LA PLACE DE L'ADJECTIF EN FRANÇAIS (IV) MÉTHODOLOGIE LEXICOMÉTRIQUE

Thierry Trubert-Ouvrard
trubert@seinan-gu.ac.jp

INTRODUCTION

Les pages suivantes, dans une forme légèrement différente, et plus brève, ont été prononcées le 29 novembre 2002 à la Maison des Sciences de l'Homme (Paris), lors du colloque « Adjectifs non-prédicatifs » organisé par le Laboratoire de Linguistique Informatique de l'Université Paris 13[1]. Complétées ici par deux annexes en supplément, elles présenteront une introduction à des méthodes de traitement lexicométrique automatique d'un corpus, et proposeront ces méthodes en particulier pour l'analyse de la place des adjectifs (désormais : adj.) en position épithète dans le groupe nominal. L'analyse lexicométrique permet en effet de répondre sous l'angle nouveau de l'exactitude statistique, et non plus selon des analyses de *morceaux choisis* ou des représentations intuitives et subjectives de la langue. Elle répondra par exemple à des interrogations morphologiques telles que : les épithètes se trouvent-elles soit avant ou soit après le nom (désormais : N) qu'elles qualifient, ou ont-elles une position ambivalente ? Quelle est la fréquence de ces positions, quels sont ces adj. et quels sont ces N ?

Il faudra pour cela se munir d'outils linguistiques efficaces qui permettront de faciliter l'observation de cette problématique pour la résoudre.

---

[1] Thierry Trubert-Ouvrard : *Inventaire des adjectifs épithètes selon leur position anté- ou post-nominale et concordance entre les analyses lexicométriques et prédicatives,* colloque «Les adjectifs non-prédicatifs», 27-28 novembre 2002, organisé par le LLI (CNRS - UMR 7456) à la Maison des Sciences de l'Homme Paris Nord, 4 rue Croix Faron, La Plaine St-Denis, 93210.

Nous présenterons ici deux outils linguistiques semi-automatique et automatique et leurs règles permettant de dégager une observation fiable de la langue selon les méthodes statistiques.

1. VOLUME DU CORPUS

1.a. Quantité des occurrences

Une quantité minimale d'occurrences d'un même adj. sera nécessaire pour une observation exacte des fréquences de sa position post- ou anténominale. Si l'on considère par exemple l'adj. *immense* en position épithète, dans un corpus d'œuvres de fiction contemporaine, l'inventaire d'une cinquantaine d'occurrences révèle que sa position anté-nominale est majoritaire (dont la fréquence est de quatre-vingt-seize pour cent). Cependant, en poursuivant le dépouillement d'une centaine d'occurrences supplémentaires (on obtient alors cent cinquante occurrences), on observe que la fréquence de l'antéposition nominale de cette épithète diminue, pour atteindre quatre-vingt-quatre pour cent (c'est-à-dire une baisse relative de douze points). Cette diminution est significative de la marginalité des résultats d'une analyse limitée à cinquante énoncés, son observation s'avère insuffisante en nombre. On devra poursuivre le décompte des occurrences en se donnant pour objectif la parfaite stabilisation de l'indice de fréquence de la position de l'adj. par rapport au N. Ce but sera atteint dans le cas de l'adj. *immense* à partir de six cent cinquante occurrences observées. L'indice de fréquence se stabilise alors à quatre-vingt-six pour cent dans le cas de l'antéposition au N (et à quatorze pour cent dans celui de sa postposition). Que l'on poursuive le décompte de cent occurrences supplémentaires ou de cinq cents, le résultat ne varie plus : l'indice de fréquence est parfaitement stable du point de vue statistique.

Cette observation de la position statistiquement évolutive de l'adj. épithète *immense* est-elle marginale ou conforme à une règle générale alors découverte (pour la fiction contemporaine) ? La quantité minimale de six cent cinquante occurrences observées est-elle nécessaire à l'obtention d'un indice de fréquence représentatif du corpus étudié ? Afin de répondre à cette question nous avons observé le comportement d'autres adj. (pour les critères de choix des adjectifs pris comme point de départ à cette

étude, nous prions le lecteur de se reporter à la première partie de notre étude précédente[2]).

Le dépouillement d'autres adj. en position épithète aura permis d'observer qu'en règle générale la stabilisation de l'indice de fréquence de leurs positions anté- et post-nominales s'opère entre quatre cents et six cents occurrences dénombrées. Ainsi, dans le cas de l'adj. *précaire,* toujours en position épithète, l'indice est stable à partir de quatre cent vingt occurrences[3], dans celui de l'adj. *profond,* à partir de six cents.[4]

Une première conclusion s'esquisse déjà devant cette importante quantité d'occurrences nécessaire (entre quatre cents et six cents) pour connaître la fréquence moyenne de la position nominale d'une épithète spécifique : la nécessité d'avoir accès à un traitement lexicométrique automatique ; l'observation de la langue en lisant ou en écoutant parler les gens ne saurait être suffisante à une tentative de généralisation de la tendance de celle-ci.

1.b. Choix du corpus

Afin de parvenir à des résultats fiables en ce qui concerne la stabilisation de la fréquence de la place de l'adj. dans le groupe nominal, il apparaît donc qu'un corpus vaste est souhaitable à une étude lexicométrique, vu le nombre important d'occurrences nécessaire à ces résultats. Si l'étude d'adj. dont la fréquence est élevée, comme les adj. *petit* ou *grand,* ne saurait rencontrer de trop grandes difficultés (un corpus com-

---

[2] Thierry Trubert-Ouvrard : «La place de l'adjectif en français (II) - Étude sur corpus avec *immense* associé à d'autres adjectifs dans le GN», cet article est accessible également sur notre site Internet à l'adresse suivante : http://www.seinan-gu.ac.jp/~trubert/adjectif3.html.

