i>	0.8	0.4	-51
<i>Indéfinis</i>	0.9	1.0	+5
Adverbes	6.3	11.5	+82
Prépositions	15.6	11.0	-30
Conjonctions	5.1	7.0	+36
<i>Coordination</i>	3.0	3.0	-0
<i>Subordination</i>	2.0	3.9	+94
Mots étrangers	0.21	0.01	-79

La loi générale se lit aisément. L'oral sur-emploie le groupe verbal. Les verbes, pronoms et adverbes connaissent une hausse moyenne de 55%. Autrement dit, quand ils utilisent 100 constituants du groupe verbal à l'écrit, les Français en mobilisent 155 quand ils parlent. Les pronoms personnels connaissent une véritable explosion. A l'inverse, les constituants du groupe nominal (noms propres et noms communs avec les déterminants et les prépositions) connaissent un mouvement inverse et pratiquement équivalent.

Cette fuite devant le nom se manifeste même au sein des verbes puisque les formes verbales les plus proches du "substantif" (participe présent et infinitif) sont sous-employées. C'est particulièrement vrai du participe présent qui semble bien être un "temps de l'écrit". Remarquons au passage que l'excédent considérable des participes passés peut s'expliquer par les conditions particulières de certains entretiens qui portaient sur des événements passés. Par exemple, la quasi-totalité des entretiens sur la négociation collective ont eu lieu après la fin des discussions et la question essentielle était la suivante :

"Pouvez-vous me raconter de façon assez détaillée ce qui s'est passé à partir de la formation conjointe et jusqu'à maintenant ?"

Naturellement, il faut replacer les mouvements spécifiques à l'une ou l'autre des parties du discours dans le tableau d'ensemble. C'est ce que nous allons faire pour la subordination et la coordination en français oral. Le tableau IV appelle évidemment une objection : la classification

est sommaire par rapport aux mécanismes de la langue. Mais il s'agit ici d'obtenir une vue synthétique et globale. A un certain niveau de généralité, l'alternative est simple : si un groupe de mots est posé comme le principal du propos, on peut lui subordonner d'autres groupes de mots à l'aide d'un pronom *relatif* ou d'une *conjonction de subordination* ; en revanche, si l'on place plusieurs mots ou propositions sur le même plan, on utilisera la *coordination*. Le tableau IV suggère que ce système n'est pas utilisé de la même manière à l'écrit et à l'oral.

III. Les pronoms relatifs en français oral et écrit

Globalement, les pronoms relatifs semblent présenter à peu près les mêmes densités à l'oral et à l'écrit, mais cela cache des emplois et des équilibres assez différents (tableau V). Un sur-emploi considérable du démonstratif est la première caractéristique du français oral. La seconde caractéristique est une fuite manifeste devant "dont" ou devant les constructions avec "lequel" (auquel, duquel, etc), probablement parce que les structures de phrases correspondantes sont trop complexes et peu usuelles. Enfin, le pronom "que" réserve une surprise : globalement la densité d'emploi est la même. Est-ce à dire qu'on en fait le même usage à l'oral et à l'écrit ? Pour le savoir, on utilise les tables de concordance. Certes, la consultation d'une telle masse est assez malcommode (plusieurs dizaines de milliers de lignes pour les deux corpus), mais il est possible de trier les contextes avant et arrière. Le tableau VI résume les principales conclusions.

Tableau V.
Les principaux pronoms pouvant entrer une construction relative

Pronoms	Français écrit (‰)	Français oral (‰)	Ecart (oral/écrit en %)
ce	0.72	2.68	+272
dont	0.11	0.02	-84
lequel	0.05	0.02	-55
que (pro)	0.60	0.60	-0.1
qui	0.88	1.03	+17

En français oral, les emplois relatifs (ie pour "subordonner une proposition "relative" à un groupe nominal) représentent moins de 4 emplois sur 10. Cette proportion semble plus que double dans le français écrit (proportion obtenue par sondages : nous n'avons pas eu le temps d'opérer un recensement exhaustif)... Cette baisse considérable est logique puisque le pronom "que" suit normalement un groupe nominal et que le groupe nominal pèse beaucoup moins lourd à l'oral. Il faut d'ailleurs noter l'importance des tournures "familières" du type : "chose que", "truc que", etc. Ces expressions fonctionnent un peu comme des "déictiques" : elles rattachent le propos à un objet déjà désigné, par la parole ou par le geste, ou au thème traité dans l'entretien...

