

De vs des* devant les noms précédés d'épithète en français : le problème de *petit

Itsuko Fujimura¹, Mitsumi Uchida², Hiroshi Nakao³

¹ Université de Nagoya – 4648601 Nagoya – fujimura@cc.nagoya-u.ac.jp – Japon

² Université féminine d'Osaka – 5900035 Sakai – Japon

³ Université de Aïchi – 4418522 Toyohashi – Japon

Abstract

This study discusses the alternation of the French articles *de* and *des* in noun phrases with a preposed adjective. Traditionally, it has been argued that the choice is conditioned by register and by the strength of the collocation between the adjective and the noun. In our earlier study, however, it was shown that these two factors fall short of a conclusive description of the observed facts. Notably, the particular tendencies displayed by *petit(e)s* could not be accounted for. For a comprehensive characterization of the alternation, we carried out surveys using very large corpora. Based on statistical analyses, we will demonstrate that *petit* apparently has no special preference for composing compound words and then point out two other relevant factors: (1) lexical properties of the adjective; (2) semantico--functional status of the adjectival in the discourse. These two factors together constitute “weight” of the adjectival and affect the choice of the accompanying article: the lighter the weight, the more easily *des* is used; the less light, the more *de* is likely. While the notion of grammatical weight is regarded as discrete in Abeillé & Godard (1999, 2000), our notion of weight here is of a gradual nature and therefore can provide a relative characterization of every type (lexically) or token (in discourse) of adjectivals. It will be concluded that *petit* is located on the lowest end of the lexical scale of adjectives and tends to appear after *des*, which marks the “lightness” of the following adjective.

Résumé

En analysant les corpus de très grande taille, nous observerons d'abord quelques faits qui n'ont guère été abordés dans la littérature, notamment le degré de cooccurrence particulièrement fort entre *des* et *petit(e)s* dans le choix entre *de* et *des* devant les noms précédés d'épithète en français. Nous démontrerons ensuite, à partir d'études statistiques, que les critères explicatifs traditionnels que sont le niveau de langue et le degré de collocation entre l'adjectif et le nom, ne suffisent pas pour expliquer ces phénomènes. Nous proposerons en revanche deux autres facteurs, l'un étant la caractéristique lexicale des adjectifs, l'autre, la nature sémantico-fonctionnelle de l'épithète dans le discours. Ces facteurs constituent le degré de « poids » ou le degré d'importance informative de l'adjectif dans l'usage ; plus l'épithète est « légère », plus *des* est choisi, et moins elle l'est, plus *de* est sélectionné. Alors que la notion de « poids grammatical » proposée dans Abeillé & Godard (1999, 2000) est de nature discontinue, la nôtre se place sur une échelle continue. C'est la méthode même de l'étude statistique qui nous oblige d'avancer cette dernière caractérisation. Avec cette caractérisation de « poids », nous pouvons traiter les facteurs lexicaux et discursifs de la même manière. Nous concluons que *petit* est situé à l'extrémité de l'échelle lexicale et tend à apparaître précédé de *des*, qui fonctionne comme marqueur de la « légèreté » de l'épithète qui le suit.

Mots-clés : petit, épithète, poids, collocation, français, IM, mot composé, t-score, z-score, article, weight

1. Introduction

Lorsqu'un adjectif au pluriel est antéposé par rapport au substantif, on a le choix en français entre *de* et *des* comme article indéfini, par exemple, *de bonnes conditions* et *des petits chiens*. Nous avons étudié statistiquement ce phénomène dans Fujimura et al., 2004 à paraître, du

point de vue aussi bien historique que stylistique en recourant à des corpus de très grande taille, et proposé plusieurs conditions qui déterminent cette variation.

Nous y avons non seulement constaté le facteur du niveau de langue et la question des mots composés, qui ont été répétés à maintes reprises dans la littérature relative à cette alternance, mais aussi découvert de nouveaux facteurs tels que la liaison phonétique : s'il y a une liaison entre l'adjectif et le nom, on note une tendance claire à éviter *des*. Dans cet article, nous allons examiner deux autres conditions, l'une étant la distribution disparate de *de* et de *des* suivant les adjectifs, et l'autre, la quantité d'informations que l'adjectif antéposé transmet dans le discours. Nous nous concentrerons surtout sur la question de *petit* qui présente une affinité extrêmement forte avec *des* par rapport à d'autres adjectifs.