[3] Thierry Trubert-Ouvrard : «La place de l'adjectif en français (III) - À propos de la catégorie et de l'épithète *précaire*», cet article est accessible également sur notre site Internet à l'adresse suivante : http://www.seinan-gu.ac.jp/~trubert/adjectif5.html.

[4] Thierry Trubert-Ouvrard : «Dimensions et particularités fonctionnelles syntaxiques de la position de l'adjectif épithète en français contemporain», cet article est accessible également sur notre site Internet à l'adresse suivante : http://www.seinan-gu.ac.jp/~trubert/adjectif6.html.

prenant quinze mille adj. serait suffisant[5]), l'analyse d'adj. moins fréquents ou rares sera plus compliquée à réaliser. Les très grands corpus comme *Frantext*[6], ou le *Gallica*[7] de la Bibliothèque Nationale de France, seront alors satisfaisants pour cette discipline, ils obligeront toutefois à une étude limitée aux langues d'hier, et sans possibilité de discernement des époques. L'Internet dans son ensemble, grâce aux moteurs de recherche comme *Google*[8], constitue un corpus encore plus vaste, mais là non plus ne sauraient être évaluées les époques diverses, ni les trop nombreux genres et niveaux de langues. Un tel niveau de généralisation de la langue prise dans son ensemble diachronique et synchronique peut-il aboutir à une quelconque analyse scientifique ?

Nous préférons recommander l'utilisation d'un corpus dont la source textuelle est connue, maîtrisée, et cadrée sur les plans temporel et du genre, et accessibles sur CD-ROM (ou sur l'Internet) comme les revues scientifiques, juridiques, *etc.,* les ouvrages de vulgarisation scientifique, les textes journalistiques, ou les dictionnaires. L'*Encyclopædia Universalis*[9] comprend environ vingt-huit millions de mots, *l'Encyclopédie de l'État du Monde*[10], deux millions et demi de mots, *Le Monde*[11], *Le Monde Diplomatique*[12] (en deux versions : archives années 1987-1997, neuf millions

---

[5] On pourra lire à ce sujet Marc Wilmet sur la fréquence relative des adjectifs qualificatifs épithètes : sur un total de 29 016 occurrences avec un adjectif dans le GN, les épithètes *grand* et *petit* sont très fréquentes : respectivement 1 034 et 1 139 fois.
WILMET, Marc (1980) : « Antéposition et postposition de l'épithète qualificative en français contemporain ».
WILMET, Marc (1981) : « La place de l'épithète qualificative en français contemporain. Étude grammaticale et stylistique ».

[6] http://www.atilf.fr/frantext.htm

[7] http://gallica.bnf.fr/

[8] http://www.google.fr/

[9] CD-ROM *L'Encyclopædia Universalis,* ou sur l'Internet : http://www.universalis.fr/

[10] CD-ROM *L'État du Monde, Annuaire économique et géopolitique mondial 1981-1998,* Éditions La Découverte, ou sur l'Internet : http://www.etatdumonde.com/

[11] http://www.lemonde.fr/

[12] CD-ROM *Le Monde Diplomatique,* ou sur l'Internet : http://www.monde-diplomatique.fr/

de mots, et archives années 1978-2004, vingt millions de mots), *Le Grand Larousse Illustré*[13], les différentes versions du *Robert électronique*[14], *etc.* En revanche se posera dans ces cas un obstacle majeur à l'analyse de la langue représentative du français contemporain : l'observation restera limitée à une langue très spécialisée.

En ce qui concerne les corpus de langue orale comme les *corpus d'Orléans,* de *Tours, d'Auvergne,* ou le corpus *LANCOM*[15], ils sont hélas quantitativement trop limités pour répondre à nos besoins. Leur accès est uniquement *via* Internet, mais sans possibilité de téléchargement, par le biais d'une concordancier (programme de recherche de concordances qui permet la recherche d'un mot ou d'une séquence de mots) offrant un nombre maximum de neuf cent quatre-vingt-dix-neuf concordances. Cette quantité est suffisante selon notre estimation du minimum indispensable à la représentativité de la langue, néanmoins le résultat de nos recherches n'est pas satisfaisant. Tous corpus réunis, ce qui oblige à répéter une même opération plusieurs fois car l'accès aux divers corpus est compartimenté, le résultat final était seulement de huit occurrences pour l'adj. *immense* (qui n'est pourtant pas un adjectif rare), cent trente-deux occurrences pour l'adj. fréquent *joli,* et sept cent soixante-six occurrences pour l'adj. le plus fréquent : *grand.* Pour comparaison : le corpus de fiction contemporaine que nous avons mis en œuvre propose respectivement pour ces adj. mille deux cent vingt-cinq, cinq cent vingt-sept, et sept mille six cent trente occurrences. Comme nous l'avons observé ultérieurement, les critères pour une recherche significative de la place des adj. dans le groupe nominal ne sont pas remplis.

