



Faculteit Letteren & Wijsbegeerte

Timotheus Vermote

*L'opposition massif-comptable:  
flexibilité et modélisation*

*Etudes de corpus, enquêtes d'acceptabilité et  
expérience d'amorçage en français et en néerlandais.*

Proefschrift voorgelegd tot het behalen van de graad van  
Doctor in de taalkunde

2014



Pour Li et pour mon père

Promotor    Prof. dr. Peter Lauwers  
                 Vakgroep Taalkunde

Decaan       Prof. dr. Marc Boone  
Rector       Prof. dr. Anne De Paepe

Nederlandse vertaling:

Het onderscheid telbaar-ontelbaar: flexibiliteit en modelvorming



Faculteit Letteren & Wijsbegeerte

Timotheus Vermote

# *L'opposition massif-comptable: flexibilité et modélisation*

*Etudes de corpus, enquêtes d'acceptabilité et  
expérience d'amorçage en français et en néerlandais.*

Proefschrift voorgelegd tot het behalen van de graad van  
Doctor in de taalkunde

2014



## Remerciements

Parmi toutes les personnes qui ont contribué à la création du présent travail, mon directeur de thèse, Peter, mérite une mention spéciale. Au cours des quatre années qu'a duré ma recherche, sa jeunesse d'esprit, son dynamisme et sa disponibilité n'ont cessé d'être une inspiration et une source intarissable de motivation. Il a été de bon conseil à chaque étape de ma progression, et un redoutable sparring-partner alors que j'apprenais à forger et décrire mes propres idées. Grâce à lui et son exemple, je commence à comprendre ce que signifie d'être spécialiste en général, et linguiste en particulier. Merci !

Merci aussi à Lisette, ma chère et tendre, qui donne du soleil à mes matins et dessine chaque soir le sourire au bord de mes lèvres. Merci pour le soutien, pour les heures passées à m'écouter raconter ma recherche. Merci pour la compagnie. Merci pour les rires. Merci pour les bleus.

Merci à tous les collègues, linguistes et psycholinguistes, qui ont contribué de près ou de loin à l'aboutissement de ma recherche. Merci à Johan De Caluwe et à Rob Hartsuiker pour les débats passionnants. Merci à Georges Kleiber et aux autres « masscolexistes » pour les commentaires encourageants. Merci aux collègues de l'Université de Caen pour leur accueil chaleureux. Merci à Gudrun Vanderbauwhede, qui m'a permis d'organiser mon expérience d'amorçage à l'Université de Mons. Merci de tout cœur à Ludovic De Cuyper pour son aide inestimable dans l'analyse statistique de mes données.

Merci à Chantal, du secrétariat de linguistique française, pour son amabilité et sa fiabilité.

Merci à « Jeunesse & Santé Namur » de m'avoir permis de me changer les idées, de me déguiser, de courir, de chanter, de vendre du rêve – et de ne perdre ainsi ni la tête, ni ma motivation. Et merci à Maxx, parce que c'est le président.





## Préface

Avant d'arriver à l'Université de Gand, en octobre 2010, et de commencer ma recherche sur l'opposition entre massif et comptable, le sujet n'avait que très peu attiré mon attention. Je n'en ai comme principal souvenir pré-académique que mes cours d'anglais de secondaire, où Monsieur Martens nous apprenait l'emploi approprié de *much* et *many*, de *less* et *fewer*. Je me rappelle avoir pensé alors que l'anglais était une langue bien curieuse, pour distinguer ainsi entre des noms dénombrables et indénombrables – une excentricité qui, assurément, n'avait pas d'équivalent en français. J'étais bien loin d'imaginer que non seulement, je me trompais carrément – mais qu'en plus, je ferais un jour précisément de cette distinction le sujet d'une thèse qui m'occuperait pendant quatre années, et deviendrait mon premier domaine de spécialisation.

Dans le parcours qui m'aura mené à mon doctorat, mes années estudiantines à la Faculté des Lettres de la K. U. Leuven ne constituent pas une étape surprenante : l'on peut se douter que c'est dans l'étude des langues que sera née mon envie d'approfondir la linguistique. Pourtant, la trajectoire dont le présent ouvrage est un premier aboutissement n'était pas toute tracée. Avant que l'opportunité de la recherche ne s'ouvre – ne s'offre – à moi, je me destinais à une carrière de comédien : j'avais pris les renseignements, j'avais visité les portes ouvertes, j'allais m'inscrire à l'examen d'entrée.

Était-ce *de la* chance, *une* aubaine, *le* destin ? Un message de Peter, alors encore pour moi Monsieur le professeur Lauwers, reçu le 27 avril 2010 (je l'ai encore !), aura permis de chambouler mes plans. (Je n'ai pas eu, depuis, la moindre raison de le regretter.) Me voilà donc, quelques jours après l'arrivée de la missive, à la Faculté de Lettres et Philosophie de l'Université de Gand, pour la toute première fois de ma vie, en costume, prêt à être reçu. J'avais commencé à réfléchir, entre-temps, à cette opposition entre massif et comptable... C'est vrai qu'on dit *des sous* mais *de l'argent* – et qu'il y a dans le langage de ma génération une concurrence pour désigner le même référent entre *des thunes* et *de la thune*. C'est vrai qu'on dit *des filles*, mais aussi – dans l'excitation des préparatifs d'une soirée – *Allez Burton, tu viens au Koogai ce soir ? Y aura d'office de la goumiche* ! Qu'on me passe, par licence académique, le peu de sophistication de l'emploi : ce sont là bel et bien les exemples que j'ai employés lors de mon premier entretien avec

Peter Lauwers et Dominique Willems, pour montrer que j'avais cerné le sujet de la recherche pour laquelle je me portais candidat.

L'usage du mot *goumiche* n'y aura peut-être pas été pour grand-chose – toujours est-il que quelques semaines plus tard, je prenais mes quartiers au Muinkkaai, à Gand, en tant que jeune chercheur, doctorant en linguistique. J'avais découvert, pendant mes années à Leuven, ce qui distinguait *une bière* de *plusieurs bières* – notamment au travers de l'étude anthropologique de mon propre comportement et de celui de mes condisciples, les soirs de sortie. A présent, voilà qu'il me fallait réfléchir à la différence langagière entre *une bière* et *de la bière*, et écrire sur le sujet.

Je ne saurais décrire à quel point j'ai adoré le statut de doctorant – dont le caractère éphémère est tant le charme que la tragédie. D'un côté, je rêverais de pouvoir poursuivre à durée indéterminée ma vie d'étudiant-chercheur dans les mêmes conditions : la liberté presque sans limites, l'excitation naïve qui accompagne chaque nouvelle idée, l'autonomie et l'auto-discipline, le luxe des périodes moins productives, le plaisir d'apprendre et de contribuer au domaine qui me passionne sans le poids des propres idées plus anciennes ni des responsabilités administratives. D'un autre, je me réjouis bel et bien qu'après quatre années de préparation, la première page de mon engagement académique se tourne. L'opposition massif-comptable gardera pour toujours une place de choix sur mon cv académique, mais je suis content de pouvoir lui prédire pour l'avenir une compagnie plus variée.

Non pas que je veuille prétendre être lassé du sujet, dont l'omniprésence dans mon esprit ces quatre dernières années se serait transformée en aversion. Mon oreille de linguiste est peut-être plus sensible qu'une autre à présent aux effets de l'alternance morphosyntaxique, mais pas tellement plus, je pense, qu'elle n'était de toute façon à l'affût de toute production intéressante – et pas au point d'en être exaspérée. A vrai dire, mon sujet de thèse ne m'aura pas tant déformé que certaines personnes de mon entourage, à présent invariablement aux aguets pour me faire remarquer, avec un grand sourire satisfait, chaque syntagme nominal massif ou comptable plus ou moins créatif entendu au détour de n'importe quelle conversation : « Eh Tim, elle a dit *créer DU lien entre les gens*. » « Haha, oui, merci Maxx, j'te citerai dans ma préface ! »

## Liste d'abréviations

ANS      Algemene Nederlandse Spraakkunst

NTU      Nederlandse Taalunie



## Liste des Tableaux

Tableau 1	Liste des effets de sens comptables en termes d'objets (individualisés ou collectifs) des noms de matériaux étudiés en français.	147
Tableau 2	Liste des effets de sens comptables en termes d'objets spécifiques des noms de matériaux étudiés en néerlandais.	158
Tableau 3	Estimations de notre modèle par régression ordonnée à effets mixtes de l'acceptabilité des noms de fruits et de légumes en syntaxe massive, en français, intégrant I comme facteur continu fixe, ainsi que « item » et « participant » comme interceptions aléatoires. (*Variance et (Déviation Standard), † Valeur <i>p</i> basée sur le test du rapport de vraisemblance.)	204
Tableau 4	Estimations de notre modèle par régression ordonnée à effets mixtes de l'acceptabilité des noms de fruits et de légumes en syntaxe massive, en néerlandais, intégrant I comme facteur continu fixe, ainsi que « participant » comme interception aléatoire. (*Variance et (Déviation Standard), † Valeur <i>p</i> basée sur le test du rapport de vraisemblance.)	210
Tableau 5	Estimations de notre modèle par régression logistique à effets mixtes de la production, en syntaxe massive ou comptable, des noms de fruits et de légumes dans notre expérience d'amorçage, intégrant l'amorce comme facteur fixe (effet de l'amorce massive co-hyponyme, et de l'amorce massive de contrôle, en comparaison avec notre état initial), ainsi que « item », « participant » et « phrase » comme interceptions aléatoires. (*Variance et (Déviation Standard), † Valeur <i>p</i> basée sur le test du rapport de vraisemblance.)	228
Tableau 6	Estimations de notre modèle par régression logistique à effets mixtes de la production, en syntaxe massive ou comptable, des noms de fruits et de légumes dans notre expérience d'amorçage, intégrant l'amorce, l'indice « I » (biais massif de chaque item, calculé en corpus) et l'indice « P » (biais massif par phrase, mesuré dans un Cloze test indépendant) comme facteurs fixes, ainsi que « item », « participant » et « phrase » comme interceptions aléatoires. (*Variance et (Déviation Standard), † Valeur <i>p</i> basée sur le test du rapport de vraisemblance.)	230
Tableau 7	Liste des phrases employées dans notre expérience pour éliciter la production d'un nom de fruit ou de légume – classées par ordre d'attirance pour le massif (tel que mesurée à l'aide d'un Cloze test, N = 61).	233



## Liste des Figures

- Figure 1 Illustration du modèle à activation du lexique de (Levelt *et al.* 1999: 4, figure 2). Légende originale : « Fragment of the lexical network underlying lexical access. The feedforward activation spreading network has three strata. Nodes in the top, conceptual stratum represent lexical concepts. Nodes in the lemma stratum represent syntactic words or lemmas and their syntactic properties. Nodes in the form stratum represent morphemes and their phonemic segments. Also at this level there are syllable nodes. » 74
- Figure 2 Modèle plus précis du niveau lemmatique de (Levelt *et al.* 1999), tel que proposé par (Pickering et Branigan 1998: 635, fig.1). Les deux nodules supérieurs représentent les deux possibilités de construction du datif en anglais, et sont partagés par les différents verbes en question, illustrés ici par les nodules *give* ‘donner’ et *send* ‘envoyer’. Ces deux nodules sont sous-spécifiés quant à leurs différents traits syntaxiques, eux-mêmes représentés dans d’autres nodules partagés (ici, au bas de la figure). 76
- Figure 3 Aperçu de l’interface de *WebCorp* (nous avons ajouté la numérotation des points). 90
- Figure 4 Sélection des noms de matériaux à étudier. 94
- Figure 5 Sélection des noms de boissons à étudier. 94
- Figure 6 Sélection des noms de fruits et de légumes à étudier. 95
- Figure 7 Prompteur de commande affiché au lancement de *SN\_FR.pl*. Le texte *fichier\_à\_analyser.txt* est fictif, et indique l’endroit où il faut écrire le nom du fichier TXT qui servira d’input au programme. 99
- Figure 8 Liste des items de contrôle. 106
- Figure 9 Profils distributionnels des 29 noms de fruits et de légumes étudiés en français. Les emplois massifs sont représentés en rouge, les emplois comptables en bleu. 110
- Figure 10: Profils distributionnels de nos noms de fruits et de légumes en français, après exclusion d’*ail*, *ciboulette* et *calabasse*. L’axe des ordonnées ne présente que les 25 % inférieurs. 111
- Figure 11 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais. Les emplois massifs sont représentés en rouge, les emplois comptables en bleu. L’axe des ordonnées s’étend de 0 à 40 %. 126

Figure 12 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais, avec distinction entre emplois massifs correspondant effectivement à un massif en français ( <i>massif_oui</i> ), et emplois massifs correspondant à une forme alternative neutre ( <i>massif_gen</i> pour l'article défini générique, <i>massif_non</i> pour les autres).	134
Figure 13 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais, en excluant les occurrences massives ne correspondant pas univoquement à une occurrence massive en français, afin de fournir un point de départ pertinent pour la comparaison avec la Figure 10.	135
Figure 14 Profils distributionnels de nos 30 noms de matériaux, en français.	140
Figure 15 Profils distributionnels de nos 30 noms de matériaux en français, avec détail sémantique des emplois comptables (lectures en termes de sortes et d'objets, collectives ou non).	145
Figure 16 Profils distributionnels de nos 34 noms de matériaux, en néerlandais.	150
Figure 17 Profils distributionnels de nos 34 noms de matériaux en néerlandais, avec détail sémantique des emplois comptables (lectures en termes de sortes et d'objets).	156
Figure 18 : Profils distributionnels de notre sélection de 36 noms de boissons, en français.	162
Figure 19 Profils distributionnels de notre sélection de noms de boissons, en néerlandais.	167
Figure 20 Ebauche d'un modèle stratifié du lexique, inspiré de (Levelt et al. 1999), avec représentation du massif (M) et du comptable (C) comme nodules partagés au niveau lemmatique, reliés d'une part aux différents lemmes nominaux (LEMME_1 à 3), et d'autre part à certains nodules conceptuels.	177
Figure 21 Zoom sur la strate lemmatique, montrant que les terminaisons C et M que nous postulons sont en fait elles-mêmes des réseaux cohérents de différents nodules combinatoires (comme par exemple celui du pluriel, ou de la combinaison avec <i>un(e)</i> pour le comptable), avec lesquels chaque lemme n'est pas obligatoirement relié de manière égale. Même cette représentation plus détaillée est du reste une simplification : le nodule pluriel correspond sans nul doute lui-même à plusieurs contextes combinatoires distincts.	180
Figure 22 Détail d'une connexion supra-lexicale, entre le nodule sémantique ('substance de construction') et le nodule combinatoire du niveau lemmatique M. Cette connexion s'établit par généralisation, en conséquence de la fréquente activation simultanée des deux nodules concernés.	184
Figure 23 Ebauche de modélisation de l'organisation lexicale des noms de boissons. Les lectures disponibles sont celles en 'portion standard', 'sorte' et 'liquide potable', chacune liée à l'une ou l'autre option combinatoire. Pour <i>vin</i> , la lecture comptable conditionnée est fortement concurrencée par la composition avec <i>verre de</i> .	187
Figure 24 Aperçu de l'organisation lexicale des noms de boissons en néerlandais, illustrée par les items <i>bier</i> 'bière', <i>kir</i> 'kir' et <i>koffie</i> 'café'. La possibilité de l'emploi massif dans le sens de 'liquide potable' est disponible pour	



- les trois items (et est l'emploi de base pour *bier* et *koffie*, mais pas pour *kir*). Pour la lecture en 'portion standard', les options de l'emploi comptable et du diminutif sont disponibles, mais jouissent de rapports différents aux trois lemmes concernés. Pour *bier*, ce sera toujours le diminutif qui l'emportera, tandis que pour *koffie*, les deux alternatives sont disponibles de manière sensiblement équivalente. Notons que la connexion entre les nœuds DIM et C indique, quant à elle, que le diminutif implique automatiquement l'emploi comptable. 191
- Figure 25 Représentation du mécanisme du Trieur, responsable de la lecture taxonomique d'un syntagme comptable. Dans le cas de noms habituellement massifs, le comptable pointe de manière claire vers la lecture en 'sortes', tandis que pour les noms habituellement comptables, il y a concurrence avec la lecture 'objet individuel', plus saillante. 194
- Figure 26 Enquêtes d'acceptabilité: échelle de Likert à 7 points telle qu'explicitée dans les consignes données aux participants. Pour les questionnaires en néerlandais, nous avons bien entendu utilisé une traduction. 199
- Figure 27 Acceptabilité des emplois massifs des noms de fruits et de légumes (en rouge), comparée à ceux des noms de contrôle (en vert), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 202
- Figure 28 Acceptabilité des emplois de nos noms de fruits et de légumes, en syntaxe comptable (bleu) et massive (rouge), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 203
- Figure 29 Corrélation positive entre l'acceptabilité d'un nom de fruit ou de légume en syntaxe massive (telle que mesurée dans notre enquête « A ») et le biais massif I calculé pour chaque item sur la base de nos données de corpus, en français. Plus le biais massif est élevé, plus l'emploi est jugé comme acceptable. [X : indice « I », Y : note d'acceptabilité.] 205
- Figure 30 Acceptabilité des noms de fruits et de légumes en syntaxe massive, dans un contexte phrastique neutre (rouge clair) et orienté en termes de « consommation culinaire » (rouge foncé), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 206
- Figure 31 Acceptabilité des noms de fruits et de légumes en néerlandais, dans deux types de syntagmes : emploi nu (orange) et emploi avec *wat* (jaune). Nombre de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 207
- Figure 32 Acceptabilité des emplois massifs des noms de fruits et de légumes (en rouge), comparée à ceux des noms de contrôle (en vert), en néerlandais. Nombre de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 208
- Figure 33 Acceptabilité des emplois de nos noms de fruits et de légumes, en syntaxe comptable (bleu) et massive (rouge), en néerlandais. Nombre

- de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 209
- Figure 34 Corrélation positive entre l'acceptabilité d'un nom de fruit ou de légume en syntaxe massive (telle que mesurée dans notre enquête « A ») et le biais massif I calculé pour chaque item sur la base de nos données de corpus. Plus le biais massif est élevé, plus l'emploi est jugé comme acceptable. [X : indice « I », Y : note d'acceptabilité.] 210
- Figure 35 Acceptabilité des noms de fruits et de légumes en syntaxe massive, dans un contexte phrastique neutre (rouge clair) et orienté en termes de « consommation culinaire » (rouge foncé), en néerlandais. Nombre de participants = 29. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.] 211
- Figure 36 Dispositif expérimental. En orange : sélection de 54 images, disposées face à chaque participant. En rouge : fiches-consigne (en alternance, une phrase incomplète à lire puis une image ou, pour le complice dans le cas d'une interaction critique, une réponse préétablie). En vert : la pile de cartes sélectionnées et mises sur le côté. Dans l'exemple illustré ici, #C devra à présent répéter la phrase complète (*Dans la sauce, on pourrait mettre des/du concombre(s)* – en fonction de la réponse de #N) puis retrouver, parmi les cartes disposées devant lui, celle représentant un/du/des concombre(s). 217
- Figure 37 Image utilisée dans notre expérience pour éliciter la production d'un SN contenant le nom *carotte*. *A priori*, la présentation n'exclut ni ne nécessite l'usage de la syntaxe massive ou comptable. 219
- Figure 38 Image utilisée dans notre expérience pour éliciter la production d'un SN contenant le nom *oignon*. *A priori*, la présentation n'exclut ni ne nécessite l'usage de la syntaxe massive ou comptable. 219
- Figure 39 Liste des 24 phrases employées dans notre expérience d'amorçage pour la production de SN comportant un nom de fruit ou de légume (contextes profilant explicitement le domaine de la nourriture). Dans une interaction critique, #C aurait l'une de ces phrases à lire, et #N devrait la compléter sur la base d'une image d'un fruit ou d'un légume. 221
- Figure 40 Tableur utilisé pour noter nos observations au cours de notre expérience (généré de manière automatique). Les réponses données par chaque participant à tour de rôle sont notées dans la colonne RESP. Dans ce cas, l'on note aussi dans les deux colonnes suivantes le type de syntaxe (massive ou comptable) produit, et le lemme en question. 224
- Figure 41 Message affiché pendant nos trois jours d'expérimentation sur la porte du local du département de néerlandais, à la Faculté de Traduction et d'Interprétation de l'Université de Mons. 225
- Figure 42 Nombre de productions de noms de fruits et de légumes en syntaxe comptable (bleu) et massive (rouge), en fonction des trois conditions d'amorçage. Nombre total d'observations = 585. 227
- Figure 43 Modélisation, dans le lexique mental, des noms *haricot*, *carotte*, *raisin* et *pomme*. Chaque nom a des préférences distributionnelles propres quant à l'emploi massif ou comptable, représenté par des connexions

plus ou moins fortes entre les lemmes correspondants et les nodules combinatoires M et C. Chaque lemme est également associé à un concept lexical (unitaire), représenté par une image. Ces concepts sont connectés (plus ou moins fortement, en fonction de nos connaissances encyclopédiques) à différents nodules sémantiques, activables en contexte, et eux-mêmes associés directement aux nodules combinatoires M et C.



# Table des matières

## Introduction 1

1.1	Du temps et des idées .....	1
1.2	Cadre théorique et choix méthodologiques .....	2
1.3	Structure de la thèse .....	5

## Partie 1 : Définition du sujet et état de la question ..... 7

### Chapitre 1 L'opposition massif/comptable ..... 9

1.1	Introduction : tour d'horizon des questions.....	9
1.2	Critères morphosyntaxiques.....	11
1.2.1	En français .....	11
1.2.2	En néerlandais.....	14
1.3	Le lien entre référence massive et comptable, et réalité.....	17
1.4	La dénotation massive et comptable .....	22
1.4.1	La dénotation homogène : critères sémantiques.....	24
1.4.2	La quantification massive et comptable .....	32
1.4.3	Le massif-comptable au-delà du domaine nominal .....	35
1.5	La représentation de l'opposition massif-comptable dans le système linguistique.....	38
1.5.1	La vision lexicaliste .....	39
1.5.2	La flexibilité de l'opposition massif/comptable .....	42
1.5.3	La vision grammaticale.....	50
1.5.4	Conclusions préliminaires.....	55
1.6	Le point de vue de l'acquisition .....	56

### Chapitre 2 Le sens, et la question de la polysémie..... 59

2.1	Introduction : multifonctionnalité du signe linguistique.....	59
2.2	Kleiber 1999 : approche lexicale polysémiste .....	60
2.3	Willems 2013 : approche monosémiste .....	63
2.4	Multifonctionnalité et massif-comptable : pistes .....	66

### Chapitre 3 Modèles du lexique mental ..... 69

3.1	Introduction : limites de la vision traditionnelle.....	69
3.2	Le lexique mental en psycholinguistique .....	71
3.2.1	Caractéristiques générales.....	71
3.2.2	Le modèle de (Levelt <i>et al.</i> 1999), ainsi que ses extensions.....	73
3.2.3	Le niveau sémantique .....	78
3.3	Retour au massif-comptable : pistes .....	80
<b>Partie 2 : Etudes de corpus .....</b>		<b>83</b>
<b>Chapitre 4</b>	<b>Principes de la méthode.....</b>	<b>85</b>
4.1	Introduction .....	85
4.2	Présentation du corpus et de l'outil <i>WebCorp Live</i> .....	86
4.2.1	Interface.....	87
4.3	Sélection des champs et des items à étudier .....	90
4.3.1	Champs sémantiques .....	90
4.3.2	Items.....	92
4.4	L'analyse quantitative.....	96
4.4.1	Principes généraux .....	96
4.4.2	Programmation d'outils informatique avec <i>Perl</i> .....	97
4.4.3	Analyse semi-automatique en français.....	98
4.4.4	Analyse semi-automatique en néerlandais .....	102
4.5	L'analyse qualitative .....	104
4.6	Analyse de quelques items de contrôle .....	105
<b>Chapitre 5</b>	<b>Les noms de fruits et de légumes .....</b>	<b>109</b>
5.1	Les noms de fruits et de légumes en français.....	109
5.1.1	Analyse distributionnelle.....	109
5.1.2	Analyse sémantique .....	115
5.2	Les noms de fruits et de légumes en néerlandais .....	125
5.2.1	Analyse distributionnelle.....	125
5.2.2	Analyse sémantique : réalisation contextuelle des emplois massifs .....	136
5.3	Conclusions.....	137
<b>Chapitre 6</b>	<b>Les noms de matériaux .....</b>	<b>139</b>
6.1	Les noms de matériaux en français.....	139
6.1.1	Analyse distributionnelle.....	139
6.1.2	Analyse sémantique .....	141
6.2	Les noms de matériaux en néerlandais .....	149
6.2.1	Analyse distributionnelle.....	149
6.2.2	Analyse sémantique .....	153
6.3	Conclusions.....	159
<b>Chapitre 7</b>	<b>Les noms de boissons .....</b>	<b>161</b>
7.1	Les noms de boissons en français.....	161

7.2	Les noms de boissons en néerlandais .....	165
7.3	Conclusions.....	171
<b>Partie 3 : Vers une modélisation.....</b>		<b>173</b>
<b>Chapitre 8</b>	<b>La double représentation de l'opposition massif/comptable .....</b>	<b>175</b>
8.1	Un modèle à deux niveaux .....	175
8.1.1	Le niveau lexical, ou lemmatique.....	178
8.1.2	Le niveau supra-lexical, ou conceptuel.....	183
8.1.3	Différences entre le français et le néerlandais.....	189
8.2	Une approche lexicaliste améliorée.....	191
8.3	Prédictions et prolongements .....	195
<b>Chapitre 9</b>	<b>Le statut lexical de l'opposition massif/comptable : enquêtes d'acceptabilité .....</b>	<b>197</b>
9.1	Introduction : hypothèses .....	197
9.2	Dispositifs.....	199
9.3	Résultats des enquêtes pour le français .....	201
9.3.1	Enquête « A ».....	201
9.3.2	Enquête « B ».....	205
9.4	Résultats des enquêtes pour le néerlandais.....	207
9.4.1	Enquête « A ».....	207
9.4.2	Enquête « B ».....	211
9.5	Discussion et conclusions .....	212
<b>Chapitre 10</b>	<b>Le statut supra-lexical de l'opposition massif/comptable : expérience d'amorçage .....</b>	<b>215</b>
10.1	Introduction .....	215
10.2	Dispositif expérimental.....	216
10.2.1	Concept général.....	216
10.2.2	Stimuli : items, phrases, amorces, <i>fillers</i> .....	220
10.2.3	Fiches-consignes.....	223
10.2.4	Organisation de l'expérience .....	224
10.3	Résultats .....	226
10.3.1	Absence de l'effet attendu d'amorçage structurel .....	226
10.3.2	Intégration des indices d'attraction massive : items et phrases .....	228
10.4	Discussion et conclusions .....	231
10.4.1	Précédence du facteur lexical et supra-lexical sur l'amorçage .....	231
10.4.2	Structure du niveau supra-lexical .....	233
<b>Partie 4 : Conclusions et prolongements.....</b>		<b>239</b>
<b>Chapitre 11</b>	<b>La question du sens.....</b>	<b>241</b>

<b>Chapitre 12</b>	<b>Le système linguistique comme réseau complexe</b>	<b>247</b>
12.1	La représentation de l'opposition massif-comptable	247
12.2	Massif-comptable : pistes pour l'avenir	249
12.3	D'autres phénomènes à modéliser	250
<b>Bibliographie</b>		<b>253</b>
<b>Appendices</b>		<b>267</b>
12.4	Etudes de corpus	267
12.4.1	Détail des recherches effectuées sur <i>WebCorp Live</i>	267
12.4.2	Résultats quantitatifs (chiffres absolus)	271
12.5	Enquêtes d'acceptabilité	275
12.5.1	Enquête « A » : flexibilité des noms de fruits et de légumes	275
12.5.2	Enquête « B » : influence du contexte	276
12.5.3	Enquêtes : liste de <i>fillers</i>	277
12.6	Expérience d'amorçage	282
12.6.1	Images de fruits et de légumes	282
12.6.2	Amorces	283
12.6.3	Fillers	284



# Introduction

## 1.1 *Du temps et des idées*

La plupart des ouvrages et articles concernant l'opposition entre massif et comptable commencent de la même manière : en rappelant (de manière plus ou moins explicite) que cette opposition n'est pas si difficile à cerner, du moins à l'intuition, et que quelques exemples suffisent pour l'évoquer à l'esprit. Ainsi, dans la présente introduction, nous ne contreviendrons pas à cette règle tacite, et nous nous contenterons de quelques illustrations emblématiques, afin de planter le décor :

- (1) J'ai acheté une voiture.
- (2) Il y a des chaises partout.
- (3) J'ai vu deux chats ce matin.
  
- (4) J'ai bu un peu d'eau.
- (5) J'ai du sable dans mes chaussures.
- (6) Il y a du vent ce matin.

On sent bien, intuitivement, que quelque chose distingue ces deux courtes séries d'exemples. Ainsi, deux éléments cruciaux peuvent d'emblée être mis à l'avant-plan : 1) il semblerait qu'il existe, en français, deux types de noms, et 2) ces deux types de noms se distinguent par le marquage morphosyntaxique dont ils s'accommodent (types de déterminants, variation en nombre). La tradition aura voulu que l'on appelle ces noms respectivement 'comptables' (en vertu, probablement, de leur affinité avec les numéraux : *trois enfants, une voiture*) et 'massifs' (puisqu'on a dit que leurs référents typiques se présentaient 'en masse' : *de l'eau, de l'air, du beurre*).

Une présentation sommaire de notre sujet, au détour d'une conversation de bistrot, pourrait fort bien s'arrêter là. La discussion en profondeur des tenants et aboutissants de cette distinction, dans ses aspects morphologiques, syntaxiques, sémantiques ne ferait, quant à elle, que commencer. En effet, si la distinction s'opère sur la base de propriétés distributionnelles, que dire alors des exemples suivants :

- (7) Je ne bois que deux eaux : la Vittel et la Spa.
- (8) La Jeep s'est embourbée dans les sables du Sahara.
- (9) Demain matin, nous aurons un vent très froid sur la côte.
  
- (10) Le parking est bondé, il y a de la voiture à perte de vue.
- (11) Ce que l'on trouve chez Ikea, ça, c'est de la chaise de qualité !
- (12) Après le passage du camion, il y avait du chat sur toute la largeur de la route.

L'une des solutions, sans encore entrer dans les détails, est de dire que les exemples (1) à (6) représentent des emplois 'de base', tandis que les phrases (7) à (12) illustrent des usages, d'une manière ou d'une autre, 'dérivés'. Cependant, l'on peut aussi imaginer des exemples où la distinction entre emplois de base et dérivé n'est probablement pas aussi évidente à établir :

- (13) J'ai mangé du chocolat. J'ai mangé quelques chocolats.
- (14) J'ai mis de l'œuf dans la salade. J'ai mis des œufs dans la salade.
- (15) Attache-les avec de la corde. Attache-les avec une corde.

En outre, l'on sent bien que le sens de chacune des deux variantes de nos énoncés (13) à (15) n'est pas tout à fait le même. L'on pourra donc s'interroger sur les manières de décrire cette différence. Et si l'on décide de formuler les critères référentiels de l'un et l'autre type de syntaxe en termes d'individualisation et d'homogénéité, que dira-t-on des noms tels que *mobilier* – dont le référent présente bien des individus (des *meubles*), tout en s'accommodant de la syntaxe massive (*J'ai acheté du mobilier neuf*) ?

Ainsi, il en faut peu pour lever – ne fût-ce qu'encore très timidement – le voile sur la nature de l'opposition massif-comptable, et les défis qu'elle pose au linguiste. Décrire la représentation de cette opposition dans le système linguistique, et ses corollaires sémantiques : voilà le chemin dans lequel nous nous engagerons, tout au long du présent ouvrage – en nous inscrivant fièrement dans la lignée des nombreux chercheurs qui nous y ont précédé.

## 1.2 Cadre théorique et choix méthodologiques

En effet, nous ne sommes certainement pas le premier à nous frotter à la question : l'opposition massif-comptable était déjà évoquée dans la première moitié du siècle précédent, par (Damourette et Pichon 1950) par exemple, ou par (Jespersen 1924) pour ce qui est de l'anglais, et elle a continué depuis lors à faire couler régulièrement de l'encre. Cependant, il semblerait qu'à plus d'un niveau, les défis théoriques que pose notre sujet restent entiers (nous entrerons dans les détails dans notre état de la

question). Nous en voulons pour preuve le regain d'intérêt récent pour le sujet, qui aura donné lieu à plusieurs publications emblématiques ces dernières années (Nicolas 2002a, Chierchia 2010, Massam 2012), pour ne pas parler des multiples publications en revue, qui ont fait avancer les choses tout en restant loin, encore, d'épuiser le sujet.

Voilà pourquoi nous désirons, à notre tour, nous pencher sur la question de l'opposition entre massif et comptable. Tout au long du présent ouvrage, nous identifierons certaines des questions qui restent sans réponse, et y proposerons nos propres solutions. Ce faisant, notre objectif ne sera pas tant de faire *du neuf avec du vieux*, que de jeter sur l'opposition massif-comptable un regard réellement nouveau. Nous entendons dès lors tirer profit de manière maximale des développements les plus récents en linguistique, et dans ses disciplines sœurs, du point de vue de l'outillage tant empirique, que conceptuel.

Plus précisément, du point de vue du cadre théorique, nous nous inscrirons dans la lignée des approches cognitivistes et *usage-based* du langage (Langacker 1987, 1991, Ungerer et Schmid 1996, Schmid 2010). De manière plus ou moins explicite, nous adhérons donc aux grands principes énoncés, par exemple, par (Croft et Cruse 2004) ou (Tummers *et al.* 2005) : 1) nous donnerons la priorité à l'usage, comme fenêtre sur l'organisation interne de la connaissance linguistique (comme l'illustrent nos méthodes de récolte de données, cf. *infra*), 2) nous ne revendiquons pas *a priori* de distinction stricte entre compétence linguistique et performance langagière (un modèle approprié de la compétence linguistique doit pouvoir refléter toutes les richesses de l'usage), et 3) nous nous méfions de la *rule-list fallacy* (voir aussi (Langacker 1990:264)), c'est-à-dire de l'illusion que la langue est organisée comme un ensemble de règles d'une part, et d'une liste d'items d'autre part (en réalité, notre connaissance linguistique est organisée à partir de notre expérience, et les schémas qui semblent être des règles sont plus globalement des généralisations, plus ou moins productives, à partir d'occurrences spécifiques rencontrées ; voir par exemple de manière célèbre (Bybee 1988)).

En outre, dans notre recherche de modélisation du lexique, nous entendons tirer profit de la longue tradition psycholinguistique sur le sujet. Comme nous le montrerons dans notre état de la question, il nous semble qu'une théorie satisfaisante de l'opposition massif-comptable implique un modèle adéquat du lexique (en tant qu'entité cognitive), or c'est là peut-être le principal sujet sur lesquels la psycholinguistique a une longueur d'avance sur la linguistique générale. Faire se rencontrer les deux traditions, en particulier en montrant l'intérêt d'enrichir d'une touche de psycholinguistiques les approches en linguistique générale, sera ainsi l'un de nos objectifs.

D'un point de vue méthodologique, et en accord avec notre cadre théorique, nous favoriserons les données langagières réelles comme point de départ de notre réflexion théorique. Ainsi, une fois passé l'exercice de l'état de la question, nos réflexions propres se construiront sur la base de diverses études de corpus, de deux enquêtes d'acceptabilité, et d'une expérience d'amorçage structurel – dans une combinaison de

méthodes répondant aux plaidoiries de, par exemple, (Gilquin et Gries 2009). Quant aux études de corpus, elles sont de plus en plus courantes dans la recherche linguistique (Arppe *et al.* 2011), et nous entendons tirer profit de cette tradition bourgeonnante. Nos enquêtes, menées selon un design quasi-expérimental, permettront quant à elle d'objectiver les intuitions natives (le premier amour de la linguistique moderne) en leur donnant une base intersubjective et contrôlée statistiquement (cf. (Schütze 1996)). Pour finir, notre expérience d'amorçage, qui clôturera notre parcours réflexif, se voudra le reflet du rapprochement que nous aurons engagé avec la tradition psycholinguistique. Tout au long de ce parcours, nous appliquerons une structure en 'hélice', telle que décrite par (Tummers *et al.* 2005), où formulation d'hypothèses de plus en plus spécifiques et récolte ciblée de données se succèdent dans une démarche d'affinement progressif.