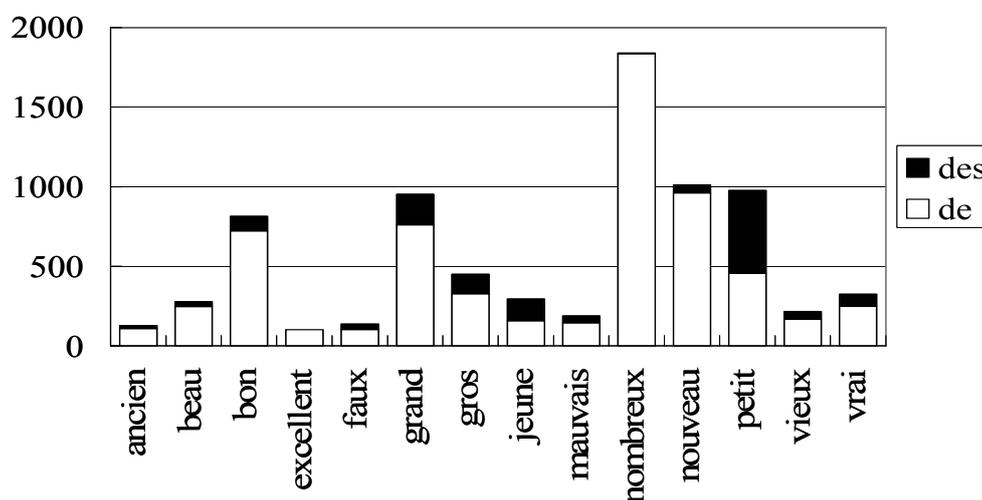
La question du choix de l'article *de* ou *des* devant l'adjectif a un rapport étroit avec la tendance générale à l'antéposition ou à la postposition de cet adjectif par rapport au nom. Togeby dit par exemple : « Seulement une dizaine d'adjectifs courts et fréquents sont régulièrement antéposés dans la langue de la conversation, qui les fait accompagner alors de l'article partitif complet (= l'article *des*) : *C'est des bonnes nouvelles...* » (Togeby, 1982). Or, dans la littérature récente de la linguistique générale, la notion de « poids grammatical » a été proposée pour résoudre entre autres la question de la position de l'épithète en français (Abeillé et Godard, 1999, 2000) ; l'adjectif est « léger » s'il est constamment antéposé au nom, tandis qu'il est « non-léger », s'il est toujours postposé. On pourrait donc supposer, en leur empruntant ces termes fort convenables, que *des* a tendance à être choisi si l'adjectif est « léger » et que *de* est sélectionné quand l'adjectif est « non-léger ». D'après les auteurs de ces termes, les adjectifs « légers » et les « non-légers » constituent deux catégories lexicales distinctes et discontinues. Pour nous, par contre, la notion de « poids » devrait être sur une échelle graduelle et continue. En définissant le « poids » comme « importance informative » ou « saillance informative », nous démontrerons que *petit* est l'adjectif le plus « léger » parmi tous les adjectifs français.

2. Adjectifs et *de* vs *des*

2.1. Aperçu général

Nous voyons dans la figure 1 les occurrences en nombre réel de *de* et de *des* qui précèdent les « ADJ + NOM ». Les données ont été recueillies à partir des corpus de la 2^{ème} moitié du 20^e siècle (71,6 millions de mots au total), que nous présenterons ci-après.

Figure 1 : *De* vs *des* et adjectifs dans les textes contemporains



Le taux de *des* pour chaque adjectif est : 53.3% (*petit*), 46.4% (*jeune*), 27.2% (*gros*), 24.8% (*faux*), 23.1% (*vrai*), 23.0% (*mauvais*), 22.7% (*vieux*), 20.2% (*grand*), 14.8% (*ancien*), 11.2% (*bon*), 10.5% (*beau*), 4.8% (*nouveau*), 0.2% (*nombreux*), 0% (*excellent*). Les « *de/des* + ADJ + NOM » les plus fréquents sont : *des petits problèmes, des jeunes filles, de gros problèmes, de faux papiers, de vrais morceaux, de mauvaises conditions, de vieux journaux, de grandes chances, d'anciens membres, de bonnes conditions, de beaux yeux, de nouvelles aventures, de nombreux cas et d'excellentes conditions*. On y trouve une disparité importante d'un adjectif à l'autre. La tendance de cooccurrence entre *des* et *petit* est remarquablement forte, tandis que *des* n'est guère compatible avec *nouveau, nombreux* et *excellent*. D'autres adjectifs se situent entre les deux. Quant à *jeune* qui est autant utilisé avec *des* que *petit*, nous le mettons de côté parce que c'est un adjectif qui ne qualifie en principe que les animés avec un pouvoir combinatoire limité.

Cette hétérogénéité de répartition de *de* et *des* demande une explication. Il y avait peu de grammairiens et linguistes qui l'avaient envisagée, ce phénomène n'ayant même pas été bien observé auparavant.

2.2. corpus et données

Pour mener à bien cette étude, nous avons établi d'abord une base de données constituée d'environ 20 000 exemples avec « *de /des* + (ADV) + ADJ + NOM au pluriel » recueillis dans les corpus mentionnés dans le tableau 1 ci-après. Ce sont toujours des séquences avec une trentaine d'adjectifs tels que *beau, bon, grand, nouveau, petit, excellent etc.*, qui se trouvaient à suivre directement un lemme des verbes : *avoir, être* et *prendre*, et les prépositions : *à, avec, dans, par, pour* et *sur*. Nous y avons aussi inclus des exemples qui étaient sujets des verbes : *avoir, être* et *prendre*. Nous avons manuellement examiné tous les exemples pour ne conserver que ceux qui sont appropriés à notre recherche. Nous avons ainsi exclu tous les usages de *de* induits de la négation (ex. *on n'a pas eu de grandes discussions sérieuses*), tous les homographes (ex. *la page de la liste odésienne est sur des nouvelles intéressantes*), toutes les occurrences de *de* ou *des* en tant que préposition (+ article) (ex. *les pièces sont de petites dimensions ; pour ce qui est des belles routes, ...*) etc.. L'examen des données nous enseigne que la fonction grammaticale du SN dans la phrase ne joue pas un rôle pertinent pour le choix entre *de* et *des*. Bien que notre base de données ne représente qu'une sous-classe du phénomène, cela n'entraîne pas de conséquence erronée pour notre but.