1.c. Forme du corpus

Préférablement à l'utilisation de corpus déjà existants dont l'intérêt et

---

[13] CD-ROM *Le Grand Larousse Illustré ou sur* l'Internet : http://www.larousse.fr/grand_larousse_illustre/

[14] CD-ROM *Le Robert électronique* ou sur l'Internet : http://www.lerobert.com/editions-electroniques/

[15] Ils sont accessibles à l'adresse Internet suivante du site *Elicop* (Etude LInguistique de la COmmunication Parlée) de l'Université Catholique de Louvain : http://bach.arts.kuleuven.be/elicop/

la fiabilité sont limités en ce qui concerne l'analyse lexicométrique, comme il l'a été observé précédemment, on pourra se doter d'un corpus numérisé original, à partir de textes choisis en fonction de ses propres critères de recherche. Un tel corpus aura l'avantage d'être extensible car pourront y être ajoutés régulièrement d'autres textes. Si la langue écrite contemporaine est l'objet d'étude, on prendra soin de choisir des textes postérieurs à la seconde guerre mondiale, car la langue y comporte moins d'expressions vieillies comme c'est souvent le cas avant 1945. Ce corpus sera constitué grâce aux procédés de la scanographie de livres, qui permet de les numériser en mode *texte* (le mode *image* est inutilisable pour notre traitement lexicométrique[16]), grâce aux échanges de fichiers *texte* avec d'autres chercheurs[17] poursuivant un mode d'analyse similaire, et au téléchargement de textes accessibles sur l'Internet ou sur CD-ROM.

C'est ce type de corpus que nous avons mis en œuvre : l'un des nôtres, celui d'œuvres de fiction, s'élève[18] à plus de sept millions et demi de mots (plus de quarante-trois millions de signes). Il est constitué de deux cent soixante-sept textes[19] écrits de 1945 à 2002, qui sont des essais, des romans, des nouvelles, des pièces de théâtre et des scénarios de cinéma (on ne saurait considérer ces deux derniers genres comme de l'oral mais comme de l'écrit *oralisé,* nous le classons donc dans le corpus de *fictions*). On pourra consulter la liste des œuvres choisies et leurs références dans l'annexe B éditée en supplément de ces pages. Ce choix pour un corpus d'œuvres de fiction a été motivé par la grande richesse linguistique qu'il présente, à la fois des styles et des thèmes, afin de pouvoir travailler sur une langue diversifiée, contrairement aux autres corpus écrits mentionnés précédemment, tout en conservant la rigueur chronologique de la contem-

---

[16] Le mode image, fac-simile de l'original, conserve l'aspect visuel du document original, alors que le mode texte est le résultat d'une saisie du document et permet une recherche par mot dans le texte.

[17] Je profite de ces pages pour remercier profondément le Professeur Makio Nishimura de l'Université Seinan Gakuin qui m'a ainsi transmis une grande partie de son corpus.

[18] (au moment où cette étude est mise sous presse : février 2005)

[19] (qui tombent sous la loi du code de la propriété intellectuelle (le droit d'auteur) mais qui ne sont utilisés et accessibles qu'uniquement dans le cadre personnel de ma recherche académique ; ils ne font l'objet d'aucune diffusion)

poranéité de la langue.

2. TRAITEMENT DU CORPUS EN SEGMENTS RÉPÉTÉS

On pourra utiliser un logiciel[20] qui génère une concordance en *segments répétés* d'une unité textuelle à observer dans son contexte (image 1) ; cela permettra de repérer dans le corpus analysé de nombreuses unités composées de plusieurs formes répétées dans le même ordre et à différents endroits. Certaines de ces unités s'avèrent être récurrentes, pouvant alors se soumettre aux méthodes statistiques utilisées pour analyser la distribution des formes simples, d'autres unités sont spécifiques et uniques (hapax). Le concordancier montre tous les *segments répétés* d'une *forme-pôle* en contexte, permettant un retour systématique au corpus et à l'environnement immédiat de la forme, comme nous le verrons dans les exemples ci-après[21]. Nous observerons les qualités et les avantages de l'outil lexicométrique qu'est le concordancier, et particulièrement les diverses fonctions du logiciel *Lexico3* dans son utilisation de recherche d'une épithète.

Image 1: Extrait d'une concordance autour de la forme-pôle *seul* dans le corpus *fiction* (logiciel *Lexico3* réalisé par l'équipe universitaire SYLED-CLA2T de Paris 3

[20] Nous choisissons le logiciel *LEXICO3* pour notre explication de possibilités offertes par un concordancier ; il a été réalisé par l'équipe universitaire SYLED-CLA2T de Paris 3 Sorbonne nouvelle (sur l'Internet : http://lexico3.no-ip.org/).

[21] Tous nos exemples de concordances sont tirés de notre corpus original de fiction contemporaine.

Sorbonne nouvelle)

Comme dans cet exemple de la forme-pôle *seul,* on pourra créer des concordances avec tout adj. pour *forme-pôle ;* nous observerons ci-après les modalités nécessaires à une recherche d'un adj. avec ce concordancier (l'image ci-avant est brute, pour les exemples suivants les images seront transformées en format texte et dans un tableau pour une lecture plus facile).