Concernant notre sujet, à défaut d'être parfaitement exhaustifs (une prétention que nul ne doit avoir), nous veillerons à examiner l'opposition massif-comptable sous un maximum de coutures, et dans un maximum de ses dimensions – abordant ainsi tant ses aspects (morpho)syntaxiques que sémantiques, ainsi que son apparition en discours. Ce faisant, nous rappelons que la triade sémiotique traditionnelle de Morris (Morris 1938) doit rester un dispositif heuristique enrichissant, et non un obstacle au développement unifié de la recherche linguistique (cf. (Rastier 1990)). Notre approche sera donc multimodale, et multidimensionnelle.

Pour finir, tout comme la langue n'existe pas sans contexte discursif, de même un phénomène linguistique n'existe pas sans un système dans lequel s'exprimer. Ainsi, nous faisons le choix d'étudier le massif-comptable au travers de son ancrage dans deux langues modernes : le français et le néerlandais. L'examen d'au moins deux langues nous semble indispensable pour ne pas confondre, dans nos résultats, les éléments à généraliser et les particularités d'une langue spécifique. De ce point de vue, le choix d'une langue romane et d'une langue germanique se justifie également – l'éloignement génétique, fût-il limité, supposant déjà de potentielles différences typologiques. Le français sera notre objet en ce qu'il est aussi notre langue d'expression académique, de même que notre langue maternelle. Le néerlandais, quant à lui, jouit tant de notre maîtrise quasi-native, que du statut de langue officielle de l'institution dans laquelle nous rendons la présente thèse. Ainsi, la comparaison systématique du français et du néerlandais sera l'un des fils rouges de notre travail, au travers d'un traitement méthodologique et analytique équivalent, dans la mesure du possible (en particulier pour les études de corpus et les enquêtes).

## 1.3 Structure de la thèse

Le présent ouvrage est composé de quatre parties.

La première constitue notre état de la question, où nous examinons principalement la littérature qui nous précède au sujet de l'opposition massif-comptable (Chapitre 1), en évoquant aussi la question de la polysémie et de la multifonctionnalité du signe linguistique (Chapitre 2), ainsi que la modélisation théorique du lexique mental (Chapitre 3).

La seconde partie est consacrée à nos études de corpus. Après un chapitre introductif présentant notre méthodologie (Chapitre 4), nous étudions tour à tour le domaine des noms de fruits et de légumes (Chapitre 5), des noms de matériaux (Chapitre 6), et des noms de boissons (Chapitre 7).

Dans notre troisième partie, nous proposons une modélisation de l'opposition massif-comptable dans le lexique mental, permettant de tenir compte de nos différentes observations de la partie précédente (Chapitre 8). Nous testons ensuite différentes prédictions de notre modèle, au travers d'enquêtes d'acceptabilité (Chapitre 9) et d'une expérience d'amorçage (Chapitre 10).

Pour finir, dans notre quatrième et dernière partie, nous concluons notre étude et évoquons des pistes quant à son prolongement. Ainsi, nous synthétisons d'abord ce que notre recherche nous a appris au sujet du sens et de la polysémie (Chapitre 11), avant de montrer comment notre modèle pourrait être peaufiné de même qu'appliqué à d'autres domaines de la linguistique (Chapitre 12).



## **Partie 1 : Définition du sujet et état de la question**

Dans cette partie, nous commençons par examiner en détail les critères de surface qui permettent de distinguer le massif du comptable, tant en français qu'en néerlandais. Ensuite, nous explorerons les différents défis théoriques que pose notre opposition, et ce que nos prédécesseurs en ont dit. Ce faisant, nous identifierons les points consensuels, de même que ceux qu'il reste à trancher – lançant ainsi les pistes de réflexion que nous suivrons dans les parties suivantes. Nous introduirons également deux autres sujets, liés indirectement mais néanmoins de manière irrévocable à la question du massif-comptable, à savoir la question du sens et de la polysémie d'une part, et la question de la modélisation du lexique mental d'autre part.





# Chapitre 1

## L'opposition massif/comptable

### 1.1 Introduction : tour d'horizon des questions

S'il l'on en croit l'abondante littérature sur le sujet, l'opposition qui nous intéresse est un phénomène complexe et fascinant. Au-delà de son apparente simplicité intuitive, elle pose au chercheur de nombreux mystères, si bien qu'il ne faut pas parler de « la » question du massif-comptable, mais bien « des » questions. Celles-ci feront l'objet du présent chapitre.

Tout d'abord, il s'agira de définir les caractères de surface qui donnent à notre opposition son expression. Nous l'avons vu dans les exemples (1) à (6) : les appellations 'massif' et 'comptable' semblent correspondre à certains types de déterminants et, plus largement, d'environnements syntagmatiques et morphosyntaxiques spécifiques – et il s'agira pour nous d'établir desquels il s'agit. Nous essayerons donc de dire ce qui, d'un point de vue morphosyntaxique, constitue *du comptable* et *du massif*. Nous ferons cela tant pour le français que pour le néerlandais, ce qui nous permettra d'emblée d'examiner en quoi, sur ce point, nos deux langues se ressemblent, et se distinguent.

Ensuite, nous nous tournerons vers les trois grands casse-têtes théoriques posés par l'opposition massif-comptable, sur lesquels nos prédécesseurs se sont penchés. Pour chaque question, nous définirons le problème et examinerons de manière critique les solutions qui y ont été proposées.

La première question sera celle du lien entre la référence massive et comptable, et la réalité. Dans quelle mesure peut-on dire que notre opposition morphosyntaxique reflète des propriétés référentielles ? De manière subsidiaire, il s'agira également d'aborder la question de l'arbitraire de la grammaire : les traits morphosyntaxiques sont-ils assignés de manière aléatoire ?

Ensuite, nous nous interrogerons sur la nature sémantique de la référence massive et comptable. Intuitivement, il semble clair que massif et comptable ont une dénotation

différente – et nous examinerons les critères et modèles qui ont été proposés pour exprimer cela.

La troisième question concernera, quant à elle, la représentation de l'opposition massif-comptable dans le système linguistique. Il s'agira ici de voir à quel niveau elle doit être située : s'agit-il d'une opposition lexicale (entre deux types de noms), ou d'une opposition purement grammaticale ? Au travers de cette question, nous parlerons aussi de la 'flexibilité' (ou 'élasticité') de l'opposition : le fait que certains noms, selon le contexte, peuvent s'employer tantôt comme des comptables, tantôt comme des massifs.

Nous clôturerons notre tour d'horizon par un dernier sujet de recherche, plus en marge des précédents : celui de l'acquisition de l'opposition massif-comptable, dans l'apprentissage du langage. Cette branche – héritée très explicitement de la tradition psycholinguistique – n'est pas sans de nombreux mérites, mais nous nous y attarderons un peu moins, n'ayant pas nous-même travaillé directement sur le sujet. Elle nous permettra néanmoins, le moment venu, d'étayer certaines de nos thèses sur les rapports entre le niveau formel et conceptuel du langage.

Notons que dans la littérature sur le massif-comptable, ces différentes questions ont joui tantôt d'une certaine autonomie, tantôt ont été traitées ensemble – voire confondues. Ainsi, les typologies des approches et modélisations du massif-comptable proposées par divers auteurs, à différents moments, reflètent généralement soit un traitement sélectif de l'un ou l'autre aspect de notre sujet, soit une tentative de synthèse qui, malheureusement, encourt le risque de confondre ce qui n'a pas lieu de l'être (nous y reviendrons). Quoi qu'il en soit, la manière de traiter l'un de nos casse-têtes aura généralement une incidence directe sur l'approche d'un autre critère. Ainsi, par exemple, la manière de décrire le sens du massif et du comptable ne peut se faire sans référence au rapport avec le réel, de même que les trouvailles concernant l'apprentissage de l'opposition pourront fournir des indices précieux quant à sa représentation dans le système linguistique. Dès lors, plus qu'un sujet modulaire, l'opposition massif-comptable apparaîtra bien vite comme un phénomène multidimensionnel – tel qu'on le trouve décrit dans (Joosten 2003), qui propose une typologie des modèles qui essaye de prendre en compte de manière combinée les différentes questions que nous avons évoquées.

## 1.2 Critères morphosyntaxiques

### 1.2.1 En français

Parler de l'opposition massif-comptable en termes de critères distributionnels est à la fois l'approche la plus intuitive, et la plus ancienne (Weinreich 1966, Galmiche 1986, Kleiber 1999, Riegel *et al.* 2009). Un simple recours à l'introspection démontre d'emblée des différences intéressantes : pour ne prendre que deux exemples au hasard, l'on a facilement *une voiture* et *trois voitures*, de même que *du beurre* et *un peu de beurre*, mais pas de manière évidente les équivalents inverses – ??*un peu de voiture*, ??*trois beurres*. Les déterminants *un* et *trois* de nos exemples semblent donc avoir quelque chose en commun, qui n'est pas partagé par le partitif *du* ou la locution déterminative *un peu de*, qui eux-mêmes semblent devoir être groupés de quelque façon. L'on pourra dès lors distinguer deux types de contextes syntagmatiques, baptisés respectivement 'comptables', et 'massifs'. Ceux-ci ont déjà été listés souvent (pour différentes langues, du reste) – comme par exemple dans (Nicolas 2002a: chapitre 1, 1. a.), où on retrouve ceux du français.

Ainsi, dans les premiers, les comptables, nous retrouvons : l'article indéfini *un(e)* (*une voiture*), les numéraux (*deux voitures*, *cent voitures*), les déterminants singuliers tels que *aucun(e)* et *chaque* (*chaque voiture*), les déterminants pluriels tels que *plusieurs* et *certain(e)s* et les constructions quantificatrices telles que *un grand nombre de* ou *des milliers de* (*des milliers de voitures*). Dans les seconds, les massifs, nous retrouvons essentiellement l'article partitif (*du beurre*) et la locution déterminative *un peu de* (*un peu de beurre*). (Au sujet d'*un peu de*, voir les travaux de (Martin 1969) et (Ducrot 1972), qui ne touchent cependant pas à l'opposition massif-comptable, ainsi que (Doetjes 2001).)

(Notons, à ce stade, que si dans nos exemples illustratifs, nous utilisons systématiquement les noms *voiture* et *beurre* (nous aurions aussi pu choisir *enfant* et *dentifrice*, ou des centaines d'autres possibilités), c'est parce que ceux-ci nous semblent être deux candidats prototypiques au statut de noms respectivement comptable et massif. Néanmoins, dire que pour autant certains noms sont massifs et d'autres sont comptables au niveau du lexique, c'est encore une autre question : nous la traiterons lorsque nous aborderons la représentation de l'opposition dans le système linguistique. Que certains déterminants peuvent être massifs ou comptables, en revanche, est une observation relativement consensuelle.)

Revenons aux déterminants. Il en existe aussi qui en eux-mêmes ne sont ni comptables, ni massifs, mais signalent néanmoins le comptable en combinaison avec un nom au pluriel, et le massif avec un nom au singulier. (Ils se combinent ainsi exclusivement avec des noms dont la référence vérifie le critère sémantique de cumulativité, voir *infra*, et nous pourrions dès lors les appeler 'déterminants cumulatifs')

– une appellation d’autant plus intéressante que ces déterminants ‘cumulent’ deux possibilités d’emploi.) Il s’agit principalement de locutions de quantification telles que *peu de, beaucoup de, trop de, combien de* (*beaucoup de voitures, trop de beurre*). Notons qu’en français moderne, la distinction entre singulier et pluriel ne s’entend généralement plus au niveau du nom (à quelques exceptions près, comme *cheval-chevaux*), mais elle reste identifiable via les liaisons et l’accord du verbe (ne remettant pas en doute la pertinence des déterminants cumulatifs pour distinguer entre massif et comptable) : *beaucoup de grands (z) enfants, beaucoup de beurre est (\*sont) produit en France*. D’autre part, il semblerait que certaines constructions du type *Dét N de N* imposent la même condition de cumulativité : comparons par exemple *un sac de billes* (pluriel, comptable) à *un sac de sable* (singulier, massif). Les noms qui imposent cette condition sont déjà mentionnés par (Nicolas 2002a: note 5), et étudiés plus abondamment par (Grinevald 2004), qui les appelle ‘termes de mesures’. Ceux-ci constituent une classe ouverte qui inclut des noms désignant des mesures conventionnelles (*kilo, litre*), des contenants (*verre, boîte*), des formes (*tranche*) ou des arrangements (*pile, tas*). En français (tout comme en anglais), tous ces noms ne se combinent, à l’instar de nos déterminants cumulatifs, qu’avec des éléments dont la référence est cumulative (comme le remarque aussi (Doetjes 2012)).

Suivant l’exemple des déterminants et contextes cumulatifs, l’on pourrait également compter dans les formes typiquement comptables toutes celles du pluriel (*les, des, ces, mes, tes*, etc.). Cependant, à ce niveau, une remarque s’impose. Il est vrai qu’il relève du comptable que de varier entre singulier et pluriel (*Paul a une voiture ; Marc à deux voitures ; les voitures de Paul et Marc*), tandis qu’en général, le massif ne se rencontre qu’au singulier. Il existe cependant en français au moins certains emplois qui répondent aux critères du massif, mais qui exhibent néanmoins les atours formels du pluriel – l’on a affaire alors à des cas de pluriel lexicalisé (pour une étude détaillée du pluriel lexical, voir (Acquaviva 2008)). L’exemple célèbre est celui d’*épinards* : s’il est bien au pluriel (*mange tes épinards ; ces épinards sont délicieux*), il se combine néanmoins volontiers avec les déterminants du massif (*un peu d’épinards ; voir aussi (Doetjes 2001)*) et non avec ceux du comptable (*\*quatre épinards, \*chaque épinard(s)*). (Ici, si le partitif ne fonctionne pas, c’est uniquement pour des raisons d’accord en nombre. A noter que certains auteurs ont plaidé pour considérer l’article *des* comme un article partitif pluriel, plutôt que comme le pluriel de l’article indéfini *un*. Nous n’entrons pas de ce débat, dont (Leeman 2004: 136-149) offre une excellente mise au point.)

Plus précisément, il a été observé que les déterminants du pluriel n’étaient pas équivalents quant à la mesure dans laquelle ils marquaient l’individualisation référentielle (et ainsi la comptabilité), et qu’ils pouvaient être classés le long d’une échelle allant de l’indéfini *des*, marquant la sommation sans profiler du tout d’individus, jusqu’au numéraux (*deux, trois, quatre*) responsables d’une individualisation forte – en passant par *quelques, plusieurs, et divers*, par ordre croissant de profilage des individus (voir (Bosveld-De Smet 2000) ; voir aussi la récente étude distributionnelle de (Lauwers

2014, à paraître), qui établit parmi les noms au pluriel lexicalisé différentes classes, sur la base de la combinaison avec ces différents déterminants). C'est ce rapport à l'individualisation qui explique que l'on puisse dire sans peine *des épinards* ou *quelques épinards*, un peu moins facilement *?plusieurs épinards*, et pas du tout *\*quatre épinards* (hors réinterprétation sémantique). (Pour la différence entre *quelques* et *plusieurs*, voir (Gréa 2008), qui distingue les deux déterminants dans une perspective cognitive et montre que les SN avec *quelques* – et non avec *plusieurs* – dénotent des collectifs, ce qui explique que des deux, seul *quelques* puisse se combiner avec les pluriels massifs.)

Ainsi, l'on formulera le critère morphosyntaxique de nombre avec quelque précaution : lorsqu'il est grammatical, c'est-à-dire qu'il découle effectivement d'une transformation à partir du singulier, le pluriel est bel et bien caractéristique du comptable. (Dans la phrase *Jean a une Porsche et une Mercedes, mais ces voitures ne sortent jamais du garage*, le nom *voiture* est effectivement employé de manière comptable.) Lorsqu'il est lexicalisé, par contre, il peut être la marque d'un massif pluriel (voir (McCawley 1975, Ojeda 2005) ; voir aussi (Wilmet 2007: §§152-153), qui dissocie l'opposition singulier-pluriel de l'opposition massif-comptable, en parlant respectivement d'une opposition continu-discontinu et massif-numératif).

Ajoutons qu'il conviendra d'user de la même précaution au sujet des noms qui semblent être invariablement singulier – une condition apparemment nécessaire pour identifier le massif (à l'exception des cas de massif pluriel, donc), mais pas suffisante. Ainsi, si l'on regarde les exemples de noms sans pluriel que donnent (Grévisse et Goosse 2011: §507) dans leur célèbre *Bon Usage*, l'on n'aura pas tât fait de les qualifier tous de massifs, sans distinction. Ainsi, si par exemple le caractère invariable des noms d'arts et de sports semble bien s'accommoder d'un caractère propice au massif (*la peinture, faire de la peinture ; le tennis, faire du tennis*), il n'en est pas de même des noms de points cardinaux (*le Nord ; \*du Nord*) ou des infinitifs nominalisés (*le boire et le manger ; \*du boire*).

Mentionnons encore que si certains déterminants et contextes syntagmatiques peuvent signaler un emploi massif ou comptable, d'autres (non moins nombreux ou courants) sont quant à eux aveugles à la distinction. Ainsi en est-il de tous les articles et déterminants définis singuliers, comme *le/la, ce/cette, mon/ma*, etc. (*la voiture, le beurre ; ma voiture, mon beurre*), qui effectuent une saisie externe sur le référent et renvoient ainsi à des totalités et non à des partitions. Nous les appellerons les 'déterminants neutres'. Par cela, nous entendons seulement qu'en surface, au niveau du SN, ils ne permettent pas de distinguer entre emplois massifs et emplois comptables. Cela ne signifie pas nécessairement que leur présence neutralise la distinction – puisque celle-ci peut réapparaître plus loin au travers de la reprise pronominale, comme l'a montré Galmiche, voir (16) et (17), repris de (Galmiche 1989: 65) – mais seulement qu'elle ne permet pas de l'identifier.

(16) Le sang avait coulé, il y en avait par terre.

(17) \*Le livre était tombé, il y en avait par terre.

Notons au passage que les exemples ci-dessus nous permettent en outre d'identifier la reprise pronominale comme un autre critère morphosyntaxique permettant de signaler un emploi massif ou comptable : à l'instar des déterminants cumulatifs, la reprise avec *en* n'est possible que pour un emploi pluriel comptable, ou un emploi singulier massif.

### 1.2.2 En néerlandais

En néerlandais, l'on utilise couramment les étiquettes *telbaar* 'comptable' et *ontelbaar* ou *niet-telbaar* 'non comptable', pour désigner le comptable et le massif – tandis que l'appellation *stofnaam* 'nom de matière' est réservée aux 'noms communs désignant des concepts concrets, indéterminés en forme et quantité' (définition fournie par la *Nederlandse Taalunie* sur son site <http://taaladvies.net/>, accès le 24 avril 2014, traduction personnelle ; suivent quelques exemples : *hout* 'bois', *ijzer* 'fer', *olie* 'huile', *water* 'eau', *zuurstof* 'oxygène') – un sous-ensemble donc de ce qui serait, en français, des 'noms massifs'. En soi, au niveau de la définition intuitive de notre opposition, le néerlandais ne se distingue pas fondamentalement du français. Ainsi, les briques de construction morphosyntaxiques disponibles dans notre seconde langue pour exprimer le massif et le comptable sont assez comparables à celles du français – à quelques différences notables près, que nous évoquerons ci-après.

Pour commencer, le déterminant indéfini *een* 'un(e)', les numéraux, de même que les déterminants tels que *elk(e)/ieder(e)* 'chaque' indiquent, tout comme en français, un emploi comptable (*een auto*, *drie auto's*, *elke auto* 'une voiture, trois voitures, chaque voiture'). Il en va de même pour le pluriel qui, avec les mêmes restrictions quant à son caractère grammatical, signale également un emploi comptable (*de auto's* 'les voitures') – si l'on accepte, donc, de maintenir en marge de ce critère les cas de pluriel lexicalisé (comme par exemple *onkosten* 'frais', qui – à l'instar de son équivalent en français – présente les marques morphosyntaxiques du pluriel, mais n'est pas pour autant dénombrable).

Pour le massif, la situation varie quelque peu. D'une part, l'on trouve deux déterminants correspondant *grosso modo* au sens de *un peu de* : *een beetje* et *wat* (*een beetje boter*, *wat boter* 'un peu de beurre'). (Notons cependant que si le sens de *wat* correspond bien à celui d'*un peu de*, son comportement syntaxique est sensiblement différent, puisqu'il se combine aussi avec le pluriel – ce qui en fait plutôt un déterminant cumulatif, d'après notre terminologie : *wat knikkers* 'quelques billes, *wat kersen* 'quelques cerises'.) Au niveau des définitions, le *Van Dale* (dictionnaire de référence du monde néerlandophone) propose pour *beetje* 'kleine hoeveelheid' ('petite quantité') et pour *wat* 'ter aanduiding van een onbepaalde, kleine hoeveelheid' ('pour indiquer une quantité petite et indéterminée'). Pour cette seconde entrée, l'on nous propose en outre aux

côtés de *een beetje*, le synonyme *enig*. Nous pouvons ainsi également intégrer *enig*, qui est défini comme ‘een zekere mate of hoeveelheid van (een zaak)’ (‘une certaine mesure ou quantité de (quelque chose)’) parmi les déterminants signalant typiquement un emploi massif en néerlandais (*dit heeft enige tijd gevergd* ‘ceci a demandé un certain temps’). Pour être complet, notons néanmoins que malgré le raccourci du *Van Dale* de les présenter comme synonymes, les déterminants *enig* et *wat* n’ont pas tout à fait la même distribution, puisqu’on peut dire *wat boter* ou *wat melk* ‘un peu de beurre’, mais pas \**enige boter*, \**enige melk* avec la même facilité.

Ensuite, l’équivalent néerlandais le plus proche du partitif français serait l’emploi nu : *boter* ‘du beurre’, comme dans *Jan doet boter op zijn boterham* ‘Jean met du beurre sur sa tartine’. Une précision concernant cet emploi nu, et sa correspondance avec le partitif, est néanmoins nécessaire. Si le partitif en français ne peut entrer que dans des lectures massives spécifiques (et jamais génériques, un domaine réservé de manière célèbre à l’article défini – voir (Kleiber 1990) –, que nous avons classé parmi les déterminants neutres), le générique en néerlandais des noms réputés massifs s’exprimera bel et bien à l’aide d’un nom nu lui aussi – contrairement au générique des noms comptables, qui comme en français s’accommodera volontiers de l’article défini (voir aussi (Lauwers et Vermote 2014, à paraître)). Comparons les exemples (18) et (19) à leurs traductions, en (20) et (21).

- (18) La voiture est le moyen de transport des temps modernes.
- (19) Le beurre est la crème barattée du lait.
- (20) De auto is / Auto’s zijn het vervoermiddel van de nieuwe tijd.
- (21) (Ø) Boter is de gekarnde room van melk (*Wikipedia*)

(Notons qu’en (20), l’emploi en néerlandais de l’article défini singulier pour indiquer le générique (d’un nom comptable) a potentiellement quelques relents de gallicisme (comme nous le signale Johan De Caluwe, communication personnelle) – le pluriel lui étant peut-être préférable. Dans ce cas, observons que le générique sera bel et bien exprimé au moyen d’un pluriel indéfini (*Auto’s zijn...* mais pas \**De auto’s zijn...* en lecture générique), de la même manière que la lecture générique des noms massifs requiert un emploi nu, donc indéfini également – comme dans (21). En français, en revanche, l’indéfini et le générique ne font pas bon ménage. Nous ignorons si cette différence entre nos deux langues a déjà été étudiée de manière systématique.)

Ces deux emplois du massif ‘nu’, que nous appelons donc ‘spécifique’ et ‘générique’, existent de manière tout à fait similaire en anglais, par exemple, où (Ter Meulen 1981) distingue respectivement les emplois prédicatifs (*My tooth is filled with gold* ‘Ma dent est remplie d’or’) et les emplois nominaux (*Gold is an element* ‘L’or est un élément’). (Laissons de côté l’origine de sa terminologie.) Dans le second cas, (Cheng *et al.* 2008) utilisent quant à eux l’appellation de *mass kind reading* ‘lecture massive en type’. Evidemment, l’existence de ce contraste n’enlève rien au caractère bien massif d’un syntagme nu en

néerlandais (qu'il soit spécifique ou générique) – mais nous devons en tenir compte au moment de comparer le français et le néerlandais dans nos études de corpus. (Voir aussi (Lønning 1987: 3), qui appelle à la prudence dans l'examen des emplois nus d'une part pour leur caractère potentiellement générique, comme nous venons de le dire, et d'autre part parce qu'il est des langues où il est compatible avec le comptable, comme en norvégien ou en danois ; voir aussi (Müller 2013).)

Quant aux contextes syntagmatiques cumulatifs, distinguant massif et comptable par le biais du nombre grammatical, ceux du néerlandais sont comparables à ceux du français : *veel* 'beaucoup de', *weinig* 'peu de', *hoeveel* 'combien', etc. (*veel auto's*, *veel boter* 'beaucoup de voitures, beaucoup de beurre'). Notons qu'à ce niveau, le néerlandais est similaire au français, et se distingue de l'anglais qui a pour équivalents sémantiques à beaucoup de nos déterminants cumulatifs des déterminants spécialisés, comme *much*, *little* ou *less* pour un emploi massif, et *many*, *few* et *fewer* pour un emploi comptable, respectivement avec le sens de 'beaucoup de', 'peu de' et 'moins de' (l'opposition entre *much* et *many* est d'ailleurs un critère cité systématiquement dans la littérature anglo-saxonne pour distinguer massif et comptable, voir par exemple (Allan 1980, Gillon 1999)). Un autre contexte cumulatif en néerlandais, mentionné par (Le Bruyn *et al.* 2011, ms.), est *vol* 'plein de' : *de parking stond vol auto's* 'le parking était plein d'autos', *het tafelkleed hing vol boter* 'la nappe était toute recouverte de beurre'.

Quant à la combinaison avec les termes de mesure ((Grinevald 2004), voir *supra*), elle est un peu particulière, car là où le français ou l'anglais utilisent une construction prépositionnelle, le néerlandais préfère bien souvent une construction partitive directe, c'est-à-dire la juxtaposition de deux items nominaux entretenant une relation partitive, sans intervention d'une préposition intermédiaire (Vos 1993) : *een glas water* 'un verre d'eau', *een doosje lucifers* 'une boîte d'allumettes', etc. Notons que toutes les constructions partitives directes ne doivent pas être considérées comme discriminantes entre massif et comptables : un sous-groupe d'entre elles, construites avec ce que (Vos 1993) appelle un *Quantifier Noun* 'Nom Quantificateur' ne s'accommodent que du comptable pluriel (*een aantal auto's* 'certaines voitures, litt. un nombre de voitures') – jamais du massif. La deuxième catégorie de (Vos 1993), celle construite avec un *Container Noun* 'Nom de Contenant', répond bien quant à elle à notre critère : selon l'usage au singulier ou au pluriel du second nom, la construction relèvera soit du massif (*een potje boter* 'un petit pot de beurre'), soit du comptable (*een zak knikkers* 'un sac de billes'). (Si pour ce dernier exemple, nous n'utilisons plus notre fameuse *voiture*, c'est seulement parce que nous n'imaginons pas de contenant suffisamment grand pour en contenir plusieurs...) Notons par ailleurs que le terme de 'nom de contenant' reste un peu limitatif : il s'agit bien d'inclure ici tous les 'termes de mesure' de (Grinevald 2004) : mesures conventionnelles (*kilo* 'kilo', *liter* 'litre'), contenants (*glas* 'verre', *doos* 'boîte', *pot* 'pot'), formes (*sneedje* 'tranche') et arrangements (*hoop*, *stapel* 'tas, pile').



Pour finir, le néerlandais possède également certains déterminants qui restent neutres par rapport à l'opposition massif-comptable. Comme pour le français, il s'agit essentiellement des articles et déterminants définis singuliers : *de-het* 'le-la', *die-deze-dit-dat* 'ce-cette', *mijn-jouw-zijn* 'mon-ton-son', etc.

### 1.3 Le lien entre référence massive et comptable, et réalité

*Epinards, par la force de l'usage et sans raison valable, est toujours employé au pluriel, alors que Oseille qui est un légume de même nature s'écrit toujours au singulier. On écrit et on dit : des épinards; de l'oseille. Aucune raison autre que l'usage ne justifie pareille anomalie. Le singulier, qui s'applique fort bien à Oseille, devrait être la règle dans les deux cas. (Escoffier 1922)*

L'établissement des critères morphosyntaxiques du massif et du comptable, c'est-à-dire de la nature massive, comptable, cumulative ou neutre des déterminants et autres contextes syntagmatiques d'une langue, est une opération plutôt consensuelle, qui n'a jamais fait l'objet de très grands débats (Gillon 1999: 51). (La difficulté réside plutôt dans la conséquence de cette distinction : ne s'applique-t-elle qu'aux déterminants eux-mêmes, ou doit-elle permettre d'identifier, par extension, des propriétés lexicales ? Nous traiterons de cette question plus loin.) Les choses se corsent, par contre, lorsque l'on s'interroge sur le « sémantisme » du comptable et, plus encore, du massif. C'est l'objet du point présent, et du suivant.

Une première question qui a été posée est celle du lien qu'entretiennent le massif et le comptable, avec d'éventuelles propriétés réelles, ontologiques. C'est la question qui est posée explicitement dans (Nicolas 2002a: chapitre 2), qui rappelle les différentes positions sur le sujet. Ainsi, si l'on considère des exemples prototypiques de noms massifs et de noms comptables, tels que *eau*, *sable* et *or* d'une part, et *arbre*, *chat* et *table* de l'autre (nous reprenons ici les exemples de (Nicolas 2002a) lui-même), l'on ne peut qu'observer que les référents des premiers dénotent des substances que l'on pourrait dire 'homogènes', tandis que les seconds ont des référents 'hétérogènes' (quant aux tenants et aboutissants de cette distinction, traduite en termes sémantiques, nous renvoyons au point suivant). L'on pourrait dès lors supposer que c'est précisément cette distinction ontologique, entre substances homogènes et individus hétérogènes, qui est codée par l'opposition massif-comptable. Or, pareille vision 'objectiviste extrême' doit bien vite être abandonnée. A ce sujet, au moins quatre arguments sont avancés systématiquement, parfois seuls, parfois ensemble (cf. (Palmer 1971, Mufwene 1984, Gillon 1999, Nicolas 2002a, Joosten 2003)) :

- 1) Une même réalité peut être exprimée au travers de la syntaxe massive dans une langue, et comptable dans une autre. C'est le cas célèbre de *cheveux*, comptable en français, contre *hair* et *haar*, massifs en anglais et en néerlandais, de même que la fameuse opposition *meuble(s)*, comptable en français, contre *furniture*, massif sans contrepartie comptable en anglais.
- 2) Au sein d'une même langue, une même réalité peut être dénotée par diverses expressions, dont l'une pourra être massive et l'autre comptable (*meubles* contre *mobilier* ; *feuilles* contre *feuillage*).
- 3) Alors que les propriétés ontologiques sont supposées stables, certains noms s'utilisent aussi facilement en syntaxe massive qu'en syntaxe comptable (*une corde-de la corde*, *un œuf-de l'œuf*, *des chocolats-du chocolat*).
- 4) Certains noms à usage habituellement massif dénotent des référents qui sont très clairement non pas des substances homogènes, mais des individus hétérogènes (*du mobilier*, *de la marmaille*, *de la pharmacie*, *de la faune*, *de la marchandise*, etc.).

Que nos exemples du quatrième point répondent généralement à un terme comptable synonymique (*mobilier-meubles*, *marmaille-enfants*, *pharmacie-médicaments*), pourrait nous mettre sur la voie d'une version modérée de la position 'objectiviste' : que la syntaxe massive et comptable ne correspond pas directement à une opposition ontologique, mais plutôt à une manière de 'conceptualiser' des référents. L'idée derrière cette position est que ce n'est pas le monde réel et ses propriétés qui nourrissent la sémantique, mais bien notre perception et interprétation de ceux-ci – qui impliquent bien souvent une certaine 'réorganisation du monde', dans les termes de (Martin 1989: 40) Nous appellerons cela la vision 'conceptualiste'. Ainsi, l'homogénéité intrinsèque de référents massifs prototypiques, tels que *boue*, *eau* ou *fer*, servirait de modèle pour un schéma cognitif appliqué à son tour à des référents hétérogènes, mais présentés néanmoins de manière homogène au travers de la langue. Pour (Whorf 1956), les noms comptables (qu'il appelle 'individuels') dénotent des corps que l'on conçoit linguistiquement comme ayant des limites définies, tandis que les noms massifs dénotent des continuums sans limites intrinsèques. C'est là la position largement récupérée par la tradition linguistique cognitive (Langacker 1987, Taylor 2002). De manière comparable, certains ont exprimé cette idée en termes d'individualisation et de non-individualisation – la conceptualisation exempte d'individus étant alors donnée comme définitoire pour le massif (Quine 1960, Wierzbicka 1988a, Wisniewski *et al.* 1996). Notons que (Nicolas 2002a), quant à lui, rejette cette position 'objectiviste modérée', car elle lui semble non conciliable avec l'existence de noms à usage massif dénotant des collections d'objets (tels que *mobilier*). (Nous reviendrons néanmoins sur cette question de la dénotation homogène et de ses critères d'identification sémantique, y compris avec les noms de collections, dans le point suivant.)

Ainsi, la version forte ou ‘extrême’ de la position ontologique n’est pas défendable, et il faut au moins lui préférer une version modérée, en termes de conceptualisation. Or, dès lors que la réalité n’est plus notre point d’ancrage, mais bien une certaine vision de celle-ci, se pose aussitôt la question subsidiaire de la motivation : l’application de la syntaxe massive ou comptable à un certain référent est-elle motivée, voire prédictible ou, au contraire, totalement arbitraire ?

Certains chercheurs, en particulier par le passé, ont plaidé pour un arbitraire total, dans la plus pure tradition bloomfieldienne, qui prône une séparation totale entre syntaxe et signification (voir (Bloomfield 1933)). Ainsi, si (Whorf 1956) donne une définition précise de ce que représente, pour lui, le sens du massif et du comptable (comme nous venons de le dire), il s’empresse d’ajouter que cette distinction est appliquée de manière largement aléatoire aux items du lexique car, dit-il :

*Rather few natural occurrences present themselves as unbounded extents; ‘air’ of course, and often ‘water, rain, snow, sand, rock, dirt, grass’. We do not encounter ‘butter, meat, cloth, iron, glass’, or most ‘materials’ in such kind of manifestation, but in bodies small or large with definite outlines. The distinction is somewhat forced upon our description of events by an unavoidable pattern of language. (Whorf 1956: 141)*

‘Relativement peu d’occurrences naturelles se présentent comme des étendues non limitées ; ‘air’, bien entendu, et bien souvent ‘eau, pluie, neige, sable, pierre, boue, herbe’. Nous ne rencontrons pas [cependant] ‘beurre, viande, tissu, fer, verre’ ni la plupart des ‘matériaux’ dans pareilles manifestations, mais bien sous forme de corps petits ou grands, aux contours définis. La distinction est quelque peu imposée à notre description des événements, par un schéma linguistique inévitable.’ (traduction personnelle)

Dans le même ordre d’idée, (Gleason 1969: 224) observe que toute une famille de noms désignant en anglais des types de pantalons (*pants, shorts, slacks, etc.*) sont pluriels, mais se refuse à expliquer ce schéma par quelque justification sémantique. De même, (Palmer 1971) et (Ware 1979) signalent respectivement que les distinctions grammaticales ne sont pas supposées avoir de correspondants sémantiques directs, et que la variation entre emplois massifs et comptables (par exemple, pour désigner une même réalité : *leaves* ‘feuilles’ vs. *foliage* ‘feuillage’) est telle qu’elle ne peut être expliquée – autrement que par le caractère arbitraire de la grammaire. Pour finir, bien plus récemment, Pelletier a adopté une approche similaire, se déclarant sceptique quant au lien entre l’opposition massif-comptable en langue, et quelque distinction ‘ontologique ou métaphysique’ que ce soit (Pelletier 2012: 26).