En début de phrase, le pronom "que" sera interprété comme exclamatif ou interrogatif mais les emplois interrogatifs peuvent se trouver n'importe où dans la phrase, par exemple : "qu'est-ce que" (le second "que" étant une conjonction) ;

Tableau VI. Les emplois du pronom "que" dans le français oral

Construction	Effectifs
Derrière une ponctuation :	
Début de phrase	490
Autres ponctuations	1181
<i>dont interrogation</i>	60
Autres débuts de phrase ou proposition*	82
Total :	1813 (14%)
Constructions non strictement relatives :	
"ce que"***	5033 (38%)
"ça que"	303
"qu'est-ce que"****	804
"celui que"	156
"quoi que", "qui que"	115
Autre pronom + que	25
Total	6436 (48%)
Groupe nominal + que	
<i>dont "chose que"</i>	348
<i>"truc que"</i>	81
"que... et que...", "que... ou que..."****	253
Total emplois strictement relatifs :	5172 (39%)
Total emplois pronominaux de que :	13421 (100)

* Du type : "Donc *que* faire ?"

** sauf dans la construction "est-ce que" où "que" est conjonction

*** seul le premier que est pronom

**** lorsque le premier "que" est pronom.

Le tableau VI souligne l'importance de "ce que" et "ça que". A eux seuls, ces deux emplois pèsent 40% du total, c'est-à-dire quantitativement plus lourd que tout les emplois "relatifs" discutés ci-dessus. Le pronom "que" est donc d'abord employé derrière un démonstratif (omniprésent dans la communication orale). En quoi cette construction peut-elle être considérée comme "relative" ? L'examen des concordances montre que la majorité des emplois obéit à trois types de constructions :

- "(voilà) **ce que** (je peux dire, faire, penser)"
- "(c'est) **ce que** (je pense, je dis...)"
- parfois "ce que" intervient au début d'une phrase, voire d'une réponse, pour rattacher le propos qui suit à ce qui a été dit avant, par le locuteur, voire par l'interlocuteur.

Dans ces trois cas, il faut interpréter "relatif" dans un sens beaucoup plus étendu que celui des grammaires traditionnelles. Ces constructions n'introduisent plus une proposition relative — à moins de considérer que le pronom joue à lui seul le rôle d'une principale ? — elles soulignent un lien "relatif" (logique) entre plusieurs propos tenus avant ou après, soit par le locuteur, soit par son interlocuteur...

Enfin, dernière remarque concernant le pronom "que" : les grammaires consultées font un grand cas des emplois "fautifs" du pronom "que", là où, selon les auteurs, il faudrait "dont" ou "lequel". L'on sent souvent percer chez ces auteurs — même les mieux intentionnés — une pointe d'agacement envers ce "français populaire". En fait, la lecture des concordances montre que les "fautes" sont rares. Si "dont" et "lequel" reculent considérablement, c'est d'abord parce qu'à l'oral, tous les Français (et pas seulement le "peuple"), évitent les constructions lourdes et affectées. Cependant, il est juste de mentionner qu'à l'oral, "que" est parfois employé pour "lequel". Par exemple, "il y a des petits bruits *qu'on* est *pas* habitué". La "faute" se voit à la

lecture. Pourquoi ne choque-t-elle pas à l'oral ? Pris dans le flux de l'entretien, l'auditeur d'une phrase comme celle-ci la complète spontanément. Il comprend que l'on veut dire : "des bruits que nous (ne) sommes pas habitués (à entendre)", et non pas : "des bruits *auxquels* nous ne sommes pas habitués". Nous retrouverons plus bas, cette part d'implicite naturelle à l'oral. Ne devraient-on pas mieux prendre en compte ces phénomènes courants ? Ne faudrait-on pas être plus tolérant avec les pratiques dominantes et cesser de les stigmatiser comme des "fautes" ?

"Qui" est l'autre grand pronom à l'oral. Là encore, ses emplois strictement relatifs sont minoritaires. Sa fonction interrogative est nettement développée à l'oral du fait de la prépondérance des groupes verbaux appelant un "qui" en début de subordonnée. Pour le reste, l'examen des concordances montre un système comparable à celui que nous venons de décrire et notamment un poids considérable de : "ce qui" et "ça qui", "celui qui", etc.

Au fond à l'oral, les pronoms relatifs fonctionnent comme des sortes de "déictiques" et non pas comme de véritables "subordonnants", au moins au sens formel que donnent les grammaires françaises au chapitre de la "subordination"... La même conclusion semble devoir être tirée de l'examen du système des conjonctions à l'oral.

IV. Les conjonctions en français oral et écrit

Le tableau VII donne le détail de ces mouvements pour chacune de ces conjonctions.