Tableau 1 : Corpus de base

Période	Genre	Détail	Année	Nombre de mots (milliers)
20 ^e s-2	Hansard	Les interventions en français dans le débat au Parlement Canadien, parlé officiel	1986-1988	3 300
	journal	<i>Le Monde</i> (Wordbanks Online)	1997 et 2001	13 000
		<i>Libération</i> (Wordbanks Online)	1992-1993	1 500
	revue	<i>Actuel</i> (Wordbanks Online)	1990-1992	2 000
		<i>Marie Claire</i> (Wordbanks Online)	1990-1993	3 100
	Forum de discussions (FD)	323 Forums de discussions abonnés, écrit non officiel sur les réseaux électroniques	10/2000	21 200
	roman	Frantext base catégorisée	1951-2000	15 600
traité ou essai	11 900			

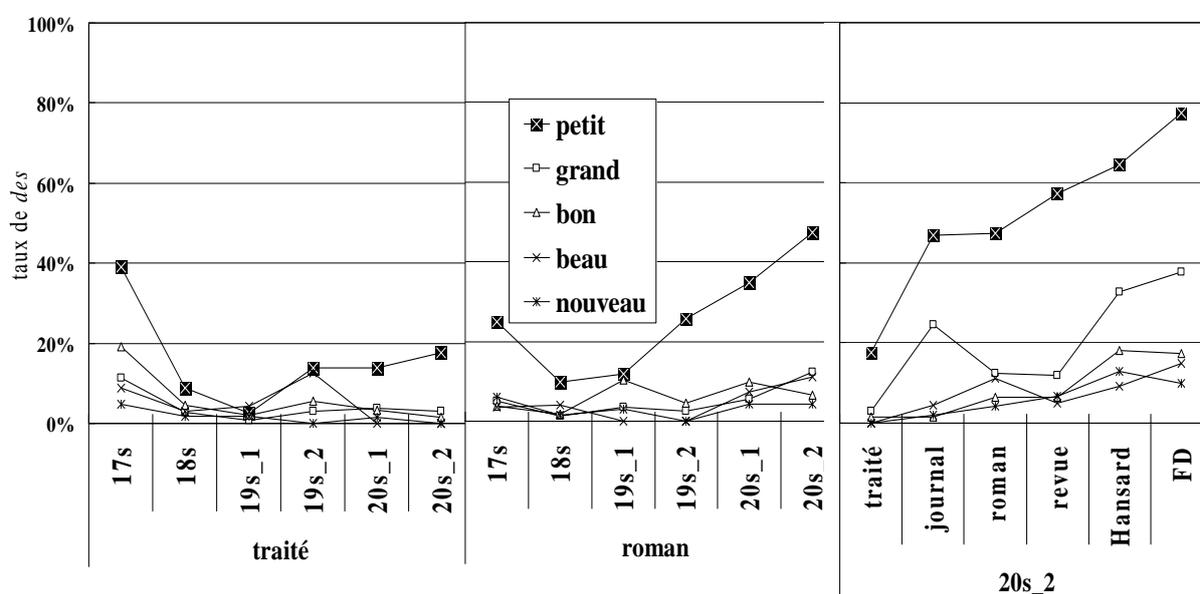
20° s-1	roman	Frantext base catégorisée	1901-1950	20 600
	traité ou essai			13 500
19° s-2	roman	Frantext base catégorisée	1851-1900	14 500
	traité ou essai			4 400
19° s-1	roman	Frantext base non-catégorisée	1801-1850	12 600
	traité ou essai			8 100
18° s	roman	Frantext base non-catégorisée	1701-1800	12 000
	traité ou essai			11 200
17° s	roman	Frantext base non-catégorisée	1601-1700	3 800
	traité ou essai			7 000
		total		179 300

3. Petit et « poids lexical »

3.1. Petit à travers l'histoire et les genres

Dans cette section, nous allons examiner *petit* de plus près. Nous montrons tout d'abord avec le figure 2 que *petit* garde la spécificité d'être plus étroitement lié avec *des* que les autres adjectifs : *grand*, *bon*, *beau*, *nouveau*, dans toutes les périodes et dans tous les genres depuis le début de l'histoire de ce phénomène. Dans les figures, la ligne indiquant *petit* est constamment au-dessus des autres lignes, c'est-à-dire que *petit* a plus d'affinité avec *des* que les autres depuis toujours. Cette remarquable caractéristique de *petit* mérite de sérieuses recherches.

Figure 2 : Taux de *des* et adjectifs dans toute l'histoire et dans tous les genres



3.2. Petit et degré de collocation

Togebly est un rare linguiste qui a fait remarquer cette forte tendance de cooccurrence, en disant que « *petit* forme souvent avec le substantif une sorte de mot composé » (1982 : 52). Nos données n'ont cependant pas relevé la spécificité de *petit* en tant que base de mots composés.

Nous avons créé, dans le but de l'étude des mots composés, une deuxième base de données constituée d'environ 3500 bigrammes (= 15000 tokens). Ce sont des « *ancien, beau, bon, grand, nouveau et petit + NOM au pluriel* » précédés de « (*à/ avec/ dans/ derrière/ devant/ entre/ malgré/ par/ parmi/ pour/ selon/ sur/ suivant/ sous*) + (*de | des*) », recueillis à partir des périodiques contemporains d'environ 277 millions de mots : *Libération* (1995-2000) , *Le Monde* (1987-1992), *Le Point* (1995-2000), *Le Monde Diplomatique* (1984-1998) et *La Tribune* (1995-1999) distribués par CDROM-SNi. Nous avons choisi ces textes afin d'obtenir des données statistiques fiables. Ils sont aussi homogènes que gigantesques et faciles à traiter contrairement aux corpus en ligne tels que Frantext. Etant donné que ces textes ne sont pas étiquetés et que l'usage des analyseurs comme Cordial, FDG (Functional Dependency Grammar), TreeTagger ne nous semblait pas, après essai, avantageux pour notre but, nous avons restreint le contexte pour identifier le plus exactement possible les séquences qui correspondent à « ART + ADJ + NOM au pluriel ».