2.a. Largeur du contexte

Il est possible d'augmenter la largeur du contexte, c'est-à-dire de modifier le nombre des signes avant et après le mot recherché, afin d'observer le texte plus en avant ou après, comme l'illustre le tableau 1 ci-dessous :

Tableau 1 de concordances :

Largeur par défaut (40 signes) :		
ce dernier quart de siècle , il fut le	**seul**	ministre de l ' éducation nationale à savoir
sés . à côté du plus gros se dressait ,	**seul**	miracle , un pilier oblong de pierre de
e le faisait presque jamais - devant le	**seul**	miroir que contenait la ferme . elle y obser
recroquevillée sur son petit lit , son	**seul**	mobilier . après avoir contemplé un moment
lleux d ' être avec un homme , mais son	**seul**	modèle était djamal . . . blaise sera -
Largeur réduite (15 signes) :		
cle , il fut le	**seul**	ministre de l '
s se dressait ,	**seul**	miracle , un pilier
ais - devant le	**seul**	miroir que contenait
petit lit , son	**seul**	mobilier . après
omme , mais son	**seul**	modèle était djamal
Largeur augmentée (80 signes) :		
. nous verrons qu ' au cours de ce dernier quart de siècle , il fut le	**seul**	ministre de l ' éducation nationale à savoir tenir tête et aux " gourous
nt couchés trois pins , carbonisés . à côté du plus gros se dressait ,	**seul**	miracle , un pilier oblong de pierre de saga (1) . comment lui seul
ur chambre et se plaça - elle ne le faisait presque jamais - devant le	**seul**	miroir que contenait la ferme . elle y observa longuement un visage inconnu
etrouver mon calme . elle était recroquevillée sur son petit lit , son	**seul**	mobilier . après avoir contemplé un moment l ' oreiller et les draps brodés
nsé : comme ça doit être merveilleux d ' être avec un homme , mais son	**seul**	modèle était djamal . . . blaise sera - t - il à la hauteur ? sera - t

On remarque que l'option d'un contexte large permet d'observer celui-ci en amont et en aval du segment répété afin d'effectuer une analyse séman-

tique approfondie. Ainsi les cinquante-cinq concordances de «grand homme» dans notre corpus ont montré que, après analyse sémantique de tout le contexte, *grand* était prédicatif dans soixante pour cent des occurrences (et non-prédicatif dans quarante pour cent) — ces résultats sont contraires aux analyses des grammaires qui proposent uniquement comme référence *grand homme* non-prédicatif (*grand* ayant comme sens *éminent*).

L'option d'un contexte étroit (quinze signes) possède quant à lui l'avantage de repérer rapidement les segments répétés recherchés, pour les compter par exemple.

2.b. Tris *avant* et *après*

On peut également effectuer deux tris alphabétiques des différents contextes relatifs à la même *forme-pôle,* en plus de l'ordre chronologique d'apparition des occurrences de la *forme-pôle* dans le texte (sans tri alphabétique). Ces tris s'ordonnent de deux manières différentes : en fonction de l'ordre alphabétique de l'occurrence qui précède la *forme-pôle* (tri avant) ou de l'ordre alphabétique de l'occurrence qui suit la *forme-pôle* (tri après). Ces trois présentations sont illustrées dans le tableau 2 ci-dessous :

Tableau 2 de concordances :

Ordre du texte :		
d ' heure trop tôt . à côté de lui , un	**gros**	bonhomme était assis , habille à la mode
dagogie , mais il ignorait son nom . le	**gros**	bonhomme lui murmura à l ' oreille : «
n revolver , il écarte la haie avec son	**gros**	ventre : une paysanne est accroupie serrant
la soeur du chef de village , agita de	**gros**	bras noirs , et la foule fit silence . la
e vue : une montagne de fourrure . " un	**gros**	e grizzly " , murmura fedeler . l ' ours se
Tri avant :		
la soeur du chef de village , agita de	**gros**	bras noirs , et la foule fit silence . la
dagogie , mais il ignorait son nom . le	**gros**	bonhomme lui murmura à l ' oreille : «
n revolver , il écarte la haie avec son	**gros**	ventre : une paysanne est accroupie serrant
e vue : une montagne de fourrure . " un	**gros**	grizzly " , murmura fedeler . l ' ours se
d ' heure trop tôt . à coté de lui , un	**gros**	bonhomme était assis , habille à la mode
Tri après :		
d ' heure trop tôt . à côté de lui , un	**gros**	bonhomme était assis , habille à la mode
dagogie , mais il ignorait son nom . le	**gros**	bonhomme lui murmura à l ' oreille : «
la soeur du chef de village , agita de	**gros**	bras noirs , et la foule fit silence . la
e vue : une montagne de fourrure . " un	**gros**	grizzly " , murmura fedeler . l ' ours se
n revolver , il écarte la haie avec son	**gros**	ventre : une paysanne est accroupie serrant

Le tri *avant* permet l'observation du mot avant la *forme-pôle,* par exem-

ple la rescherche d'un article défini ou indéfini, d'un adj. possessif, etc., susceptible de modifier la position anté- ou post-nominale de l'épithète. Le tri *après* aide la recherche d'un N donné après la *forme-pôle* pour comparer ou compter les occurrences d'un même segment répété adj.-N.

2.c. Différences morphologiques

Les différentes morphologies que possède un même adj. (masculin-singulier, masculin-pluriel, féminin-singulier et féminin-pluriel) ne peuvent être recherchées simultanément. Nous savons qu'un adj. comme *beau* ou *profond* possède quatre formes (respectivement *beau, beaux, belle* et *belles ; profond, profonds, profonde et profondes*) : la recherche des concordances avec *beau* ou *profond* demande donc quatre recherches successives, comme l'illustre le tableau 3 ci-dessous :

Tableau 3 de concordances :