Tableau VII. Récapitulatif des conjonctions à l'écrit et à l'oral

Conjonctions	Français écrit (‰)	Français oral (‰)	Ecart (oral/écrit en %)
car	0.46	0.05	- 88
cependant	0.23	0.01	- 95
comme	3.01	3.38	+ 12
donc	0.83	4.39	+ 428
et	22.14	14.23	- 36
lorsque	0.32	0.15	- 52
mais	3.80	8.39	+ 121
ni	0.49	0.21	- 58
or	0.11	0.03	- 69
ou	2.11	2.75	+ 31
parce que	0.50	5.79	+ 1 059
puisque	0.31	0.50	+ 59
quand	1.38	5.58	+ 303
que	12.36	18.90	+ 53
quoique	0.05	0.03	- 36
si	2.61	4.60	+ 77
sinon	0.03	0.38	+ 1071
tandis que	0.16	0.09	- 43

En ce qui concerne la subordination, les mouvements internes sont de très grande ampleur et, comme pour les pronoms, ils tendent à la simplification : "lorsque" et "tandis que" ou "quoique" n'occupent plus qu'une place marginale. En fait, le système se réduit à : que, parce que, quand et puisque.

Les concordances montrent l'importance des petites locutions conjonctives du type "à condition que". Avec la difficulté que ces locutions sont parfois difficilement distinguables d'un groupe nominal "libre". Par exemple, dans le corpus on rencontre : "le fait que vous

mentionnez" (pronom) et "le fait que vous veniez..." (conjonction). Il est donc essentiel que l'automate chargé de l'étiquetage ne soude pas automatiquement ces locutions. Elles peuvent ensuite aisément être retrouvées grâce aux concordances.

Les conditions particulières à l'oral expliquent en partie ces mouvements. L'oral est toujours une discussion, une explication ("parce que", "puisque") ou un récit ("quand") et comporte de nombreuses marques d'énonciation (dire que, penser que, faire que...).

Au total, la croissance de "que" est exactement corrélée à l'augmentation du groupe verbal et, au sein de celui-ci, des verbes pouvant introduire une subordonnée. C'est même une autre caractéristique qui permet de détecter avec certitude un texte oral spontané.

En ce qui concerne la coordination, la stabilité globale du groupe des coordonnants est la résultante de mouvements contradictoires de forte ampleur.

Même si son poids reste considérable à l'oral, la conjonction "et" pâtit clairement de la faible densité du groupe nominal. En effet, l'examen des concordances montre que la grande majorité de ses emplois sont du type "nom et nom" ou "nom adjectif et adjectif". C'est pourquoi son recul est pratiquement de même ampleur que celui du couple substantif + adjectif. Autrement dit, la copule est l'instrument privilégié pour la construction du groupe nominal complexe et ce groupe est plutôt une caractéristique du français écrit (tout comme l'usage "soutenu" du "et" dont cette phrase donne un exemple).

Trois autres conjonctions semblent bien appartenir au français écrit : car, ni et or. Pour "car" et "or", il s'agit même plutôt d'un registre soutenu (on en trouve surtout dans le corpus littéraire), ce qui explique leur quasi-disparition à l'oral.

Trois conjonctions connaissent un mouvement inverse dont l'ampleur compense globalement les reculs considérables que nous venons de signaler : "ou" et surtout les conjonctions "mais" et "donc" qui explosent littéralement.

Un cas limite : "donc"

Le cas le plus intéressant est probablement offert par "donc". La théorie souligne habituellement que ce mot peut avoir au moins deux fonctions différentes :

(1) conjonction établissant un lien logique entre deux éléments de la phrase (A -> B) ;

(2) adverbe apportant une nuance à un verbe ("entrez donc") ou à un groupe verbal : "il est donc idiot". La "fonction adverbiale" est considérée comme particulièrement nette quand le mot est inséré entre l'auxiliaire et le participe passé ("il est donc arrivé") ;

A titre anecdotique, on peut aussi le voir employé comme un nom : "ce "donc" qui nous occupe actuellement"...

Remarquons que :

- on peut supposer qu'à l'oral (2) est en augmentation sensible puisque l'oral se caractérise par un suremploi du verbe et des adverbes...

- si effectivement, dans les corpus représentatifs, du français écrit et oral, il apparaissait que les emplois adverbiaux de "donc" sont fréquents et nettement disjoints de la conjonction, il faudrait modifier la nomenclature des dictionnaires et les grammaires afin d'avoir deux entrées ("donc, adv." et "donc, cj."). Au passage, nous serions également obligés d'introduire une homographie de plus dans notre logiciel d'étiquetage...

Quel est "donc" le poids relatif de (1) et de (2) dans le corpus "français oral" ?