A ce stade, il nous apparaît cependant comme évident que la vision en termes d’un arbitraire total (illustrée notamment par la citation toute profane que nous avons placée en exergue du présent point) est tout aussi peu réaliste que la vision ontologique extrême. Nous en prendrons pour preuve 1) que l’on peut bel et bien faire certaines

généralisations ontologiques limitées concernant l'expression par le massif ou le comptable de certains référents (par exemple, tous les noms désignant des matériaux ou des liquides sont à usage massif : *de la terre, du bois, du béton ; de l'eau, du jus, de l'huile* ; voir aussi (Nicolas 2002a: note 33)) ) et que, dans une perspective plus globale, l'étude récente de langues présentant non pas deux classes de noms, mais trois (gallois), voire quatre (dagaare, cf. (Grimm 2012)), a mené à établir un classement des entités concrètes (sur une « échelle d'individualisation ») permettant de prédire en partie l'organisation morphosyntaxique, en termes de massif-comptable, des noms devant les désigner (voir les travaux de Grimm), 2) que malgré de claires différences inter-linguistiques, il existe aussi entre les langues des similarités indéniables quant au statut massif ou comptable d'équivalents de traduction (ainsi, les généralisations ontologiques limitées mentionnées ci-avant s'appliquent-elles aussi au néerlandais ou à l'anglais ; plus largement, l'on observe que les langues qui possèdent une opposition entre massif et comptable s'accordent généralement sur les prototypes de chacune des catégories, comme le rappelle (Landman 2011: 3)), 3) que malgré l'exception notable des noms de collections, l'association (intuitive) entre syntaxe comptable et individus d'une part, et syntaxe massive et substance d'autre part, n'est pas sans fondement. Il s'avère d'ailleurs, comme l'ont démontré (Kiss *et al.* 2014) pour l'anglais, que les noms de collections (qu'ils appellent *object-mass nouns*, à l'instar de (Barner et Snedeker 2005) ou (Bale et Barner 2012)), toujours évoqués comme contre-exemples sur ce point, sont très fortement minoritaires dans le lexique, dont ils ne constitueraient que 0,30% (une conclusion basée sur l'examen d'un total d'environ 14 000 items lexicaux).

Or, la position 'arbitraire' connaît son extrême inverse en l'ouvrage de (Wierzbicka 1988a), qui se pose en fervente opposante de la syntaxe asémantique de Bloomfield. Elle critique par exemple la position de (Whorf 1956), en disant que sa conclusion, erronée selon elle, n'est qu'un avatar de sa propre définition : pour Wierzbicka, le trait définitoire du massif n'est pas l'absence de limites intrinsèques, comme le prétend Whorf (effectivement, toute occurrence réelle d'une substance sera toujours limitée, au moins dans l'espace, voir aussi (Kleiber 2011), cf. *infra*), mais l'absence – pour une raison ou une autre – d'individus « clairement perceptibles » au niveau des entités dénotées. Elle étoffe son approche d'une analyse hautement raffinée du lexique de l'anglais, dans lequel elle distingue 14 classes formelles (définies donc sur la base de propriétés morphosyntaxiques), toutes associées – de manière totalement motivée, selon elle – à une certaine « signification de classe » (*class meaning*) (Wierzbicka 1988b: 555-560).

Or, tout attrayante que soit l'idée d'une justification sémantique systématique et totale de toute distinction linguistique formelle, la position de (Wierzbicka 1988b), en étant aussi extrême que celle qu'elle contredit, présente également le flanc à la critique. Ainsi, force sera de constater que les descriptions sémantiques proposées sont *post-hoc*, et n'ont pas de réel pouvoir de prédiction (tout en étant difficilement falsifiable). Ainsi, si certaines explications sont convaincantes (la distinction entre *rice* 'riz', massif, et *peas*

'petits pois', comptable, découlerait du mode de consommation de ces deux aliments : une fois cuits, dans notre assiette, les grains de riz sont généralement agglomérés sous forme de bouillie, contrairement aux petits pois, toujours bien distinguables ; *contra* (Joosten 2003) qui donne précisément cet exemple pour démontrer les faiblesses de l'analyse de (Wierzbicka 1988b)), d'autres ne résistent pas une seconde au contrôle inter-linguistique : les arguments avancés pour justifier le caractère strictement massif de *bread* ou *hair*, ou le caractère pluriel de *chives* tombent à l'eau aussitôt que l'on s'aperçoit qu'ils s'appliquent tout autant aux référents des équivalents français, *pain*, *cheveux*, et *ciboulette*, qui pourtant n'ont pas les mêmes propriétés morphosyntaxiques (*pain* étant flexible, *cheveux* comptable et *ciboulette* massif). L'on devra donc conclure que si plaider l'arbitraire total de l'opposition massif-comptable n'est pas judicieux, l'alternative d'une « prévisibilité *post-hoc* » est tout aussi peu adéquate.

Ainsi, dire que l'opposition massif-comptable n'est pas arbitraire ne signifie pas qu'elle doit être entièrement prévisible. En cela, nous donnerons raison à (Desagulier 2012: 207), lorsqu'il dit qu'à sa connaissance, « aucun modèle ontologique de l'opposition massif-comptable basé sur la physique ou la logique ne peut prédire de manière satisfaisante le caractère morphosyntaxique des noms » (traduction personnelle, légèrement paraphrasée). Néanmoins, comme le rappelle (Lakoff 1987), si la langue n'est ni totalement arbitraire, ni totalement prévisible, elle est généralement motivée. C'est ce que la 'vision conceptualiste' mentionnée ci-dessus, interprétée dans le cadre d'une sémantique référentielle (voir les travaux de Kleiber, par exemple (Kleiber 1994, 1999)), permet d'expliquer. Ainsi, le sens d'une expression linguistique se définit, très largement, comme son référent, c'est-à-dire l'entité perceptuelle et/ou conceptuelle associée à l'élément du monde extralinguistique qu'elle désigne. (Cet élément réel est parfois lui-même appelé 'référent' (Kleiber 1999: 16), mais il nous semble préférable de définir le référent en termes de perception ou de conceptualisation, pour nous démarquer de toute approche 'ontologique' ou 'objectiviste extrême'.) Or, selon qu'un nom s'emploiera comme massif ou comme comptable, la manière dont nous concevrons son référent sera différente – une variation liée à divers facteurs de motivation (de nature ontologique, fonctionnelle, contextuelle : nous reparlerons de ces facteurs plus tard). La question qui se pose alors est celle de la nature de cette « différence de conceptualisation ». C'est ce que nous examinons au point suivant.

## 1.4 La dénotation massive et comptable

Aussi vrai qu'il existe deux types de SN distingués par leur caractère morphosyntaxique massif ou comptable, de même il est relativement consensuel (si l'on exclut les approches bloomfieldiennes, aujourd'hui largement dépassées) de prétendre qu'à ces deux types de caractères correspondent deux manières distinctes d'appréhender les référents. Décrire cette distinction en termes sémantiques ne s'est cependant pas avérée une mince affaire. Nous avons déjà évoqué dans le point précédent quelques propositions en termes de présence ou d'absence d'individualisation ou de limites intrinsèques, mais nous verrons ici que les options suggérées sont plus nombreuses encore.

Les intuitions sur le sujet sont pourtant relativement unanimes : au comptable correspondraient des référents individualisés, et au massif des référents imaginés comme des substances, sans individus. Dans le cas du domaine comptable, cette proposition ne pose pas de réel problème : effectivement, tous les noms incontestablement comptables désignent des individus. Or, c'est pour le massif que le bât blesse : le défi majeur, dans la traduction en termes proprement sémantiques de notre distinction, est qu'elle doit pouvoir s'appliquer à tous les noms à usage massifs. Cependant, un groupe de ces noms, déjà mentionné ci-avant, semble plus réticent que les autres à se laisser décrire d'un point de vue sémantique (car leurs référents sont, ontologiquement parlant, les plus éloignés des substances) – il s'agit des noms à emploi massif désignant, justement, des individus.

Disons-en quelques mots en guise d'introduction au point présent. A vrai dire, l'on peut distinguer en français au moins trois types de cas où un SN massif a pour référents un ensemble d'individus. Considérons quelques illustrations :

- (22) Kim aime manger du riz. J'ai fait mettre du gravier dans l'allée de mon garage.
- (23) Jean vend de la vieille ferraille. Marie a acheté du mobilier neuf.
- (24) Cette année, il y a du touriste à Paris !

Au sens strict, les noms comme *riz* ou *gravier* dans (22) sont bel et bien des noms à usage massif désignant des 'objets' distincts – en l'occurrence, des grains de riz, et des petits cailloux. Ils ne rentrent pourtant pas dans la classe des termes que (Nicolas 2002a) appelle 'noms de collection', et ne sont généralement pas considérés comme réellement problématiques pour la définition sémantique du massif. Ce qui les distingue des cas en (23), c'est que les individus qu'ils désignent sont, à l'échelle de notre perception humaine, relativement petits et identiques ((Kleiber 1997: 333-334) parle d'individus 'homogènes'), à « faible saillance référentielle » (Joosten 2003) – ce qui fera que l'on s'accorde généralement à les associer aux noms de substances, moyennant cette précision. (A vrai dire, l'on fait de même pour les massifs même les plus typiques,

puisque la science sait aujourd'hui que même l'eau est composée de molécules et d'atomes, voir ci-dessous.) Ainsi, ce n'est pas ce genre de cas qui pose fondamentalement problème.

Les Romains s'empoignent cependant lorsqu'il est question des noms tels que ceux en (23) – qui s'emploient en syntaxe massive, mais désignent néanmoins des objets qui ne sont ni identiques, ni même suffisamment petits que pour perdre leur caractère discernable dans un contexte perceptuel normal (contrairement aux molécules d'eau, ou même aux grains de riz). Ceux-ci se sont vu désigner par une terminologie très variée, que nous évoquons ici sans ambition d'exhaustivité : 'non-taxonomic supercategories' (Wierzbicka 1988b), 'collective mass nouns' (Krifka 1991, Doetjes 2012), 'noms de masse à référence hétérogène' (Wiederspiel 1992, Kleiber 1997), 'count-mass nouns' (Doetjes 1997), 'mass superordinates' (Wisniewski *et al.* 1996), 'noms massifs collectifs' ou 'noms massifs dénotant des collections' (Nicolas 2002a, Lammert 2010), 'aggregate terms' (Huddleston et Pullum 2002), 'unsorted stuff nouns' (Müller et Oliveira 2004), 'object-mass nouns' (Barner et Snedeker 2005, Bale et Barner 2012), 'fake mass nouns' (Chierchia 2010: 110), 'atomic mass nouns' (Rothstein 2010, Pires de Oliveira et Rothstein 2011), 'neat mass nouns' (Landman 2011). (Notons que certaines de ces terminologies, en particulier celles se contentant d'évoquer la notion d'objet ou d'atome, bien que généralement utilisées pour évoquer les cas de (23), devraient s'appliquer aussi à ceux de (22). Relevons aussi que le terme *collective noun* 'nom collectif' est peut-être mal choisi, car l'on a appelé aussi comme ça les noms à emploi comptable tels que *essaim*, *armée*, *meute*, *groupe*, etc., voir (Bloom et Kelemen 1995). Or, si certains considèrent bel et bien les noms tels que *mobilier* comme des collectifs (voir par exemple (Dölling 1991, Lammert 2010)), cette position n'est pas entièrement consensuelle (voir (Joosten 2006), qui parle d'« agrégats » ; ou Fasciolo, communication personnelle : « *une table c'est une partie et non un membre de mobilier* »).) Notons que dans la tradition anglo-saxonne, l'on pourra entendre également (de manière peut-être plus informelle) l'appellation de 'furniture-nouns', en référence à *furniture* 'mobilier', le membre possiblement le plus emblématique de la catégorie. Pour ce qui nous concerne, dans ce qui suit, nous reprendrons le terme de 'nom (massif) de collection', en référence à la terminologie de (Nicolas 2002a) tout en évitant l'appellation non neutre de 'nom collectif'.

Pour finir, l'on a encore le cas de (24). Ici aussi, il est évident que le SN massif, *du touriste*, ne désigne pas une substance, mais fait bien référence à un ensemble d'individus. Une caractéristique majeure distingue cependant cet exemple des deux précédents : alors que l'emploi massif semble naturel pour *riz*, *gravier*, *mobilier* ou *ferraille*, il semblerait bien que pour *touriste*, il s'agisse d'un emploi quelque peu 'hardi'. Il s'agit plus précisément d'un exemple de ce que (Galmiche 1989) a appelé le mécanisme du « Multiplicateur Universel ». Nous y reviendrons lorsque nous traiterons de la flexibilité de l'opposition massif-comptable. Néanmoins, à ce stade, toujours est-il que

ce type d'usage du massif existe, et qu'il faudra pouvoir en rendre compte dans notre examen sémantique.

L'on voit donc de quel ordre est la principale difficulté pour ce qui est de définir le sens du comptable et, plus encore, du massif. Dans ce qui suit, nous explorerons les tentatives et propositions faites par nos prédécesseurs, en les groupant selon deux types d'approche. D'une part, nous évoquerons la question de la nature de la dénotation massive et comptable, en termes de propriétés et de structure référentielles, liée à diverses terminologies et mise en évidence par différents tests. Ensuite, nous nous aventurerons sur un terrain plus formaliste, en abordant la question du massif-comptable par la lunette de la quantification, et des modèles formels proposés pour expliquer les propriétés référentielles identifiées.

### 1.4.1 La dénotation homogène : critères sémantiques

Au cœur même de sa terminologie asymétrique (comme la décrit (Joosten 2003)) réside la difficulté traditionnelle de la description sémantique du massif et du comptable – où le terme de 'massif' renvoyant à une propriété référentielle (voire ontologique) s'oppose à celui de 'comptable', qui est au mieux morphosyntaxique, et au pire purement pragmatique. (Que l'on ne 'compte' pas les grains de sable sur une plage relève plus d'une question de perspective que d'impossibilité absolue...) L'intuition est néanmoins très forte, et plutôt consensuelle, que la dénotation du comptable et du massif diffèrent dans leur plus profonde nature. On la retrouve au travers de la multitude de terminologies proposées, à un moment où un autre, pour tenter de la capturer (voir (Kleiber 2014, à paraître-a)) : 'discontinu-discret vs. continu-non discret', 'comptable-dénombrable vs. non comptable-non dénombrable', 'hétérogène vs. homogène' (voir (Jackendoff 1991, Kleiber 1997), *i.a.*), 'borné-avec limites vs. non borné-sans limites' (Langacker 1987, Taylor 2002), 'discret vs. compact vs. dense' (proposée par (Culioli 1999)), 'atomique vs. non atomique' (dans le domaine de la quantification, voir (Rothstein 2010) pour une approche critique).

Traditionnellement, deux critères ont été proposés pour décrire les propriétés sémantiques des référents massifs : celui de « référence cumulative », et celui de « référence distributive » (la combinaison des deux étant parfois appelée « référence homogène »). Avant d'explicitier ce dont il s'agit, signalons d'emblée que ces critères, en particulier le second, sont hautement polémiques – non pas qu'ils soient délicats à définir (au contraire, ils ont été décrits de manière plutôt consensuelle par tous leurs commentateurs), mais bien parce que la réussite de leur application dépend entre autre des choix théoriques posés quant à la question du rapport entre référence et réalité (cf. point précédent). En outre, en ce qu'ils impliquent des tests imaginaires d'addition et de prélèvement, certains auteurs ont noté qu'ils n'étaient utiles que pour traiter de la



référence des noms concrets (cf. (Joosten 2003), mais voir néanmoins (Van de Velde 1996) et (Nicolas 2002a: chapitre 2, 2. c.) quant à l'application métaphorique de la notion d'homogénéité aux abstraits). Dans ce qui suit, nous en donnerons les caractéristiques sans présumer de leur pertinence, puis examinerons de manière critique leur utilité.

#### 1.4.1.1 La référence cumulative

Le premier critère suggéré pour décrire le sémantisme du massif est celui de référence « cumulative » : un N réfère de manière cumulative si et seulement si pour deux éléments pouvant être désignés par N, leur somme peut également être désignée par N. Proposé initialement par (Quine 1960: 90-91), il a été repris et commenté abondamment par ses successeurs (voir par exemple (Bunt 1979, Roeper 1983, Moltmann 1997)). Ainsi, si j'ai *de l'eau* dans une tasse, et *de l'eau* dans un verre, la somme de ce qui est contenu dans mes deux récipients, c'est aussi *de l'eau* : *eau* serait ainsi un nom à référence cumulative. D'autre part, si j'ai *un sac* dans ma main gauche, et *un sac* dans ma main droite, la somme de ce que je transporte n'est pas *un sac*, mais bien *deux sacs* (ou *des sacs*) : le nom *sac*, quant à lui, ne réfère donc pas de manière cumulative.

(On voit ici déjà une première difficulté du test : si le critère est formulé en termes de N, son applicabilité – en français – ne peut être contrôlée qu'au travers de SN : dans notre premier exemple, c'est bien le syntagme *de l'eau* qui réfère cumulativement, et non le N *eau*. Le critère a cependant été formulé en premier en anglais, où l'emploi nu d'un nom correspond environ à notre usage avec le partitif (voir aussi ce que nous avons dit sur le néerlandais). Dans cette langue, le critère peut donc aussi sembler se vérifier pour N : *this cup contains water* 'cette tasse contient de l'eau', *this bottle contains water* 'cette bouteille contient de l'eau', *the sum of both is also water* 'la somme des deux est également de l'eau'.)

Dans son applicabilité, le critère de « cumulativité » ne pose pas de problème : l'on peut aisément imaginer deux référents séparés, puis leur somme, et s'interroger intuitivement sur la dénomination de ces différentes entités. Il s'applique aussi sans problème à nos noms de collections : s'il y a *du mobilier* dans la chambre de Marie, et *du mobilier* dans celle de Jean, alors ensemble, les deux chambres contiennent aussi *du mobilier*. Il a néanmoins été remis en doute comme critère de reconnaissance du massif, en ce qu'il ne s'applique pas seulement aux noms à usage massif, mais également aux noms au pluriel (ou, du moins, aux SN pluriels indéfinis). Par exemple, si *des voitures* roulent sur l'autoroute en direction de Gand, et que dans l'autre sens, *des voitures* sont bloqués dans les bouchons en direction de Bruxelles, la somme de ces véhicules occupant la E40, ce sont aussi *des voitures*. Les conséquences à tirer de cette observation varient : si pour (Mufwene 1981), cela signifie que les noms pluriels indéfinis sont en réalité bien des termes massifs, pour (Gillon 1999) c'est la preuve de l'inutilité et de l'inaptitude totale du test. Cette seconde considération nous semble néanmoins abusive :

il n'en reste pas moins vrai que seuls les noms qui ont été appelés massifs peuvent référer de manière cumulative dans leur emploi au singulier.

#### 1.4.1.2 La référence distributive

Le second critère traditionnel évoqué pour décrire la nature de la référence massive est celui de la « distributivité » ou « divisibilité » référentielle : un (S)N réfère de manière distributive si et seulement s'il peut s'appliquer à toute partie de ce à quoi il s'applique. Ainsi, si j'ai *de l'eau* dans une bouteille, et que j'en verse la moitié dans un verre, et l'autre moitié dans un autre, le contenu de chaque verre sera à son tour *de l'eau*. Ce critère a été mis en avant par (Cheng 1973: 286-287), en réponse à Quine, qui lui-même se refusait à le considérer comme définitoire du massif, car « il est des parties d'eau, de sucre ou de mobilier trop petites pour compter comme de l'eau, du sucre ou du mobilier » et qu'en outre, « ce qui est trop petit pour compter comme du mobilier n'est pas trop petit pour compter comme de l'eau ou du sucre » (Quine 1960: 99, traduction personnelle). Cette observation, que (Moltmann 1998) décrit comme « le problème des parties minimales », est celle qui divise jusqu'à ce jour ceux qui se sont intéressés à la question. Autrement que le test de référence cumulative, il est vrai que celui de la distributivité pose un certain nombre de questions délicates, que nous examinons dans ce qui suit.

Premièrement, il y a ce que Quine lui-même observait : si l'on se tourne vers les référents « réels » des N à usage massifs, force est de constater qu'à force de divisions, l'on finira toujours par tomber sur une partie tellement petite que le N ne pourra plus s'y appliquer. Ainsi en est-il très certainement des noms de collections, comme *mobilier* (la moitié d'une chaise, disons son dossier, ce n'est plus *du mobilier*), mais aussi des référents les plus emblématiques du massif (même si on ne les voit pas, on sait qu'au niveau moléculaire, une molécule d'*eau*, c'est encore *de l'eau*, mais que ce SN ne s'applique plus aux atomes d'hydrogène et à l'atome d'oxygène qui la compose). Pour Quine, mais aussi, par exemple, pour (Pelletier 2012), ceci suffit à rendre le critère de référence distributive parfaitement caduc pour décrire le sémantisme du massif.

Néanmoins, cette critique présuppose que l'univers sémantique d'une langue devrait être rigoureusement identique au monde extérieur, ou du moins le refléter exactement – une position apparentée à ce que nous avons dit ci-dessus de la vision « objectiviste extrême » du lien entre massif-comptable et réalité, et intenable pour les mêmes raisons. Voilà pourquoi d'autres auteurs, comme (Bunt 1979) ou (Lønning 1987: 8), ont argumenté qu'il ne fallait pas confondre « faits empiriques » et « faits de langue ». Si l'on sait aujourd'hui que l'eau est constituée d'atomes (qui eux-mêmes ne sont pas *de l'eau*), il s'agit là d'une connaissance scientifique, encyclopédique, qui n'a qu'une incidence indirecte sur le sens que nous mettons derrière le mot *eau*. La manière dont nous percevons son référent, et exprimons cette conception en langue, n'en est pas modifiée

fondamentalement par cette connaissance (d'ailleurs, notons que le caractère grammatical massif de noms comme *eau* sera toujours acquis bien avant que l'on apprenne, à l'école, quoi que ce soit de la composition atomique du monde matériel). Ainsi, l'on pourra répéter (cf. point précédent) que la langue, au travers de l'opposition massif-comptable, reflète non pas des propriétés ontologiques, mais des propriétés référentielles – au sens du référent comme représentation perçue, conceptualisée. Dans pareille vision conceptualiste, l'on doit pouvoir accepter que la langue nous permette d'exprimer *de l'eau* ou *du sable* en faisant conceptuellement abstraction des parties minimales dont sont constitués l'eau et le sable, en les traitant « comme si » leur référence était distributive. (Sur le lien entre sémantique et connaissance du monde, voir aussi notre point 3.2.3 ci-dessous.)

Cependant, il y a d'autres problèmes avec le test de divisibilité. S'il peut s'accommoder, moyennant la correction que nous venons de faire, des référents dont la structure perçue est homogène, il devient plus délicat à appliquer dès que l'on a affaire à des référents plus complexes. L'on trouve ici les questions, sensiblement similaires, du *driehoekjesbehang* 'papier peint à motifs triangulaires' (Landman 2011) et de la *tarte aux fraises* (Kleiber 1997). Le premier exemple est évoqué par (Landman 2011) pour illustrer que, selon lui, le test de distributivité est inutile. Si je prélève, sur *du papier peint à motifs triangulaires*, un morceau (quelle que soit sa taille) qui ne comporterait aucun triangle, ce que j'obtiens n'est plus *du papier peint à motifs triangulaires* (mais au mieux, simplement *du papier peint*). De manière similaire, si je prélève un morceau qui contienne exactement un triangle, celui-ci pourrait toujours bien être désigné par le SN massif *du papier peint à motifs triangulaire* (ce qui, en néerlandais, se traduit donc pas un seul mot : *driehoekjesbehang*), mais ne pourrait plus lui-même être divisé en deux parties pouvant prétendre toutes deux à ce statut. Il n'est donc pas même nécessaire de descendre jusqu'au niveau atomique pour mettre notre test en difficulté. Pour Landman, c'est la preuve que le critère de divisibilité n'est pas pertinent.

Le second exemple, celui de la *tarte aux fraises*, est sensiblement similaire, et est bien connu de la littérature francophone : la fraise que je prélève sur *de la tarte aux fraises*, est-elle elle-même encore *de la tarte aux fraises*? (Kleiber 1997) mentionne cette question non pas pour remettre en doute le test de divisibilité en lui-même, mais pour rappeler que son application est conditionnée par des limites pratiques. Pour Kleiber, ces limites sont du même ordre que celles évoquées pour les molécules d'*eau*, et la condition à respecter pour l'application du test de distributivité est que l'on considère, dans l'exercice de division, non pas toute partie d'un tout, mais uniquement les parties de même nature :

En soustrayant une partie de N à une partie continue non délimitée de N, je ne puis enlever qu'une partie qui est elle-même une partie continue non délimitée de N. Il faut évidemment que ce soit une partie de N, si je fais éclater cette partie, la division homogène connaît l'échec, comme avec la fraise prélevée sur la tarte de

fraises ou encore l'hydrogène prélevée sur H<sub>2</sub>O qui, évidemment, n'est plus de l'eau, ou encore le morceau du manche du marteau qui n'est plus de la quincaillerie. (Kleiber 1997: 335)

Pour Kleiber, les noms massifs réfèrent donc bien de manière distributive, ce que l'on peut vérifier par le test du prélèvement, à condition de respecter la condition précitée. Nicolas, quant à lui, formule une condition similaire, lorsqu'il précise que le test de divisibilité référentielle n'est pas supposé s'appliquer à toute partie méréologique d'un tout, mais bien à « toute partie relative au nom N, ou N-partie » (Nicolas 2002a: chapitre 3, 1). Ces précisions résolvent du même coup la question de la référence comptable, qui elle n'est pas censée être distributive. Ainsi, de la même manière qu'en prélevant une fraise sur *de la tarte aux fraises* il me reste toujours *de la tarte aux fraises*, si j'arrache un essuie-glace à *une voiture*, il me restera toujours *une voiture* (fût-elle abîmée). Néanmoins, en suivant les conditions de Kleiber et Nicolas, ceci n'est pas une manière valide d'appliquer le test de distributivité : la portion prélevée doit faire partie elle-même de N. Or, si la tarte aux fraises peut être, par exemple, coupée en deux morceaux qui seront tous deux *de la tarte aux fraises*, pour notre voiture, il ne sera jamais possible de prélever un morceau auquel s'appliquerait le prédicat *une voiture*, tout en gardant pour reste quelque chose qui soit également *une voiture*.

Ainsi, moyennant quelques précisions quant à son application, le test de distributivité se révèle utile pour aider à comprendre la nature de la dénotation massive. Cependant, il est une fois de plus mis à mal dès lors que l'on considère que le caractère de divisibilité de la référence est également vrai de certains groupes de noms à usage comptable, décrits par (Langacker 1990: 70) ou encore (Kleiber 1997: 328-332) comme ayant une « structuration interne homogène ». Il s'agit de noms tels que : *chose* (Kleiber 1987) ; *morceau, tas, partie ; moment, instant, durée ; étendue ; tache, bip, son, flash ; corde, clôture, caillou ; etc.* Effectivement, si je prélève quelque chose à *un morceau*, tant ce que j'obtiens que ce qui me reste seront à leur tour *un morceau* ; et la partie d'*un son* ou d'*un bip*, c'est aussi *un son* ou *un bip* (fussent-ils plus courts) ; de même, si l'on coupe *une corde* ou *un caillou* en deux, n'obtient-on pas deux fois *une corde* ou deux fois *un caillou* ? Sans entrer dans les détails de l'application du test à ces différents référents (la division d'*une tache* étant peut-être moins évidente que celle d'*une clôture*), force est de constater qu'ils présentent bel et bien, à l'instar des référents massifs, la propriété de référence distributive. Pour certains auteurs, comme par exemple (Gillon 1999), cette particularité est la preuve que la référence distributive est une propriété non pertinente – car elle ne permet pas de distinguer de manière définitive le massif du comptable. Cependant, à nouveau, pareille généralisation est abusive : elle appelle seulement à une précision. Ainsi, comme le précisent (Langacker 1990) ou (Kleiber 1997), le test de distributivité (ou de cumulativité, du reste) ne définit pas la nature de la référence massive, mais seulement une de ses facettes : celle de « structuration homogène ». Or, cette homogénéité n'est pas exclusive au massif : certains référents comptables, comme nous

venons de le voir, peuvent partager cette caractéristique. Le massif, quant à lui, se définit dès lors par la combinaison de deux propriétés référentielles : la structuration homogène, et l'absence de limites, ou de bornage intrinsèque. C'est principalement vers ce second aspect du sémantisme référentiel massif que nous nous tournons dans le point suivant.

### 1.4.1.3 La question des limites et des occurrences

Les propriétés de référence cumulative et distributive, associées (avec les réserves que nous venons de décrire) au massif, ne sont pas encore le fin mot de notre histoire. Si en combinaison, elles définissent la propriété de « référence homogène » (Bunt 1979, Ter Meulen 1981: 123, Bunt 1985, Moltmann 1997) ou de « structuration homogène » (Langacker 1990, Kleiber 1997), celle-ci, comme nous venons de l'évoquer, ne suffit pas à elle seule à définir le sémantisme du massif, car elle est partagée par certains référents comptables. Or, notre intuition reste intacte qu'il y a bel et bien une différence sémantique entre le massif et le comptable – y compris entre expressions massives et expressions comptables à référence homogène. Considérons la situation suivante (adaptée très librement des exemples donnés par (Kleiber 1997: 331)) : en dégustant un hamburger, j'ai fait couler *de la moutarde* sur mes vêtements, et il y a à présent *une tache* de sauce sur ma chemise, et une autre sur mon pantalon. Tant *moutarde* que *tache* ont des référents structurés de manière homogène. Cependant, si je peux dans cette situation parler de *la moutarde sur mes vêtements*, je ne pourrai pas parler de *la tache* (je devrai dire *les taches*) pour identifier les témoins de ma maladresse.

Cette distinction s'explique par une seconde propriété des référents massifs (qui les distingue bel et bien des comptables) : celle d'être « dépourvus de limites intrinsèques » – décrite abondamment dans la tradition cognitive (Langacker 1987, Taylor 2002). Cette propriété se définit le mieux par défaut, c'est-à-dire par opposition au comptable, dont il est dit qu'il fournit, par le sémantisme nominal même, 'un principe de distinction des individus' (Strawson 1959), ou un 'principe individuant intrinsèque' (Galmiche et Kleiber 1996: 35). Autrement dit, si le référent d'une expression comptable est limité dans un certain domaine (par exemple, dans l'espace), cette limitation est intrinsèque à son sémantisme : elle fait partie du sens de l'expression. Cela signifie que ce qui compte comme une occurrence de *voiture* est déterminé par le nom *voiture* lui-même (et est une question de diverses propriétés, entre autres de forme). Par opposition, le référent d'une expression massive n'est pas, quant à lui, pourvu de pareille limitation intrinsèque : ses limites ne font pas partie de son sens, et n'apparaissent qu'au travers de son occurrence réelle.

Cette idée a parfois été traduite en termes d'une opposition entre « continu » et « discontinu » : les référents comptables, puisqu'ils possèdent des limites intrinsèques, sont nécessairement discontinus, en ce que la catégorisation par le comptable implique

l'identification successive d'occurrences délimitées et distinctes. C'est par exemple en ces termes que se trouve formulée la définition de (Riegel *et al.* 2009: 170) : « Les noms comptables renvoient à des segments discontinus de la réalité (*pilote, avion, orage*) que l'on peut dénombrer (...). Les noms massifs dénotent des substances continues (...). » Cependant, la pertinence de cette terminologie a récemment été remise en question par (Kleiber 2011: 3.2), qui appelle à « remettre les pendules de la continuité à l'heure de la discontinuité en rappelant que toutes les occurrences, massives comme comptables, sont discontinues ou, autrement dit, ont des limites ou ont des bornes. » En effet, si j'ai un verre rempli d'eau devant moi, l'occurrence d'*eau* qui s'y trouve n'est pas moins bornée que celle de *verre*. Ainsi, prétendre que les référents d'expressions massives seraient strictement continus équivaldrait au paradoxe de leur refuser toute forme d'occurrence – ce qui n'est évidemment pas une solution souhaitable. (Kleiber 2011) réexplique dès lors en termes d'occurrences ce que l'on entend par l'absence typiquement massive de bornage intrinsèque : qu'une « situation d'occurrence » donnée ne peut jamais comporter qu'une, et une seule, occurrence d'un référent massif – puisque c'est par cette situation même que l'occurrence est créée. L'on en arrive ainsi à une inversion du vocabulaire habituel (Kleiber 2011: 3.3) : si la structuration interne des référents massifs est homogène, leurs occurrences, elles, varient sans cesse et sont donc « hétérogènes » – à l'inverse des comptables, dont la structuration est plus typiquement hétérogène (sauf pour les exceptions que nous avons évoquées), mais les occurrences homogènes – car répondant à un même principe intrinsèque d'identification.

Le point de vue des occurrences est une nouvelle approche prometteuse, car elle permet de capturer adéquatement la distinction entre référence massive et comptable, y compris pour nos noms de collections à usage massif (*mobilier, pharmacie, etc.*) : si dans un salon se trouvent deux chaises, un sofa et une armoire, j'aurai là précisément une occurrence de *mobilier*. Si j'enlève une des chaises, et que je la remplace par deux tables et un fauteuil, rien ne change : j'aurai toujours dans mon salon une seule et unique occurrence de *mobilier*. Voilà comment (Kleiber 2011) propose de comprendre la référence massive, et son absence de bornage. Pareille approche nous permet, du reste, de faire sens de la différence entre *moutarde* et *tache* évoquée ci-avant : la sauce étalée à deux endroits sur mes vêtements constitue bien deux occurrences de *tache* (expression comptable, aux bornes intrinsèques), mais une seule et unique occurrence de *moutarde* (d'où l'obligation d'une référence au singulier : *la moutarde sur mes vêtements*, et non *\*les moutardes sur mes vêtements*).

Cette caractéristique de la référence massive fournit aussi du même coup une explication quant à l'impossibilité de soumettre les référents massifs au dénombrement : celui-ci implique nécessairement un comptage d'occurrences homogènes (c'est-à-dire suffisamment identiques), dans une situation donnée (si je dis avoir *deux voitures* dans le garage, j'entends bien sûr « deux occurrences de *voiture* ») – une opération non pertinente pour le massif, puisque le compte total d'occurrences

équivaldra toujours strictement à un. (Sauf si un principe distinctif autre que la seule situation d'occurrence force la différenciation, comme une distinction qualitative de type taxonomique : *Je ne bois que deux eaux – la Spa et la Chaudfontaine*, voir plus loin.)

Dans certains cas, les occurrences, distinctes et hétérogènes, de différentes situations d'occurrence peuvent néanmoins se voir évoquées ensemble, sans se trouver combinées au travers d'une nouvelle situation qui les engloberait. C'est là, selon nous, une belle preuve de la validité de l'approche occurrence, qui se voit illustrée dans (25) – adapté d'un exemple de (Landman 2011: 4) :

(25) Tant le café de ma tasse que le café de mon thermos sont froids, à présent.

Ici, deux occurrences de *café* se voient identifiées séparément, et combinées à la manière de deux objets distincts, ce qui est reflété par la forme verbale *sont* et l'adjectif *froids*, tous deux au pluriel. Notons que (Landman 2011), qui nous inspire cette remarque, ne s'inscrit néanmoins pas dans la même lignée que (Kleiber 2011) : pour lui, si l'on ne peut dénombrer les référents massifs, ce n'est pas parce que leurs occurrences sont toujours strictement uniques, mais parce qu'ils sont au contraire composés d'une infinité de « blocs de construction » se chevauchant (toute portion imaginable d'une occurrence pouvant prétendre au statut de pareil bloc, ce qui donne lieu à une opulence infinie ne pouvant être comptée) – voir aussi ci-après.

Dans ce qui précède, nous avons donc évoqué deux propriétés sémantiques cruciales pour la description de la référence massive : celle de structuration homogène, et celle de non bornage intrinsèque (traduit en termes d'occurrences). Avant de clôturer le présent point, il ne nous reste plus qu'à dire un mot du rapport entre ces deux propriétés. Comme nous l'avons vu, c'est bien la seconde qui est réellement définitoire pour le massif. La structuration homogène des référents massifs, quant à elle, est cependant une conséquence de cette absence de limites intrinsèques (ce qui en fait, *nota bene*, une condition nécessaire bien que non suffisante du massif). Ce lien de causalité est mentionné, et expliqué par (Kleiber 1997) : si un tout présente une structuration interne hétérogène, cela signifie qu'il est constitué de parties distinctes, et donc délimitées, qui elles-mêmes sont des touts, et qui font hériter leurs limites à l'ensemble qu'elles forment. Or, un référent massif n'ayant pas de limites intrinsèques, sa structuration ne peut jamais se concevoir que comme homogène.