Afin de faire une comparaison avec leur degré d'affinité avec *des* (= taux de *des*) , nous avons donné à chaque bigramme les informations concernant le degré de collocation entre l'adjectif et le nom : l'Information Mutuelle, le t-score et le z-score, calculées sous Excel par nous-mêmes (cf. Barnbrook, 1996, Oaks, 1998). Nous ne savions pas si le statut de mot composé était mesurable avec ces scores, mais sans aucun bon dictionnaire disponible définissant des mots composés français, nous n'avions pas d'autre moyen pour vérifier notre thèse. Les bigrammes avec les scores de collocation les plus élevés étaient : *petits boulots, bons offices, petits riens, belles empoignades, bonnes volontés* (IM), *grandes entreprises, grands groupes, nouvelles technologies, grandes villes, grandes lignes* (t-score), *petits boulots, grandes surfaces, anciens combattants, nouvelles technologies, bons offices* (z-score). Voici les expressions développées avec lesquelles nous avons calculé les scores :

$$IM(x,y) = \log_2 \frac{F(x,y) * \text{Nombre total de mots}}{F(x) F(y)}$$

$$t\text{-score}(x,y) = \frac{F(x,y) - \frac{F(x)F(y)}{\text{Nombre total de mots}}}{\sqrt{F(x,y)}}$$

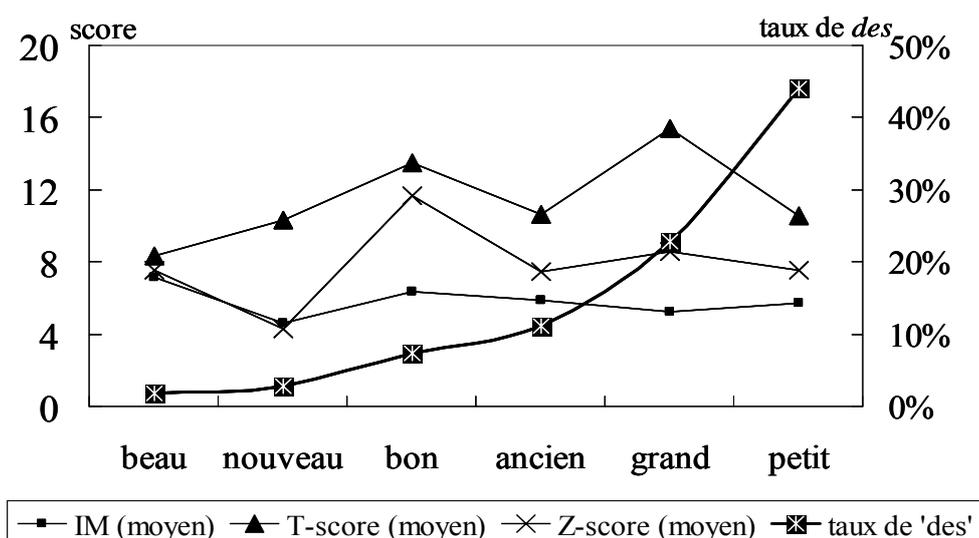
$$z\text{-score}(x,y) = \frac{F(x,y) - \frac{F(x)F(y)}{\text{Nombre total de mots}}}{\sqrt{F(x)F(y)}}$$

Nombre total de mots = 277 millions

Les résultats démontrent tout d'abord que les scores de degré de collocation entre l'adjectif et le nom sont en corrélation bien que faible avec le taux de *des* dans la totalité des bigrammes : les coefficients de corrélation de Pearson entre les taux de *des* et l'IM, le t-score et le z-score sont 0.255, 0.210, et 0.285 respectivement (n = 994). Pour ce calcul, nous avons pris en compte seuls les bigrammes qui apparaissaient plus de 2 fois dans l'environnement défini ci-dessus et, en même temps, plus de 9 fois au total dans toute circonstance dans le corpus : soit 994 types de bigramme (= 11301 tokens). On constate que le choix de *de* ou de *des* est conditionné par le degré de collocation entre l'adjectif et le nom, comme toutes les grammaires l'avaient indiqué à l'unanimité.

Ces scores n'expliquent cependant pas la diversité entre les adjectifs quant au choix de *de* ou de *des*, ainsi que le montre la figure 3. Le degré d'affinité entre l'adjectif et *des* indiqué par la ligne « taux de *des* » n'a de relation avec aucun des scores de collocation, ni avec l'IM, ni avec le t-score, ni avec le z-score. L'affinité forte entre *des* et *petit* n'est pas attribuable au figement étroit entre *petit* et le nom qui le suit.

Figure 3 : Taux de *des*, adjectifs et scores de collocation



Des a une très forte affinité avec *petit* dans toutes les périodes et dans tous les genres. Or, le degré de collocation entre l'adjectif et le nom est en corrélation avec le taux de *des*. Nos données ne relèvent pas néanmoins la particularité de *petit* en tant que base de mots composés.