Nous pouvons répondre à cette question grâce aux concordances. Dans le corpus oral, il y a 9 936 occurrences de "donc". Les tris sur les mots placés immédiatement devant et derrière clarifient considérablement les choses. Les tableaux ci-dessous présentent les principaux emplois que l'on pourrait considérer comme "adverbiaux" avec leurs effectifs puis ceux qui sont, sans contestation, des conjonctions.

Tableau VIII. Emplois "adverbiaux" de la conjonction *donc*

Derrière :	Effectifs
- avoir	87
- être	67
- dire	51
- aller	13
- faire	12
- falloir	12
- autres verbes	51
Total adverbess	298 (3%)

Tableau IX. Emplois coordonnants de la conjonction *donc*

Position :	Effectifs
Derrière une ponctuation faible	3 750
En tête de phrase	2 934
Fin de phrase :	
- <i>trois points</i>	448
- <i>autres</i>	298
"et donc"	434
"alors donc"	85
"voilà donc"	29
"puis donc"	27
"après donc"	13
...	
Total conjonctions	9 639 (97%)

La conjonction présente une très grande mobilité dont ne rend pas totalement compte le tableau IX ci-dessus. En fait, la majorité des emplois de cette conjonction montre qu'en français, elle joue un rôle de "macro-coordonnant". C'est la raison pour laquelle on en trouve tant en début de phrase. Mieux : près d'une centaine de fois, on rencontre "donc" en début de réponse. On peut interpréter cet emploi particulier comme :

"(Vous me demandez A) **donc** je vous réponds B"

L'avantage est évident : on ne répète pas A ; on suppose implicitement acquis entre les participants à l'interlocution tout ce qui précède la conséquence que l'on va présenter.

En définitive, tous les emplois "adverbiaux" peuvent être interprétés ainsi. Quand, dans l'une des enquêtes citées, un enquêté déclare "mon chauffage est donc déréglé", il n'emploie pas "donc" pour renforcer le groupe verbal — comme le ferait un adverbe du type "bien" ou "vraiment" à la place de "donc" —, il tire la conclusion logique des phrases précédentes qui décrivent les dysfonctionnements de son appareil A l'oral, "donc" placé derrière un verbe renvoie toujours au schéma (A -> B), c'est un coordonnant, pas un adverbe.

L'examen des concordances de "mais" montre que, dans la majorité des cas, le fonctionnement est sensiblement le même, simplement on pose une contradiction logique au lieu d'une déduction logique.

A cette fonction, on rattache tous les cas où "donc" ouvre une phrase. Au milieu d'un propos, commencer une période par "donc" consiste à dire à son auditoire : "de tout ce qui précède (A), je tire logiquement ce qui suit (B)". L'annexe III en donne plusieurs exemples.

Si l'on y ajoute tous les "donc" placés derrière une ponctuation faible (ie au sein d'une même période, pour lier deux propositions), on a plus des deux tiers des emplois de "donc" dans le français oral ! Ils lient ensemble plusieurs arguments complexes en établissant entre eux un lien

logique, ce que ne fait pas, ou plus difficilement, la copule "et", d'où la désaffection relative qui la frappe à l'oral.

En définitive, la communication orale fonctionne suivant un principe d'économie. Avec les petites conjonctions "donc" et "mais", on peut indiquer à l'interlocuteur que l'on présuppose partagées certaines choses et s'épargner de les dire. Commencer une réponse par "donc" consiste à dire à l'interlocuteur : "puisque vous me posez cette question, voici ma réponse". En revanche, ouvrir sa réponse par "mais", c'est lui dire : "vous venez de dire A, je ne suis pas d'accord". A elle seule, cette fonction d'embrayage entre les propos des deux interlocuteurs pèse plus lourd que tous les "donc" placés entre l'auxiliaire et le participe passé !

Ainsi s'expliquent les très nombreux "et donc". Ces constructions sont d'ailleurs beaucoup plus nombreuses que les emplois prétendument adverbiaux de "donc". Or toutes les grammaires du français indiquent que "les conjonctions (de coordination) ne peuvent se cumuler entre elles" (par exemple : Arrivé & Al, p. 191). Elles en font même un trait caractéristique de la conjonction en français. En fait, notre corpus montre que cette "loi" ne vaut pas à l'oral, au moins pour "et" ou "mais" qui peuvent être combinés avec "ni" et "donc".

Pour compléter le tableau de la coordination en français, il faudrait encore examiner les ponctuations. En effet, dans la transcription de l'oral, la virgule correspond à une pause pas trop marquée suivie d'une reprise sans variation notable du ton. A l'examen, il semble qu'on peut pratiquement toujours remplacer cette ponctuation faible par une conjonction du type "et", "ou", "donc", sans altérer le sens de la phrase.