Recherche de «beau» :		
mon bébé ! - mais c ' est de la terre ! - mon	**beau**	bébé ! mais où es - tu , mon ange , mo
s acquises récemment lui laisseront encore un	**beau**	beau bénéfice . confiant . sérieux . nature
hambre l ' irritait à chaque fois . malgré le	**beau**	mobilier rustique français , une décor
Recherche de «beaux» :		
at , la pelle remplaçait le fusil , d ' où de	**beaux**	paragraphes dans les hebdomadaires ,
) : dolorès , elle est gentille et elle a des	**beaux**	seins , mais elle va trop vite . marc
lque chose de très juvénile . et puis il a de	**beaux**	yeux . natacha . malheureusement je n
Recherche de «belle» :		
des copies ? des moulages ? tu bricoles . la	**belle**	affaire ! jusqu ' à quand ? jusqu ' à
aré " sa tenue " civile " sous l ' uniforme .	**belle**	affiche de film romantique et musical
rison pour cinq ans . et leur faire payer une	**belle**	amende de surcroît . darcy fut persua
Recherche de «belles» :		
rouleau : mais elles sont derrière nous , les	**belles**	années , pendant la guerre , on écri
des raisonnements comme ça , on se prépare de	**belles**	années , tu vois ! rouleau : mais el
née fine et jaunâtre . nadal qui bâtissait de	**belles**	villas sur la corniche , au milieu d

En revanche un adj. comme *immense* ou *jeune* possède seulement deux formes (respectivement *immense* et *immenses ; jeune* et *jeunes*) puisque les formes masculine et féminine sont morphologiquement identiques ; on n'effectuera donc que deux recherches dans ce cas, comme l'illustre le tableau 4.

Tableau 4 de concordances :

Recherche de «immense» :		
bougie à la main, formèrent une	**immense**	chaîne humaine silencieuse à Munich le
à une multitude de possibilités,	**immense**	champ ouvert à la virtuosité. Le théoric
il contemple le paysage noyé, l'	**immense**	ciel gris, les arbres que fouettent le
Recherche de «immense» :		
unique pour se transformer en d'	**immenses**	cimetières indifférenciés... Le seul m
noirs s'élèvent vers le ciel. D'	**immenses**	cités grouillantes de vies et de lumiè
les oeufs de ses amours dans les	**immenses**	espaces blanchâtres de poussière, de s

2.d. Différences fonctionnelles

La recherche pourrait sembler simplifiée avec un adj. dérivant d'un N comme *albinos* ou *orange* qui ne possède qu'une seule *forme,* mais dans ce cas les résultats comportant à la fois des N et des adj., leur lecture pour éliminer les N parasites est en fait plus longue à effectuer (voir le tableau 5 ci-après). Dans leur grande majorité les adj. sont dérivés de N ou de verbes, mais ceux-ci correspondent aux adj. les moins fréquents dans la langue (on pourra consulter la liste de tous les adj. en français et des mots pouvant avoir une fonction adjectivale dans l'annexe A éditée en supplément de ces pages).

Tableau 5 de concordances :

Recherche de «orange» :		
ancing my love lovée dans la lumière mon	**orange**	mon coquelicot d ' hier ma pervenche
train avec un immense plumet de cheveux	**orange**	, le visage maquillé à la dracula et
: - à ce soir ! sans prendre le jus d '	**orange**	et les tartines grillées que vous lui
Recherche de «albinos» :		
te un instant . . . " c ' est marcel l '	**albinos**	, le tapeur . en attendant dans l '
iscuter , vient de lui " avancer " , l '	**albinos**	cherche déjà (d ' un regard vague
. salut , marcel . (c ' était le copain	**albinos**	venu lui emprunter de l ' argent ,

Les adj. les plus fréquents sont les épithètes originales[22] —voici les vingt premières dans l'ordre décroissant de fréquence en position épithète : *petit, grand, jeune, seul, vieux, nouveau, gros, bon, beau, long, plein, vrai, mauvais, propre, pauvre, ancien, léger, bas, cher et joli*). Par conséquent

[22] Lire à ce sujet : Thierry Trubert-Ouvrard : «Position des 20 épithètes les plus fréquentes en français».

la recherche de concordances est facilitée avec des adj. fréquents ; il faudra cependant toujours trier les occurrences d'adj. en position épithète de celles d'adj. en position attribut. En effet l'analyse des concordances s'effectue manuellement en lisant les résultats, le concordancier ne comporte pas de moyen automatique de sélection des adjectifs seuls, et non des N (ni des adj. en position épithète seulement et non des attributs *a fortiori,* voir le tableau 6). Il faut donc lire tous les contextes des segments répétés de l'adj. recherché pour éliminer les occurrences qui ne répondent pas aux critères demandés. Dans la recherche des fréquences d'un adj. épithète, cela demande donc la lecture d'un très grand nombre de lignes du concordancier pour obtenir entre cinq cents et six cents occurrences, minimum quantitatif requis comme nous l'avons observé en 1.a.

Tableau 6 de concordances :

Recherche de «joli» :		
ul mangeant un gâteau .) puis ce fut un	**joli**	bout de femme désinvolte qui tenait le co
) est située au plan - de - la - tour ,	**joli**	petit village à l'intérieur des terres ,
tre . jim . la vie ? neutre ? c'est très	**joli**	, et surtout très logique . plan sur cath
ère , si on t ' écoute . . . tu vas être	**joli**	à l ' arrivée . . . elle va se payer . .
r de paris . il faut que ce soit un très	**joli**	souvenir , tu me le promets . je suis déj
assourdi de cris , de " oh ! que c ' est	**joli**	" répétés et traduits en vingt langues ét

2.e. Concordances adj. + N

Dans le cas d'une recherche de concordances d'un adj. anté- ou postposé à un N spécifique, on devra opérer quatre recherches : adj.-N et N-adj. aux formes singulier et pluriel, comme dans le tableau 7 ci-dessous.