3.3. Contraste entre *petit* et *grand*

Pour confirmer avec plus de sûreté que la question de *petit* est bien attribuable à sa propre nature et non à l'environnement où cet adjectif apparaît, nous avons ensuite observé le phénomène dans des conditions plus restreintes. Nous avons étudié, dans les corpus périodiques, l'occurrence de *de* et de *des* à côté des deux adjectifs antonymes : *petit* et *grand*, suivis d'un de ces noms : *actionnaires*, *affaires*, *appartements*, *banques*, *cahiers*, *centres*, *chefs*, *choses*, *compagnies*, *écrans*, *éditeurs*, *entreprises*, *établissements*, *films*, *firmes*, *fonds*, *groupes*, *hôtels*, *investisseurs*, *lunettes*, *maisons*, *marchés*, *moyens*, *opérations*, *pays*, *programmes*, *projets*, *quantités*, *restaurants*, *rôles*, *sacs*, *salles*, *séries*, *sociétés*, *structures*, *surfaces*, *tables*, *tentes*, *travaux*, *usines*, *valeurs* et *villes*. Il s'agit des noms qui peuvent se

combiner librement avec *petit* et *grand* et qui sont effectivement apparus plus de 3 fois, associés avec chacun de ces deux adjectifs dans nos corpus.

Le résultat indique que même dans ces conditions limitées, *petit* présente une tendance forte de cooccurrence avec *des*, contrairement à *grand* qui a une tendance de cooccurrence avec *de* ($X^2= 11.038$, $dl = 1$, $p < .01$).

Tableau 2 : *de* vs *des* et *petit* vs *grand* (occurrences en nombre réel)

	<i>de</i>	<i>des</i>
<i>grand</i>	712	302
<i>petit</i>	366	223

Voici un exemple qui représente bien le contraste entre *petit* et *grand* :

- (1) Or, cette relation entre la taille et le taux de mortalité est probablement un simple biais statistique, lié au fait qu'un taux calculé sur des petits effectifs est beaucoup plus fluctuant que quand il est calculé sur de grands effectifs. (*Le Monde*, 06 /10/97)

et quelques autres auxquels on va revenir un peu plus bas :

- (2) Pour tenter d'accroître ses parts de marché, la firme de Cupertino a signé depuis le début de l'année des accords de licence avec des petites sociétés comme Power Computing, Radius, ainsi qu'avec le japonais Bandai. (*La Tribune*, 07/09/95)
- (3) Il va y parvenir de façon spectaculaire: en quelques mois, il passe un accord de licence avec RSA Data Security - une firme spécialisée dans l'encryptage -, jette les bases d'un standard pour les serveurs sécurisés, fait considérablement évoluer le langage HTML (2) du Web et, surtout, signe des accords avec de grandes sociétés comme Bank of America, le groupe MasterCard, ainsi que le consortium CommerceNet qui regroupe les grands noms de l'industrie informatique. (*Libération*, 21/4/95)

3.4. Caractéristique lexicale de *petit* : légèreté

Nous avons démontré que la forte tendance de cooccurrence entre *des* et *petit* est seulement explicable avec une spécificité lexicale de cet adjectif. Ce phénomène n'est pas abordable en recourant à d'autres principes linguistiques, mais il est immanent au mot même de *petit*, qui est, d'après nous, l'adjectif « le plus léger » parmi ceux que nous avons examinés.

Cette caractéristique de *petit* n'est pas sporadique. Nous pouvons en effet trouver, autour de *petit*, d'autres phénomènes qui s'accordent bien avec la notion de « poids léger » définie comme « importance informative ». Nous donnons ici, évoqués par Fujita (2001), trois arguments qui soutiennent notre thèse que *petit* est un adjectif « très léger ».

3.4.1. Unidimensionnalité

Le premier est la question de l'unidimensionnalité de l'échelle *petit* / *grand* proposée dans Rivara, 1993. Selon l'auteur, les adjectifs antonymiques *bon* / *mauvais*, *beau* / *laid*, *nouveau* / *vieux* etc. sont bipolaires en ayant deux échelles indépendantes l'une de l'autre, alors que *petit* et *grand* portent une même échelle en commun : celle de grandeur. Sur cette échelle, *petit* est orienté vers le néant, tandis que *grand* l'est vers l'infini. On peut facilement dire que

le pouvoir descriptif et qualificatif de *petit* est souvent restreint par rapport à celui de *grand* comme on peut l'observer dans les exemples 2 et 3 ci-dessus. L'adjectif de l'exemple 3 serait paraphrasable en *important* ou *extraordinaire*, par contre, celui de l'exemple 2 signifierait *peu important* ou *ordinaire*.

3.4.2. Diminutif

Le deuxième argument concerne le diminutif. En français, *petit* a souvent une valeur hypocoristique qui est en général exprimée dans d'autres langues par un suffixe ajouté à un mot (Herisson, 1956, Delhay, 1996). L'épithète *petit* dans cet usage n'implique aucune référence à la taille (ex. *Bonjour à ta petite famille*). On peut dire qu'il s'agit d'une déficience sémantique, qui devrait aboutir à une dégradation syntaxique. Si ce genre de changement est surtout permis à *petit* et non à d'autres adjectifs, ce serait parce que le mot *petit* a une propriété adéquate à cette déviation, qui est d'après nous le caractère d'être « léger ».