Au-delà, ce sont les structures de phrases qui devraient être prises en compte. Cela est maintenant possible puisque l'étiquette attachée à chaque mot permet de reconstituer finement ces structures. Les programmes informatiques sont en cours de développement, mais un autre exposé serait nécessaire pour évoquer ces questions.

Conclusions

La conclusion essentielle apparaît clairement : par rapport à l'écrit, **le français oral se caractérise par une simplification du vocabulaire et de la syntaxe. Certains mots outils — principalement pronoms, adverbes et conjonctions — servent non pas surtout à construire des phrases complexes, comme à l'écrit, mais d'abord à établir des liens logiques entre les propos de chacun des interlocuteurs et avec les objets qu'ils désignent par la voix ou par le geste.**

Naturellement, ce travail soulèvera des objections. Par exemple, dans nos corpus, les individus qui parlent et qui écrivent ne sont pas les mêmes ; la comparaison entre les textes n'est pas totalement justifiée car les conditions dans lesquelles ils ont été produits sont trop différentes ; enfin, le français "littéraire" pèse trop lourd dans le corpus de l'écrit et n'est pas représentatif de la manière dont nos contemporains écrivent réellement... Pour neutraliser l'influence respective du genre, du style personnel, de la situation et du thème, il faudrait par exemple, prendre le même homme, lui demander des textes sur un sujet donné et, parallèlement, recueillir ses propos spontanés sur le même thème. Alors nous pourrions voir apparaître les véritables différences existant entre oral et écrit. Eh bien ! ce cas existe dans mes corpus (Labbé, 2002b) : il s'agit du général de Gaulle qui rédigeait toujours soigneusement ses discours et conférences de presse... sauf en décembre 1965, où il a improvisé une série d'émissions télévisées évidemment sur les mêmes questions politiques. Nous avons comparé ces deux corpus : ils présentent exactement les caractéristiques qui viennent d'être esquissées : même augmentation considérable du groupe verbal — spécialement les verbes "faire", "dire", "aller", "falloir", de la première personne et du pronom démonstratif "ce" —, augmentation tout aussi nette de "chose" et "fait"... ; recul des conjonctions propres à l'écrit ainsi que de la copule "et",

renforcement notable des occurrences de "mais" et "donc"... Nous avons là une véritable expérience de laboratoire qui valide l'essentiel des conclusions que nous venons de présenter.

Quelle est la portée de tout cela ? Il s'agit simplement de joindre modestement notre voix à toutes celles qui souhaitent que l'on développe les outils indispensables pour l'étude scientifique de la langue : enquêtes d'usage (Hug, 2002) et vastes corpus étiquetés. Naturellement, ces enquêtes et ces corpus remettront en question bien des choses. Aujourd'hui, toutes les grammaires et tous les dictionnaires sont plus ou moins calqués sur les ouvrages du même type parus antérieurement, d'où le poids excessif d'un passé sans doute idéalisé. Pour le reste, les auteurs se fondent sur leur "sentiment linguistique" : les exemples sont forgés pour les besoins de la démonstration, ou tirés des "grandes oeuvres". De telle sorte que ces ouvrages offrent une représentation du français assez éloignée de la langue telle qu'elle se parle. Si le modèle colle encore à peu près pour les œuvres littéraires ou les grands textes politiques contemporains, on peut se demander s'il en est de même pour des genres écrits moins soutenus (presse, textes techniques, correspondance...) Pour répondre à cette question, il faudra constituer de grands corpus représentatifs de ces différents genres.

Evidemment, je n'ai pas l'ambition de combler des lacunes aussi considérables. Je ne peux que présenter des expériences à mon échelle et avec le matériel qui se trouve à ma portée. Des corpus "représentatifs" et étiquetés — comme ceux qui existent pour la plupart des grandes langues — ne peuvent être réalisés que par plusieurs équipes importantes, coopérant pendant une longue période autour d'un projet commun et dans le respect mutuel. Cela paraît inconcevable étant donné l'état des sciences humaines en France. Mais, surtout, il faudrait que l'on abandonne l'idée selon laquelle la langue française est l'affaire des seules "autorités" et non pas le trésor commun à tous ceux qui la parlent.