A notre sens, ce rapport est crucial pour comprendre que des ensembles d'objets distincts (comme des grains de riz, des meubles, ou même parfois des touristes, cf. (24)) puissent être conçus, linguistiquement, en faisant abstraction du niveau des individus : par le biais de la syntaxe massive (*du riz, du mobilier, du touriste*), la langue se refuse de fournir un principe de distinction intrinsèque d'occurrences, et signale ainsi le caractère non borné de ces référents, ce qui oblige qu'on leur prête conceptuellement une structuration homogène, c'est-à-dire dépourvue de parties délimitées – et donc d'individus.

## 1.4.2 La quantification massive et comptable

Cette question des individus est aussi au cœur d'une toute autre manière d'aborder la dénotation du massif et du comptable : celle de la quantification, et des modèles formels, algébriques, de la référence. Il ne s'agit pas là d'approches déconnectées de ce que nous avons dit précédemment : généralement, les sémanticiens plus formalistes que nous mentionnerons ici évoquent bel et bien (parfois de manière critique), les critères de référence cumulative et distributive que nous avons présentés. Ce qui les caractérise, néanmoins, c'est que leurs études ne visent plus tant à décrire le sémantisme du comptable et du massif dans tous ses détails et subtilités, mais plutôt à fournir un modèle formel de la dénotation nominale (sous forme de structure mathématique), souvent sur un fond de sémantique vériconditionnelle, visant essentiellement à expliquer les propriétés de quantification du massif et du comptable – en ce compris, pour ceux qui les acceptent, la cumulativité et la distributivité. Entrer dans le détail de ces formalismes dépasserait grandement tant l'objectif du présent ouvrage que notre domaine d'expertise. Cependant, il nous semble important d'évoquer les principales tendances en la matière.

Pour ce qui est du comptable, comme le soulignent (Barner et Snedeker 2005) ou (Bale et Barner 2009) dans leurs typologies des modélisations, la situation est plutôt consensuelle : une expression comptable dénoterait un « ensemble d'individus » – une hypothèse qui remonte à (Quine 1960). Ainsi, dans une perspective où la valeur sémantique d'un énoncé doit correspondre à ses conditions de vérité, l'on pourra affirmer avec raison *Ceci est une pomme* si l'objet désigné est un individu appartenant à l'ensemble dénoté par le nom *pomme*. Dans la terminologie mathématique (utilisée par (Link 1983), puis abondamment récupérée), l'on dira que toute expression comptable dénote un « sup-demi-treillis avec des parties minimales » (*join-semi-lattice with minimal parts*) : un ensemble partiellement ordonné (en l'occurrence, par la relation d'inclusion) qui, pour tout sous-ensemble fini non vide, possède un supremum (*join*).

Concernant le massif, comme l'on s'y attend, la situation est plus partagée. L'intuition de (Quine 1960) que, contrairement à la dénotation comptable, la dénotation massive serait quant à elle exempte d'individus, est reprise par (Link 1983), qui définit la dénotation de toute expression massive en termes d'un « sup-demi-treillis sans parties minimales ». De manière similaire, (Roepfer 1983) et (Lønning 1987) proposent un modèle sémantique basé sur une algèbre de Boole (*Boolean algebra*, un treillis distributif borné et complété), permettant de traiter la quantification comptable et massive dans un même cadre – la première supposant l'existence d'individus, la seconde étant caractérisée par une absence d'individualisation (pas de parties minimales).

Avec (Bunt 1985) et (Gillon 1999), l'on s'éloigne un peu de cette distinction claire entre dénotations avec et sans individus. Ainsi, pour (Bunt 1985), si les comptables dénotent des individus, les massifs, quant à eux, auront tantôt une dénotation avec,



tantôt une dénotation sans individus – en fonction des propriétés ontologiques des référents concernés, telles que nous pouvons les observer par une « inspection du monde » (voir (Bale et Barner 2009: 219-220)). C'est ainsi que se distinguent, par exemple, les dénnotations de noms tels qu'*espace* et *mobilier*. (Gillon 1992, 1999) traduira la vision de (Bunt 1985) en disant que là où les dénnotations comptables ont des parties minimales, les dénnotations massives sont, pour leur part, « non spécifiées » (*unspecified*) quant à ce qu'elles comportent de telles parties minimales.

Cette vision sera à son tour récupérée par Barner et ses collègues (Barner et Snedeker 2005, 2006, Bale et Barner 2009), et étoffée par les résultats de diverses expériences, concernant entre autres les *dual nouns* (les noms flexibles, s'accommodant supposément avec une égale facilité de la syntaxe massive et comptable, comme *stone* 'pierre' ou *rope* 'corde', voir 1.5.1 ci-dessous) dont l'emploi flexible, même lorsque les entités qu'ils désignent dans le monde ne varient pas, remet en question le fait que ce soit une « inspection du monde » qui déciderait de l'individualisation. Au contraire, comme le proposent (Barner et Snedeker 2005), tout reviendrait à une seule et unique propriété grammaticale, « *+individual* », codée par la syntaxe comptable, et lexicalisée dans certains cas (en particulier dans le cas des noms massifs de collections comme *furniture* 'mobilier'). Cette vision est formalisée par (Bale et Barner 2009), qui présentent un modèle de la quantification massive et comptable en termes de racines nominales (*nominal roots*) individualisées ou non, et de deux fonctions grammaticales : une fonction comptable projetant les racines non individualisées sur des dénnotations individualisées, et une fonction massive projetant une racine individualisée ou non sur une dénnotation de même nature (*identity function*). Les diverses dénnotations ainsi obtenues sont décrites sous formes d'ensembles et de demi-treillis, dans la tradition de Link et de sa terminologie.

Ensuite, nous trouvons encore une troisième alternative quant à la description de la dénnotation massive : celle que non pas certains, mais bien tous les noms massifs dénoteraient des ensembles d'individus. Cette vision a été proposée par (Chierchia 1998) : pour lui, tant les noms massifs que comptables dénotent des sup-demi-treillis avec des parties minimales. En est prise pour preuve, entre autres, la propriété de référence cumulative (voir *supra*), qui est partagée par le massif et le pluriel. Ainsi, dans pareille vision, les noms massifs de collections, dénotant des individus ontologiquement saillants, ne constituent plus des exceptions parmi les massifs, mais plutôt la règle (notons que dans (Chierchia 2010), on trouve le contrepied de cette proposition), et les paires nominales telles que *meubles-mobilier* sont vus comme ayant une dénnotation identique. La différence entre les parties minimales du comptable et du massif est donnée en termes de stabilité : si dans le premier cas, elles sont stables et prédéfinies, dans le second elles sont « vagues ». Ce que ce caractère vague implique est précisé dans (Chierchia 2010), néanmoins la définition elle-même reste vague et ne permet pas, selon la critique formulée par (Landman 2011: 7-15), de distinguer réellement pourquoi le

comptable peut être compté, mais pas le massif. En outre, dire que des paires comptable-massif quasi-synonymiques telles que *meubles-mobilier* ou, en anglais, *carpets-carpeting* ‘tapis’ ou *curtains-drapery* ‘rideaux’ ont la même dénotation n’explique pas certaines de leur caractéristiques sémantiques, comme leur comportement face aux prédicats réflexifs (cf. le problème de (Gillon 1992)). Comme le montre (Landman 2011), les énoncés anglais (26) et (27) – se traduisant en français comme ‘Les tapis et les rideaux se ressemblent’ – ne sont pas équivalents d’un point de vue sémantique. En effet, si le premier autorise plusieurs interprétations du prédicat (lectures distributive ou cumulative), le second ne permet que la lecture où l’ensemble formé par les rideaux (*the drapery*) ressemble à l’ensemble formé par les tapis (*the carpeting*) :

(26) *The curtains and the carpets resemble each other.*

(27) *The drapery and the carpeting resemble each other.*

Dans un cadre différent, (Nicolas 2008) propose également de se concentrer, dans la description de la dénotation massive, sur le point commun entre cette dernière et le pluriel. Choisisant de s’éloigner de la traditionnelle logique des prédicats, il décrit la particularité du massif comme étant celle de pouvoir référer à « plusieurs choses en un » (*several things at once*), ce qu’il invite à définir dans le cadre de la « logique plurielle » (*plural logic*) que (Schein 2006) ou (Rayo 2006) ont utilisée pour la description de la dénotation plurielle.

Pour finir, mentionnons encore le modèle proposé par (Landman 2011), qui se distingue de tous les autres en proposant une sémantique nominale nouvelle, qu’il appelle *Iceberg Semantics*, non plus en termes d’atomes, mais de « générateurs ». Ainsi, la dénotation des noms sont des « ensembles générés » (*generated sets*), par des générateurs qui soit se chevauchent (*overlapping generators*, massif), soit ne se chevauchent pas (*non-overlapping generators*, comptable). Pour les premiers, deux possibilités sont encore distinguées : soit les générateurs qui se chevauchent génèrent des ensembles dont les parties minimales elles-mêmes se chevauchent également (*mess mass nouns* : les noms de substance), soit ils génèrent des ensembles dont les parties minimales ne se chevauchent pas (*neat mass nouns* : nos noms massifs de collections). Cette description est assortie d’une formalisation détaillée, et les propriétés sémantiques des diverses dénotations « générées » sont traduites selon des formules mathématiques inspirées de la théorie des ensembles.

Ceci clôture notre bref tour d’horizon de quelques propositions de description de la dénotation massive, dans le cadre des théories de la quantification – caractérisées par un contexte plus formaliste (que nous nous sommes bien gardés de décrire en détail). Nous retiendrons que trois grandes positions se partagent le terrain théorique, selon que la dénotation massive est dite 1) dépourvue de parties minimales, 2) tantôt pourvue, tantôt dépourvue de parties minimales (sous-spécification), et 3) toujours pourvue de parties minimales. Notons par ailleurs que c’est la première de ces visions – celle du

massif comme marqueur de la non-individualisation – que l'on pourrait rapprocher le plus de l'approche cognitiviste-conceptualiste de la question, centrée autour de la notion de « référence homogène », que nous avons décrite en détail ci-avant (voir point 1.4.1).

Nous retiendrons principalement trois choses de tout cela. Premièrement, que nous ne tenterons pas, dans le présent ouvrage, de présenter à notre tour une formalisation des dénотations massives et comptables en termes de logique formelle et/ou d'objets conceptuels et outils terminologiques mathématiques. Ce domaine n'est pas le nôtre, et nous choisirons donc de garder à l'esprit, dans la suite du présent travail, les définitions sémantiques d'inspiration cognitivistes, basées sur les notions d'homogénéité (conceptualisée) et d'occurrences.

Deuxièmement, que sans remettre en question l'intérêt des formalisations en sémantique vériconditionnelle, les modèles mentionnés ici nous semblent parfois relever d'un tel degré d'abstraction qu'ils ne nous paraissent plus refléter de manière évidente les propriétés réelles de l'objet d'étude qu'est la langue. Ainsi, lorsque l'on parle de « racines nominales individualisées », qui seront « projetées sur des dénотations de même nature » par la syntaxe massive (Bale et Barner 2009), l'on en viendrait à se demander si l'on parle de syntaxe, de sémantique, ou des deux – tandis que la nature de cette « individualisation », en termes référentiels, reste un mystère. De manière similaire, décrire toute dénотation massive en termes d'un ensemble d'individus (Chierchia 1998) nous semble plus un artéfact de la formalisation que la réflexion d'une réalité linguistique. Cette aliénation du potentiel descriptif et cognitif de la linguistique ne nous semble pas souhaitable.

Pour finir, le débat au sujet de l'individualisation et des « parties minimales » de la référence massive, semble découler – comme nous le disions déjà au point précédent – d'une réelle tension entre perception conceptuelle et cognitive, et réalité ontologique. Le fait que Gillon dérive l'individualisation des référents massifs, autrement sous-spécifiés, de notre observation du monde est une piste intéressante, mais appelle à ce que l'on intègre de manière satisfaisante la question des « connaissances encyclopédiques » dans celle du sémantisme du massif-comptable – et de la langue en général. Nous reviendrons à cette question dans les chapitres suivants, lorsque nous traiterons de la question du sens lexical, et de la modélisation du lexique.

### **1.4.3 Le massif-comptable au-delà du domaine nominal**

Si dans ce qui précède, nous avons appliqué des critères de cumulativité, de divisibilité, d'homogénéité, et des considérations concernant le bornage (point 1.4.1), ainsi que des modèles de quantification formels (point 1.4.2), à des référents nominaux, il conviendra que nous disions – avant de clôturer cette partie sur la dénотation massive et comptable

– quelques mots de leur application à d'autres niveaux du système linguistique. Ainsi, par similarité avec le domaine nominal, il a aussi été proposé dans la littérature d'appliquer l'opposition massif-comptable (ou un de ses avatars) aux adjectifs, ainsi qu'aux verbes (états et événements, cf. (Vendler 1957)).

En l'occurrence, c'est bien des aspects sémantiques de l'opposition dont il est question. En effet, les critères morphosyntaxiques nominaux (combinaison avec le déterminant partitif contre combinaison avec le pluriel et les numéraux, etc., voir point 1.2) ne peuvent s'appliquer sans plus aux adjectifs ou aux verbes. C'est ce qui justifie que l'on aborde ce point ici.

#### 1.4.3.1 Le massif-comptable, et les adjectifs

Pour les adjectifs, les considérations sur le sujet remontent à (Quine 1960), qui propose de distinguer entre adjectifs massifs et comptables sur la base du critère de cumulativité. Ainsi est-il possible de distinguer, selon lui, un adjectif tel que *rouge*, qui répond au critère de référence cumulative (et est ainsi un massif), d'un adjectif tel que *sphérique*, qui n'y répond pas (et est donc comptable). En effet, si la somme de deux objets rouges se laisse toujours décrire par l'adjectif *rouge*, ce n'est pas le cas de deux objets sphériques – dont la somme ne sera pas, quant à elle, *sphérique*.

Comme observé par plusieurs commentateurs (Pelletier 1975: 460-461, Bunt 1985: 16, Kleiber 1989: 269), le critère de Quine est cependant trop limitatif, puisqu'il oblige à placer dans deux catégories différentes les adjectifs *léger* et *lourd* (seul le second répondant au critère de cumulativité), alors qu'intuitivement, on les décrirait comme étant du même ordre. (Bunt 1985) propose dès lors que pour être massif, un adjectif doit répondre à la fois au critère de cumulativité, et à celui de divisibilité, c'est-à-dire référer de manière homogène – ce qui permet toujours de mettre *rouge* du côté des massifs, et *sphérique* ainsi que tant *lourd* que *léger* du côté des comptables. (Bunt 1985: 203) justifie la pertinence de cette proposition par ce qu'il appelle le « principe de combinaison homogène » (voir aussi (Kleiber 1989: 270-271)) : que seuls les adjectifs massifs peuvent modifier de manière restrictive les noms massifs (cf. le caractère malformé de *\*du sable lourd* ; la modification d'un nom massif par un adjectif comptable serait possible uniquement en cas de modification collective, générique, ou discrète).

(Kleiber 1989), quant à lui, qui présente d'abord les propositions de Quine puis de Bunt, plaide *contre* l'extension de la distinction massif-comptable aux adjectifs. Pour cela, différents arguments sont évoqués, tournant essentiellement autour de l'application des tests de référence homogène : en réalité, ceux-ci ne s'appliquent jamais directement aux adjectifs, mais ne peuvent s'utiliser que par le biais d'un référent nominal – ce qui empêche que l'on appelle homogène la référence de quelque adjectif en lui-même. Dans la foulée, le principe de combinaison homogène de Bunt est réfuté, et remplacé par un « principe de structuration partitive », qui explique les mêmes données

distributionnelles en n'opérant non pas une distinction entre deux types d'adjectifs, mais en imposant sur toute combinaison « adjectif + nom massif » une condition de « structure individuanse ». Ainsi, la question du statut massif-comptable des adjectifs est une question débattue, mais néanmoins tranchée par (Kleiber 1989) en défaveur d'une extension de l'opposition au domaine adjectival – une position qui, à notre connaissance, n'a pas été réfutée explicitement depuis. Du reste, ou peut-être pour cette raison, la question ne semble plus avoir tracassé les linguistes ces dernières années.

#### 1.4.3.2 Le massif-comptable, et les prédicats verbaux

Concernant les verbes, la plus célèbre classification remonte aux travaux de (Vendler 1957), qui distingue différentes classes de prédicats verbaux sur la base des différents types de schémas temporels (*time schemata*) qu'ils présupposent. D'une part, il identifie les éventualités stables, sans changement apparent, qu'il appelle les « états » (par exemple, *Ce livre est amusant*, *Paul est grand*). A ceux-là, il oppose les « événements », qui se divisent eux-mêmes en trois catégories : les activités, les accomplissements, et les achèvements. (Concernant ce dernier, le mot français n'est pas tout à fait équivalent à l'anglais *achievement*, mais nous l'utiliserons néanmoins, à la suite de (Récanati et Récanati 1999) et (Nicolas 2002a).) Les « activités » se distinguent en ce qu'elles n'ont pas de but intrinsèque, qu'elles n'ont pas de finalité (à l'instar, par ailleurs, des états) – elles sont dites « atéliques » (*Jean aime peindre*). Les « accomplissements » et les « achèvements », quant à eux, sont téliques : ils ont un but. Ils se distinguent à leur tour l'un de l'autre par rapport à leur rapport au temps : si les accomplissements se déroulent dans la durée (*Jean est en train de construire une maison*), les achèvements quant à eux ne se réalisent qu'à leur point final (*Jean a retrouvé sa clé*). Cette classification a été souvent commentée, critiquée, améliorée (voir par exemple (Récanati et Récanati 1999)), néanmoins notre objectif présent ne nécessite pas que nous entrions plus dans les détails.

Ce qui nous intéresse, c'est que cette classification évoque le domaine aspectuel, auquel les outils de description sémantique du massif-comptable ont parfois été appliqués. Ainsi, comme le signale et le discute (Khallouqi 2003), l'appareil analytique sémantique de (Link 1983), en termes de treillis, appliqué d'abord aux individus, a été récupéré par exemple par (Bach 1986) et (Krifka 1998), pour la modélisation de la sémantique des événements.

Plus précisément, un parallèle est souvent tracé entre la notion aspectuelle de télicité, et la sémantique du massif-comptable ((Mourelatos 1978, Borillo 1989, Krifka 1989, Alexiadou 2011) ; voir aussi (Franckel *et al.* 1989) dans un cadre plus culiolien) – les prédicats téliques étant dit apparenté au comptable, les atéliques au massif. Ces derniers semblent en effet référer, d'une certaine façon, de manière homogène : sans entrer dans les subtilités de l'application des tests, si un prédicat atélique est vrai d'un intervalle de

temps donné, alors il est également vrai de toute partie méréologique de cet intervalle (cf. divisibilité), de même que s'il est vrai de deux intervalles donnés, alors il sera vrai également de leur somme (cf. cumulativité). Ainsi, les états et les activités « s'inscrivent dans une durée homogène » (Récanati et Récanati 1999: 171). En revanche, parce qu'un prédicat télique implique quant à lui un but, intrinsèque à son sémantisme, ces mêmes tests ne s'appliquent pas : les accomplissements et les achèvements dénotent ainsi « des transitions, c'est-à-dire des changements d'état introduisant une discontinuité, une rupture » (Récanati et Récanati 1999: 171). Cette application commune des traits [±homogène] au domaine nominal et verbal a été examinée en détail par (Asnès 2004), qui entend identifier par ce biais les similarités entre quantification et flexion.

Pour finir, concernant le domaine nominal en lui-même, le parallèle entre (a)télicité et référence homogène-hétérogène s'est surtout avéré utile pour discuter le statut massif ou comptable des noms déverbaux. L'observation centrale est ainsi que les noms désignant des états et des activités – atéliques – seraient massifs (*de l'amour, du désordre ; de la marche, du travail*), tandis que ceux nommant des accomplissements et achèvements – téliques – seraient comptables (*la préparation d'un examen, \*de la préparation d'un examen ; une apparition, une trouvaille*). (Voir (Nicolas 2002a) pour un examen critique.) En anglais (voir (Englehardt 2000) et (Borer 2005), *i.a.*), ceci permet de prédire que les nominalisations en *-ing*, qui forment des noms à usage massif, ne soient possibles qu'avec les prédicats atéliques, en particulier les activités (*the sinking of the ship* 'le naufrage du bateau' vs. *\*the arriving of the train* 'l'arrivée du train', cf. (Alexiadou 2011)).

## 1.5 La représentation de l'opposition massif-comptable dans le système linguistique

Dans ce qui précède, nous avons évoqué l'opposition massif-comptable en termes de ses corollaires morphosyntaxiques et sémantiques – démontrant, s'il le fallait, la pertinence de la distinction (au moins comme outil descriptif) – néanmoins, il reste une question centrale à aborder : où peut-on localiser cette distinction, dans le système linguistique ?

Dans la mesure où certains déterminants et autres contextes syntagmatiques sont massifs, et d'autres comptables (voir Chapitre 1, point 1.2), il est relativement consensuel d'attribuer au moins aux SN ces propriétés. Cependant, les intuitions quant à ce que les noms comme *eau* et *farine* sont, d'une certaine manière, différents de ceux comme *voiture* et *chaise*, a fait dire – sans grande polémique jusque récemment – que l'opposition massif-comptable distinguait aussi, plus fondamentalement, (au moins) deux types de noms. Dans ce cadre, la flexibilité de certains N, susceptibles d'apparaître

parfois dans une syntaxe ne correspondant pas à leur type profond, a été expliquée en termes de « transferts » (plus ou moins systématiques) ou de « coercion ».

Cette position intuitive et traditionnelle, que nous appellerons « lexicaliste », s'est néanmoins vu mettre à mal par des auteurs plus récents qui, accentuant l'importance de la flexibilité de l'opposition massif-comptable (qui la distinguerait rigoureusement des oppositions « vraiment » lexicales telles le genre formel), ont proposé que massif et comptable n'étaient pas, de fait, des propriétés lexicales, mais plutôt des marques syntaxiques apparaissant uniquement au niveau des syntagmes nominaux – laissant à tous les noms un statut morphosyntaxique sous-déterminé dans le lexique. Nous appellerons ceci la position « grammaticale ».

Dans ce qui suit, nous évoquerons d'abord la vision lexicaliste, qui nous servira de point de départ pour examiner ensuite la flexibilité de notre opposition – traitée traditionnellement en termes de « transferts », ou du moins de rapports systématiques (si l'on veut éviter une interprétation processuelle) entre emplois comptables et massifs de certains noms. Après cela, nous nous tournerons vers la vision grammaticale, que nous examinerons et commenterons à son tour, avant de dresser quelques conclusions préliminaires sur la question.

### 1.5.1 La vision lexicaliste

L'approche traditionnelle, qui situe notre opposition dans le lexique, repose essentiellement sur l'observation courante et maintes fois répétée que les locuteurs d'une langue ont, généralement, des intuitions fortes et convergentes quant au statut massif ou comptable de bon nombre de noms. Toutes considérations de flexibilité mises à part, il semble intuitif de classer *beurre* ou *eau* dans la première catégorie, et *arbre*, *voiture* ou *chaise* dans la seconde. En outre, comme l'observait (Galmiche 1989: 65), les effets de l'opposition massif-comptable peuvent s'observer même en l'absence d'un marquage morphosyntaxique explicite, par exemple au travers de la reprise pronominale avec *en* – voir énoncés (16) et (17), reproduits par commodité comme (28) et (29) ci-dessous :

(28) Le sang avait coulé, il y **en** avait par terre.

(29) \*Le livre était tombé, il y **en** avait par terre.

C'est bien signe que fondamentalement, il doit y avoir quelque chose qui distinguerait *sang* de *livre*, au niveau lexical. Outre Galmiche, parlons encore parmi les lexicalistes de (Gillon 1999), qui cite les traits lexicaux [+CT] (*countable* 'comptable') et [+PL] (pluriel) (interagissant pour définir dans le lexique les noms comptables, les massifs, les pluriels massifs, ainsi que les pluriels internes), de (Nicolas 2002a), qui identifie deux sous-classes de noms, les fondamentalement massifs, et les fondamentalement comptables,

de (Cheng *et al.* 2008), qui plaident explicitement contre une « *mass-(or-unmarked)-in-the-lexicon theory* » (c'est-à-dire contre la vision grammaticale, évoquée *infra*), ou encore de (Landman 2011) et (Zhang 2013), qui décomposent la distinction massif-comptable en deux traits lexicaux, respectivement [ $\pm C$ ] et [ $\pm N$ ], et [ $\pm Numerable$ ] et [ $\pm Delimitable$ ].

Il est intéressant de noter que si ces différents modèles supposent tous l'existence de traits lexicaux sous-tendant l'opposition massif-comptable, la teneur plutôt syntaxique ou sémantique de ces traits n'est pas toujours équivalente – ni même toujours bien définie. Ainsi, les traits de (Gillon 1999) ou (Zhang 2013) sont clairement syntaxiques, dans la mesure où ils correspondent à des restrictions combinatoires : [ $\pm CT$ ] correspond à la combinaison possible du nom avec les déterminants propres au massif ou au comptable, [+PL] est responsable de l'accord au pluriel (*The police are...*), [ $\pm Numerable$ ] a trait à la possibilité pour un nom de se combiner directement à un numéral, et [ $\pm Delimitable$ ] correspond aux possibilités de combinaison avec un « modificateur délimitatif » (indiquant une forme, une taille, ou une autre forme de limitation). Les critères de (Landman 2011), quant à eux, sont à mi-chemin : [ $\pm C$ ] correspond bel et bien à la possibilité de combinaison avec un numéral (un critère syntaxique), tandis que [ $\pm N$ ] évoque la nature de la dénotation, en termes de parties minimales qui se chevauchent [-N] ou non [+N] (un critère sémantique). (Nicolas 2002a) lui-même assume pleinement d'aborder l'opposition massif-comptable comme une distinction combinant des aspects syntaxiques et sémantiques. D'autre part, le plaidoyer de (Cheng *et al.* 2008) concerne essentiellement la dénotation des noms, et est donc essentiellement sémantique : il s'agit de l'individualisation des référents, dont il est démontré qu'elle ne dépend pas nécessairement, contrairement aux propositions de la vision « grammaticale », de quelque marquage syntaxique préalable. De même, c'est cette question du caractère lexical ou non du trait sémantique d'homogénéité qui était au cœur de la présentation SLE2013 de (Müller 2013) – qui pour sa part observait que le statut de l'homogénéité varie d'une langue à l'autre (étant codée lexicalement en danois, mais pas en espagnol).

Bien entendu, que critères sémantiques et syntaxiques se combinent parfois lorsqu'il s'agit du massif-comptable n'est pas un souci en soi : le sujet est éminemment complexe, et doté de nombreuses facettes. Pour notre part, il s'agira seulement d'être attentif à ne pas mélanger les deux points de vue – mais plutôt à montrer comment ils interagissent. Il s'agira aussi, si nous parlons de traits lexicaux, d'être précis quant à la nature formelle ou référentielle que nous leur attribuons.

En effet, lorsqu'il s'agit de définir pour l'opposition massif-comptable des traits supposés présents dans le lexique, la question de leur place le long de l'axe syntaxe-sémantique, de même que plus généralement leur nature, n'est pas anodine. Si l'idée traditionnelle d'un lexique aux entrées « tagguées » a la vie dure, l'on trouve dans la littérature également des propositions plus nuancées ou progressives, ouvrant des pistes intéressantes. Ainsi, (Allan 1980) – qui précisément ne se réclame pas d'une approche lexicaliste, voir ci-dessous – propose de situer l'opposition massif-comptable



au niveau du SN, mais d'accorder aux N des « préférences de comptabilité calculables », sur la base des possibilités de combinaison avec une batterie réfléchie d'environnements phrastiques. Or, à notre sens, ce genre de préférences calculées n'est pas moins une forme de marquage lexical que quelque trait binaire. En d'autres termes, l'attraction d'un nom pour le massif ou le comptable, en tant que fonction de son comportement dans différents types de contextes choisis, est autant que d'autres traits formels une propriété de l'item lexical – ce qui rend la proposition d'(Allan 1980) plus lexicaliste qu'elle ne voudrait s'en réclamer. Mentionnons aussi, dans un cadre bien différent, l'essai de typologie lexicale de (Flaux et Van de Velde 2000), où les classes de noms établies à partir de critères qui se veulent pourtant pour beaucoup sémantiques regroupent, finalement, des items sur la base de propriétés combinatoires (donc syntaxiques) communes. Notons le parallèle avec l'idée originale d'(Allan 1980) : plutôt que d'en découler seulement, les caractéristiques distributionnelles des noms ne seraient-elles pas ce qui fonderait leur caractère lexical ? Nous reviendrons très certainement sur cette piste.

Quoi qu'il en soit, si la vision lexicaliste de l'opposition massif-comptable, par son caractère très intuitif, a de quoi séduire, il n'en reste pas moins vrai qu'elle doit pouvoir, pour être satisfaisante, apporter une explication à l'existence d'une certaine flexibilité : l'observation que bien des noms, dans divers contextes, puissent apparaître parfois en syntaxe massive, parfois en syntaxe comptable (sans nier pour autant l'existence d'une attraction préférentielle pour l'un ou l'autre type d'environnement). Sur ce sujet, l'approche lexicaliste avance généralement une solution à deux niveaux : soit, un nom polysémique peut être lexicalisé dans ses deux interprétations, soit son interprétation préférentielle peut se voir détournée au travers de mécanismes de transfert (plus ou moins systématiques). La différence entre les deux cas a essentiellement été décrite en termes d'acceptabilité intuitive. Ainsi, le nom *agneau* serait un polysème, lexicalisé d'une part comme comptable pour désigner un animal (*Cette brebis a donné naissance à deux agneaux*) et d'autre part comme massif pour désigner de la viande (*On nous a servi de l'agneau et des pommes de terre*) – en ce qu'il paraît tout aussi naturel dans ses deux emplois (voir (Kleiber 1999, Nicolas 2002a)). D'autre part, dans notre énoncé (12), reproduit ici comme (30), l'emploi massif de *chat* semble plus créatif :

(30) Après le passage du camion, il y avait du chat sur toute la largeur de la route.

Il reste néanmoins légitime, bien que moins habituel, car associé à un glissement de sens spécifique et prévisible (un « broyage » conceptuel – et aussi littéral, en l'occurrence). Dans la littérature, les différentes possibilités de pareils glissements de sens, associés au passage du comptable au massif ou inversement, ont été généralement décrits à l'aide de machines métaphoriques – que nous citerons et expliquerons au point suivant.

Notons néanmoins encore, avant de passer à la suite, que si cette distinction entre polysémie lexicalisée et mécanismes de transfert productifs semble légitime ou du

moins attrayante, la frontière entre les deux procédés – supposée stricte (puisqu'on ne peut imaginer, d'après les définitions évoquées, de cas intermédiaires) – reste un problème : sur quelle base faut-il trancher ? En particulier, le traitement de la flexibilité des co-hyponymes doit nous intriguer. Si l'on a dit qu'*agneau* était un terme polysémique, l'on dira probablement que dans le cas de (31), *kangourou* est bien utilisé de manière détournée – car l'on ne saurait supposer, suivant nos intuitions de fréquence et nos connaissances des habitudes culinaires francophones, que pour ce nom, l'usage massif soit lexicalisé.

(31) Lors de mon récent voyage en Australie, j'ai mangé du kangourou.

Néanmoins, si parler de fréquences en la matière n'est déjà pas anodin (car à notre connaissance, cette question n'a jamais jusqu'à présent été traitée de manière quantitative), l'intrigue va plus loin : opérer ainsi une distinction stricte entre l'emploi d'*agneau* (relevant de polysémie) et de *kangourou* (relevant d'un transfert) oublie d'honorer le très évident parallèle qui existe tout à la fois entre [*manger*] de l'*agneau* et [*manger*] du *kangourou*, deux cas d'un 'lexème utilisé en contexte massif pour désigner la viande de l'animal qu'il désigne lorsqu'il est utilisé comme comptable'.

En outre, parler de transfert suppose bien sûr que certains usages reflètent des emplois de base, et que d'autres soient dérivés – or nous n'avons pas connaissance de réels critères formels et systématiques, autres que l'intuition et le bon sens, qui auraient été avancés pour les distinguer. Dès lors, tous les cas sont susceptibles d'être polémiques : l'on aurait envie de dire que *chêne*, en tant que nom massif désignant du bois, est 'dérivé' du nom comptable *chêne*, désignant un arbre – mais pourquoi ? A nouveau, peut-être s'agit-il d'une question de fréquence, mais dans ce cas, il restera à le démontrer. Nous devons bien sûr garder toutes ces questions à l'esprit lors de notre propre analyse. Mais passons à présent à l'examen de ces fameux mécanismes de transfert, que nous avons annoncés ci-avant.

### 1.5.2 La flexibilité de l'opposition massif/comptable

Comme nous le disions, bon nombre de nos prédécesseurs ont observé des régularités quant aux effets de sens liés aux emplois massifs et comptables (plus ou moins créatifs) de certains groupes de noms, et celles-ci se sont vues expliquées en termes de machines métaphoriques : le Broyeur Universel, le Multiplicateur Universel ou le Broyeur Qualitatif pour les transferts du comptable au massif, le Trieur Universel et le Conditionneur Universel pour ceux du massif au comptable. Nous les explorons tour à tour dans ce qui suit.

### 1.5.2.1 Du comptable au massif

Notre manie du machinisme remonte à (Pelletier 1975), qui lance le terme de *Universal Grinder* ‘Broyeur Universel’ – qui sera ensuite maintes fois récupéré. Notons que pour Pelletier, le Broyeur est bel et bien une machine littérale (fût-elle imaginaire), capable de broyer n’importe quel objet jusqu’à une masse homogène. (On n’en est donc pas encore à la machine métaphorique illustrant un mécanisme sémantique.) Il propose de l’inventer pour justifier que n’importe quel référent, désigné habituellement par un nom en syntagme comptable, une fois physiquement broyé, pourra être désigné par ce même nom en syntaxe massive. Ainsi, après son passage entre les lames de la machine infernale, *un chapeau* deviendrait *du chapeau*, *un livre* ne serait plus que *du livre*, etc. Pour Pelletier, cet exercice d’imagination permet précisément de démontrer que n’importe quel nom est en réalité flexible, et que situer l’opposition massif-comptable au niveau du lexique n’est donc pas pertinent.

Par la suite, la proposition de Pelletier s’est vu maintes fois critiquée. (Galmiche 1989) observe ainsi qu’*un livre* broyé, ce serait plutôt *du papier* que *du livre* (de même que d’une statue broyée, il ne resterait non pas *de la statue*, mais bien *de la pierre* – voire *de la caillasse*). En outre, il identifie aussi plusieurs groupes de référents résistants au traitement par le broyage littéral : les noms de mesure (*kilo*, *centimètre*), les noms de divisions (*classement*, *paragraphe*) et les noms d’événements ponctuels (*chute*, *éclair*, *clin d’œil*). L’on imagine en effet mal comment *un centimètre* ou *un clin d’œil* pourraient être soumis à un déchiquetage physique systématique. (Yi 2012 <in progress>) formule des critiques similaires, et signale encore comme candidats problématiques au broyage des noms comme *trou*, *vallée*, *rivière*, *galet*, *rocher* ou encore *goutte d’eau*, *goutte d’huile* et *glaçon*.

L’appellation de « Broyeur Universel » a néanmoins été récupérée précisément dans le contexte d’une approche lexicaliste de l’opposition massif-comptable, pour nommer le dispositif sémantique de transfert du comptable vers le massif, lorsqu’un référent autrement comptable est présenté conceptuellement de manière homogène – c’est-à-dire sous forme d’une substance de quelque nature, dérivée du référent habituel de l’usage comptable. Pour certains auteurs (comme (Landman 2011), qui parle d’une opération sémantique de « fission », ou pour (Kleiber 1999)), ce Broyeur est limité aux emplois réellement créatifs, comme celui du malheureux félin étalé sur toute la largeur de la route, dans notre exemple (30), ou encore – si l’on reste dans le domaine de la maltraitance animale – dans (32) ci-après, inspiré de (Landman 2011: 58) :

(32) Il y a du Doberman dans la salade.