3.4.3. Le moindre petit + NOM

Le troisième argument repose sur l'expression pléonastique : « *le moindre petit NOM* ». Cette tournure est fréquente et productive en français ; on a relevé 91 cas du « *le moindre petit NOM* » contre 2824 « *le moindre NOM* » dans le FD (ex. *Je n'ai pas la moindre petite information à vous donner !* (FD, 04/09/2000)). Bien qu'il y ait des cas où le « *petit NOM* » dans la tournure constitue un mot composé (ex. *...dans la société du XIX^e siècle où le moindre petit bourgeois a au moins une bonne.*), ce n'est pas le cas de tous les exemples. Il nous semble que *petit* y perd son « poids » jusqu'à ce que la suite *moindre petit* ne soit plus redondante à l'oreille. À l'opposé, on n'entend jamais dire « *le meilleur bon NOM* », ni « *le pire mauvais NOM* », excepté dans les cas de mots composés comme *la meilleure bonne foi*.

On peut expliquer ce phénomène de la même manière que les faits observés ci-dessus. Le « poids » de *petit* étant très faible, cet adjectif devient donc non encombrant, il est facile de ne pas tenir compte de son existence sémantique et formelle. La comparaison des exemples suivants démontre bien que l'information transmise par le mot *petit* est minimale.

- (4) Dans 99 % des cas, explique-t-on, c'est une fausse alerte. Mais les gens ne veulent pas prendre le moindre petit risque, car on n'en revient pas si le 1 % qui reste n'est pas faux. (*Le Monde*, 25/09/97)
- (5) Les banques ne veulent plus prendre le moindre risque pour l'argent de leurs clients et exigent systématiquement le cautionnement. (*Le Monde*, 26/11/97)

4. « Poids discursif » ou quantité d'informations de l'épithète

Après avoir examiné le « poids lexical » de *petit* parmi d'autres adjectifs, nous allons passer à l'examen de la relation entre le « poids discursif » des épithètes et le choix de l'article *de* ou *des*. Notre hypothèse est la suivante : plus l'information que l'épithète véhicule est abondante, plus *de* est choisi ; moins elle l'est, plus *des* est sélectionné. L'épithète dans les mots composés se situe à une extrémité de cette échelle, car elle n'a aucune information à transmettre indépendamment du nom.

4.1. Description des grammairiens

Cette caractérisation de *de* et de *des* n'est pas une nouveauté. Au contraire, c'était une vue répétée dans la littérature, comme l'indiquent les citations suivantes.

Elle (=la langue) différencie ainsi la qualification préalable, notoire, lexicalisée, qu'exprime parfois l'épithète antéposée (*des grands frères, des grandes sœurs, des petits pains*, etc.) de la qualification nouvelle, transitoire, voire prédicative, qui est son rôle dans bien des cas (*Elle a de grands yeux*, c'est-à-dire : Elle a les yeux grands). (GLLF : 260)

Sans l'article (= *de*), l'épithète a toute sa valeur propre, toute sa force de caractérisation accusée : *de beaux pays, de grands chirurgiens*. Avec l'article (= *des*) (...), la caractérisation s'efface un peu, ... (Le Bidois et Le Bidois, 1967 : 85).

Notre apport original consiste dans le point de vue de la gradualité. Cette notion permet de réunir divers facteurs sur une même échelle même s'ils paraissent hétérogènes à première vue.

4.2. Centre d'information

Le centre d'information est selon nous une autre manifestation du « poids non-léger ». Dans l'exemple suivant, où *de* est exceptionnellement employé devant *petits* dans le texte oral (1 cas de *de* contre 30 cas de *des* dans ce corpus du français parlé), *petits* joue le rôle de centre d'information dans le SN : *petits libraires*. Le mot *libraires* est dans ce dernier une répétition, à savoir une information superflue puisqu'il a déjà été dit en tant que sujet de la phrase copulative :

- (6) les FNAC n'existaient pas les choses comme ça bon les l- les libraires étaient de petits libraires indépendants elle existe d'ailleurs toujours je pense la maison Gibert non (Giron, S., 2001, Corpus Allier, PHYMO~54 (6,10 - 6,12))

Nous avons également attesté la préférence de *de* dans cet environnement : « NOM_i être (*de/des*) ADJ NOM_j (*i = j*) » dans les corpus du tableau 1 par exemple dans :

- (7) Les artistes vivent souvent dans l'irréalité et à cause de cela, sont de grands artistes. (Marie-Claire)
- (8) Si on considère que plus de 90 p. 100 des entreprises au Canada sont de petites entreprises, (Hansard)
- (9) ... ramasser les légumes, aller à l'herbe pour les lapins, nourrir les poules..... les maisons sont de petites maisons comme la mienne, ... (Triolet.E. *Le premier accroc coûte deux cents francs*)

4.3. Renforcement adverbial

Nous traitons en dernier lieu le renforcement de l'épithète par un adverbe. On peut facilement postuler que la présence d'un adverbe ajoute une information de plus ; on peut aussi estimer qu'elle rend la phrase plus complexe ou « plus lourde », en empruntant ce dernier terme à Arnold et al., 2000. L'épithète devenant « moins légère », la possibilité du choix de *de* augmente.

Cette tendance est clairement démontrée dans le tableau 3, qui relève les occurrences de *de* et de *des* devant les adverbes *fort, moins, plus, si, tout, très et trop* dans les corpus : 2^{ème} moitié du 20^e s, présentés dans le tableau 1. S'il y a un adverbe entre l'article et l'adjectif, *des* n'est guère sélectionné, mais *de* est de fait presque toujours choisi ($X^2= 38.4887$, $dl = 1$, $p < .01$).