Bibliographie

- Arrivé M., Gadet F., Galmiche M. (1986) : *La grammaire d'aujourd'hui. Guide alphabétique de linguistique française*, Paris, Flammarion.
- Berger G., Leselbaum N. (dir.) (2002) : *La prévention des toxicomanies en milieu scolaire : éléments pour une évaluation*, Montpellier, CNDP.
- Bergeron J.-G., Labbé D. (2000) : "L'évaluation de la négociation raisonnée par les acteurs. Une analyse lexicométrique", (XVI^e Congrès de l'AISLF, Québec, juillet 2000) Reproduit dans Bernier C. et Al : *Formation, relations professionnelles à l'heure de la société-monde*, Paris-Québec, L'Harmattan - Presses de l'Université Laval, 2002, 239-252.
- Bergeron J.-G., Bourque R. & White F. (2003) : « Empirical Assessment of an Interest-based Bargaining Training Program in Labor-Management Relations », à paraître dans *Relations industrielles*.
- Blanche-Benveniste C. & Al. (1964) : *Grammaire Larousse du français contemporain*, Paris, Larousse.
- Brugidou M., Labbé D. (2000) : *Le discours syndical français contemporain (CFDT, CGT, FO en 1996-98)*, Paris, EDF - Division Recherche et Développement.
- Burnard L. (1995) : *Users Reference Guide for the British National Corpus*, Oxford, Oxford University Computing Service.
- Crowdy S. (1993) : "Spoken Corpus Design", *Literary and Linguistic Computing*, 8-4, 259-266.
- Douglas F. M. (2003) : "The Scottish Corpus of Texts and Speech : Problems of Corpus Design", *Literary and Linguistic Computing*, 18, 1-2, 23-37.
- Hug M. (2002) : *L'enquêtes d'usage linguistique*, Paris, Champion, 2002.
- Kucera K. (2002) : " The Czech National Corpus : Principles, Design and Results", *Literary and Linguistic Computing*, 17-2, 245-258.
- Labbé D. (1990) : *Normes de saisie et de dépouillement des textes politiques*, Grenoble, Cahier du CERAT.
- Labbé C. & Labbé D. (1994) : "Que mesure la spécificité du vocabulaire ?", Grenoble, CERAT. Repris dans : *Lexicometrica*, 3, 2001.
- Labbé D. (2001) : "Normalisation et lemmatisation d'une question ouverte. Les femmes face au changement familial", *Journal de la Société Française de Statistique*, 142-4, décembre 2001.
- Labbé D. (2002a) : *Analyse des représentations du confort électrique à partir d'un corpus d'entretiens*, Rapport pour le GREST-EDF, Grenoble, CERAT, juin 2002.
- Labbé D. (2002b) : "Le général de Gaulle en campagne", Communication aux III^e Journées de l'ERLA, *Aspects linguistiques du texte de propagande*, Brest, 15-16 novembre 2002.
- Labbé D., Monière D. (2003) : *Le vocabulaire gouvernemental. Canada, Québec, France (1945-2000)*, Paris, Champion.
- Leselbaum J., Labbé D. (2002) : "Lexicographie assistée par ordinateur. Signification de "Banque" dans le vocabulaire économique", in Morin A. et Sébillot P. (eds) : *VI^e Journées Internationales d'Analyse des Données Textuelles*, IRISA-INRIA, 2002, tome II, 447-456.
- Nelson G. (1997) : "Standardizing Wordforms in a Spoken Corpus", *Literary and Linguistic Computing*, 12, 2, 9-85.
- Pionchon S. (2001) : *Les Françaises et la politique*, Thèse pour le doctorat de science politique, Institut d'Etude Politique, Grenoble.
- Wagner R.-L. & Pinchon J. (1962) : *Grammaire du français classique et moderne*, Paris, Hachette.

Annexe I. Les corpus

Le « français oral »
(voir en bibliographie les ouvrages correspondants)

Les Français(es) et la politique (Pionchon) :
32 entretiens : 345 752 mots, 6 540 vocables différents
La négociation raisonnée au Québec (Bergeron & Labbé, 2000, Bergeron & Bourque, 2003)
61 entretiens : 409 225 mots, 6 591 vocables différents
La prévention des toxicomanies en milieu scolaire (Berger & Leselbaum) :
15 entretiens : 92 992 mots, 4 255 vocables différents
Confort électrique EDF réalisé par les sociologues du Grets en six enquêtes (Labbé, 2002) :
201 entretiens : 1 270 307 mots, 10 904 vocables différents
Questions ouvertes dans un sondage auprès des femmes divorcées réalisé par l'INED
(Labbé, 2001) :
3000 enquêtés : 56 107 mots, 2 786 vocables différents
Questions ouvertes dans un sondage auprès des citoyens belges sur la droite et la gauche :
1000 enquêtés : 22 294 mots , 1 706 vocables différents
Divers :
6 entretiens : 115 494 mots, 4 922 vocables différents
Total transcriptions de l'oral :
322 entretiens et deux sondages : 2 264 498 mots, 16 809 vocables différents