Pour d’autres, comme par exemple (Nunberg et Zaenen 1992), l’appellation de broyage métaphorique s’applique aussi plus largement aux cas où, pour une raison ou pour une

autre, un référent comptable devient homogène – sans nécessairement qu’un moulinage particulièrement fin doive être impliqué. Ce serait le cas, par exemple, dans (33) ou (34) :

- (33) Cette préparation contient du concombre.
- (34) Je porte du lapin, car ça fait chic (= de la fourrure de lapin).

L’opération sémantique de broyage, même dans cette interprétation large, ne couvre cependant pas encore toutes les possibilités interprétatives de l’emploi massif d’un nom autrement comptable. Considérons les exemples suivants, inspirés de (Galmiche 1989) :

- (35) Cet été, il va y avoir du touriste à Namur.
- (36) Il y a de la voiture à perte de vue dans le parking.
- (37) Ces skieurs ivres surestiment leurs propres aptitudes : il va y avoir de la chute !
- (38) Les investisseurs se sont échangés de la poignée de main tout au long du cocktail.

Dans ces différents cas, l’intégrité physique des référents évoqués n’est nullement mise en péril : *les touristes* qui arpenteront cet été la Rue de Fer et la Place de l’Ange seront bien en un seul morceau, de même qu’on peut difficilement imaginer que *des chutes* ou *des poignées de main* se retrouvent littéralement broyées, de quelque façon que ce soit.

Dès lors, pour ces types d’exemples, (Galmiche 1989) propose d’introduire, pour le français, un second mécanisme sémantique de transfert du comptable au massif : celui du « Multiplicateur Universel ». Dans le cas du Multiplicateur, l’homogénéité conceptuelle supposée par l’emploi massif passe par le reniement de l’individu suite non pas à sa destruction, mais à sa sommation indéfinie. Notons par ailleurs que ce mécanisme est bien spécifique au français. En néerlandais ou en anglais, on le retrouve uniquement dans un emploi très restreint, limité aux noms d’animaux qui font l’objet de la chasse ou de la pêche (voir (Nicolas 2002b: 7, Lauwers et Vermote 2014, à paraître)). En français, le mécanisme est sensiblement plus productif – bien qu’il reste emblématique d’un français quelque peu familier (Cohen et Zribi-Hertz 2012).

Les exemples ci-après illustrent quant à eux un troisième et dernier mécanisme de passage du comptable vers le massif, également propre au français familier, lorsqu’une évaluation qualitative est opérée :

- (39) Ca, c’est de la bagnole !
- (40) Ca, c’est pas de la baraque de prolos !
- (41) Ce nouveau film, c’est de la bombe !

En néerlandais, ce mécanisme n’est pas disponible du tout (*\*Dat is pas auto (om over naar huis te schrijven) !*). Pour le français, ces types d’emplois sont étudiées par (Desagulier 2012), qui analyse le mécanisme en termes d’une fonction de conversion qualitative – et non quantitative (comme c’est le cas pour les transferts généralement étudiés) – identifiant par la syntaxe massive un sujet avec « l’archétype de la catégorie dénotée par le prédicat nominal ».

Ce mécanisme a été appelé parfois « Broyeur Qualitatif », par analogie avec le Broyeur (Quantitatif) évoqué ci-avant (par Kleiber notamment, communication personnelle, qui ne revendique cependant pas particulièrement cette appellation). Notons que si les constructions du type *Ce N, c'est du N* (CDN) et, plus encore, *Ca, c'est du N* (CCDN) permettent bel et bien, le cas échéant, d'employer en syntaxe massive des noms habituellement utilisés comme comptables, elles ne sont pas limitées à ces candidats-là : on pourrait avoir tout autant (42) ou (43), dans une lecture « qualitative » à notre sens tout équivalente aux exemples précédents :

(42) Ca, c'est du mobilier de qualité !

(43) Ca, c'est du beurre de ferme !

Il n'est pas étonnant dès lors que l'analyse de (Desagulier 2012) s'inscrive dans un contexte de Grammaire des Constructions (*Construction Grammar, CxG*, voir notamment (Goldberg 2006)) et que dans son analyse, le trait massif propre à la construction soit hérité par un item qui lui-même n'est pas spécifié comme massif ou comptable dans le lexique. On ne se trouve donc pas, chez Desagulier, dans une approche lexicaliste, comme en témoigne du reste la citation suivante :

Si, en tant que linguistes, l'on doit se fier au comportement grammatical, peut-on dire que certains noms soient principalement [+comptable] et d'autres principalement [+massif] ? Il existe un consensus général pour dire que non. (Desagulier 2012: 207, traduction personnelle)

Desagulier plaide donc contre la vision lexicaliste, que nous présentions ci-dessus. Comme nous espérons que notre explication l'a montré, parler en la matière de « consensus général » est cependant très clairement abusif ! Il est vrai, néanmoins, que l'approche grammaticale n'est pas sans partisans. Nous l'évoquerons dans un instant, une fois que nous aurons présenté les mécanismes de transferts sémantiques répondant aux précédents : ceux concernant les passages du massif au comptable.

### 1.5.2.2 Du massif au comptable

Le machinisme, auquel nous devons déjà le Broyeur et le Multiplicateur Universels, s'est vu tout aussi productif dans la littérature pour expliciter les effets de sens liés à l'usage comptable d'un nom autrement massif. Considérons d'abord les exemples suivants :

(44) Je ne bois que deux eaux : la Vittel et la Spa.

(45) Ce marchand ne vend que des beurres d'origine française.

(46) Nous vendons différents terreaux naturels et enrichis.

(47) L'on trouve sur cette plage un sable très fin.

(48) Paul entrouvrit la lucarne, et aussitôt un air irrespirable remplit la pièce.

Les emplois de ce type se sont vu désignés comme l'œuvre d'un « Trieur Universel » (*Universal Sorter*, (Bunt 1985: 11)) : une machine sémantique fournissant, sur la base d'un référent massif, un référent individualisé en vertu d'une division en 'sortes'. C'est ce que d'autres ont appelé « l'interprétation taxinomique » (Kleiber 2011) ou encore la *kind reading* 'lecture en types' (De Belder 2011) d'un nom. Il s'agit d'un mécanisme qu'on retrouve aussi bien en français, en anglais ou encore en néerlandais (49)-(50) – quoique cette dernière langue y soit un peu moins propice, vu les restrictions quant à l'emploi au pluriel de la plupart des noms fondamentalement massifs (voir (Lauwers et Vermote 2014, à paraître)).

- (49) Chianti is een rode wijn uit Toscane.  
'Le Chianti est un vin rouge de Toscane.'
- (50) Het Belgische Chimay-bier is een authentiek trappistenbier. [Google]  
'La bière belge de Chimay est une bière trappiste authentique.'

Si l'on creuse un peu, l'on verra encore que le mécanisme de triage peut être sous-tendu par deux réalités : soit une distinction encyclopédique (c'est-à-dire objective, ou intersubjective), soit une distinction contextuelle. Dans le premier cas, illustré dans (44), (45) et (46), l'existence de la division en sortes est extralinguistique et indépendante de la situation d'énonciation : que je les boive ou non, et que j'en parle ou non, la Vittel et la Spa sont deux types d'eau. Les usages « encyclopédiques » du Trieur sont dès lors plus ou moins équivalents du point de vue sémantique à la paraphrase avec *sorte(s) de*. Ainsi, aux énoncés (44) et (46) répondent par exemple les énoncés (51) et (52) :

- (51) Je ne bois que deux sortes d'eau(x) : la Vittel et la Spa.  
(52) Nous vendons différentes sortes de terreaux, naturels et enrichis.

D'autre part, le triage peut également avoir pour base une individualisation contextuelle, liée à la situation d'énonciation. Dans ce cas, la paraphrase avec *sorte(s) de* n'est pas toujours satisfaisante, et le nom nécessite un modificateur, responsable de la différenciation qualitative en contexte (*très fin* dans (47), *irrespirable* dans (48)), sans quoi l'énoncé n'est pas acceptable (53) :

- (53) \*L'on trouve sur cette plage un sable.

A notre sens, l'usage « contextuel » du Trieur s'explique facilement dans le cadre d'une approche occurrence de l'opposition massif-comptable ((Kleiber 2011), voir point 1.4.1.3 *supra*), comme l'individualisation linguistique forcée de l'occurrence unique d'un référent massif, dans une situation donnée. (Ce qui expliquerait, du reste, pourquoi ce Trieur contextuel est généralement à l'œuvre au singulier, cf. (Lauwers et Vermote 2014, à paraître).) Cette individualisation est rendue pertinente par l'ajout du modificateur, qui identifie comme particulière l'occurrence de la situation en question – une

particularité qui s'exprimera linguistiquement par le passage à la syntaxe comptable. Conserver la syntaxe massive dans pareil contexte n'est pas agrammatical :

(54) L'on trouve sur cette plage du sable très fin.

Cependant, il y a une nuance : dans (54), avec le partitif, il s'agit d'une partie, éventuellement mais pas nécessairement maximale, du *sable* de la plage qui serait *très fin*, tandis que dans (47), avec le déterminant comptable indéfini singulier, seule l'interprétation totale est considérée (*L'on trouve sur cette plage un sable très fin ~ Le sable de cette plage est très fin*).

C'est ce même mécanisme de Trieur contextuel que l'on retrouve dans les énoncés impliquant des noms de propriétés, comme dans les exemples (55) et (56), que nous empruntons à (Kleiber 2014, à paraître-b).

(55) Marie a une patience d'ange.

(56) Paul éprouve une tristesse douce et mélancolique.

Ici, l'alternance avec le partitif n'est pas possible, car la propriété – au travers de la nominalisation – est qualifiée dans son ensemble, sans qu'une qualification partielle soit envisageable (contrairement à notre exemple du *sable sur cette plage*, dont seule une partie pourrait être *très fin*). De même, tout emploi autre qu'avec l'article indéfini singulier est exclu (*\*Marie a deux patiences d'ange*), car la modification du nom de propriété (*d'ange*) dépend autant de l'occurrence support (*Marie*) que la propriété elle-même (*la patience*) – rendant l'occurrence de l'ensemble (*patience d'ange*) obligatoirement unique. (Kleiber 2014, à paraître-b), à qui l'on doit cette analyse, parle en la matière de « Conditionneur Qualitatif », par analogie avec le conditionneur quantitatif que nous évoquerons ci-après. En parlant dans ce cas de « Trieur contextuel », nous entendons pour notre part marquer plus clairement l'apparement entre ces exemples, concernant les noms de propriétés, et ceux du type (47), (48), ou encore (57) ci-dessous, qui nous semblent bel et bien des exemples à rapprocher de ce qui s'est dit du Trieur. ((Kleiber 2014, à paraître-b) donne lui-même ce dernier énoncé, tiré de (Kupferman 2000), comme exemple parallèle dans le domaine des concrets des énoncés (55) ou (56) dans celui des propriétés.)

(57) Le lac avait ce jour-là une eau plutôt verdâtre.

Signalons encore que l'explication du Trieur en termes d'occurrences permet aussi de rapprocher de ce mécanisme les exemples comme (25) ci-dessus, reproduit ici en (58) :

(58) Tant le café de ma tasse que le café de mon thermos sont froids, à présent.

Dans ce cas, l'emploi de *café* avec l'article défini (un déterminant neutre, nous l'avons vu, quant à l'opposition massif-comptable) empêche que l'on parle ici réellement de transfert du massif vers le comptable. Néanmoins, il y a bien dans cette situation une

individualisation de deux occurrences de *café*, par le biais des modificateurs génitifs (*de ma tasse* et *de mon thermos*) – à l’instar des cas de triage contextuel. Dans cette situation, il y a donc « deux » cafés qui sont froids : celui de ma tasse, et celui de mon thermos – comme en témoigne l’accord du verbe au pluriel.

Ce traitement occurrenceiel expliquerait aussi, pour finir, les restrictions qui pèsent sur le Trieur Universel (encyclopédique, et contextuel), qui ne peut bel et bien s’appliquer aisément qu’aux noms fondamentalement massifs (voir (Kleiber 2011: note 21) : si j’ai mangé *deux bananes*, *trois pommes* et *une orange*, peut-on dire que j’ai mangé *trois fruits* ?) : le Trieur est un mécanisme qui récupère les limites occurrenceielles contextuelles d’un référent pour en faire, provisoirement, des limites codées par la syntaxe (comptable) – or seuls les référents massifs ont des limites uniquement contextuelles, et non intrinsèques.

Voilà pour la question du Trieur Universel. Celui-ci ne couvre pas encore, cependant, l’ensemble des emplois comptables des noms supposés fondamentalement massifs. Pour cela, nous devons faire appel à une seconde machine, susceptible de rendre compte des emplois suivants (où un nom supposé massif est utilisé en syntaxe comptable pour désigner soit une portion conventionnelle, soit un objet typique fait de la matière désignée par son référent habituel) :

- (59) Je prendrai deux bières et un coca.
- (60) Tante Isabelle nous a offert des chocolats.
- (61) Faute de ceinture, Pierre a attaché son pantalon avec une corde.
- (62) J’ai noté son nom sur un papier.

Le mécanisme mettant ainsi en rapport un nom supposé massif, et son interprétation comptable en terme d’unité ou de portion conventionnelle, a été appelé le *Universal Packer* (Bunt 1985), *Universal Packager* (Jackendoff 1991) ou, en français, le ‘Conditionneur Universel’ (Galmiche 1989). L’on trouve aussi le terme de *rule of portioning* ‘règle de conditionnement’ dans (Copestake et Briscoe 1995), ou l’appellation « fission » dans (Landman 2011) – qu’il décrit comme une opération sémantique d’ajout d’un classificateur nul (*une bière* sous-entend *~un verre de bière*). (De Belder 2011) parle, quant à elle, de la *unit reading* ‘lecture en unités’ d’un nom.

Le domaine des noms de boissons, illustré en (59), est le plus célèbre terrain d’application de ce mécanisme, du moins en français et en anglais. L’on mentionnera dans ce contexte l’étude comparative de (Wiese et Maling 2005), qui examine la question du « conditionnement » linguistique des boissons dans trois langues germaniques (anglais, allemand et islandais). Le domaine culinaire est également représenté, avec des noms comme *chocolat* en (60), ou encore *une lasagne* par rapport à *de la lasagne*, *une tartiflette* par rapport à *de la tartiflette*, *une pizza* par rapport à *de la pizza* etc. (Notons que dans ces derniers cas, l’idée du Conditionneur s’applique uniquement en ce que le référent comptable est une unité conventionnelle, dont la substance constitutive serait



désignée par le référent massif. Parler ici d'un transfert du massif vers le comptable, sous-entendant que l'emploi comptable serait dérivé de l'emploi massif, est pour autant loin d'être une évidence. En l'occurrence, l'emploi massif *de la pizza* semblerait même plutôt dérivé du comptable. C'est l'une des difficultés posée par la notion de « transfert » dans l'approche lexicaliste : nous y reviendrons.) Pour finir, on retrouve aussi des exemples à rapprocher du Conditionneur dans le domaine des noms de matériaux (comme dans (61) et (62), par exemple).

Par rapport au Trieur, le Conditionneur (en français) se distingue particulièrement en ce qu'il n'est pas réellement productif (d'un point de vue synchronique) : si comme nous venons de le voir le Trieur connaît un emploi « contextuel » permettant de le convoquer à volonté dans une multitude de situations, l'usage du Conditionneur quant à lui repose crucialement et exclusivement sur la disponibilité encyclopédique, c'est-à-dire prédonnée, d'un référent délimité dénoté par l'emploi comptable. Les usages comptables classés sous la bannière du Conditionneur relèvent ainsi du conventionnel, et sont par là même lexicalisés (voir aussi (Lauwers et Vermote 2014, à paraître)). Ceci fait en réalité du Conditionneur bien moins un mécanisme de transfert productif en contexte du massif vers le comptable qu'un outil descriptif permettant de grouper *a posteriori* des emplois comptables entérinés comparables. Nous aurons l'occasion de le montrer au travers de nos études de corpus, en particulier de celle des noms de matériaux.

Le néerlandais, quant à lui, se distingue ici du français en ce qu'il possède un morphème diminutif productif, « *-[C]je* », remplissant précisément pour les noms massifs la fonction sémantique de singulatif, c'est-à-dire celle du Conditionneur (voir (Bunt 2009), *i.a.*). (Que le diminutif soit propice à la lecture en portion ou unité n'est pas sans explication : en évoquant une contrainte de taille, il impose automatiquement au référent une délimitation. Notons du reste que son emploi singulatif n'est certainement pas limité aux seuls noms massifs, puisqu'ils permet aussi de former des noms comptables à partir de racines verbales, adjectivales ou même prépositionnelles : *een weetje* 'un détail intéressant (litt. petit savoir)', *een blondje* 'un blondinet', *een groentje* 'un novice, un bleu (litt. un petit vert)', *een uitje* 'une sortie (cf. préposition *uit*)'.) Ainsi, si dans cette langue certains noms massifs peuvent bien, comme en français, être employés directement en syntaxe comptable pour désigner une portion (comme *een koffie* 'un café (au sens d'une tasse de café)' ou *een brood* 'un pain', par exemple), d'autres nécessiteront l'emploi du diminutif : *een biertje* 'une bière' (et non \**een bier* pour désigner la portion, du moins pas en néerlandais de Flandre ; voir aussi Chapitre 7), *een snoepje* 'un bonbon' (à partir du nom massif *snoep* 'friandises'), *een papiertje* 'un (bout de) papier', etc. Notons que dans ces cas-là, le diminutif n'a pas (ou plus) trait à quelque petitesse du référent, pas plus qu'il ne déclenche d'effet hypocoristique : il signale seulement que l'on parle d'une portion ou d'une unité spécifique, comme d'un verre de bière ou d'un morceau de confiserie. Le rapport à la lexicalisation reste, cependant, bien visible, puisque la nécessité de l'ajout du diminutif pour une lecture « conditionnée » varie d'un

nom à l'autre (d'indispensable, comme pour *bier* 'bière', à superflu, comme pour *koffie*), de même que l'effet de sens associé (de strictement singulatif à purement diminutif et/ou hypocoristique, voir (Bakema 1998)).

Pour les noms plus abstraits, (Bale et Barner 2009) mentionnent encore, en passant, deux autres mécanismes mettant en lien selon eux des noms supposés massifs et leur emploi comptable : l'*individuation-by-cause* 'individualisation-en-termes-de-cause' et l'*individuation-by-instance* 'individualisation-en-termes-d-exemple'. Pour le premier cas, ils citent l'exemple de *fear* 'peur, crainte' auquel correspondrait le comptable *fears*, désignant les causes de la peur en question. Sur ce point, nous sommes indécis quant à la réussite de la traduction du mécanisme en français. Quand on dit *A force de courage, j'ai pu surmonter mes peurs*, est-ce que le SN comptable *mes peurs* désigne bien, spécifiquement, les 'causes' de ma peur ? C'est une question à creuser. Ensuite, pour le second cas, (Bale et Barner 2009) donnent les exemples de *thoughts* 'pensées' à partir du massif *thought* 'pensée', ou encore de *judgments* 'jugements' à partir du massif *judgment* 'jugement'. A nouveau, la félicité du passage au français nous semble quelque peu douteuse : si nous avons bien dans notre langue les noms à usage *a priori* comptable *pensée* (*J'ai eu une petite pensée pour Lisette, Je suis perdu dans mes pensées*) ou *jugement* (*Il a porté un jugement hâtif*), il ne nous semble pas que ces emplois puissent être dits soutendus par quelque emploi massif plus fondamental. Mentionner *de la pensée* nous semblerait même plutôt marqué, par rapport aux emplois comptables. (Nous laisserons ici la question en suspens, étant donné que nous nous sommes limité dans notre recherche essentiellement au domaine des noms concrets.) Quoi qu'il en soit, les deux mécanismes mentionnés s'expliquent mieux si l'on considère la position de (Bale et Barner 2009) sur la question de la localisation de l'opposition massif-comptable : pour eux, à la suite de (Borer 2005), tous les noms sont fondamentalement massifs dans le lexique – et tout usage comptable découle dès lors d'un mécanisme de transfert. En cela, ils sont à classer du côté non pas de la vision lexicaliste, mais bien de la vision grammaticale de notre opposition – une position que nous explorons au point suivant.

### 1.5.3 La vision grammaticale

Dans ce qui précède, nous avons détaillé la vision lexicaliste du massif-comptable, qui localise notre opposition au niveau lexical, c'est-à-dire au niveau de N. A celle-ci répond une vision grammaticale, qui sous-détermine l'opposition dans le lexique, en ne la faisant apparaître qu'au niveau de la grammaire, c'est-à-dire au niveau de SN. Cette position se résume parfaitement dans la citation de (Desagulier 2012: 207), que nous donnions ci-avant : pour les tenants d'une vision grammaticale, il n'existe ni noms principalement massifs, ni noms principalement comptables.

Dans la littérature récente, cette vision est particulièrement associée à l'ouvrage de (Borer 2005). Pour elle, toutes les racines nominales, dans toutes les langues, sont dépourvues de marquage massif-comptable – car ne sont considérés comme lexicaux que les traits syntaxiques invariables, qui ne peuvent être outrepassés ou ignorés (quant au bien-fondé de pareil raisonnement, nous aurons quelques réserves que nous exprimerons plus loin). En outre, la lecture massive des noms est l'interprétation par défaut, en ce qu'elle implique une référence dénuée d'une structure de division (*dividing structure*). La lecture comptable des noms, quant à elle, ne peut qu'être dérivée syntaxiquement, par l'application de la tête fonctionnelle *Div°* (exprimée par exemple par le pluriel, ou l'article indéfini *a* 'un(e)'), qui fournit précisément à la référence une telle structure de division. (La proposition ne dit rien, par contre, de l'éventuelle fonction du partitif dans une langue comme le français, où l'emploi nominal nu est généralement exclu et ne constitue jamais un SN massif explicite.) Ce modèle est récupéré notamment par (De Belder 2011), qui à côté du trait [Div] argumente aussi pour l'existence d'un trait syntaxique [Size] – l'interaction des deux étant supposée expliquer l'existence de trois lectures des syntagmes nominaux : *mass* ([-Div], [-Size]), *count kind* ([+Div], [-Size]) et *count unit* ([+Div], [+Size]). (Bale et Barner 2009) reprennent eux aussi, dans leur étude de la dénotation massive et comptable, le modèle de (Borer 2005), tout en le peaufinant sur certains points de détail (le modèle de Borer « *with important differences in the details* », p. 233).

Il est à noter que l'idée de base sous-tendant la position grammaticale, notamment que l'opposition massif-comptable ne se situe qu'au niveau des syntagmes nominaux (et non des noms), permettant dès lors une flexibilité syntaxiquement absolue, n'est pas nouvelle. Dans la littérature en français, on la retrouve aussi tôt que dans (Damourette et Pichon 1950), dont le passage sur le sujet a été très souvent cité (par exemple par (Galmiche 1989) ou (Nicolas 2002a: 7)). Parce qu'il est particulièrement représentatif, nous ne dérogerons pas à la règle tacite, et le donnons à notre tour (à noter que l'appellation « numérative » correspond à notre « comptable ») :

Ce qui fait la beauté du système de quantité de la langue française de nos jours, c'est son extrême souplesse et son absolue généralité. C'est après avoir bien médité et posé la question, et après avoir observé la langue tant parlée qu'écrite, que nous osons affirmer ici qu'il n'est pas de substance nominale qu'il soit interdit à un locuteur plus ou moins hardi de concevoir soit comme numérative, soit comme massive. (Damourette et Pichon 1950: 414)

Signalons que malgré les prétentions sous-jacentes de fournir une approche grammaticale avant la lettre, la précision quant à la « hardiesse » nécessaire pour la production de certains emplois aura surtout donné de l'eau au moulin des lexicalistes (voir (Galmiche 1989)). Cependant, on retrouve des idées toutes similaires chez (Gleason 1965), qui considère également que l'opposition massif-comptable ne divise pas

réellement le lexique en deux classes, puisque tout nom – dans le bon contexte – est susceptible d’être employé tantôt comme comptable, tantôt comme massif. Il donne en guise d’illustration le célèbre exemple de *Johnny*, la termite, qui mange *du livre*, mais pas *de l’étagère* (*Johnny is very choosy about his food. He will eat book, but he won’t touch shelf*, voir (Gleason 1965: 136)). C’est dans ce contexte qu’interviennent (Pelletier 1975) et son Broyeur Universel : par l’invention de sa machine infernale, Pelletier entend justifier que pour tout nom, tant le sens massif que le sens comptable doivent être donnés lexicalement, puisque tout référent habituellement comptable peut devenir massif si on l’imagine réduit en bouillie. Seul le contexte d’emploi spécifique distingue donc entre massif et comptable, et non le lexique.

Tout à fait dans la même veine, si ce n’est un rien plus timidement, (Weinreich 1966) précise qu’à son sens, le trait [ $\pm$ count] n’est pas réellement à attribuer aux noms, mais bien aux SN. C’est là, nous semble-t-il, une proposition identique à celle de (Parsons 1970), qui observe que ce sont les usages des noms (« *occurrences of words, or something of the sort* », p. 137) et non les noms eux-mêmes, qui peuvent être dits massifs ou comptables. Dans le même contexte, (Sharvy 1978) émet l’hypothèse la plus osée : que l’anglais pourrait bien n’avoir aucun nom proprement comptable – le caractère comptable ne pouvant apparaître que par le biais d’un classificateur vide (*empty classifier*). C’est là le prédécesseur direct du modèle de (Borer 2005), et l’une de ses sources d’inspiration les plus notables.

Les différentes propositions énumérées ci-dessus ont de commun qu’elles attribuent le trait massif-comptable aux syntagmes nominaux, et non aux noms. Un corollaire potentiel d’une pareille approche est de lier les possibilités combinatoires (éminemment flexibles) des noms à une échelle de « comptabilité » : les contextes syntagmatiques ne seraient non pas ou massifs, ou comptables, mais plutôt plus ou moins fortement orientés dans un sens ou dans l’autre, donnant une indication (quantifiable) du caractère préféré des noms avec lesquels ils se combinent. C’est précisément ce que propose (Allan 1980), que nous mentionnons déjà ci-dessus, en suggérant une batterie de tests permettant de classer les noms selon une échelle de préférence, allant du plus massif au plus comptable – avec toutes les nuances possibles entre les deux. Plus précisément, il propose que le massif, qu’il appelle *uncountable* ‘non comptable’, soit l’option de base pour tous les noms, et que les préférences identifiables quant à la syntaxe massive ou comptable se calcule en termes de « degré d’attirance pour le comptable » (*degree of countableness*, p. 554), indiqué par la possibilité ou non de se trouver dans certains environnements. Ce qui maintient (Allan 1980) du côté de la vision grammaticale, c’est que pour lui, les préférences morphosyntaxiques qu’il calcule ne revêtent aucun caractère lexical (l’option de base étant le massif pour tous les noms). Pourtant, l’option d’attribuer au sein du lexique une certaine pertinence à des données distributionnelles nous paraît très intéressante (et est d’ailleurs annoncée comme prometteuse par (Joosten 2003)) – et nous y reviendrons dans notre propre analyse.

Parmi d'autres versions de la vision grammaticale, caractérisées donc par l'absence de marquage lexical de l'opposition massif-comptable, il y a encore (Desagulier 2012), que nous avons déjà cité, et dont la particularité est qu'il se situe dans le cadre théorique de la Grammaire des Constructions. Dans pareille approche, le trait massif-comptable est situé à un niveau de complexité structurelle plus élevé encore que le seul SN, puisqu'il se voit prédit par la construction dans laquelle le nom apparaît – une approche particulièrement utile, il est vrai, pour traiter des cas comme *Ce livre, c'est de la bombe* ou *Ca, c'est du livre de grande qualité!* (qui sont le sujet de (Desagulier 2012)). Pour finir, nous devons mentionner encore (Pelletier 2012), dont la contribution est intitulée de manière à susciter immanquablement la curiosité : « Lexical nouns are both +MASS and +COUNT, but they are neither +MASS nor +COUNT » – « Les noms lexicaux sont à la fois +MASSIF et +COMPTABLE, mais ils ne sont ni +MASSIF, ni +COMPTABLE ». La seconde partie de la phrase correspond, dans sa vision, à l'opposition massif-comptable en termes syntaxiques : suivant (Allan 1980) et (Borer 2005), il attribue la propriété syntaxique de massif ou de comptable non pas aux N, mais aux unités grammaticales plus longues (comme le SN). Syntaxiquement, les noms sont donc sous-déterminés dans le lexique. La particularité de (Pelletier 2012) est cependant qu'il propose une seconde opposition massif-comptable, différente de la première, qui serait quant à elle sémantique, et pour lesquels la grande majorité des noms ne sont non pas sous-déterminés mais au contraire sur-spécifiés, en ce qu'il possèdent les deux traits. Traduits dans les termes formels que nous avons introduits au point 1.4.2, la dénotation de pour ainsi dire tous les noms consisterait donc en l'union d'un ensemble d'individus, et d'un semi-treillis (atomique ou non). (On constatera que cette proposition est tout à fait dans le prolongement de (Pelletier 1975), où il était déjà avancé que tout référent nominal comptable, parce qu'immanquablement candidat au traitement du Broyeur Universel, se devait d'être codé sémantiquement à la fois comme hétérogène, et comme homogène.)

Nous l'avons dit, le choix d'une approche grammaticale de l'opposition massif-comptable vise généralement à résoudre la question de la flexibilité du trait. Si l'approche lexicaliste, en qualifiant les noms de massifs ou de comptable directement dans le lexique, doit pouvoir expliquer (par des mécanismes de transfert) l'existence d'emplois dérivés outrepassant le trait lexical, l'approche grammaticale peut se passer de tout cet arsenal : la flexibilité est *built-in*, c'est-à-dire qu'elle fait partie intégrante du système. Pour (Pelletier 2012), c'est une position indispensable, car « presque tous les noms peuvent être employés des deux manières » et « dans un très, très grand nombre de cas, les deux significations paraissent tout aussi naturelles » (p. 24, traduction personnelle). C'est là une assertion forte, qui nous semble cependant très peu judicieuse, et intuitivement erronée : même s'il est vrai qu'une grande marge de créativité est possible dans l'usage massif ou comptable des noms (voir la remarque de (Damourette et Pichon 1950), ci-dessus), il n'en reste pas moins vrai que ces emplois se distinguent bien

souvent, précisément, par leur caractère plus ou moins naturel ou plus ou moins étrange ou « hardi ». L'on peut bien sûr se répéter cent fois notre fameuse phrase *Après le passage du camion, il y avait du chat sur toute la largeur de la route*, jusqu'à ce qu'elle nous paraisse entièrement naturelle – mais l'intuition originale persiste que *du chat* est un emploi marqué, par rapport à l'emploi habituel comptable. C'est, à notre sens, quelque chose que la théorie linguistique sur le sujet doit pouvoir expliquer. Il existe du reste bien d'autres oppositions linguistiques qu'une certaine flexibilité ne rend pas pour autant caduque : si je peux dire *Paul est un César*, il n'en est pas moins vrai que *César* reste, fondamentalement, un nom propre, et que l'opposition lexicale entre noms communs et noms propres reste pertinente. De même, si je peux dire, avec quelque poésie, *Ce matin, le soleil sourit*, le statut fondamentalement inanimé de *soleil* n'est pas pour autant nié, pas plus que l'opposition sémantique fondamentale animé-inanimé (nous empruntons cette remarque à Fasciolo, communication personnelle).

Or, contrairement à ce que la citation de (Pelletier 2012) laisserait croire, la tradition grammaticale de l'opposition massif-comptable n'a pas toujours été aveugle à pareilles distinctions intuitives et inter-subjectivement partagées. (De Belder 2011) propose néanmoins de les reléguer en-dehors du système linguistique, en les expliquant à partir de nos connaissances encyclopédiques (sujettes à la flexibilité, contrairement au système langagier formel, qui lui ne connaît aucune exception) : en essence si, hors contexte, *?les quatre eaux* nous paraît un syntagme curieux, c'est seulement parce que nous ne connaissons pas de référent à quoi le rattacher. Si un tel référent nous était connu, l'énoncé n'aurait plus rien de problématique.

Cependant, cette solution n'est pas convaincante, car elle ne permet nullement d'expliquer pourquoi l'emploi intuitivement le plus naturel de *raisin* (*du raisin*, massif) différencierait de celui de *groseille* (*des groseilles*, comptable). A moins que ce que (De Belder 2011) appelle « connaissances encyclopédiques » ne concerne pas les référents, mais bien l'usage des mots qui les désignent ? Dans ce cas, on commence déjà à se rapprocher d'une vision lexicaliste, et la question réapparaît de la nature précise de cette connaissance, et de sa représentation dans ... le lexique.

Pour nous, la principale faiblesse de l'approche grammaticale est dès lors de refuser d'expliquer, dans la langue, l'existence intuitive et très largement reconnue de différences lexicales entre différentes sortes de noms (fondamentalement massifs, fondamentalement comptables et, pourquoi pas, fondamentalement flexibles), en écartant ce fait comme une donnée extralinguistique, non pertinente pour le linguiste ou syntacticien (De Belder 2011) ou en le niant tout bonnement (Pelletier 2012). Dans ce contexte, lorsqu'il est dit que l'opposition massif-comptable n'est pas lexicale, il nous paraît soudainement très peu clair ce que l'on est supposé entendre par là, et ce que cette affirmation nous apprend sur la langue en général. Car il est évident que différents noms se comportent différemment par rapport à l'opposition massif-comptable – que ce soit en anglais, en français ou en néerlandais. Et la pirouette théorique de (Borer 2005)

ne résout rien : dire que l'opposition n'est pas lexicale parce qu'elle est flexible, et que par définition les traits flexibles ne sont pas lexicaux, est au mieux une observation triviale, et au pire un raisonnement circulaire. Elle dévoile surtout, selon nous, l'absence criante d'une définition appropriée et précise de ce qu'est le lexique – sans laquelle toute discussion quant au statut lexical ou non de notre opposition devient automatiquement stérile.

#### 1.5.4 Conclusions préliminaires

Par notre comparaison des visions lexicales et grammaticales, nous avons mis en évidence une question cruciale concernant l'opposition massif-comptable : celle de sa localisation et de sa représentation dans le système linguistique. Cette question s'est montrée difficile à résoudre, car elle fait face au caractère flexible de l'opposition – ce qui a fait douter certains de son caractère lexical. Or, si les alternatives grammaticales ne s'avèrent pas, à notre sens, concluantes, il reste tout autant des questions à résoudre du côté des approches lexicalistes. Pour résumer, un modèle satisfaisant de la représentation, dans la langue, de l'opposition massif-comptable doit pouvoir tenir compte, à notre sens, des faits suivants :

1. Les locuteurs d'une langue ont des intuitions fortes et convergentes quant au statut massif ou comptable de bon nombre de noms (hors contexte) ;
2. Beaucoup de noms présentent néanmoins un caractère morphosyntaxique flexible, s'accommodant tantôt du massif et tantôt du comptable, dans certains contextes, et avec certains effets de sens (l'une des options apparaissant alors éventuellement comme plus marquée que l'autre) ;
3. Ces effets de sens sont généralement systématiques, et apparaissent cohérents au sein de certaines classes sémantiques, indépendamment des préférences individuelles de tel ou tel item.

Pour ce dernier point, nous avons en tête les généralisations qui ont été faites en termes de machines de transfert, de même que le parallèle entre des sens comme ceux d'*agneau* et de *kangourou* dans *J'ai mangé de l'agneau* et *J'ai mangé du kangourou* (voir au point 1.5.1).

Dans la mesure où nous supposons que les préférences reconnaissables de tel item ou de tel autre pour l'un des deux types de syntaxe ne se peut se passer de quelque encodage ou spécification dans le système linguistique – de quelque marquage lexical, donc – il sera également crucial de définir la nature de cet encodage, et ce à deux niveaux : premièrement, quant à son caractère syntaxique (c'est-à-dire combinatoire) ou sémantique (c'est-à-dire référentiel), et deuxièmement – dans la mesure où cet encodage est effectivement combinatoire – quant à son corollaire distributionnel précis. Par ce second point, nous entendons, à l'intérieur de la perspective *usage-based* dans

laquelle nous nous inscrivons, les caractéristiques de l'input linguistique responsable de tel ou tel marquage. L'on peut ainsi se poser la question suivante : le statut morphosyntaxique d'un nom dans le lexique découle-t-il simplement des emplois dans lesquels il est attesté, ou faut-il intégrer au marquage lexical également des considérations de fréquence ?