Tableau 7 de concordances :

Recherche de «changement profond» :		
Rabin . . . Ils attendent un	**changement profond**	dans la position israélienne
nt dans la vie du cinéaste un	**changement profond**	: sa femme , Letizia , le qui
ise du XIVe siècle , connu un	**changement profond**	dans ses structures sociales
Recherche de «changements profonds» :		
s continuent de craindre des	**changements profonds**	qui menaceraient leurs privi
téraire subissait encore des	**changements profonds**	et répétés , il s ' avéra au
spect nouveau , entraîna des	**changements profonds**	dans la conscience et la psy

Recherche de «profond changement» :
e Bourbon (Louvre) témoignent de ce **profond changement** par l ' ampleur du
nthétiseurs , ordinateurs . . .), un **profond changement** s ' est produit .
conquête par Prithvi Narayan amena un **profond changement** social . Les Gurkh
Recherche de «profonds changements» :
et l' Andhra Pradesh , a traduit de **profonds changements** . Dans ce qui fut
votants est un signe annonciateur de **profonds changements** . De fait , les é
/ La vie littéraire connaît aussi de **profonds changements** . L ' esprit de c

On pourra compter le résultat des occurrences N-adj. ou adj.-N afin de calculer un pourcentage et de comparer telle séquence avec une ou des autres pour observer la position d'un adj. spécifique avec un N spécifique. Par exemple «grande personne» apparaît soixante-dix sept fois (sur six mille quatre cent dix-huit occurrences de *grand*), mais «grand amour» trente-quatre fois, la première séquence pourra sembler plus figée que la seconde.

3. ANALYSE AUTOMATISÉE AVEC UN CORPUS CODÉ

3.a. Codage en banque de données

On a vu précédemment que le concordancier ne reconnaissait pas automatiquement les fonctions adjectivale, nominale, adverbiale ou verbale d'une même forme, ni les fonctions épithète ou attribut d'un adj., ce manque nécessitant une lecture manuelle des occurrences et une importante perte de temps passée à cet exercice. Des analyseurs syntaxiques comme *Cordial 8*[23] sont capables de catégoriser syntaxiquement chaque unité textuelle du corpus avec une grande précision. Dans le cas qui nous intéresse, le logiciel effectuera le codage des N et des adj.—cependant il ne peut pas non plus distinguer la position attribut de la position épithète de l'adj. Le catégoriseur *Cordial 8* créera une étiquette «ADJ» (*TAG*) pour chaque adj., indiquant également le genre et le nombre de celui-ci.

Cordial 8 présente l'intérêt majeur d'ajouter le *lemme* de chaque adj. identifé (le lemme est identique à la forme du masculin singulier, par ex-

[23] CD-ROM «Cordial 8 Analyseur» élaboré par Synapse Développement ; sur l'Internet :
http://www.synapse-fr.com/Cordial_Analyseur/Presention_Cordial_Analyseur.htm

emple le lemme de «belles» est «beau»). Les formes au masculin, au féminin, au singulier et au pluriel d'un même adjectif étant également identifiables par leur lemme unique, une seule recherche par adj. suffira pour un logiciel conçu pour rechercher des épithètes. Les quatre formes de genre et de nombre y étant traitées simultanément, offrant alors l'avantage d'un gain de temps inestimable, nous utiliserons ce codage syntaxique pour la réalisation du logiciel.

3.b. Conception d'un logiciel original

Le logiciel *Adjectif* que nous avons réalisé peut lancer la recherche d'un adj. spécifique (par exemple *petit* dans l'image 2 ci-après) en identifiant et comptabilisant uniquement les occurrences de celui-ci dans ses fonctions adjectivale (grâce à l'étiquette «ADJ» de l'analyseur syntaxique) et épithète. En effet, le logiciel identifie et sélectionne seulement les occurrences de l'adj. recherché qui sont suivies ou précédées d'un N (*cf.* la mention *matched* dans l'image 2). Notons aussi que les antépositions et les postpositions nominales sont comptées séparément, comme l'indique le résultat de la recherche de *grand* dans l'image 3 :

Image 2 : Extrait du résultat de *petit* dans le corpus *fiction*
(logiciel *Adjectif* réalisé par H. Mashima et T. Trubert de Seinan Gakuin University)