Le français écrit

• Vocabulaire presse économique 1996-98 (Leselbaum & Labbé, 2002) :
1077 textes : 1 517 228 mots, 32 829 vocables différents
• Vocabulaire syndical 1996-98 (Brugidou & Labbé)
336 textes : 216 759 mots, 8455 vocables différents.
• Premiers ministres canadiens, français et québécois depuis 1945 (Labbé & Monière, 2003)
143 textes : 610 258, 10720 vocables différents.
• Présidents de la République française : (Labbé 1990 et 2002)
147 textes : 507 142 mots, 10 115 vocables différents.
• Littérature française :
XVIIe 78 textes : 1 084 779 mots 10 604 vocables différents
XVIIIe 61 textes : 822 713 mots 14 958 vocables différents
XIXe 76 textes : 1 121 674 mots 17 941 vocables différents
Xxe 92 textes : 1 460 850 mots 24 770 vocables différents
Total corpus écrit
2 010 textes : 7 341 403 mots 59 367 vocables différents

* Les discours des Premiers ministres, des présidents de la République, le corpus "théâtre du XVIIe siècle" et divers extraits d'œuvres littéraires peuvent être obtenus auprès de : Oxford Text Archives (ota.ahds.ac.uk). Outre les textes originaux, on peut accéder à leurs versions étiquetées, ainsi qu'à divers outils logiciels et documents de recherche.

Annexe II. Le vocabulaire spécifique du français oral

Vocables significativement suremployés au seuil de 1 pour mille
(Classement par catégories grammaticales et spécificité décroissante)

Verbes : être, avoir, faire, dire, aller, falloir, pouvoir, savoir, voir, vouloir, mettre, penser, prendre, arriver, trouver, passer, parler, croire, travailler, payer, partir, demander, changer, chauffer, connaître, essayer, marcher, comprendre, vivre, rentrer, régler, appeler, expliquer, revenir, servir, commencer, sortir, ouvrir, utiliser, poser, arrêter, occuper, obliger, fonctionner, toucher, couper, allumer, installer, remettre, acheter, consommer, rencontrer, tourner, fermer, laver, intéresser, discuter, aider, coûter, amener, dépendre, importer, retrouver, exister, négocier, vendre, programmer, gérer, baisser, envoyer, éteindre, préférer, habiter, refaire, boire, construire, téléphoner, enlever, intervenir, voter, dépenser, isoler, convaincre, supporter, signer, compliquer, convenir, satisfaire, durer, repasser

Substantifs : chose, gens, problème, heure, fait, temps, fois, jour, chauffage, niveau, négociation, femme, façon, maison, moment, mois, travail, personne, côté, enfant, client, partie, rapport, système, truc, exemple, tout, électricité, part, question, facture, gaz, radiateur, compte, solution, point, besoin, mari, chambre, température, chaleur, cas, convecteur, journée, syndicat, installation, matin, hiver, soir, eau, suite, idée, comité, semaine, pièce, début, appareil, petit, compteur, degré, froid, formation, parti, argent, convention, départ, salle, famille, appartement, machine, confort, consommation, logement, attention, fonction, été, sens, cuisine, chaud, différence, bain, courant, impression, table, information, fenêtre, époque

Adjectifs : vrai, petit, bon, sûr, pareil, important, électrique, chaud, raisonné, gros, différent, difficile, cher, syndical, politique, froid, rouge, normal, facile, clair, collectif, capable, intéressant, simple, donné, creux, traditionnel, évident, patronal, content, dur, autre, technique, pratique, neuf, personnel, agréable, énorme, individuel, positif, précis, correct, partiel, demi, négatif, frais, pire, rayonnant, électronique, violent, socialiste, ouvrier, municipal, scolaire, conjoint, logique, objectif, gentil, sympa, performant, chauffant, automatique, absent, porteur, gagnant, bête, compliqué, super, compétent, communiste

Pronoms : ce, on, je, il, ça, qui, y, ils, le, moi, en, tout, quoi, rien, autre, leur, quelqu'un, celui-là, nous autre, même, eux autres, vous autre

Adverbes : pas, plus, oui, là, non, bon, bien, même, peu, alors, puis, très, aussi, enfin, peut-être, beaucoup, tout, toujours, vraiment, moins, trop, déjà, comment, maintenant, jamais, ici, assez, souvent, surtout, mal, pourquoi, justement, effectivement, tellement, plutôt, finalement, juste, autrement, ensemble, forcément, complètement, combien, là-dessus, là-bas, cher, normalement, actuellement, dehors, là-dedans, simplement, exactement, absolument, évidemment, pratiquement, dedans, certainement, relativement, énormément, d'accord, sûrement, directement, automatiquement, dessus, carrément, facilement, ci, franchement