Pour finir, la question de la représentation de l'opposition massif-comptable, même en tenant compte de tous les faits que nous avons listés, ne peut se résoudre sans que ne soit redéfini, de manière spécifique et satisfaisante, ce que l'on entend par « lexique ». Ce que la confrontation entre la vision lexicaliste et grammaticale a surtout permis de mettre en avant, c'est l'absence soit d'une définition consensuelle, soit d'une définition précise de cette notion. Dans notre travail, nous creuserons donc la question, en étendant notamment la main vers les acquis de la recherche psycholinguistique, où les résultats de l'étude détaillée du lexique mental, ces trente dernières années, et en particulier l'avènement de « modèles à activation » (*activation models*, voir au Chapitre 3), nous fourniront des pistes cruciales.

## 1.6 Le point de vue de l'acquisition

Pour clôturer notre tour d'horizon de la recherche effectuée au sujet de l'opposition massif-comptable, il nous reste un dernier domaine à explorer, dont nous n'avons encore rien dit : celui de l'acquisition. Si ce champ d'investigation est plutôt en marge des préoccupations de la linguistique générale, et se voit généralement classé du côté des approches psycholinguistiques, nous entendons qu'il peut néanmoins nous donner des indications cruciales quant à l'organisation de notre connaissance linguistique, et du même coup de la représentation dans notre lexique mental de la distinction. Dans cette partie, nous présentons donc dans les grandes lignes les principaux résultats en la matière – afin de pouvoir plus commodément y faire référence par la suite.

Une idée fondatrice de la recherche sur l'acquisition du langage, liée à celle de déterminisme linguistique, est la notion de *syntactic bootstrapping* : l'idée que l'acquisition de distinctions syntaxiques (par exemple entre les différentes parties du discours) aide à façonner la compréhension des distinctions sémantiques correspondantes (Brown 1957). Dans le cadre de l'étude de l'opposition massif-comptable, pareille vision a été défendue par (Quine 1960), qui suggérait que l'apprentissage d'une distinction entre noms massifs et noms comptables aiderait l'enfant à apprendre la distinction ontologique entre objets et substances. Cette position a cependant été remise en doute et réfutée par la suite, au travers d'arguments logiques comme des résultats de diverses expériences (Soja *et al.* 1991, Soja 1992, Barner et



Snedeker 2005). Principalement, il a été montré à de nombreuses reprises que les jeunes enfants maîtrisaient la distinction ontologique entre objets et substances bien avant de maîtriser celle toute morphosyntaxique entre massif et comptable. (Ce qui n'exclut pas, en revanche, que l'hypothèse de *syntactic bootstrapping* reste intéressante dans d'autres domaines, proches : ainsi, (Sarnecka *et al.* 2007) ont montré par exemple que les enfants apprenant une langue faisant la distinction entre singulier et pluriel, comme l'anglais ou le russe, étaient plus rapides à apprendre le sens des concepts *un*, *deux* et *trois* que les apprenants du japonais, où le marquage du nombre n'existe pas.)

Répondant à l'hypothèse de *syntactic bootstrapping*, on trouve l'hypothèse inverse appelée *semantic bootstrapping* : l'idée qu'au contraire, ce seraient des distinctions sémantiques fondamentales qui guideraient la création des catégories grammaticales (Macnamara 1982). En ce qui concerne notre sujet, l'opposition entre objets et substances serait ainsi à l'origine de l'opposition entre comptable et massif – une correspondance qui disparaîtrait ensuite avec l'âge (pour permettre les emplois plus marqués, tels que *meublier* – où un nom massif désigne des objets).

Pour évaluer cette proposition, plusieurs recherches ont examiné spécifiquement les mécanismes d'apprentissage de l'opposition massif-comptable, afin de déterminer l'importance de divers facteurs dans son acquisition. Ainsi, (Gordon 1985) plaide explicitement contre la vision de (Macnamara 1982), en démontrant de manière expérimentale que la représentation de l'opposition ne diffère pas entre enfants et adultes, et qu'aussitôt que l'âge de 2 ans (le plus jeune âge où il est possible de le tester), elle n'est pas encodée en termes d'une distinction entre objets et substances. Au contraire, dès le plus jeune âge, la distinction apparaît comme premièrement formelle, puisque son acquisition repose cruciallement sur l'identification de *syntactic cues*, de signaux syntaxiques, tandis que les *semantic cues*, signaux sémantiques, jouent un rôle bien plus limité. Presque exactement au même moment, (Gathercole 1985) fait des observations toutes similaires, dans son étude de l'apprentissage de *much* et *many*, dont elle conclut que l'opposition massif-comptable est apprise d'abord sur base distributionnelle – les facteurs sémantiques n'entrant en ligne de compte que bien plus tard, vers l'âge de 8 ans, lorsque sont aussi observés les premiers emplois flexibles. Ces résultats sont compatibles avec ceux de (Subrahmanyam *et al.* 1999), qui étudient l'influence des facteurs ontologiques, perceptuels et syntaxiques sur l'acquisition de nouveaux mots désignant des entités physiques, et plus précisément sur la généralisation de leur référence à partir de la forme ou de la matière. (C'est une question cruciale en recherche sur l'acquisition du lexique : si un enfant apprend les mots *cube* et *pâte*, il doit savoir que le premier s'applique à différents référents ayant la même forme mais éventuellement une autre texture, et le second à différents référents ayant la même texture mais éventuellement une autre forme.) A partir de trois expériences, (Subrahmanyam *et al.* 1999) observent que dans leur groupe d'enfants de 3 ans, ce sont en premier les traits ontologiques et perceptuels qui guident la

généralisation référentielle : les mots nouveaux utilisés pour désigner des objets (formes angulaires bien définies, matière dure) sont généralisés à d'autres objets de même forme, ceux utilisés pour désigner des substances (forme de présentation vague, matière fluide) à d'autres occurrences de la même substance – indépendamment de la syntaxe employée. Ce n'est qu'à partir de 4-5 ans que l'emploi morphosyntaxique (massif, comptable ou neutre) commence à jouer un rôle, en permettant d'outrepasser les préférences perceptuelles – un emploi en syntaxe massive augmentant par exemple la probabilité d'une généralisation à partir de la matière, même dans le cas d'un référent objet. A l'instar des résultats de (Gordon 1985) et (Gathercole 1985), ceci indique que l'opposition morphosyntaxique entre massif et comptable est d'abord apprise de manière combinatoire, et que la correspondance éventuelle avec l'opposition perceptuelle entre substances et objets n'interviendrait que plus tard (sous la forme de généralisations progressives entre propriétés référentielles et traits syntaxiques, à partir de régularités statistiques, voir (Samuelson et Smith 1999)).

Ces résultats sont particulièrement intéressants, car ils donnent une indication de l'importance de deux facteurs cruciaux concernant l'opposition massif-comptable : d'une part, la « distribution », c'est-à-dire la fréquence à laquelle nous rencontrons un item dans tel ou tel type de syntaxe, et ensuite le « sens » qui (et c'est intéressant) n'intervient que plus tard dans l'apprentissage, mais module l'usage une fois les données distributionnelles intégrées. Nous ne manquerons pas de faire référence à ces notions au moment de nous pencher sur notre propre modélisation de l'opposition.

## Chapitre 2

# Le sens, et la question de la polysémie

### 2.1 Introduction : multifonctionnalité du signe linguistique

Le Chapitre 1 comportait une introduction aux acquis ainsi qu'aux principales questions à résoudre concernant l'opposition massif-comptable et sa théorisation en langue. Il constitue dès lors la portion principale de la définition de notre objet d'étude et de notre état de la question, qui sont au cœur de la première partie du présent travail. Cependant, avant d'attaquer nos propres données et résultats, nous devons poser le cadre quant à deux autres sujets, immanquablement liés à notre objet de recherche : d'une part la question du sens et de la polysémie, de l'autre la question de la modélisation du lexique. Le premier de ces deux sujets fera l'objet du chapitre présent.

Considérons quelques énoncés :

- (63) La brebis a donné naissance à trois petits agneaux.
- (64) A Pâques, dans ma famille, nous mangeons toujours de l'agneau.
- (65) Après le passage du camion, il avait de l'agneau partout sur l'autoroute.
  
- (66) Dans toutes les soirées estudiantines, on sert de la bière à un prix imbattable.
- (67) Maxx était si fatigué qu'après avoir bu deux bières, il s'est endormi sur la table.
- (68) Je ne connais que deux bières françaises : la 1664 et la Kronenbourg.

Laissons pour ce qu'elle est notre méconnaissance totale de la tradition brassicole française – car nous avons bien conscience que la 1664 étant brassée par le groupe Kronenbourg, ces deux noms ne désignent même pas réellement deux bières différentes. Dans les exemples (63) à (65) d'une part, et (66) à (68) d'autre part, les items *agneau* et *bière* sont utilisés à chaque fois avec une valeur (ou fonction) sémantique différente : *agneau* désigne tour à tour 'un animal, le petit de la brebis', 'une sorte de viande, issue de cet animal' et 'le résultat du broyage maladroit ou cruel d'un ou plusieurs de ces animaux', tandis que *bière* désigne respectivement 'un breuvage brassé',

‘une portion de ce breuvage’ et ‘une sorte de ce breuvage’. A ce stade, ne disons encore rien de ce que cette variation signifie (en termes de référence) – constatons seulement qu’un même item lexical se retrouve utilisé dans divers contextes avec différentes valeurs, et que cette variation est intimement liée à la variation morphosyntaxique, entre emploi massif et comptable. Ceci nous montre que notre étude de l’opposition massif-comptable ne peut se passer d’un examen de cette question de la « multifonctionnalité du signe linguistique » ou « polysémie » (la première appellation étant plus prudente, la seconde peut-être plus courante mais reposant déjà sur des prises de position théoriques plus précises). Cette question est évidemment extrêmement vaste, et a déjà fait l’objet d’innombrables études – rendant hors de propos toute prétention présente de lui accorder un traitement exhaustif. Néanmoins, nous devons nous pencher sur certains de ses aspects – dans la mesure où ils sont pertinents pour notre sujet (comme le laissent entendre nos exemples (63) à (68)). Dans ce qui suit, nous prendrons comme points de départ principaux l’ouvrage sur la polysémie de (Kleiber 1999) et l’article sémiotique de (Willems 2013).

## 2.2 Kleiber 1999 : approche lexicale polysémiste

(Kleiber 1999) réalise une étude très détaillée du phénomène de polysémie, comprenant une profonde réflexion sur la question du sens, une redéfinition de la polysémie en tant que phénomène linguistique et descriptif, un examen critique des différentes théories sur la question, et de nombreuses pistes de traitement – auxquelles il nous est impossible de faire honneur dans le présent ouvrage. Au cœur de l’ouvrage se trouve la profession d’adhérence à une conception référentielle de la sémantique : pour (Kleiber 1999), le sens ne peut être considéré indépendamment d’une relation de référence, liant le contenu sémantique de telle ou telle forme linguistique à une réalité extralinguistique – à un « monde extérieur ». Il s’oppose dès lors fermement à une vision strictement différentielle et aréférentielle du sens, comme on la retrouve dans les travaux de Culioli, Franckel & Paillard ou encore Rastier : pour Kleiber, la langue est immanquablement en lien avec notre perception de la réalité, que nous prenons comme étant la réalité elle-même.

En cela, (Kleiber 1999) défend aussi une sémantique des sens lexicaux encodés, stables et fixes. Il s’oppose à la vision constructiviste et contextualiste de la sémantique, défendue par l’école de Culioli, qui voudrait que chaque sens encodé ne soit qu’un indice schématique et abstrait d’un sens immanent toujours reconstruit en contexte. C’est vrai, tout sens n’est pas stable – mais il ne faut pas pour autant refuser un statut lexical à ses aspects qui le sont. Ainsi, pour Kleiber, si certains sens atteignent un certain degré de

conventionalité, qui peut se mesurer à leur degré de partage intersubjectif, alors il faut leur accorder une place dans le système linguistique, et les décrire en tant que tels.

Du point de vue de l'opposition massif-comptable, qui fait l'objet de son Chapitre IV, « Du *veau* comptable au *veau* massif, histoires de polysémie », cette position correspond à une vision lexicaliste traditionnelle : les noms sont lexicalisés avec une étiquette morphosyntaxique liée à leur sens, et dans les cas où un nom a plusieurs sens différents, reliés d'une certaine manière mais suffisamment distincts, il est dit polysémique – ces différents sens sont représentés séparément dans le lexique, avec la spécification morphosyntaxique correspondante. Ainsi, *veau* correspondrait à deux entrées lexicales : un *veau* comptable, désignant le petit de la vache, et un *veau* massif, désignant sa viande, de même que *lapin* correspondrait d'une part à un *lapin* comptable désignant un mammifère à longues oreilles, d'autre part à deux *lapin(s)* massifs, l'un désignant de la fourrure, l'autre de la viande. En spécifiant dans le lexique ces différents emplois, Kleiber prend aussi distance d'une certaine tendance qu'il observe à tenter d'expliquer la polysémie par une batterie de mécanismes de transferts – qu'il juge trop puissants. Pour sa part, il limiterait les cas de transfert productif à ceux comme en (65), où par exemple un référent comptable est effectivement littéralement broyé, devenant ainsi massif par transformation.

Il y a donc deux cas de figures quant à la flexibilité observée de l'opposition massif-comptable : d'une part quelques possibilités de transfert systématique, et d'autre part beaucoup de polysémie lexicale. Kleiber en veut pour preuve l'application des tests d'ambiguïté (cf. (Zwicky et Sadock 1975)), comme dans l'énoncé (69) par exemple, tiré de (Nunberg et Zaenen 1992: 390), et qu'il juge n'être « pas aussi acceptable qu'ils le pensent » (Kleiber 1999: 116) :

(69) ?Ma religion m'interdit de manger et de porter du lapin.

Si l'énoncé était acceptable, alors l'on pourrait effectivement qualifier de vague le sens de *lapin*, c'est-à-dire sous-spécifié quant à ses différents emplois. Kleiber juge, quant à lui, que l'énoncé ne fonctionne pas, ou alors seulement par zeugma, et que *lapin* est bel et bien ambigu quant à ses différentes lectures – qui sont donc lexicalisées.

En cela, il critique ouvertement la position unitaire de (Nunberg et Zaenen 1992), qui pour leur part postulent, en ce qui concerne les noms d'animaux, une unique fonction de transfert du comptable au massif (à un niveau lexicologique), dont le sens homogénéisé résultant est sous-spécifié quant à ses différents emplois, et seulement précisé plus avant en contexte (au niveau lexicographique). Ainsi, s'ils proposent l'exemple en (69) ci-dessus, c'est au contraire pour conclure à son acceptabilité – démontrant selon eux le caractère sous-déterminé du *lapin* massif. Ses différents usages, en termes de viande ou de fourrure, sont quant à eux précisés au niveau pragmatique, répondant bien à des conventions spécifiques à chaque langue (donc supposant bien un certain encodage), mais situées en-dehors du lexique. (C'est peut-être ici que le bât

blesse : pour (Kleiber 1999), l'existence d'un sens stable, partagé inter-subjectivement, implique la lexicalisation – si tout le monde sait que *du veau* s'utilise, dans tel et tel contexte, spécifiquement pour désigner 'de la viande de veau', alors ce sens doit être représenté comme tel dans le lexique. Chez (Nunberg et Zaenen 1992), l'aspect conventionnel est plus nuancé, puisqu'ils distinguent un niveau lexicologique et lexicographique. C'est là une piste intéressante : nous y reviendrons ci-après, en discutant les positions de (Willems 2013.)

Avant de poursuivre, revenons un instant sur la critique de (Kleiber 1999) à l'encontre de la vision unitaire du sens, de (Nunberg et Zaenen 1992). L'un des arguments de Kleiber est le suivant : si tout emploi massif d'un nom d'animal a pour sens une présentation « broyée » du référent, comment expliquer que précisément la lecture proprement broyée, comme dans l'exemple du camion ayant roulé sur le corps d'un malheureux agneau en (65), soit celle qui nécessite, pour être acceptable, le plus grand déploiement de détails contextuels ? Or, il nous semble que sur ce point, (Kleiber 1999) n'a peut-être pas tout à fait compris (Nunberg et Zaenen 1992) : à notre sens, l'opération de *grinding* qu'ils définissent n'est pas en termes d'un moulinage littéral, comme c'était le cas chez (Pelletier 1975), mais bien simplement une opération sémantique allant d'un individu à 'de la matière dérivée de ce dernier' (voir aussi ci-dessus, point 1.5.2). Il est vrai qu'ils donnent l'exemple du *lapin* écrasé après le passage du camion comme emblématique de ce transfert – mais uniquement dans la mesure où, effectivement, ce contexte ne nécessite pas d'inférences plus précises quant à la partie du lapin dont il est question : ce qui reste sur l'autoroute, c'est seulement de la matière indéterminée, issue du rongeur malchanceux. Ils donnent du reste deux autres arguments, qui laissent bien entendre que leur sens unitaire massif n'est pas une présentation comme 'matière broyée', mais bien simplement comme 'matière dérivée, de quelque manière que ce soit'. Premièrement, ils observent que si je dis (70), le sens du SN ne peut pas être limité à de la viande :

(70) Ma religion m'interdit de manger du lapin.

En effet, il y a fort à parier que la religion en question n'établit pas d'exceptions quant à manger des poils, des griffes ou des moustaches (Nunberg et Zaenen 1992: 390). Qu'il s'agisse de viande lorsque je déclare que j'ai *mangé du lapin* découlerait plutôt d'une application de notre connaissance des normes et habitudes quant à la consommation de nourriture, plutôt que d'une spécification sémantique lexicale.

Deuxièmement, outre l'exemple du camion, ils donnent l'énoncé suivant (71), comme tout aussi emblématique de leur transfert sous-spécifié :

(71) Ne touchez pas au lapin importé de la région de Tchernobyl.

Dans l'exemple original en anglais, on trouve l'usage nu (résolument massif, donc) de *rabbit*, que nous traduisons ici avec un article défini pour mieux rendre le côté

générique. Dans cet exemple, il est évident qu'il ne s'agit pas spécifiquement de substance lapine effectivement broyée – il s'agit simplement de 'n'importe quelle substance, dérivée d'un ou plusieurs lapins venus de Tchernobyl'.

Ce point de critique de (Kleiber 1999), qui juge paradoxal l'accès plus difficile à la lecture broyée alors qu'elle est supposée plus proche du sens général unique, n'est donc pas tout à fait pertinent, nous semble-t-il. Pour la lecture massive de *lapin*, l'interprétation en termes de viande ou de fourrure s'impose plus facilement à l'esprit, mais il peut s'agir là simplement de conventions d'usage : ces emplois sont aussi les plus fréquents, tant du point de vue linguistique, que de notre connaissance du monde. Il nous faudra pouvoir l'expliquer, bien entendu, mais ceci ne remet pas nécessairement en doute l'existence d'un sens unitaire pour *du lapin*, *de l'agneau*, ou *du veau*, à un certain niveau du système linguistique. Nous y reviendrons.

D'autre part, la méfiance de (Kleiber 1999) à l'égard des sens unitaires se comprend bien, à la lumière de sa critique des approches différentielles et constructivistes du sens qui, en regroupant sous des chapeaux d'invariants sémantiques schématiques les sens les plus divers, finissent par refuser tout encodage à des polysémies pourtant clairement identifiables inter-subjectivement. C'est en réaction à ces approches que Kleiber entend donner une place légitime dans le système linguistique à différents sens, comme 'viande' pour *du veau* ou 'fourrure' pour *du lapin*, qui sont suffisamment courants que pour mériter quelque qualité préconstruite dans le lexique.

Pourtant, le débat ne doit pas seulement viser à trancher entre accorder ou refuser un caractère encodé, préconstruit, à tel ou tel emploi. Nous venons déjà d'effleurer l'idée : que certains sens spécifiques découlent d'opérations pragmatiques secondes n'empêche pas que celles-ci soient conventionnelles également. (On peut bien sûr refuser l'appellation de pragmatique à toute opération de sens conventionnelle, comme le fait (Nicolas 2002a), mais l'on ne se trouve alors plus que dans des considérations terminologiques.) Un tableau plus nuancé, distinguant le niveau du système de celui de l'usage – tous deux sujets aux conventions, mais de natures différentes – semble en la matière un cadre propice à une meilleure compréhension des tensions entre sens pré-donnés, et sens productifs construits. Ceci nous amène à la position de (Willems 2013).

## 2.3 Willems 2013 : approche monosémiste

Dans son article, (Willems 2013) réalise une investigation sémiotique du sens linguistique, prenant comme point de départ une évaluation critique de la théorie du Lexique Génératif de (Pustejovsky 1995). A nouveau, nous n'avons pas pour objectif de faire un compte-rendu complet de ce travail : cela dépasserait largement les limites de

notre sujet. Néanmoins, il y a quelques idées cruciales à glaner qui nous seront utiles. (Willems 2013) se place lui-même du côté d'une approche monosémiste du sens, suivant l'idée fondamentale de Davidson – rejetée par ailleurs par (Pustejovsky 1998) :

*Language is the instrument it is because the same expression, with semantic features (meanings) unchanged, can serve countless purposes. (Davidson 1984: 108)*

La langue est l'instrument qu'elle est parce que la même expression, sans changement dans ses traits sémantiques (significations), peut servir d'innombrables fonctions. (traduction personnelle)

Ainsi, pour (Willems 2013), les sens encodés dans le lexique, qu'il appelle *meanings* 'significations', sont uniques et stables (d'un point de vue synchronique). Ainsi, d'après lui, le Lexique Génératif, qui pourtant se veut économique, reste trop spécifié (*overspecified*), au même titre que les *sense enumeration lexicons* 'lexiques à énumération des sens', que pourtant Pustejovsky critique. Dans une certaine mesure, la position de (Willems 2013) est donc à rapprocher des visions constructivistes de la tradition française : au niveau du système, les sens sont généraux et abstraits, c'est-à-dire sous-déterminés quant à leurs diverses réalisations possibles en contexte.

Or, (Willems 2013) rappelle que ce niveau de base est sous-jacent à un second niveau : celui du discours, ou de la langue dans son emploi – et que la distinction entre les deux niveaux ne doit pas être brouillée arbitrairement. Ainsi, dans l'absolu, les différentes interprétations contextuelles de telle ou telle forme linguistique sont toujours équiprobables (car dérivée d'un sémantisme unique et abstrait), mais elles apparaissent non équivalentes une fois réalisées en contexte, car elles y sont sujettes aux conventions de l'usage – qu'il ne faut pas confondre avec le sémantisme lexical. Ainsi, la critique à l'encontre du Lexique Génératif est qu'il échoue à faire une distinction claire entre ce qui relève de la 'signification sémantique' (*semantic meaning*) et des 'sens pragmatiques' (*pragmatic senses*), en intégrant de manière problématique à la description sémantique, des éléments de la théorie du discours. C'est notamment l'erreur des tenants du Lexique Génératif que d'oublier que les interprétations 'par défaut' sont tout autant que les autres des interprétations contextuelles – fussent-elles ancrées sur l'un ou l'autre « contexte conventionnalisé » (Willems 2013: 250). Ainsi, pour Willems, le Lexique Génératif n'est pas tant une théorie sémantique qu'une ébauche de théorie des normes d'usage du langage (*normal language use*, au sens de (Coşeriu 1952)), en ce qu'il s'applique à décrire non pas tous les sens possibles des mots, mais seulement leurs sens les plus typiques.

En conclusion, (Willems 2013) défend une position unitaire de la signification lexicale encodée, et appelle à étudier la polysémie « dans le domaine de l'interprétation linguistique, où se rencontrent les inférences pragmatiques, la connaissance extralinguistique et, en particulier, les normes d'usage » (Willems 2013: 283, traduction



personnelle, légèrement adaptée). L'on n'est pas loin de la distinction entre niveau lexicologique (où se situent les significations encodées, unitaires) et lexicographiques (où l'on peut décrire la variation spécifique, contextuelle) de (Nunberg et Zaenen 1992). D'autre part, l'approche n'est plus tellement incompatible non plus avec celle de (Kleiber 1999). L'opposition cruciale est entre ce que l'on appelle encodé lexicalement, et ce que l'on appelle dérivé de manière conventionnelle – les conventions elles-mêmes étant, si l'on ne veut pas dire « encodées » (par fausse pudeur terminologique), au moins tout autant partagées de manière intersubjective. Pour (Kleiber 1999), c'est précisément ce partage qui définit le sens lexical – or, (Willems 2013) ouvre une porte de conciliation entre approche unitaire et polysémiste du sens : en situant la première dans le lexique, et la seconde dans l'usage.

Notons que cette séparation permettrait peut-être aussi de cadrer les propositions de (Cruse 1996, Croft et Cruse 2004, Cruse 2004) concernant la décomposition des concepts sémantiques en facettes. Leur explication concerne le caractère volatile de noms tels que *livre*, qui peut s'employer tantôt pour désigner un texte (*J'ai lu un livre très intéressant*), tantôt un objet (*Un livre m'est tombé sur la tête*), ou *porte*, qui est tantôt une ouverture (*Ils sont entrés par la porte*), tantôt un pan mobile (*Paul s'est pris la porte en pleine figure*). A leur sens, il n'est pas question ici de polysémie, comme en atteste par exemple l'emploi non syllephtique de : *Ce livre est très lourd à trimballer, mais extrêmement intéressant et bien écrit !* Pour expliquer la multifonctionnalité de pareils items, ils postulent dès lors que le sens des noms puisse être décomposé en différentes facettes, pouvant être activées ensemble sous forme d'un « concept global unifié » (sans ambiguïté, comme dans *J'ai acheté un livre*), mais aussi séparément, selon le contexte. Ainsi, un énoncé tel que *J'ai lu un livre très intéressant* activerait seulement la facette 'texte' de *livre*, et non la facette 'tome, objet'. Comme le remarque (Kleiber 1999), il s'agit là d'une explication assez similaire aux « types pointés » du Lexique Génératif de Pustejovsky.

La faiblesse de la proposition, d'après (Kleiber 1999: 96-98), est la dualité des facettes : certaines seraient représentées de manière permanente dans le lexical mental, d'autres seraient générées *ad hoc*, sous l'influence du contexte. Ainsi, l'item *roman* contiendrait la facette permanente 'texte', mais pourrait aussi obtenir, en contexte, la facette transitoire 'tome, objet', qui n'a pas exactement les mêmes propriétés (Cruse 1996: 95-96). Or, Kleiber ne voit pas pourquoi, dans ce cas, les facettes ne seraient pas toutes contextuelles. Il nous semble ici que la remise au point de (Willems 2013) pourrait s'avérer utile : peut-être n'y a-t-il pour *porte* ou *livre* au niveau lexical effectivement qu'un concept unique et unitaire, et les facettes n'entrent-elles en ligne de compte qu'au niveau de l'usage ? Que certaines paraissent plus contextuelles que d'autres serait alors simplement une question de conventionalité plus ou moins grande – au même titre que pour *veau*, le sens 'viande' soit plus habituel et donc plus accessible que le sens 'matière broyée'. L'objectif de notre travail n'étant bien sûr pas de régler une fois pour toute la question de la multifonctionnalité du signe linguistique, nous n'entrerons ici pas plus

dans les détails. Néanmoins, nous referons référence aux facettes de (Croft et Cruse 2004) plus loin, lorsque nous aborderons notre propre modélisation du lexique.

## 2.4 Multifonctionnalité et massif-comptable : pistes

En tous les cas, la possible conciliation entre approche unitaire du sens lexical et approche polysémiste des sens en discours ouverte par (Willems 2013) nous semble particulièrement cruciale pour notre propre objet d'étude. Car quoi qu'on en ait dit, dans le domaine du massif-comptable, une certaine vision monosémiste de la signification sera nécessaire. Premièrement, il y a le test d'ambiguïté de (Nunberg et Zaenen 1992), en (69) : (Kleiber 1999) le présente comme peu acceptable, mais nous ne sommes pas certains qu'il ait raison. Pour nous, l'énoncé est possible, et il faudra expliquer pourquoi. En outre, l'on peut trouver ou imaginer d'autres énoncés du même type, qui remettent en doute une polysémie stricte, comme (72), qui nous a été inspiré à la lecture de (Kleiber 1999) justement, (73) que nous adaptons d'une conversation réelle entendue au détour d'une soirée scout, (74) que nous inventons mais qui ne nous semble pas poser le moindre problème d'acceptabilité, ou pour finir (75) et (76), en anglais, que nous tirons d'internet.

(72) Ce veau est très coûteux à l'élevage, mais qu'est-ce qu'il est savoureux !

(73) C'est de la bière à la pompe ? Eh bien, mettez-m-en deux !

(74) Je voulais ajouter de la tomate dans la salade, mais je n'en ai plus qu'une au frigo, ce qui ne sera pas suffisant.

(75) *I've never had lamb or veal. I don't know if I dare try it. They're too cute! (Google)*

'Je n'ai jamais mangé d'agneau ou de veau. Je ne sais pas si j'oserais goûter ça. Ils sont trop mignons !'

(76) *I would not eat dog or cat, because they are too cute. (Google)*

'Je ne pourrais pas manger de chien ou de chat, car ils sont trop mignons.'

Dans (72), c'est bien l'animal qu'on élève, mais sa viande qui est savoureuse : *veau* semble ici donc bel et bien sous-déterminé entre ses différents emplois – au-delà même de la barrière morphosyntaxique entre comptable et massif. C'est la même chose pour *bière* dans l'énoncé qui suit : la reprise par le pronom numéral *deux* laisse entendre que lorsqu'on commande *deux bières*, le sens fondamental de *bière* est le même que lorsqu'on parle de *la bière*, massive, dans le tonneau. Pareil pour le lien entre *de la tomate* et *une tomate*, dans (74). A noter que nous ne prétendons pas que l'interprétation sémantique de *de la tomate* ou *de la bière* soit identique à celle d'*une tomate* ou d'*une bière*, au contraire

(nous y reviendrons dans nos études de corpus, dont l'une est dédiée aux noms de fruits et légumes, et une autre aux noms de boissons). Mais il semblerait bien qu'il y ait, au-delà de la différence, un invariant sémantique au niveau des lexèmes *tomate* et *bière* qui sous-tende leurs divers emplois (massifs, comptables), et qu'une séparation stricte en termes de polysémie lexicale n'explique pas. Remarquons du reste qu'on n'est pas, dans (73) ou (74), dans une interprétation sylleptique telle qu'on l'aurait en (77) :

(77) \*Quand j'ai appris que l'espace interstellaire ne contenait pas d'air, j'en ai pris un vraiment étonné !

Il semblerait donc bien qu'il soit légitime de parler, en ce qui concerne nos différents items morphosyntaxiquement flexibles, d'une signification unitaire (vague, sous-déterminée) reliant les différents sens contextuels, plutôt que de différents sens lexicalisés de manière strictement séparée. Cela semble même le cas lorsque les différents sens apparaissent aussi conventionnels l'un que l'autre, comme dans le cas d'un *veau* 'animal' et du *veau* 'viande' (72), ou de la *bière* 'boisson' et une *bière* 'verre, portion' (73) – ce qui semble plaider en faveur d'une séparation en deux niveaux (lexique et discours), au sens de (Willems 2013). Il faudra en tenir compte dans nos propres propositions.

Ces constats et réflexions ouvrent différentes pistes quant à notre recherche. D'une part, dans nos études de corpus, nous serons attentifs à fournir une description précise des effets de sens associés à la variabilité morphosyntaxique des items étudiés, et des contextes dans lesquels ces différents sens apparaissent. D'autre part, nous veillerons aussi à généraliser ce qui peut l'être, en distinguant ce qui relève du sémantisme profond de ce qui est précisé de manière pragmatique, en contexte. Sur un autre plan, nous distinguerons aussi la flexibilité réelle, découlant de mécanismes productifs de transfert, de la flexibilité uniquement apparente résultant de la lexicalisation de divers emplois. Pour ces différentes questions, l'étude parallèle du français et du néerlandais s'avérera cruciale, en permettant de séparer ce qui semble généralisable de manière inter-linguistique, de ce qui est propre à l'un ou l'autre système linguistique.

Finalement, dans notre modélisation du lexique – puisque nous avons dit qu'un modèle précis du lexique était indispensable pour comprendre et cadrer la question de la représentation du massif-comptable dans le système linguistique – nous veillerons à se faire rencontrer ces différents aspects, en donnant leur place aux significations lexicales unitaires, à leurs éventuelles différentes facettes, et à leurs sens contextuels plus précis, plus ou moins conventionnels. Nous devons aussi parvenir à montrer ce que signifie l'association entre certains sens (lexicalisés ou non), et un certain emploi morphosyntaxique. Cette étape ne sera complète bien sûr qu'à l'issue de la présentation de nos données (corpus, enquêtes et expérience), puisqu'elle reposera de manière cruciale sur nos observations empiriques, néanmoins elle est préfigurée au chapitre suivant – le dernier morceau de cette première partie introductive – où nous abordons

précisément la question de la modélisation du lexique, en linguistique générale mais surtout dans la tradition psycholinguistique, dont nous entendons tirer quelques enseignements.

# Chapitre 3

## Modèles du lexique mental

### 3.1 Introduction : limites de la vision traditionnelle

Nous avons parlé, au Chapitre 1, point 1.5, de la localisation de l'opposition massif-comptable, et de sa représentation dans le système linguistique. Cette question concerne principalement celle de la présence, dans le lexique ou non, des traits massif-comptable, ainsi que la nature (syntaxique et/ou sémantique) de ceux-ci. Or, il semblerait que les acteurs de ce débat, lorsqu'ils discutent le caractère plus ou moins lexical du massif-comptable, prennent pour point de départ une vision traditionaliste du lexique, qui n'est pas vraiment bien définie. La vision structuraliste (et générativiste) d'un langage composé d'une part d'une liste de règles, d'autre part d'un stock d'items, s'est vu largement critiquée et remise en question dans les paradigmes théoriques récents (rejet de la *rule-list fallacy*, (Langacker 1990:264) ; Grammaire des Constructions, (Goldberg 2006, Goldberg et Suttle 2010)) – pourtant, c'est bel et bien la vision qui prédomine, lorsque la question est résumée à situer le massif-comptable de l'un ou l'autre côté de cette dichotomie artificielle. Et lorsqu'un certain effort est fait pour définir ce qu'on entend par lexique, la définition n'est pas satisfaisante. : nous l'avons dit, (Borer 2005) exclut par définition l'opposition massif-comptable du lexique, car elle est flexible – mais au final, cette approche ne nous apprend rien de la vraie nature et représentation de notre distinction, dans le système linguistique. Au mieux, lorsque le lexique et ses propriétés sont abordées, l'on suppose *grosso modo* que l'on sait de quoi on parle – mais au vu de la stérilité du débat entre vision lexicaliste et grammaticale, rien n'est moins sûr, et la question d'une redéfinition du lexique n'est dès lors pas anodine.

Parallèlement, comme nous l'avons vu au Chapitre 2, la question de la multifonctionnalité du signe linguistique – dont celle de la variation sémantique associée à la flexibilité massif-comptable fait partie – ne peut pas, elle non plus, se passer d'une définition précise du lexique, puisque l'on a jonglé dans les débats avec les

notions de signification lexicale plus ou moins unitaire, de sens contextuels, et autres facettes, lexicales ou non. (Willems 2013) a rappelé l'importance de distinguer, dans la description des phénomènes de polysémie, les deux niveaux d'analyse que sont celui de la langue et celui du discours – mais il reste à montrer comment à la fois la distinction et l'interaction entre représentations sémantiques lexicales et processus interprétatifs contextuels peuvent être expliqués dans un modèle cognitif global. A ce niveau aussi, la vision naïve d'un lexique sous forme d'une liste d'items avec des propriétés discrètes stables fait défaut : poussée à l'extrême, elle ne permettrait d'expliquer la polysémie qu'en termes d'une duplication ou multiplication d'entrées lexicales homophones et/ou homographes, qui ne rende compte ni de la relation entre sens liés (et potentiellement activés de manière unitaire), ni des effets de profilage partiel de certains concepts lexicaux en contexte (facettes, métonymie intégrée, cf. (Kleiber 1999)), ni même du lien entre variation morphosyntaxique (massif-comptable) et modifications sémantiques. Or, comme l'a bien résumé (Evans 2010), un modèle approprié du lexique mental ne peut se passer 1) de décrire et de représenter la distinction entre sens lexical et contextuel, et 2) de définir la nature des représentations sémantiques.