Tableau 3 : De vs des et adverbe

	<i>de</i>	<i>des</i>
sans adverbe	5503	1122
avec adverbe	251	6
<i>très</i>	173	1
<i>tout / toutes</i>	22	3
<i>plus</i>	21	2
<i>si</i>	17	0
<i>fort</i>	7	0
<i>trop</i>	6	0
<i>moins</i>	5	0

Cette tendance n'est cependant pas non plus une règle absolue. Alors que l'usage de *des* est rare avec *tout / toutes*, il n'est pas aussi exceptionnel qu'avec *très*. On peut en effet noter une différence statistiquement significative de l'occurrence de *des* avec *tout / toutes* et avec *très* ($X^2 = 14.4871$, $dl = 1$, $p < .01$). Les exemples suivants relèvent le contraste de *très* et *tout / toutes* : « *des tout / toutes* ADJ + NOM » est employé même dans le corpus écrit académique dans lequel l'usage de *des* est généralement limité comme dans (10), tandis que « *de très* ADJ + NOM » est utilisé dans le corpus parlé où l'usage de *des* est en général fréquent comme dans (11).

(10) Elle (= la diarrhée) apparaît dans un élevage sur des tout jeunes poussins de trois à six jours.
(Garcin, E. *Guide vétérinaire*)

(11) hein quand tu travailles comme ça sur les marchés tu as les tu as de très bonnes relations avec eux et puis voilà donc je crois que c'est ça qu'il apprécie voilà mmh (Giron, S., 2001, Corpus Allier, FRUIT~22 (13,16 - 14,2))

Or, on sait bien qu'il y a une disparité quant au choix de l'adverbe selon les adjectifs. *Très* se combine avec des adjectifs qualificatifs typiques, comme *grand*, *beau*, *bon*, alors que *tout* s'associe, d'un emploi beaucoup plus limité, avec des adjectifs qui ont « une limite idéale » (Hanse, 1987 : 950) comme ceux de couleur : *tout rouge*, par exemple. *Tout* est un adverbe plus notionnel et plus logique que *très*, qui est plutôt emphatique, marquant sans réserve un degré très élevé. Nous pouvons penser que cette différence sémantico-fonctionnelle est la raison pour laquelle « *des tout* ADJ NOM » est moins exceptionnel que « *des très* ADJ NOM » dans le corpus. Le rôle de *de* est de mettre en relief la présence d'un adjectif « non léger » qui a un caractère typiquement qualificatif dans la position qui le suit.

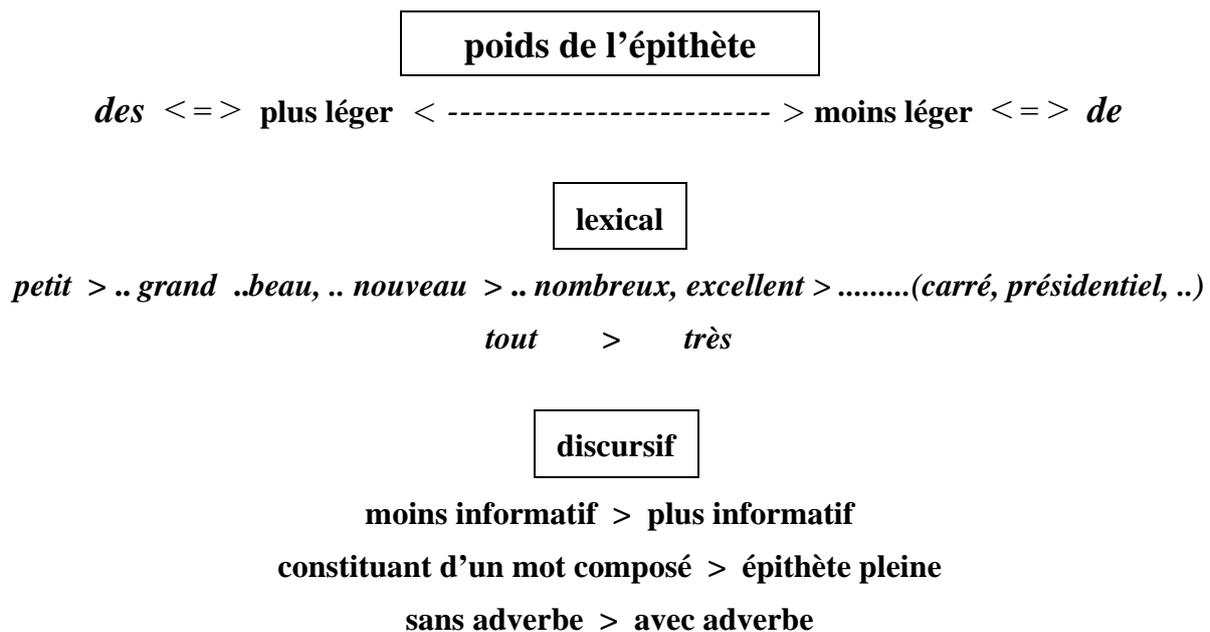
D'après nous, *tout* est un adverbe « plus léger » que *très*, contrairement à ce qu'ont proposé Abeillé et Godard (1999, 2000). Selon ces auteurs, ce sont tous les deux des adverbes de la catégorie du « poids léger » au même titre que *trop*, *assez*, *vraiment*, *peu*, à l'opposé de *politiquement*, *véritablement* ou *absolument* qui sont « non-légers ». Le critère pour cette classification réside dans la position de l'épithète : « adverbe + adjectif » par rapport au nom. Avec les adverbes « légers », l'adjectif peut rester antéposé comme dans « *de très bonnes conditions* », alors qu'avec les « non-légers », il doit être postposé comme dans « *une décision politiquement habile* ». Nous voudrions en revanche avancer pour notre part l'idée de gradualité du « poids », en recourant à la fréquence comme critère d'évaluation. Si le taux