Déterminants : tout, deux, trois, autre, même, certain, deuxième, sixième, un, notre, certain, soixante-dix, demi, zéro,

Conjonctions et prépositions : que, mais, parce que, quand, si, donc, comme, ou, après, voilà, avant, chez, puisque, pendant, sinon, soit, sauf, plein, vu, suivant, partir, strictement,

Vocables significativement sous-employés au seuil de 1 pour mille
(Classement par catégories grammaticales et spécificité décroissante)

Verbes : mourir, porter, entrer, assurer, agir, craindre, oser, courir, cesser, conduire, souffrir, poursuivre, disposer, constituer, briller, prier, demeurer, songer, jeter, offrir, régner, devenir, fuir, venger, emporter, tenir, trembler, crier, ajouter, céder, disparaître, répondre, ôter, prétendre, quitter, accroître, brûler, approcher, apparaître, traverser, devoir, glisser, achever, atteindre, rompre, troubler, promettre, affirmer, rendre, sembler, consentir, taire, punir, rire, douter, parvenir, cacher, mériter, embrasser, examiner, pleurer, échapper

Substantifs : gouvernement, amour, ciel, coeur, oeil, milliard, pays, loi, madame, marché, mer, âme, état, seigneur, million, virgule, roi, entreprise, main, visage, dieu, effet, voix, terre, grâce, sang, bras, vent, salarié, regard, honneur, nom, développement, mort, père, corps, croissance, action, prince, gloire, effort, groupe, soin, emploi, pierre, doute, ombre, amant, ministre, industrie, esprit, banque, objet, tête, flamme, vertu, crime, sort, voeu, douleur, capital, lumière, force, feu, haine, malheur, mot, cheveu, peuple, société, soleil, bonheur, nation, président, reine, bruit, réforme, dollar, empire, production, filiale, dessein

Adjectifs : grand, public, français, nouveau, noir, national, beau, européen, international, américain, doux, financier, mondial, économique, vain, fiscal, jeune, digne, heureux, industriel, entier, véritable, cruel, sombre, étranger, long, haut, profond, immobile, étrange, nécessaire, triste, immense, gris, canadien, législatif, secret, allemand, fédéral, chargé, destiné, seul, régional, pâle, unique, vieux, britannique, social, québécois, fidèle, illustre, assis, blanc, ingrat, comptable, jaune, puissant, propre, funeste, vaste, pauvre, nu, divers, mort, général, essentiel, pris, noble, généreux, éternel, prompt, léger, fort, fatal, dernier

Adverbes : ainsi, point, si, désormais, fort, lentement, aujourd'hui, bientôt, seulement, certes, toutefois, notamment, hélas, également, debout, pourtant, lors, loin, soudain, autrefois, guère, tant, encore, au-delà, auprès, doucement, d'autant, autour, assurément, environ, aisément, d'abord, aussitôt, naturellement, jadis, presque, davantage, droit, particulièrement, pleinement, au-dessus, au-dehors, naguère, brusquement, profondément, sitôt, néanmoins, çà, récemment, outre-mer, haut, hautement, promptement, conformément, par-dessus, largement, vivement, parfois, d'ores et déjà, progressivement, lourdement, fortement, longuement, tantôt, bas, nullement, étroitement, clairement

Pronoms : cela, dont, lui, vous, se, celui, lui-même, lequel, vôtre, mien, vous-même, sien, toi, celui-ci, personne, moi-même, toi-même, tien, quiconque, tel, aucun, quel, nul, eux-mêmes, nôtre, chacun

Déterminants : votre, mille, leur, le, cent, neuf, son, ce, mon, ton, quatre, notre, vingt, quel, soixante, premier, seize, cinquante, second, quarante, chaque, sept, treize, nul, trente, douze, quinze, quatorze, tel, unième, un, onze, aucun, neuvième, cinquantième, dix-neuvième, dix-huitième, quinzième, centième, seconde, trentième, seizième, vingtième, treizième, son, huitième, septième

Conjonctions et prépositions : et, sans, à, de, sur, sous, car, afin, devant, vers, dans, cependant, par, jusque, en, parmi, ni, outre, depuis, malgré, pour, selon, entre, lorsque, contre, voici, or, dès, hors, tandis que, durant, pourvu, envers, touchant, moyennant, quant, jusque-là, quoique, tandis