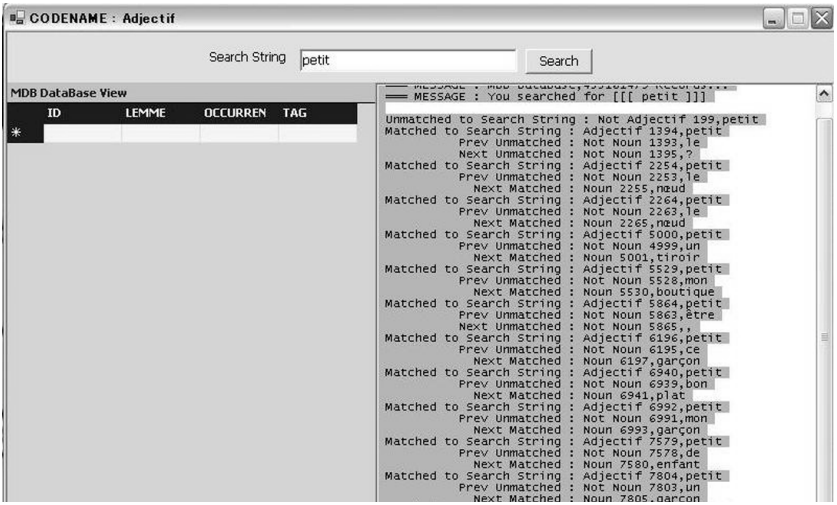
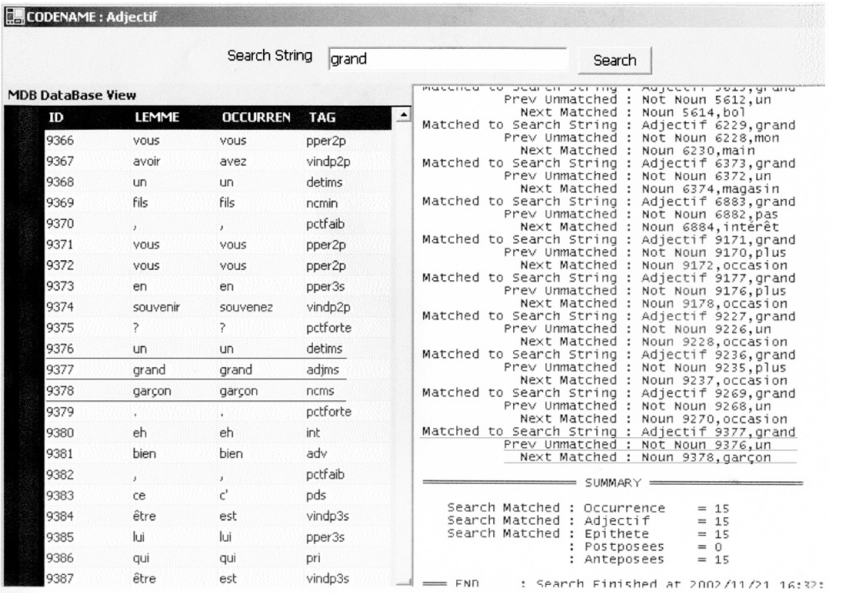


Image 3: Extrait du résultat de *grand* dans le corpus *fiction*
(logiciel *Adjectif* réalisé par H. Mashima et T. Trubert de Seinan Gakuin University)



Tout comme avec le concordancier, le logiciel permet un retour systématique au texte et à l'environnement immédiat de la forme. Dans la fenêtre de gauche, on peut retrouver le numéro d'identité (*I.D.*) attribué à chaque adj. de la fenêtre de droite, et y observer le contexte de l'occurrence ainsi que la validité de l'étiquette, comme l'illustre plus clairement sous forme de tableau l'exemplier suivant (les adj. sont marqués en caractères gras).

Tableau 8 des contextes et codages :

ID	OCCURRENCE	LEMME	TAG
114	tu	tu	pper2s
115	crois	croire	vindp2s
116	que	que	sub
117	c’	ce	pds
118	est	être	vindp3s
119	une	un	detifs
120	**mauvaise**	**mauvais**	**adjfs**
121	idée	idée	ncfs
122	?	?	pctforte
.../...			
3281693	un	un	detims
3281694	homme	homme	ncms
3281695	**seul**	**seul**	**adjms**
3281696	sur	sur	prep
3281697	qui	qui	pri
3281698	reposait	reposer	vindi3s
3281699	**toute**	**tout**	**adjfs**
3281700	la	le	detdfs
3281701	responsabilité	responsabilité	ncfs
3281702	de	de	prep
3281703	l’	le	detdfs
3281704	opération	opération	ncfs
3281705	.	.	pctforte
.../...			
3281735	c’	ce	pds
3281736	était	être	vindi3s
3281737	un	un	detims
3281738	**petit**	**petit**	**adjms**
3281739	homme	homme	ncms
3281740	vêtu	vêtir	vparpms
3281741	de	de	prep
3281742	vert	vert	ncms
3281743	**sombre**	**sombre**	**adjsig**
3281744	.	.	pctforte
.../...			
3281833	ses	son	detposs
3281834	collaborateurs	collaborateur	ncmp
3281835	se	se	pper3s
3281836	tenaient	tenir	vindi3p
3281837	à	à	prep
3281838	distance	distance	ncfs
3281839	**respectueuse**	**respectueux**	**adjfs**
3281840	.	.	pctforte

ID	OCCURRENCE	LEMME	TAG
3281886	une	un	detifs
3281887	mèche	mèche	ncfs
3281888	**noire**	**noir**	**adjfs**
3281889	et	et	coo
3281890	**lourde**	**lourd**	**adjfs**
3281891	,	,	pctfaib
3281892	en	en	prep
3281893	forme	forme	ncfs
3281894	de	de	prep
3281895	virgule	virgule	ncfs
3281896	,	,	pctfaib
3281897	ponctuait	ponctuer	vindi3s
3281898	son	son	detposs
3281899	**vaste**	**vaste**	**adjsig**
3281900	front	front	ncms
3281901	**blanc**	**blanc**	**adjms**
3281902	.	.	pctforte
.../...			
3282602	d’	d’	prep
3282603	une	une	detifs
3282604	beauté	beauté	ncfs
3282605	**pâle**	**pâle**	**adjsig**
3282606	et	et	coo
3282607	**fragile**	**fragile**	**adjsig**
3282608	,	,	pctfaib
3282609	et	et	coo
3282610	qui	qui	pri
3282611	parlait	parler	vindi3s
3282612	d’	de	prep
3282613	une	un	detifs
3282614	voix	voix	ncfin
3282615	**douce**	**doux**	**adjfs**
3282616	,	,	pctfaib
3282617	aux	au	detdpig
3282618	inflexions	inflexion	ncfp
3282619	**calmes**	**calme**	**adjpig**
3282620	,	,	pctfaib
3282621	une	un	detifs
3282622	voix	voix	ncfin
3282623	dont	dont	pri
3282624	l’	le	detdfs
3282625	**extraordinaire**	**extraordinaire**	**adjsig**
3282626	profondeur	profondeur	ncfs
3282627	rappelait	rappeler	vindi3s
3282628	celle	celui	pds
3282629	du	du	detdms
3282630	gardien	gardien	ncms
3282631	de	de	prep
3282632	nuit	nuit	ncfs
3282633	stojilkovitch	stojilkovitch	npi
3282634	.	.	pctforte