Bien entendu, il existe en linguistique générale des approches plus modernes du lexique, rompant avec l'image formaliste d'une liste stable doublée de règles dynamiques. Ainsi, les propositions de (Bybee 1988) quant à la représentation lexicale de la morphologie ont ouvert la voie à la tradition des approches *usage-based* du langage, interprétant les règles comme des généralisations à partir de l'expérience, sensibles à la fréquence, ne pouvant s'interpréter indépendamment des représentations les soutenant (voir aussi (Bybee 2006)). Dans pareille approche, ce n'est que sur la base de connexions phonologiques et sémantiques régulières entre items du lexique que peuvent s'établir des schémas productifs, qui sont moins des règles abstraites que des généralisations par analogie, et qui sont renforcés en fonction de la fréquence respective de leurs exemplaires les plus courants. En d'autres termes, le lexique n'est plus présenté comme une liste, mais comme un réseau de relations – dont émergent les règles de la (morpho)syntaxe. C'est là d'emblée une piste de réflexion intéressante, si l'on repense à la flexibilité massif-comptable, souvent décrite en termes de règles de transferts productives – dont la description n'est cependant pas encore entièrement satisfaisante, comme nous l'avons vu (avec en particulier le caractère strictement directionnel à remettre en question).

Dans un autre ordre d'idée, la recherche dans le cadre de la Linguistique Cognitive a jeté les bases d'une approche encyclopédique du sens, abandonnant l'idée d'une distinction modulaire entre sens linguistique et connaissances du monde (Evans *et al.* 2007), et ouvrant des perspectives quant à la compréhension des représentations sémantiques. Ainsi, (Fauconnier 2000) propose que les formes linguistiques ne portent pas en elles-mêmes de réelle signification, mais plutôt fournissent la clé de riches réseaux cognitifs de connaissances. Ainsi, tout mot présenterait un certain potentiel

sémantique, et son sens serait à définir comme la partie activée de ce potentiel, dans un contexte donné (voir aussi (Evans 2006)). Nous n'entrerons pour l'heure pas dans les détails à ce niveau, mais observons seulement que l'idée du sens comme l'accès à un réseau de connaissances, fondé sur notre perception et compréhension encyclopédique du monde, présente, à condition d'être bien définie, une manière intéressante de représenter la vision conceptuelle traditionnelle de la signification, qui a besoin de facettes, mécanismes contextuels et autres dispositifs d'enrichissement pour rendre compte de la multifonctionnalité du signe linguistique. Nous y reviendrons ci-après, de même qu'à la fin de notre parcours de théorisation, au Chapitre 11.

Ainsi, la tradition linguistique générale n'est pas dépourvue de profondes réflexions sur la nature du lexique et de la représentation sémantique. Cependant, à notre connaissance, le fruit de cette réflexion n'a jamais été appliqué en tant que tel à l'opposition massif-comptable où, comme nous le disions, une vision traditionaliste et/ou formaliste du lexique semble prédominer. D'autre part, à notre connaissance, la réflexion sur le sujet dans le cadre d'une Linguistique Cognitive semble indifférente aux développements, souvent parallèles nous semble-t-il, de modèles de plus en plus fins du lexique dans la tradition de recherche psycholinguistique. Or, par contraste avec la linguistique générale où de nombreux paradigmes de recherche se contentent encore de la vision naïve du lexique comme liste, les psychologues du langage ont passé ces dernières décennies à décrire notre connaissance lexicale en termes de structures cognitives de plus en plus riches et complexes, que les linguistes grammairiens gagneraient, selon nous, à exploiter. Nous les introduisons au point suivant, et montrons comment elles pourront nous aider à faire sens de nos propres données sur le massif-comptable, à l'issue de notre parcours de description empirique.

## 3.2 Le lexique mental en psycholinguistique

### 3.2.1 Caractéristiques générales

L'origine de la modélisation du lexique en psycholinguistique remonte à deux traditions complémentaires, en rapprochement constant (Levelt 1999: 223) : d'une part, celle visant à comprendre le langage par l'analyse des erreurs de la parole (*speech error tradition*), d'autre part celle ancrée dans les expériences de nommage d'images (*picture-naming*), entendant rendre compte de l'influence de divers facteurs d'amorçage sur les temps de réaction de locuteurs à divers stimuli (*chronometric tradition*). De chacune de ces traditions sont issus divers modèles, fort proches dans l'ensemble et différant dans les détails. Dans la présente partie, nous présenterons les principales propriétés

communes de ces modèles, et évoquerons aussi les quelques points majeurs sur lesquels ils ont tendance à diverger. Notre objectif en la matière n'est pas d'être exhaustif, mais de poser un cadre dans lequel nous pourrions décrire, au point suivant, le modèle de (Levelt *et al.* 1999), qui nous servira de point de départ pour nos propres réflexions théoriques ultérieures.

Un acquis majeur de l'étude du lexique en psycholinguistique est la division en deux étapes du processus de récupération lexicale (*lexical retrieval*), c'est-à-dire de production d'un mot, par exemple dans le contexte d'une phrase (voir par exemple (Bock et Levelt 1994, Vigliocco *et al.* 1999, Cleland et Pickering 2003)) : d'abord, 1) la récupération d'un « lemme », qui correspond à la représentation schématique du mot sous forme de traits sémantiques et de caractéristiques morphosyntaxiques, puis 2) la récupération d'un « lexème », c'est-à-dire la représentation phonologique associée à ce lemme. Les données en faveur de cette distinction sont nombreuses, et incluent notamment des études sur l'aphasie (Butterworth 1989), ainsi que de manière célèbre les études du *tip of the tongue phenomenon* (TOT) – le phénomène qui se produit lorsqu'on sait que l'on connaît un mot (qu'on l'a « sur le bout de la langue »), et qu'on peut même donner certaines de ses caractéristiques (notamment son genre, dans le domaine nominal), sans pour autant pouvoir produire sa forme finale (Brown et McNeill 1966, Vigliocco *et al.* 1997). Dans ce cas, il est supposé que c'est le passage de la première à la deuxième étape de la récupération lexicale qui est, pour une raison ou une autre, bloqué : un lemme est activé, mais pas le lexème associé. On trouve au moins ces deux étapes de récupération lexicale chez la plupart des auteurs sur le sujet, sous des noms divers. Ainsi, (Garrett 1975) parle de *functional level* et *positional level processing* 'traitement au niveau fonctionnel' et 'positionnel', tandis qu'on trouve par exemple chez (Levelt 1989) les dénominations de *grammatical* et *phonological encoding* 'encodage grammatical' et 'phonologique', pour désigner ces deux étapes.

En général, les modèles du lexique en psycholinguistique sont des modèles à activation. Cela signifie que la connaissance lexicale (comme du reste bien d'autres types de connaissances en sciences cognitives) est représentée sous forme d'un réseau, composé de *nodes* 'nœuds, nodules' comportant de l'information, reliés entre eux par des connexions plus ou moins fortes. Lors de la récupération lexicale, ces nodules sont activés, et l'activation suit les connexions entre nodules pour se propager, jusqu'à aboutir à la retrouvaille et à la production de l'expression voulue. Notons qu'en vertu de leurs propriétés de réseau et de connexions, ces modèles sont parfois appelés « connexionnistes » – même si récemment, comme nous le signale Rob Hartsuiker (communication personnelle), l'habitude a été prise de réserver cette appellation plus spécifique à un sous-ensemble de ces modèles, comme ceux de McRae (voir par exemple (McRae 2004)), que nous ne détaillerons pas ici. Dans ce qui suit, nous utiliserons dès lors l'appellation plus générale de « modèles à activation », pour éviter toute confusion.



Les modèles de ce type se distinguent les uns des autres le long de différents axes, c'est-à-dire selon la position qu'ils adoptent quant à diverses questions cruciales. Nous pouvons en identifier au moins quatre. Premièrement, la question de la stratification : dans certains modèles, la connaissance lexicale est organisée en différentes strates (Dell 1986, Levelt *et al.* 1999), tandis que dans d'autres toute l'information est représentée sur un seul niveau (Caramazza 1997). Deuxièmement, dans les modèles stratifiés se pose la question du flux d'activation, qui est dit tantôt unidirectionnel (Levelt *et al.* 1999), allant dans le cas de la production linguistique strictement du niveau sémantique vers le niveau morpho-phonologique, tantôt interactif et bidirectionnel (Dell 1986), permettant un feedback d'un niveau inférieur vers un niveau supérieur. Troisièmement, la question du caractère séquentiel de l'activation : celle-ci parcourt-elle les différents niveaux de représentation les uns après les autres (Levelt *et al.* 1999), ou peut-elle déborder d'un niveau à l'autre à la manière d'une cascade (Dell 1986, Peterson et Savoy 1998) ? Quatrièmement, et pour finir, la nature du niveau lemmatique a été discutée, certains le décrivant comme strictement syntaxique (Levelt *et al.* 1999), tandis que d'autres y intègrent une composante sémantique non indépendante (Kempen et Huijbers 1983, Zorzi et Vigliocco 1999).

Dans le présent travail, nous ne pourrions évidemment pas nous intéresser aux tenants et aboutissants de ces différentes questions (qui nous mèneraient bien au-delà des limites tant de notre sujet que de notre domaine de spécialisation), et notre objectif ne sera donc pas de fournir un nouveau modèle total du lexique, combinant les acquis des traditions linguistiques et psycholinguistiques dans une grande explication définitive du langage. Néanmoins, nous sommes convaincu que certains acquis des modèles psycholinguistiques, ignorés jusqu'à présent dans les théories linguistiques générales, peuvent constituer un enrichissement non négligeable de notre compréhension de différents phénomènes lexicaux et grammaticaux, notamment de l'opposition massif-comptable, restant jusqu'à présent sans réponse satisfaisante.

Pour cette raison, nous avons sélectionné le très influent modèle de (Levelt *et al.* 1999) et (Pickering et Branigan 1998) comme point de départ de notre réflexion en la matière, afin d'illustrer de manière un peu plus détaillée le fonctionnement d'un modèle à activation, et d'examiner les pistes qu'un tel modèle nous ouvre dans notre propre recherche d'une représentation appropriée du lexique, susceptible d'expliquer les propriétés lexicales, sémantiques et morphosyntaxiques de notre sujet.

### **3.2.2 Le modèle de (Levelt *et al.* 1999), ainsi que ses extensions**

Le modèle du lexique de (Levelt *et al.* 1999), issu de la tradition chronométrique de la recherche sur le lexique mental (voir ci-dessus) et développé comme modèle d'accès lexical dans la production langagière, est un modèle à activation stratifié, modulaire et

séquentiel. Il représente notre connaissance lexicale sous forme d'un réseau à trois niveaux – sémantique (*conceptual stratum*), lemmatique (*lemma stratum*), et morpho-phonologique (*word form stratum*) – dans lequel l'activation se déplace de manière strictement unidirectionnelle. Plus précisément, en ce que seuls le niveau lemmatique et le niveau conceptuel sont communs à la production et à la compréhension langagière, aucun feedback n'est possible ni prévu entre le niveau morpho-phonologique et le niveau lemmatique. (D'autres modèles, en particulier ceux issus de la *speech error tradition*, postulent des mécanismes de feedback à ce niveau, comme par exemple (Dell 1986), mais cette position est généralement considérée comme problématique, voir (Harley 2008: 423-424).) La Figure 1, tirée directement de (Levelt *et al.* 1999: 4), illustre un fragment du réseau lexical tel qu'il est représenté dans ce modèle.

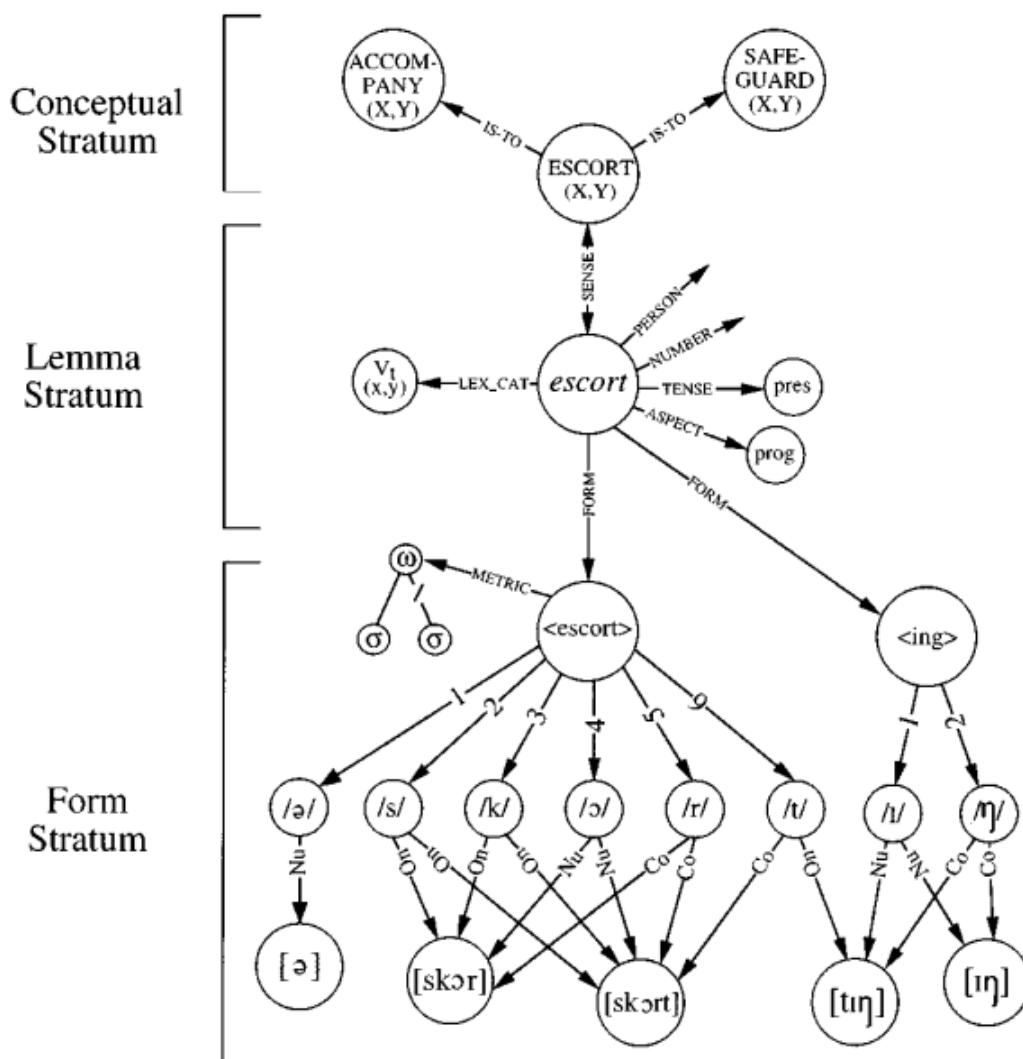


Figure 1 Illustration du modèle à activation du lexique de (Levelt *et al.* 1999: 4, figure 2). Légende originale : « Fragment of the lexical network underlying lexical access. The feedforward activation spreading network has three strata. Nodes in the top, conceptual stratum represent lexical concepts. Nodes in the lemma stratum represent syntactic words or lemmas and their syntactic properties. Nodes in the

form stratum represent morphemes and their phonemic segments. Also at this level there are syllable nodes. »

Dans le cadre de notre propre recherche, c'est moins le niveau inférieur (celui de la réalisation phonologique sous forme de syllabes et des propriétés métriques) qui doit nous occuper que le niveau, crucial en ce qui nous concerne, de la strate lemmatique. Comme nous l'avons dit, c'est là que se situent les caractéristiques morphosyntaxiques des items langagiers, plus précisément sous forme d'informations concernant : la catégorie ou partie du discours (nom, verbe), les traits (nombre, personne, temps), et les possibilités combinatoires. Ce niveau est examiné en détail par (Pickering et Branigan 1998) qui proposent, au travers de différentes expériences d'amorçage structurel (*structural priming*), une extension du modèle de (Levelt *et al.* 1999) au travers d'une précision de l'organisation de cette strate lemmatique, et en particulier du dernier point – celui des informations concernant les possibilités combinatoires des lemmes. (Que l'article de Pickering et Branigan soit daté d'avant celui de Levelt *et al.* n'est pas une incohérence : le premier est paru alors que le second était encore annoncé comme à paraître). Leur étude, concernant l'alternance dative en anglais (datif prépositionnel vs. double objet : *Mary gives a kiss to Paul vs. Mary gives Paul a kiss*), consiste en différentes expériences d'écriture (*written completion tasks*), au cours desquelles ils observent que la sélection de l'une ou l'autre construction dans l'emploi d'un verbe datif est sensible à l'amorçage (*priming*) – c'est-à-dire que la production d'une alternative sera toujours biaisée en faveur de l'alternative précédemment rencontrée – avec un fort *boost lexical* : un effet d'amorçage accru lorsque le même verbe est utilisé comme amorce et comme cible.

Partant d'une extension de (Roelofs 1992, 1993), ils interprètent leurs résultats comme la preuve que les informations syntaxiques du niveau lemmatique, dans leur cas les deux types de constructions du datif en anglais, peuvent être représentées comme des nodules uniques, partagés par tous les lemmes concernés. Leur modèle du niveau lemmatique est illustré à la Figure 2.

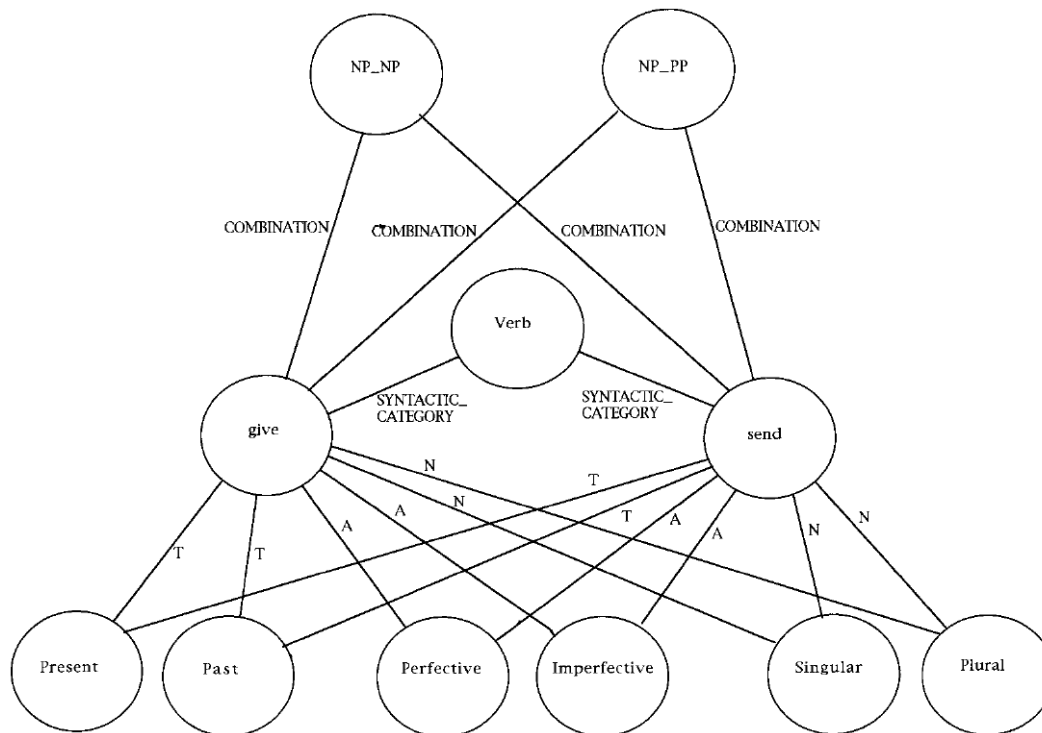


Figure 2 Modèle plus précis du niveau lemmatique de (Levelt *et al.* 1999), tel que proposé par (Pickering et Branigan 1998: 635, fig.1). Les deux nodules supérieurs représentent les deux possibilités de construction du datif en anglais, et sont partagés par les différents verbes en question, illustrés ici par les nodules *give* ‘donner’ et *send* ‘envoyer’. Ces deux nodules sont sous-spécifiés quant à leurs différents traits syntaxiques, eux-mêmes représentés dans d’autres nodules partagés (ici, au bas de la figure).

A noter que si les nodules combinatoires sont dits partagés, cela ne signifie pas qu’ils sont connectés de manière également forte à chaque lemme pertinent. Au contraire, dans le cas de l’alternance dative par exemple, tel ou tel verbe sera plus enclin à être produit dans l’un ou l’autre type de structure, ce qui peut se représenter par une connexion plus au moins forte entre le lemme et l’un ou l’autre des nodules combinatoires. (Notons du reste qu’il a été démontré que ce biais verbal avait bel et bien une réalité linguistique, encodée, en ce qu’il influence le processus d’amorçage, les variantes moins fréquentes produisant un effet de *priming* accru (Bernolet et Hartsuiker 2010).) C’est là une idée cruciale, que nous saurons récupérer pour l’opposition massif-comptable.

Par la suite, le modèle de (Pickering et Branigan 1998) a été étendu encore par (Cleland et Pickering 2003), dans un effort d’intégration des théories de représentation lexicale et des modèles de la formulation phrastique, pour intégrer aussi sous forme de nodules combinatoires au niveau lemmatique, la structure du syntagme nominal. Ainsi, (Cleland et Pickering 2003) observent des effets d’amorçage dans la production de SN de différents types utilisés pour la description d’images dans un dialogue – plus précisément dans l’alternance entre une construction *ADJ + N* (*the red square* ‘le carré rouge’) et une structure *N + SP relative* (*the square that’s red* ‘le carré qui est rouge’). Leurs

résultats démontrent, comme chez (Pickering et Branigan 1998), l'existence d'un effet d'amorçage en la matière, avec un boost lexical, doublé d'un boost en cas de similarité sémantique – c'est-à-dire, une augmentation de l'effet d'une part lorsque le nom dans l'amorce et la cible sont identiques, mais aussi, fût-ce plus modestement, lorsqu'amorce et cible entretiennent une relation de co-hyponymie.

Donnons un exemple (traduit et adapté par nos soins, à partir des données de (Cleland et Pickering 2003)). L'on doit décrire l'image d'une chèvre blanche : à cette fin, l'on peut utiliser au moins deux descriptions, à savoir soit *la chèvre blanche*, soit *la chèvre qui est blanche* (la différence entre les deux alternatives étant plus grande encore en anglais, où l'adjectif précède le nom dans la première variante). Or, ce que (Cleland et Pickering 2003) démontrent c'est que dans ce cas, la seconde possibilité – pourtant possiblement moins fréquente – sera favorisée, par rapport à une situation de contrôle, si l'on a entendu précédemment la description *le navire qui est brun*, un peu plus encore si l'on entend *le mouton qui est brun*, et carrément plus si l'on entend *la chèvre qui est brune*. C'est effet d'amorçage est expliqué en termes du modèle à activation de (Levelt *et al.* 1999), comme le résultat de la pré-activation, plus ou moins forte selon la nature du lien sémantique entre amorce et cible, d'un nodule combinatoire partagé. Que l'effet est accru en cas de co-hyponymie s'explique au travers du lien au niveau sémantique, entre les concepts 'mouton' et 'chèvre' (reliés soit directement, soit par l'intermédiaire de l'hypéronyme commun 'animal'), en vertu duquel, au moment de l'audition de *le mouton qui est brun*, sont non seulement activés le concept de 'mouton', son lemme associé, et le lien entre ce lemme et le nodule combinatoire correspondant à la construction avec sous-phrase relative, mais aussi, via le lien au niveau sémantique, celui de 'chèvre', son lemme à lui, et le lien entre ce lemme et ce même nodule combinatoire. C'est ce qui explique, d'après (Cleland et Pickering 2003), que si l'on doit produire ensuite une description d'une chèvre blanche, l'alternative avec sous-phrase relative sera favorisée plus qu'en temps normal, car elle profitera de cette pré-activation au niveau lemmatique.

Les données de (Cleland et Pickering 2003) contribuent donc à soutenir et à peaufiner le modèle à activation du lexique de (Levelt *et al.* 1999), en confirmant via les effets d'amorçage que les informations combinatoires concernant la construction de SN sont elles aussi à représenter au niveau lemmatique du lexique mental, sous forme de nodules syntaxiques partagés. (C'est là quelque chose d'extrêmement intéressant pour ce qui sera de l'opposition massif-comptable, qui est concernée de manière cruciale par la construction syntagmatique.) En outre, l'extension de (Cleland et Pickering 2003) aborde la question, non moins intéressante, de la structuration du niveau conceptuel du modèle, que nous abordons au point suivant.

### 3.2.3 Le niveau sémantique

La question de la représentation et de la structuration des concepts de notre langue et de notre pensée est extrêmement complexe et fascinante (voir (Leshinskaya et Caramazza 2014) pour un aperçu du sujet) : nous ne saurions dès lors envisager d'en faire le tour. Cependant, nous veillerons à tirer en la matière quelques enseignements cruciaux des propositions du modèle de (Levelt *et al.* 1999).

Premièrement, nous retiendrons l'organisation des concepts sous forme d'un réseau, fait de nodules d'informations reliés entre eux par des relations de similarité, de hiérarchie, etc. (c'est-à-dire des *labelled links* 'liens nommés'). Notons, à ce stade, qu'on se rapproche ici déjà sensiblement des propositions récentes de la Linguistique Cognitive, comme celles de (Fauconnier 2000) mentionnées ci-dessus, dans la mesure où l'idée du sens linguistique est conciliée avec celle d'un réseau de connaissances – nous dirons dans un instant, concernant ce point, un mot du lien possible entre connaissance sémantique et encyclopédique.

A noter que la manière même dont l'on doit se représenter les concepts de la strate sémantique n'est pas une évidence : à ce sujet, les deux visions qui s'opposent sont celles d'une représentation holistique (Fodor *et al.* 1980, Roelofs 1997), où les concepts correspondent à des représentations mentales uniques, ou d'une représentation décompositionnelle (Pinker 1989, Jackendoff 1990), où les concepts plus complexes sont représentés comme la combinaison de concepts plus basiques. Ainsi, pour reprendre l'exemple de (Roelofs 1997: 34), la question est de savoir si le sens de *mère* est représenté comme l'unité sémantique unique MÈRE(X,Y), ou sous forme d'une combinaison des concepts PARENT(X,Y) et FÉMININ(X), par exemple.

Sur ce point, (Levelt *et al.* 1999) plaident pour une représentation holistique, non décomposée des concepts lexicaux. Ils évoquent à ce sujet deux arguments principaux. D'une part, un modèle décompositionnel du sens lexical se heurte immanquablement au *hyperonym problem* 'problème de l'hypéronymie' (voir aussi (Levelt 1989)) : l'activation d'un concept via ses traits sémantiques décomposés impliquerait toujours l'activation, du même coup, des traits de tous ses hypéronymes (par exemple, l'activation des traits supposés du concept *CHIEN* implique l'activation automatique de tous les traits du concept *ANIMAL*, inclus dans le premier) – or aucune tendance systématique aux erreurs de substitution d'un nom par son super-ordonné n'ont jamais été démontrée dans la production langagière. D'autre part, il a été montré que les concepts lexicaux qui correspondraient à des ensembles plus complexes de traits sémantiques n'étaient pas plus difficile à recouvrer (Levelt *et al.* 1978) : à nouveau, ceci est difficilement conciliable avec une représentation décomposée des concepts au niveau de la strate sémantique du lexique mental. Pour notre part, nous suivrons donc (Levelt *et al.* 1999) sur ce point, et imaginerons d'abord les concepts lexicaux comme des entités non décomposées. L'idée de concepts lexicaux holistiques, doublés de connexions plus ou moins fortes avec des

concepts associés, nous parait du reste de bon augure pour aider à concilier l'indispensable vision monosémiste du sens (Willems 2013) avec la polysémie et la multifonctionnalité des items langagiers, évoquées au chapitre précédent. Nous y reviendrons, en particulier dans notre Chapitre 11.

Quant à la question du lien entre connaissances sémantiques et connaissance du monde, le modèle de (Levelt *et al.* 1999) ne dit pas grand-chose. Au mieux, la connaissance encyclopédique peut y être imaginée comme un dispositif parallèle au niveau sémantique du lexique mental, qui le nourrit sans y être identique. Cependant, les résultats de recherches récentes tendent à remettre en question la distinction stricte entre informations sémantiques et connaissances pragmatiques, encyclopédiques. Ainsi, (Hagoort *et al.* 2004) ont montré que dans l'interprétation d'un énoncé, tant nos connaissances sémantiques que nos connaissances du monde étaient mobilisées de manière automatique, et qu'une violation à l'encontre des unes ou des autres donnaient les mêmes réactions en termes de potentiel évoqué (*event-related potentiel, ERP*), tel que mesuré par un encéphalogramme. Plus précisément, comparé à la phrase de base *Les trains aux Pays-Bas sont jaunes*, tant la phrase sémantiquement incohérente *Les trains aux Pays-Bas sont amers* que la phrase cohérente, mais fautive *Les trains aux Pays-Bas sont blancs* causaient un même effet négatif appelé N400 dans le potentiel évoqué – connu pour être associé aux violations sémantiques (Kutas et Hillyard 1980). De manière similaire, (McRae et Matsuki 2009) fournissent des preuves empiriques que les connaissances du monde et des événements sont mobilisées dès les premiers instants de l'interprétation langagière, et non pas en second lieu après l'intégration des informations syntaxiques et sémantiques – plaidant pour une approche intégrée de ces différentes sources d'information dans la modélisation du lexique.

Ces pistes semblent bien pointer en direction de ce qu'annonce aussi la Linguistique Cognitive moderne : que le sens des items de la langue peut difficilement s'imaginer de manière indépendante à nos connaissances encyclopédiques des référents qu'ils désignent. Pour notre part, nous n'entendons pas trancher ici la question de la séparation définitive entre sémantique et connaissances du monde – ni arguer qu'elle serait dépassée ou inutile. En tout état de cause, nous avons évoqué plus haut, au point 1.4.1.2, qu'une approche conceptualiste de la référence massive et comptable impliquait au moins une certaine séparation entre faits de langue et faits du monde (Lønning 1987: 83), nécessaire pour expliquer qu'un nom comme *eau* puisse être massif, c'est-à-dire se comporter comme si son référent était dépourvu d'individus, alors que nous savons aujourd'hui que l'eau est composée de molécules et d'atomes. Cela ne veut pas encore dire que le sens soit nécessairement indépendant de nos connaissances du monde, mais plutôt que ces dernières ne doivent pas être comprises comme des connaissances strictement scientifiques (ce que l'appellation d'encyclopédique pourrait peut-être faire croire, à tort). Au contraire, elles impliquent toutes les informations que nous avons au sujet des concepts qui peuplent notre connaissance sémantique, y compris des

informations perceptuelle et fonctionnelle. Ainsi, nous devons nous autoriser, en suivant cette vision, à intégrer des facteurs considérés souvent comme ‘pragmatiques’ (au sens large) dans notre analyse du comportement de différents noms par rapport au massif-comptable, sans que l’on puisse nous reprocher de quitter en cela le domaine de la description linguistique : que l’énoncé *J’ai mangé de l’éléphant* paraisse moins habituel que *J’ai mangé du bœuf* a certainement à voir avec nos connaissances de ces deux sortes d’animaux (le premier ne se trouvant virtuellement jamais dans nos assiettes) – mais nous soutiendrons que ce facteur fait bel et bien partie intégrante du système linguistique, par l’incidence qu’il a sur l’organisation de notre lexique (conceptuel) mental.

### 3.3 Retour au massif-comptable : pistes

Avec cette dernière remarque, nous voilà de retour dans le domaine du massif-comptable. Nous espérons que notre discussion, certes introductive, des modèles du lexique mental, en particulier des modèles à activation de la tradition psycholinguistique, aura permis de reposer autrement la question de la localisation de notre opposition, et de sa représentation dans le système linguistique. Plus précisément, au vu des études réalisées sur la structuration du lexique, en particulier au sujet du niveau lemmatique (en relation avec le niveau conceptuel) et de son importance pour la combinatoire syntagmatique, nous avancerons que la question cruciale concernant la représentation de l’opposition massif-comptable ne doit plus être tant de sa localisation dans le lexique ou en-dehors, mais bien une description de son statut effectivement éminemment lexical – à la lumière d’une vision moderne et psycholinguistique du lexique.

Ainsi, de notre description des modèles à activation de la tradition psycholinguistique de l’étude du lexique, et celui de (Levelt *et al.* 1999) en particulier, nous retiendrons deux pistes de réflexion principales, que nous saurons rouvrir au moment d’aborder notre propre modélisation de l’opposition massif-comptable. Premièrement, du point de vue morphosyntaxique, les traits combinatoires massif et comptable peuvent être situés, sous forme de nodules partagés, au niveau de la strate lemmatique du lexique. A noter que cette idée n’est pas nouvelle. Le massif-comptable fait partie des traits généralement mentionnés lorsqu’un aperçu est donné des caractéristiques syntaxiques localisées au niveau lemmatique (par exemple dans (Cleland et Pickering 2003: 217), où l’on lit : « Noun lemma nodes obviously link to category and featural nodes (e.g., number, gender, count/mass status) », nous soulignons), et sa localisation à ce niveau du lexique ressort également de l’étude de



(Vigliocco *et al.* 1999), sur la récupération du trait massif ou comptable d'un nom en cas de TOT ou d'anomie. Pour ce qui nous concerne, l'exploration de cette proposition consistera principalement en l'étude (en corpus) des préférences distributionnelles, pour l'un et l'autre type de syntaxe, de divers items flexibles (que l'on pourra se représenter comme un lien plus ou moins fort entre les lemmes concernés, et les nodules massif et comptable), ainsi que l'examen (au travers d'enquête d'acceptabilité) de la corrélation entre cette distribution, et les jugements de locuteurs natifs.

Deuxièmement, dans ce chapitre, nous avons évoqué une proposition très claire pour une meilleure appréhension de la question du sens, par la représentation de la composante conceptuelle du langage sous forme d'un réseau complexe d'information et de relations, non indépendante *a priori* de données encyclopédiques traditionnellement associées (voire reléguées) à la pragmatique, ouvrant par là directement aussi une piste quant à la description de la flexibilité massif-comptable, éminemment modulée par des facteurs contextuels – pragmatiques. Plus précisément, il s'agira en la matière d'étudier les interactions entre le niveau conceptuel (tel qu'activé en contexte) et le niveau lemmatique où se joue la sélection du type de syntaxe employé. Sur ce point, l'opposition massif-comptable se distingue crucialement des oppositions formelles (altérant au mieux la structure de l'information) telles que l'alternance dative étudiée par (Pickering et Branigan 1998) ou l'alternance syntagmatique de (Cleland et Pickering 2003), en ce qu'à notre variabilité morphosyntaxique est associée aussi une variation sémantique (le sens de *Augustin a bu de la bière* n'étant pas identique à celui de *Augustin a bu une bière*, cf. point 1.4). Les implications sémantiques de l'alternance massif-comptable nous occuperont tout au long de nos études de corpus, tandis que les conclusions à en tirer pour la modélisation du lexique mental seront examinées principalement au Chapitre 10, par le biais d'une expérience d'amorçage.



## Partie 2 : Etudes de corpus

Notre objet d'étude étant défini, et l'état de la question étant terminé, nous passons dans cette partie à la présentation de nos études de corpus. Celles-ci constituent réellement le cœur de notre propre recherche, et la majeure partie de nos données empiriques – qui seront complétées dans la troisième partie par les résultats de quelques enquêtes d'acceptabilité, ainsi que d'une expérience d'amorçage structurel.

Afin d'aider à élucider les différentes questions à résoudre au sujet de l'opposition massif-comptable, et dans la lignée de notre approche d'une étude de la langue par l'usage, nous avons réalisé trois études de corpus concernant un total de 92 items nominaux en français, répartis sur trois champs sémantiques (fruits/légumes, matériaux, boissons). Dans ce qui suit, nous présentons d'abord les détails de notre méthodologie, de la sélection de nos items à l'analyse de nos occurrences, en passant par notre choix de corpus. Ensuite, nous consacrons tour à tour un chapitre à chacun des champs sémantiques étudiés.