de *des* est plus élevé avec *tout* qu'avec *très* et que les autres conditions sont égales, c'est parce que *tout* est « plus léger » que *très*. Il en va de même pour les adjectifs comme nous l'avons déjà affirmé : *petit* est l'adjectif « le plus léger » parmi ceux que nous avons examinés. *Excellent* et *nombreux* sont par contre supposés « les moins légers » parmi eux, bien qu'ils soient eux-mêmes « plus légers » que ceux qui sont toujours postposés, comme *carré*, *français* ou *présidentiel*.

5. Conclusion

Partis de l'observation de l'adjectif *petit* qui a une forte tendance de cooccurrence avec l'article *des* depuis le 17^e siècle jusqu'à nos jours et à travers tous les genres de texte, en passant par la réfutation de l'argument « *petit* comme base de mots composés » au moyen de quelques techniques statistiques, nous avons abouti à l'idée de « poids », notion d'une portée importante qui permet d'expliquer de nombreux facteurs pertinents de ce phénomène.

Nous avons proposé dans cet article deux sous-classes de « poids ». L'un est le « poids lexical » et l'autre, le « poids discursif ». Le premier concerne un caractère immanent du mot, tandis que l'autre repose sur le rôle sémantico-fonctionnel de l'expression dans le discours. Voici les échelles relevées comme pertinentes pour rendre compte du phénomène :



Bien que l'idée de « poids » nous ait été directement inspirée par Abeillé et Godard (1999 et 2000), la différence entre notre point de vue et le leur est claire. Pour eux, le « poids » est une notion qui sert à classer des éléments dans des catégories discontinues, tandis que pour nous c'est une échelle continue et graduelle. On peut dire que c'est la méthode même de l'étude statistique qui nous oblige à avancer cette dernière caractérisation du « poids ». Cette position n'est d'ailleurs pas vraiment nouvelle, au contraire, mais bien orthodoxe dans la littérature d'une longue tradition concernant le « poids » et l'ordre des mots (cf. Arnold et al., 2000, Wasow, 1997 entre autres). Une de nos contributions y serait l'application de cette notion dans un autre domaine que l'ordre des mots.

Nous voudrions enfin signaler une possibilité de rendre compte d'un autre facteur qui détermine le choix entre *de* et *des* : la liaison phonique, avec la notion de « poids ». S'il y a une liaison entre l'épithète et le nom, *de* est préféré, sinon, *des* l'est. Le « poids » concerné pourrait être phonétique, parce qu'avec une liaison, le mot devient plus long. Ou bien, il pourrait porter sur la quantité d'information, parce qu'avec la liaison, une information concernant la pluralité est transmise en plus. Voilà la question intéressante qui demande une exploitation.

Références

- Abeillé, A. et Godard, D. (1999) : « La place de l'adjectif épithète en français : le poids des mots », *Recherches Linguistiques*, 28 : 9-31.
- Abeillé, A. et Godard, D. (2000) : « French Word Order and Lexical Weight », in R.Borsley (ed), *The Nature and Function of Syntactic Categories, Syntax and Semantics* 32 : 325-360.
- Arnold, J., Wasow, Th., Losongco, A. et Ginstrom, R. (2000) : « Heaviness vs. Newness : The effects of complexity and information structure on constituent ordering », *Language* 76 : 28-55.
- Barnbrook, G. (1996) : *Language and Computers: A Practical Introduction to the Computer Analysis of Language*, Edinburgh U. P.
- Delhay, C. (1996) : *Il était un « petit X » : pour une approche nouvelle de la catégorisation dite diminutive*, Larousse.
- Dupré (1972) : *Encyclopédie du bon français dans l'usage contemporain*, Editions de Trévise, t.II.
- Fujimura, I. Uchida, M. et Nakao, H. (2004) : « Opposition entre *de* et *des* devant les noms précédés d'épithète en français : étude descriptive », *Acte des 3èmes journées de la linguistique du corpus*, à paraître.
- Fujita, T. (2001) : « Tokushu na saijo-kyu *le moindre* (*Le moindre* : un superlatif atypique) », *Grant-in-Aid for COE Research Report (5) : Researching and Verifying an Advanced Theory of Human Language*, Kanda University of International Studies : 177-195.
- Giron, S. (2001) : *Corpus Allier*.
- Grand Larousse de la Langue Française* (1971), t.1, Larousse.
- Hanse (1987) : *Nouveau dictionnaire des difficultés du français moderne*, Duculot.
- Herisson, Ch. D. (1956) : « Le diminutif hypocoristique « *petit* » », *le Français Moderne*, XXIV : 35-47.
- Le Bidois G. et Le Bidois. R. (1967) : *Syntaxe du français moderne, Tome 1*, Éd. A. Picard.
- Togoby K. (1982) : *Grammaire française , vol.1 : Le Nom*, Akademisk Forlag.
- Oakes, M. P. (1998) : *Statistics for Corpus Linguistics*, Edinburgh U. P.
- Rivara, R.(1993) : « Adjectifs et structures sémantiques scalaires », *Information grammaticale* 58 : 44-56.
- Wasow, Th. (1997) : « Remarks on Grammatical Weight », *Language Variation and Change* 9 : 81-105.