L'observation de chaque occurrence offre la possibilité d'analyser l'ensemble de l'énoncé et de vérifier par ailleurs la fiabilité du traitement automatique. En effet, de trop nombreuses erreurs venant parasiter la sélection de l'adj. recherché en position épithète modifieraient le résultat final et le fausseraient. Ainsi par exemple le système d'analyse du logiciel ne sélectionne pas les cas suivants dans le résultat final :

1) les prépositions *de* et *à* placées devant l'adj. *nouveau*+N indiquent un cas de locution adverbiale (*de/à nouveau*) et non celui de l'adj. *nouveau* caractérisant antéposé au N.
2) *marché* placé après l'adj. *bon* indique la locution adjectivale *bon marché* et non l'épithète *bon* caractérisant postposée au N.
3) *trouver* placé devant N+adj. (dans : «je trouve Bernard impressionnant») indique que l'adj. *impressionnant* n'est pas en fonction épithète, mais la séquence figée *trouver*+N+adj. tout comme avec *sentir* (dans : «je sentais Renaud fragile»)
4) *comme* placé après N+beau (dans : «une femme belle comme un camion») indique une séquence figée dans l'ordre N+adj. où l'adj. ne peut jamais être placé avant le N.
5) les occurrences adj.+N et N+adj. dans lesquelles les deux parties n'appartiennent pas aux mêmes catégories du genre ni du nombre.
6) *etc.*

Par ailleurs, la forte récurrence des segments répétés doit être prise en compte dans le résultat statistique, et ceux-ci réduits à une seule unité pour ne pas fausser ce résultat. Ainsi *grande personne* est une catégorie qui peut être analysée comme un lemme à elle seule, les soixante-dix sept occurrences trouvées dans notre corpus ne seront donc comptabilisées que pour une unité.

Grâce à l'identification automatisée des adj. pris uniquement dans leur fonction épithète (seulement sur un plan morphologique), cet outil linguistique offre, nous l'avons vu, l'immense avantage de remplacer le travail long et fastidieux de la lecture et de l'analyse manuelle des énoncés, quand ce travail a pour but de parvenir à des résultats pour les méthodes statistiques.

CONCLUSION

Les deux logiciels proposés dans ces pages sont utilisés pour calculer le nombre et le pourcentage des positions anté- et post-nominales des adj. de manière spécifique pour un adj. donné. Ces calculs pourraient cependant être généralisés à l'ensemble des adj. du français. Ils pourront servir aussi à analyser les éléments qui influent sur la position de l'épithète, en analysant manuellement les énoncés et en y corrélant les résultats statistiques. Ainsi par exemple, les énoncés avec l'adj. *autre* placé avant les séquences adj.-N et N-adj. pourront être décomptés afin de déterminer si l'adj. exerce une influence sur l'ordre des mots, il en ira de même avec des locutions adverbiales comme *bien de* et *bien des*. On pourra par ailleurs observer la position des adj. avant ou après le N selon la présence de pronoms possessifs ou de déterminants, afin de confirmer ou d'infirmer grâce aux méthodes statistiques les représentations linguistiques que l'on pourrait avoir : qu'il y a une tendance à *le+adj.*+N et à *un+N+adj.* par exemple.

Ces outils linguistiques pourraient enfin être fort utiles pour connaître la position des épithètes au travers d'exercices de comparaison entre la langue contemporaine de celles d'hier, entre les différents genres de l'oral et de l'écrit, entre les genres littéraire, journalistique, juridique, *etc.,* ou même entre deux auteurs spécifiques ou deux périodes distinctes d'une même source : les applications sont nombreuses.

BIBLIOGRAPHIE

Thierry Trubert-Ouvrard : «Dimensions et particularités fonctionnelles syntaxiques de la position de l'adjectif épithète en français contemporain», *in* Actes du Colloque *L'Adjectif en français et à travers les langues, Caen, juin 2001,* Presses Universitaires de Caen, 2004, pp. 279-292.

Thierry Trubert-Ouvrard : «Position des 20 épithètes les plus fréquentes en français » *in* Études de Langue et Littérature françaises, Université Seinan Gakuin, numéro 44, 2003, ISSN 0286-2409.

Thierry Trubert-Ouvrard : «La place de l'adjectif en français (III) - À propos de la catégorie et de l'épithète *précaire*» *in* Études de Langue et Littérature françaises, Université Seinan Gakuin, numéro 43, 2002, ISSN 0286-2409.

Thierry Trubert-Ouvrard : «La place de l'adjectif en français (II) - Étude sur corpus avec *immense* associé à d'autres adjectifs dans le GN» *in* Études de Langue et Littérature françaises, Université Seinan Gakuin, numéro 40, 2000, ISSN 0286-2409.

WILMET, Marc (1981) : «La place de l'épithète qualificative en français contemporain. Étude grammaticale et stylistique», *in* Revue de linguistique romane 177-178 tome 45, pp. 17-73.

WILMET, Marc (1980) : «Antéposition et postposition de l'épithète qualificative en français contemporain», *in* Travaux de linguistique 7, pp. 179-201.