# Chapitre 4 Principes de la méthode

## 4.1 Introduction

Comme nous l'annonçons dans notre introduction générale, nous avons opté dans notre examen du massif-comptable pour une approche méthodologique résolument empirique. En cela, notre désir est avant tout d'ancrer nos réflexions théoriques sur des fondements solides, constitués d'observations linguistiques réelles, systématiques, et quantifiables.

Dans la recherche en linguistique moderne, les avantages des approches empiriques ne sont plus à prouver, comme en témoigne le nombre grandissant de recherches basées sur des études de corpus (pour un état de la question relativement récent, voir (Gilquin et Gries 2009)), et la littérature foisonnante sur la question méthodologique (Tummers *et al.* 2005, Arppe *et al.* 2011). Or, pour l'étude du massif-comptable, la recherche fondamentale a trop souvent fait l'économie de la récolte de données empiriques – se fondant généralement sur les intuitions, parfois très fines il est vrai, de l'un ou l'autre spécialiste.

Fort heureusement, la situation semble avoir viré récemment en faveur du changement. Ainsi, avec leur étude distributionnelle sur le corpus anglophone UKWAC, (Katz et Zamparelli 2012) ont proposé une première analyse quantitative de l'opposition massif-comptable, examinant la question de son 'élasticité' (quels items peuvent apparaître à la fois en syntaxe massive et en syntaxe comptable ? – ce que nous appelons la 'flexibilité') et du sémantisme qui y est associé (la variation en emploi massif et comptable s'accompagne-t-elle d'un changement de sens ?). Suivant pour leur part un objectif de classement lexical sur base morphosyntaxique, (Kiss *et al.* 2014) ont quant à eux organisé l'annotation par des locuteurs natifs d'environ 14 000 noms de la langue anglaise selon le comportement dans différents contextes prédéterminés, avec pour résultat une typologie lexicale quantifiée des noms en anglais. Pour ce qui nous concerne, nous entendons nous inscrire explicitement dans cette nouvelle vague d'approches empiriques de notre question.

Plus précisément, la pièce maitresse du volet empirique de notre recherche sera constituée des résultats de trois études de corpus, que nous présenterons tour à tour dans les chapitres suivants. En réalisant ces études, notre objectif était triple. Nous désirions : 1) obtenir des données quantitatives quant à l'opposition massif-comptable, et à sa flexibilité, 2) obtenir des données qualitatives systématiques quant à la variation sémantique associée à la flexibilité massif-comptable, et 3) obtenir pour ces deux aspects, quantitatif et qualitatif, des données parallèles en français et néerlandais permettant une comparaison systématique de ces deux langues. Nous verrons qu'à partir de ces trois types de données, nous serons en mesure de jeter sur l'opposition massif-comptable un regard neuf, et d'apporter des pistes de réponses aux questions en suspens évoquées dans les chapitres précédents.

Avant cela, nous détaillerons les principes de notre méthode. Tout d'abord, nous présenterons notre corpus, puis préciserons la manière dont nous avons sélectionné les trois champs sémantiques à étudier (fruits/légumes, matériaux, boissons), ainsi qu'à l'intérieur de ceux-ci les items examinés. Ensuite, nous expliciterons comment nous avons mené les analyses quantitatives et qualitatives de nos données. Pour finir, nous dirons quelques mots au sujet d'une sélection d'items de contrôle.

## 4.2 Présentation du corpus et de l'outil *WebCorp Live*

Pour la récolte de nos données de corpus, nous avons utilisé le Web, grâce à l'application en ligne *WebCorp Live*, disponible gratuitement sur <http://www.webcorp.org.uk/live/>. Il s'agit d'une suite d'outils permettant d'accéder à la Toile comme s'il s'agissait d'un corpus, d'y effectuer des recherches ciblées à partir de mots ou syntagmes spécifiques et/ou d'expressions régulières, et de récupérer les résultats de ces recherches sous formes de concordances. *WebCorp Live* n'est pas en lui-même un moteur de recherche, mais est un outil tirant profit des moteurs de recherche commerciaux existants, comme *Google* ou *Bing*, pour obtenir des données pertinentes pour la recherche linguistique. Plus précisément, les algorithmes *WebCorp* entrent dans l'un ou l'autre moteur de recherche (au choix parmi sept possibilités) la requête introduite par l'utilisateur, visitent les sites retournés en résultat (avec un maximum dépendant du moteur de recherche utilisé, allant de 50 pour *Bing* à 200 pour *Blekkko*), et affichent toutes les occurrences ainsi repérées sous forme de concordances modulables (l'utilisateur peut sélectionner l'étendue, en nombre de caractères ou de mots, de la sélection à retenir pour chaque occurrence). L'on trouvera sur le site mentionné ci-dessus toutes les informations nécessaires quant au fonctionnement plus précis de l'outil, ainsi que les

références à diverses publications concernant sa création et son application dans divers domaines (voir notamment (Renouf 2003, Renouf *et al.* 2007)).

Vu le caractère ouvert de la Toile, les résultats obtenus via *WebCorp* pour une recherche donnée sont susceptibles de fluctuer avec le temps, s'enrichissant au fur et à mesure des ajouts les plus récents – contrairement aux résultats d'une recherche dans un corpus fermé. Cela fait de *WebCorp* un outil de recherche expérimental, selon la définition de (Gilquin et Gries 2009) : chaque recherche fournira un résultat spécifique, différent de la même recherche effectuée à un autre moment, au même titre que dans une expérience prototypique, chaque observation est susceptible d'être différente de la précédente. Notons que malgré cela, nous n'avons pour chaque item étudié effectué qu'une seule et unique récolte de données (datée). Malgré la variation des résultats en fonction du temps, nous n'avons en effet pas de raison de croire que sur la durée qu'a duré notre recherche, le profil distributionnel et sémantique de nos items ait pu changer de manière tellement significative que les conclusions que nous tirerons sur la base de données récoltées au cours des trois années écoulées (2012-2014) ne seraient plus pertinentes au moment de l'écriture du présent travail.

Ceci étant dit, l'avantage évident de l'emploi de *WebCorp*, et plus généralement de la Toile comme terrain de recherche, est qu'il permet d'obtenir facilement, pour des expressions spécifiques, un grand nombre d'occurrences réelles et récentes – ce qui n'est pas toujours le cas dans un corpus fermé, en particulier lorsque l'objet de la recherche n'est pas de fréquence élevée. En guise d'illustration, pour notre recherche de *pamplemousse(s)* (formes au singulier et au pluriel confondues, voir ci-après), nous avons trouvé grâce à *WebCorp* un total de 787 occurrences, tandis que dans *Frantext*, les formes *pamplemousse* et *pamplemousses* n'apparaissent que respectivement 41 et 43 fois. L'usage de *WebCorp* permettra donc que l'on rassemble des données réellement pertinentes pour un examen quantitatif – une option qui n'est pas envisageable lorsque le nombre d'occurrences à examiner est trop bas.

### 4.2.1 Interface

L'interface de recherche de *WebCorp Live* propose à l'utilisateur différentes options, que nous passons ici brièvement en revue, en indiquant les éventuels choix systématiques que nous avons fait pour chacune d'elles. La Figure 3, à la fin de la partie présente, donne un aperçu visuel de cette interface, avec une numérotation (que nous y avons ajoutée) suivant l'ordre des points ci-après :

1. « Search » : dans cette zone, l'on introduit l'objet de la recherche, avec la possibilité d'utiliser (pour les moteurs de recherche qui le permettent) des expressions régulières – c'est-à-dire un système de notation spécifique permettant de grouper sous une unique suite de caractères plusieurs objets de

recherche ; pour nos propres études, nous avons ainsi pu effectuer une unique recherche pour chaque item, groupant sa forme au singulier et au pluriel (par exemple, l'expression 'banane[|s]' retourne à la fois des résultats pour *banane* et *bananes*) ainsi que, pour le néerlandais, ses formes diminutives éventuelles (par exemple, '[tomaat|tomaten|tomaatje|tomaatjes]' qui cherche simultanément chacune de ces formes – et est du reste un exemple d'utilisation peu efficace mais très prudente des expressions régulières, que nous avons préconisé pour toutes nos recherches en néerlandais, par facilité) ;

2. « Case Insensitive » : l'on précise ici si l'on veut que la recherche soit sensible à la distinction entre majuscules et minuscules ; pour toutes nos récoltes de données, nous avons coché la case, indiquant de ne pas tenir compte de la casse ;
3. « Span » : permet de préciser l'étendue souhaitée des concordances fournies ; nous avons utilisé systématiquement l'option « 10 words », retournant les résultats dans un environnement comprenant dix mots avant et dix mots après l'expression recherchée ;
4. « Search API » : menu déroulant permettant de sélectionner le moteur de recherche à utiliser pour la récolte des résultats ; nous avons utilisé systématiquement *Google*, car c'était le seul pour lequel sont pris en charge à la fois les expressions régulières, et la sélection d'une langue de recherche ;
5. « Language » : permet de limiter la recherche d'occurrences aux sites dans une langue donnée ; nous avons fait systématiquement usage de cette fonctionnalité (en précisant si notre objet de recherche était en français ou en néerlandais) ;
6. « Show URLs » et « Pages » : la première option demande si l'on veut que les URL d'où sont tirés les occurrences soient affichés avec les résultats ; nous avons toujours maintenu cette option cochée ; « Pages » indique, pour le moteur de recherche sélectionné, le nombre maximal de pages qui seront consultées par les algorithmes *WebCorp* ; pour *Google*, la limite est de 64 ;
7. « One concordance line per web page » et « Site » : permettent respectivement de limiter les résultats à une seule occurrence par page visitée, et de limiter les résultats à une sélection de sites au choix ; nous n'avons jamais employé ces options, notre objectif étant précisément d'obtenir un maximum de résultats, d'un maximum de sources ;
8. « Word Filter » : permet de préciser quelles expressions doivent, ou au contraire ne peuvent pas, se trouver sur la même page que les occurrences retournées ; nous avons fait abondamment usage de cette option, afin de limiter le bruit dans nos données ; pour ce faire, nous effectuions généralement d'abord une première



recherche naïve, sans l'application de filtres ; ensuite, après un parcours des résultats et si cela s'avérait nécessaire (ce n'était pas toujours le cas), nous sélectionnions des mots-clés qui permettraient d'exclure d'emblée les pages sélectionnées ne présentant à première vue aucune occurrence pertinente (par exemple, exclusion de « peer review » dans la recherche pour *peer* 'poire', ou des pages de *twitter*, comme en exemple dans la Figure 3) – obtenant ainsi plus de résultats pertinents lors d'une seconde recherche.

Concernant ce dernier point, nous avons conservé pour chaque recherche la trace des éléments utilisés comme filtres. On pourra les trouver en Appendice, avec le détail de toutes nos recherches sur *WebCorp* (12.4.1). Notons que l'usage de cette option étant basé sur un premier coup d'œil rapide aux données (à la recherche d'éventuelles occurrences parasites prenant une place évidente dans les résultats), elle n'a pas empêché que l'analyse plus précise des résultats, après export des données, donne lieu à de nouvelles exclusions d'occurrences non pertinentes, passée entre les mailles de ce premier filet.

Pour finir, précisons que l'interface de *WebCorp Live* ne fournit pas d'option spécifique pour l'export des données récupérées en ligne, à l'issue de chaque recherche : une fois la recherche lancée, les résultats apparaissent simplement sur une nouvelle page dans le navigateur internet, avec la source de chaque occurrence sous forme d'un lien, et l'objet de la concordance en couleur. La manière la plus simple de récupérer ces données est alors simplement de les copier à l'écran (sélection, clique-droit, copier), puis de les coller à l'endroit de son choix. Pour ce qui nous concerne, nous avons systématiquement collé les résultats d'abord dans un fichier Word, en guise de backup, pour y conserver la mise-en-forme du site, puis à nouveau, après quelques manipulations techniques, dans un fichier TXT – nécessaire pour notre analyse semi-automatique (voir ci-après).

The screenshot shows the WebCorp search interface with the following elements and annotations:

- 1. Search:** A text input field.
- 2. Case Insensitive:** A checked checkbox.
- 3. Span:** A dropdown menu set to "10 words".
- 4. Search API:** A dropdown menu set to "Google".
- 5. Language:** A dropdown menu set to "French".
- 6. Show URLs:** A checked checkbox.
- Pages:** A numerical value of "64".
- 7. One concordance line per web page:** An unchecked checkbox.
- Site:** A large text input field.
- Add popular sites:** A list of site suggestions including:
  - UK Broadsheet Newspapers
  - UK Tabloid Newspapers
  - French Newspapers
  - Greek Newspapers
  - US Newspapers
  - BBC News
  - Wikipedia
  - US Academic
  - UK Academic
  - Argentina
  - Australia
  - Brazil
  - Canada
  - China
  - France
  - Germany
  - Italy
  - Japan
  - Netherlands
  - New Zealand
  - Spain
  - UK
- [clear]:** A button to clear the site input.
- 8. Word Filter:** A text input field containing "-twitter".
- Réinitialiser:** A button to reset the search.
- Search:** A button to execute the search.

Figure 3 Aperçu de l'interface de *WebCorp* (nous avons ajouté la numérotation des points).

## 4.3 Sélection des champs et des items à étudier

### 4.3.1 Champs sémantiques

Puisque c'est la question de sa flexibilité qui est au cœur des débats concernant l'opposition massif-comptable, notre choix s'est tout naturellement porté sur l'étude de différentes classes de noms présentant d'emblée, au moins à notre intuition, un profil morphosyntaxique variable, alternant entre emplois massifs et comptables. Dans ce qui précède, nous avons déjà évoqué plusieurs fois le domaine des noms comptables d'animaux, pouvant s'utiliser également de manière massive pour désigner de la viande, de la fourrure, ou toute autre substance animale dérivée. Comme ce domaine a souvent joui d'un certain monopole dans les discussions de la flexibilité massif-comptable systématique (Nunberg et Zaenen 1992, Ostler et Atkins 1992, Kleiber 1999, Nicolas

2002a), nous avons préféré ne plus nous y attarder, et avons tourné notre attention vers les trois domaines suivants : 1) les noms de fruits et de légumes, 2) les noms de matériaux, 3) les noms de boissons. Notons que dans ce qui suit, nous utilisons de manière interchangeables les appellations « champ sémantique », « catégorie nominale » ou « classe de noms », pour désigner ces trois groupes – sans présumer de l'équivalence de ces termes dans d'autres domaines de la description linguistique, ou du langage en général. Dans ce qui suit, nous justifions brièvement le choix de chaque champ sémantique, en précisant nos objectifs principaux les concernant.

Les noms de fruits et de légumes nous paraissaient intéressants, en ce qu'il s'agit là de noms que nous supposerions en premier lieu comptables (*un concombre, une banane, un poireau, un chou-fleur* ; *idem* en néerlandais) mais présentant, dans le contexte de leur consommation culinaire, des emplois massifs *a priori* tout aussi naturels (*recette avec du concombre, de la banane, du chou-fleur, etc.*), pas forcément très éloignés, du reste, des emplois massifs de noms d'animaux en termes de 'viande' – puisqu'il est question là aussi de 'matière comestible'. Du point de vue quantitatif, nous souhaitions dès lors voir de quelle manière cette flexibilité intuitive se traduirait au niveau de profils distributionnels, de même qu'identifier d'éventuelles différences entre items à ce niveau-là. D'un point de vue qualitatif, nous désirions nous pencher de plus près sur les emplois massifs des noms de fruits et de légumes, afin de mieux comprendre les facteurs (sémantiques, contextuels) favorisant leur apparition.

Pour les noms de matériaux (de construction, au sens large), la démarche était similaire, mais inversée : ici, nous avons affaire à des noms *a priori* fondamentalement massifs (*du béton, du verre, du carton* ; *beton* 'béton', *glas* 'verre', *karton* 'carton'), présentant néanmoins à la réflexion divers types d'emplois comptables (*un verre* 'récipient pour boire', *un carton* 'morceau de carton' ; *een glas* 'un verre') ou au moins pluriels (*des cuivres* 'instruments', *des sables* [mouvants]). Ainsi, pour ceux-ci, l'objectif quantitatif serait d'identifier le corrélât distributionnel d'un statut fondamentalement massif, tandis que l'objectif qualitatif serait de décrire et classer les différents types d'emplois comptables, découlant soit de mécanismes réguliers, soit d'une éventuelle polysémie lexicalisée.

Pour finir, les noms de boissons ont attiré notre attention car ils présentent d'emblée une grande flexibilité entre emplois massifs (*de la limonade, du café*) et comptables (*une limonade* 'un verre de limonade', *un café* 'une tasse de café'), sans pour autant que le rapport entre les deux emplois puisse être dit entièrement systématique (\**un vin* 'un verre de vin' vs. *une bière* 'un verre de bière'), ni simplement directionnel. Sur ce dernier point, si pour *bière* ou *limonade*, l'emploi comptable pourrait bien être dit dérivé de l'emploi massif, que dire d'items comme *kir, grog* ou *expresso*, qui sont possibles en lecture massive – *Nous avons bu du kir, du grog* – mais dont l'appellation comptable, parce que liée à la portion spécifique définitoire de la boisson, semble précéder l'autre ? Ainsi, notre objectif quantitatif serait en particulier de comparer les différents items de la

catégorie pour voir s'ils diffèrent du point de vue de leurs profils distributionnels (et examiner à quelles différences intuitives cette variation pourrait correspondre). En outre, le domaine des noms de boissons est l'endroit typique, toujours donné en exemple, de la différence du point de vue de l'opposition massif-comptable entre le français et le néerlandais, ce dernier se montrant sensiblement moins flexible, car nécessitant notamment pour certains emplois l'usage du diminutif (*\*een bier vs. een biertje* pour 'un verre de bière'). Ainsi, la comparaison entre le français et le néerlandais, présente en filigrane de nos trois études de corpus, trouverait ici possiblement son retentissement le plus intéressant.

### 4.3.2 Items

Pour nos trois études de cas, notre sélection d'items a suivi les mêmes étapes générales (avec une petite particularité pour les noms de fruits et légumes, voir ci-après) Dans un premier temps, nous avons établi pour nos trois classes de noms une liste d'items (en français) se voulant le plus exhaustif possible, arrivant de cette manière à des totaux respectifs de 116 noms de fruits et 131 noms de légumes (la plupart exotiques), 79 noms de matériaux, et 140 noms de boissons. Pour cette étape, nous avons utilisé d'une part notre intuition et notre propre connaissance du monde, et d'autre part diverses ressources encyclopédiques telles que *Wikipedia* (<http://fr.wikipedia.org/>), le *Nouveau dictionnaire analogique* de (Niobey 1979), ou encore le *Dictionnaire électronique des mots (DEM)* des Dubois (voir (Dubois et Dubois-Charlier 2010)). A noter que nous ne nous sommes pas, à ce stade, contenté uniquement d'items mono-lexicaux, puisque nos sélections originales incluaient des entrées telles que *noix de cajou*, *ail rocamboule*, *chou de Bruxelles*, ou encore *vodka orange*.

Pour chacun de nos items, nous avons ensuite procédé à une annotation incluant divers paramètres, en ce compris un indice de fréquence, une indication quant à la catégorisation sémantique plus précise, une note spécifique pour les items présentant un homonyme potentiellement gênant pour la recherche en corpus, et une traduction en néerlandais.

Concernant la fréquence, nous avons utilisé un indice récapitulatif par million de mots, se voulant la moyenne des indices par million de *Frantext* (basés sur un corpus de 247 937 581 mots), et de la base de données *LEXIQUE 3.55* (New, Pallier et Ferrand, <http://lexique.org/>, basés sur un corpus littéraire de 14,7 millions de mots ainsi que sur deux corpus de sous-titres de films et de séries totalisant 50 millions de mots). Nous avons eu accès à *LEXIQUE 3.55* via Marc Brysbaert, que nous remercions. (Notons que dans un premier temps, nous avons également pris en compte les indices de fréquence manuels de (Juilland *et al.* 1970), datant d'une ère pré-informatique, qui se sont cependant très vite avérés complètement dépassés.) Pour les items multi-lexicaux (*noix*

de cajou, pois chiche), l'obtention d'un indice de fréquence était limitée à *Frantext*, par le biais de la recherche d'occurrences – ces items étant généralement absents de *LEXIQUE*. Pour les items présentant un homonyme non pertinent (*bois, avocat, tango*, etc.), l'indice global a été retenu dans la mesure où il n'était pas possible d'y distinguer la part des différents emplois. Néanmoins, tous ces items (à l'exception d'*orange*) ont été exclus d'emblée de la sélection finale.

Pour ce qui est de la sous-catégorisation sémantique plus précise, nous entendons essentiellement : la catégorie botanique pour les fruits et légumes (*Liliaceae Allium, Prunus*, etc.), la nature physique pour les noms de matériaux (*métal, étoffe, roche*, etc.) et le type pour les noms de boissons (*bières, liqueurs*, etc.). Pour ces deux dernières catégories, nous avons également prévu une annotation taxonomique, précisant la nature hypéronymique (*métal ; boisson*), basique (*béton ; café*) ou hyponymique (*tilleul*) de nos items, si cela s'avérait pertinent.

Sur la base de ces différentes informations, nous avons établi nos trois sélections d'items, en nous appuyant de manière combinée sur les facteurs de fréquence (sélection des items les plus fréquents, afin d'obtenir des données suffisamment nombreuses) et de variation sémantique (sélection d'items variés, couvrant un maximum de sous-classes sémantiques). Pour les noms de fruits et de légumes, nous avons en outre exclu tous les noms composés (préférant par exemple *patate* à *pomme de terre*), tandis que pour les noms de boissons, au contraire, nous avons intégré certains cas de composition, afin de limiter les interférences dans les données (par la sélection finale, par exemple, d'*eau plate* ou lieu d'*eau* tout court, pour exclure tous les emplois où *eau* n'est pas utilisé pour désigner un liquide destiné à la consommation, ou de *café noir* au lieu de *café*, pour exclure les occurrences concernant le débit de boisson). Pour les noms de matériaux, nos items étaient tous mono-lexicaux. Nous avons également exclu, à rebours, les noms de fruits secs (*noisettes, amandes ; dattes, figes*), présents dans notre sélection exhaustive originale, de notre sélection de fruits et de légumes – les supposant trop en marge de la classe de noms qui nous intéressait.

De ce processus de sélection, nous avons retenu au final 30 noms de matériaux (Figure 4, analyse au Chapitre 6), et 35 noms de boissons (Figure 5, analyse au Chapitre 7). Concernant la sélection d'items en néerlandais, nous avons simplement repris la traduction la plus intuitive et la plus fréquente de ces noms, en nous aidant au besoin du *Van Dale*, dictionnaire de référence du monde néerlandophone, dans sa version de traduction Français-Néerlandais (disponible en ligne sur le réseau de l'Université de Gand). Dans certains cas, nous avons retenu pour un item français deux traductions, lorsque nous ne pouvions décider lequel des deux constituait la traduction la plus fidèle. C'est le cas par exemple de *koord* et *touw* pour *corde*, *modder* et *slijk* pour *boue*, ou encore *eik* et *eiken* pour *chêne* (*eiken* pouvant être soit le pluriel de *eik* 'chêne', soit un singulier *a priori* massif désignant 'du bois de chêne', comme dans l'énoncé attesté *Europees eiken lijkt het meest op wit Amerikaans eiken* 'C'est le (bois de) chêne européen qui ressemble le

plus au (bois de) chêne blanc américain'). Dans ces cas, nous avons récolté des données pour les deux items (tandis que nous avons distingué manuellement les deux formes homonymes de *eiken*). D'autre part, pour l'item *vinasse* parmi les noms de boissons, nous n'avons trouvé aucun équivalent satisfaisant, réduisant ainsi d'un la sélection de boissons en néerlandais. Finalement, pour la traduction de *café noir*, nous avons opté simplement pour *koffie*, ce mot n'ayant pas en néerlandais l'homonyme gênant que présente *café* en français.

#	français	néerlandais	#	français	néerlandais
1	ACIER	STAAL	16	FER	IJZER
2	ARGILE	KLEI-LEEM	17	FOURRURE	BONT
3	BÉTON	BETON	18	GRANIT	GRANIET
4	BOUE	MODDER-SLIJK	19	GRAVIER	GRIND
5	BRIQUE	BAKSTEEN	20	LAINES	WOL
6	BRONZE	BRONS	21	MARBRE	MARMER
7	CAOUTCHOUC	RUBBER	22	MÉTAL	METAAL
8	CARTON	KARTON	23	PAPIER	PAPIER
9	CHAUX	KALK	24	PLOMB	LOOD
10	CHÊNE	EIKEN-EIK	25	PORCELAINE	PORSELEIN
11	CIMENT	CEMENT	26	SABLE	ZAND
12	CORDE	TOUW-KOORD	27	TISSU	STOF
13	COTON	KATOEN	28	VELOURS	FLUWEEL
14	CUIR	LEDER	29	VERRE	GLAS
15	CUIVRE	KOPER	30	ZINC	ZINK

Figure 4 Sélection des noms de matériaux à étudier.

#	français	néerlandais	#	français	néerlandais
1	ALCOOL	ALCOHOL	19	LIMONADE	LIMONADE
2	APÉRITIF	APERITIEF	20	MILKSHAKE	MILKSHAKE
3	BIÈRE	BIER	21	MOUSSEUX	SCHUIMWIJN
4	BOISSON	DRANK	22	PANACHÉ	SHANDY
5	BORDEAUX	BORDEAUXWIJN	23	PUNCH	PUNCH
6	BREUVAGE	BROUWSEL	24	SAUTERNES	SAUTERNES
7	CAFÉ NOIR	KOFFIE	25	SHERRY-CHERRY	SHERRY
8	CHAMPAGNE	CHAMPAGNE	26	SODA	SODA
9	CIDRE	CIDER	27	SOUPE	SOEP
10	COCA	COLA	28	THÉ	THEE
11	COCKTAIL	COCKTAIL	29	TILLEUL	LINDETHEE
12	EAU PLATE	PLAT WATER	30	TISANE	KRUIDENTHEE
13	EXPRESSO	ESPRESSO	31	VIN	WIJN
14	GROG	GROG	32	VIN CHAUD	GLÜHWEIN
15	GUEUZE-GUEUSE	GEUZE	33	VINASSE	/
16	JUS DE FRUIT	FRUITSAP	34	VODKA	WODKA
17	KIR	KIR	35	WHISKY-WHISKEY	WHISKY
18	LAIT	MELK			

Figure 5 Sélection des noms de boissons à étudier.

Pour les noms de fruits et légumes (Figure 6, analyse au Chapitre 5), notre sélection finale est un peu différente, car elle est en réalité le résultat de plusieurs sous-études.

D'abord, nous avons retenu 18 items appartenant exclusivement aux sous-classes des agrumes (5), des cucurbitacées (8), et des aulx (5), pour les besoins d'une publication spécifique (voir dans (Vermote 2013)). Ensuite nous avons établi, en respectant les critères de fréquence et de variation sémantique expliqués ci-dessus, une seconde sélection de 12 noms de fruits et de légumes, à étudier en corpus, et que nous utiliserions en outre dans nos enquêtes d'acceptabilité (voir au Chapitre 9). Cela nous a mené à un total, en français, de 27 items, dont trois représentés dans nos deux sous-études. Pour la traduction en néerlandais, nous nous sommes néanmoins contenté du second groupe de 12 items – la première sélection nous paraissant moins pertinente. (Comme nous avons retenu pour *oignon* les deux traductions *ui* et *ajuin*, cela nous fait un total de 13 noms de fruits et de légumes étudiés pour le néerlandais.) Pour finir, nous avons également ajouté à notre sélection en français deux items supplémentaires au moment de réaliser notre expérience d'amorçage (Chapitre 10), faisant monter le total à 29. Que nous ayons facilement pu augmenter le nombre d'items étudiés en français, sans pouvoir suivre le même rythme en néerlandais, a tout à voir avec les outils à notre disposition pour l'analyse des données de ces deux langues, que nous présentons au point suivant.

#	<i>français</i>	<i>néerlandais</i>	#	<i>français</i>	<i>néerlandais</i>
1	AIL	/	16	HARICOT	BOON
2	BANANE	BANAAN	17	MANDARINE	/
3	CALEBASSE	/	18	MELON	MELOEN
4	CAROTTE	WORTEL	19	OIGNON	UI-AJUIN
5	CEDRAT	/	20	ORANGE	/
6	CERISE	/	21	PAMPLEMOUSSE	/
7	CHOU-FLEUR	/	22	PASTEQUE	/
8	CIBOULETTE	/	23	PATATE	AARDAPPEL
9	CITRON	/	24	POIRE	PEER
10	CITROUILLE	/	25	POIREAU	/
11	CONCOMBRE	KOMKOMMER	26	POMME	APPEL
12	CORNICHON	/	27	POTIRON	/
13	COURGE	/	28	RAISIN	DRUIF
14	ECHALOTE	/	29	TOMATE	TOMAAT
15	FRAMBOISE	FRAMBOOS			

Figure 6 Sélection des noms de fruits et de légumes à étudier.

## 4.4 L'analyse quantitative

### 4.4.1 Principes généraux

L'objectif premier de nos analyses de corpus est d'établir, pour chacun des items étudiés, un profil distributionnel reflétant son usage en contexte massif et comptable. Plus précisément, cela implique que nous analysions chaque occurrence récoltée grâce à *WebCorp* en termes du type de syntaxe – massive, comptable, cumulative ou neutre, voir point 1.2 – dans lequel le nom concerné y apparaît. L'importance des contextes indiscutablement massifs et comptables va sans dire : ils constituent l'essentiel des énoncés que nous voudrions examiner, tant par leur quantité que leurs caractéristiques sémantiques. Les contextes cumulatifs sont eux aussi cruciaux, puisqu'assimilables aux premiers ou aux seconds par le biais de l'emploi au singulier ou au pluriel du nom utilisé – une option morphologique facilement identifiable à l'écrit grâce à la marque de l'accord (dans *beaucoup de briques*, le -s final permet de déduire que l'on est dans une interprétation comptable du nom *brique*).

Pour finir, les contextes neutres sont importants à identifier, principalement dans le but de leur exclusion. En effet, ils ne fournissent pas en eux-mêmes d'indices de surface pour discriminer : dans *Le sable de cette plage est blanc*, l'article défini n'indique rien du caractère morphosyntaxique du nom *sable* (d'ailleurs, les deux reformulations *Eh bien oui, ce qu'on voit sur cette plage, c'est du sable blanc* ou *c'est un sable blanc* sont tout aussi acceptables). En outre, (Lønning 1987: 11) a évoqué le caractère délicat des descriptions définies – l'essentiel de ce que nous classons du côté des contextes neutres – par rapport à notre opposition, en ce que même avec le nom le plus typiquement massif, elles ne peuvent être dites massives elles-mêmes car elles ne répondent plus à la contrainte d'homogénéité : l'on ne peut pas dire *??Beaucoup d'eau pesait seize grammes*, par contre l'on peut dire sans problème *L'eau que Jean a bu pesait seize grammes*. Tout au long de nos études de corpus, nous avons donc laissé systématiquement de côté les emplois de noms en contexte neutre.

Puisqu'à l'emploi syntagmatique massif, comptable, cumulatif ou neutre correspondent normalement des contextes de surface spécifiques, c'est-à-dire qu'ils peuvent être distingués grâce à la nature de l'environnement direct du nom, l'analyse de nos données n'a pas nécessité un classement manuel de toutes nos phrases : elle a pu fort heureusement bénéficier des avantages du traitement informatisé, (semi-)automatisé. Considérons, au hasard, l'une des occurrences de l'un de nos noms :

- (78) [N]otre homme découvrira une boisson locale qui aurait, selon ses consommateurs, de puissantes vertus tonifiantes.



Il est évident qu'elle constitue une occurrence en syntaxe comptable de l'item *BOISSON*, comme le révèle sans équivoque le déterminant *une* qui précède ce nom. Or, cette tâche de reconnaissance à partir du contexte précédent est à la portée de n'importe quel ordinateur : nous avons dès lors pu la confier, au moins en partie, à l'un ou l'autre logiciels de notre propre confection, comme nous l'expliquons dans ce qui suit.

#### 4.4.2 Programmation d'outils informatique avec *Perl*

Pour l'analyse tant de données en français qu'en néerlandais, nous avons puisé dans nos connaissances (modestes) en informatique, et en particulier dans notre aptitude à manipuler le langage de programmation *Perl*, pour construire des outils sur mesure, correspondant précisément à ce dont nous avons besoin. Dans le présent travail, nous nous contenterons d'évoquer les grandes lignes de cette démarche, en épargnant au lecteur un maximum de considérations techniques. Des informations plus détaillées sur notre démarche concernant le français peuvent être trouvée dans (Vermote 2011, ms.).

Disons seulement quelques mots de *Perl*, et de la manière dont il faut se figurer notre démarche. *Perl* est un langage de programmation développé à partir de 1987 par Larry Wall (Wall *et al.* 2000). Il s'agit d'un ensemble de commandes que l'on peut utiliser pour composer un algorithme, qui sera ensuite lu par un interpréteur, c'est-à-dire un logiciel qui interprétera les commandes rédigées, et les effectuera. Il existe un grand nombre de langages de programmation, destinés à la création de divers types de logiciels et de fonctionnalités. *Perl*, en l'occurrence, est particulièrement conçu pour la manipulation textuelle – c'est la raison pour laquelle nous y avons fait recours dans le contexte présent. La création d'un logiciel, en soi, n'a rien de sorcier, dès lors qu'on connaît la syntaxe du langage de programmation que l'on veut employer : il s'agit seulement d'écrire, via n'importe quel outil de traitement de texte (nous avons toujours utilisé *Notepad++*, cf. <http://notepad-plus-plus.org/>), le code composant l'algorithme qui réalisera les manipulations voulues, et de sauvegarder le fichier avec la bonne extension (pour les scripts *Perl*, il s'agit de *.pl* ou *.plx*). Ensuite, moyennant l'installation préalable de l'interpréteur requis, l'application peut être lancée par un simple clic.

*Perl* présente de nombreuses fonctions, avec de manière cruciale des options de base permettant, via un script donc, d'accéder au contenu d'un ou plusieurs fichiers TXT au choix, de manipuler leur contenu (en l'examinant à partir d'expressions régulières et/ou en le modifiant, par exemple), de créer des nouveaux fichiers TXT – ou formats apparentés – et d'y inscrire du contenu, de même qu'interagir avec l'utilisateur au moyen d'un prompteur de commandes (appelé *shell*) pouvant afficher ou recevoir du texte. Sans qu'il n'ait été besoin d'évoquer la moindre caractéristique technique, on voit bien comment *Perl* peut s'avérer pertinent pour la création d'outils d'analyse de nos données : en essence, ce qu'on attend d'un tel outil est qu'il lise les fichiers contenant

nos données de corpus, y détecte certaines combinaisons de signes, et leur assigne une certaine valeur. C'est précisément ainsi que l'on peut résumer les fonctionnalités de tous les scripts que nous avons créés pour nous assister dans notre analyse. Dans ce qui suit, nous évoquons ceux que nous avons imaginés pour le français et pour le néerlandais, toujours en restant au plus loin du jargon technique, mais de manière à expliciter la philosophie de notre approche, les objectifs de nos différents outils, et les stratégies mises en place pour augmenter leur efficacité.

### 4.4.3 Analyse semi-automatique en français

L'outil que nous avons créé pour l'analyse de nos données en français est issu d'un projet dont l'envergure dépasse la simple annotation syntagmatique d'occurrences d'un nom spécifique tirées de *WebCorp*. En effet, bien au-delà de cet objectif modeste, notre script, baptisé *SN\_FR.pl*, se conçoit comme un outil de TAL à part entière, destiné à repérer dans tout texte français l'ensemble des syntagmes nominaux qui s'y trouvent, à en identifier le nom-tête, et à les classer (sur la base des déterminants et autres indices de surface) comme massifs, comptables, cumulatifs et neutres. Il se compose d'un script central, d'un fichier CSV associé constituant en essence sa mémoire lexicale – c'est-à-dire contenant des informations morphologiques et syntaxiques pertinentes au sujet de plus de 200 000 entrées (un nombre en constante augmentation, car cette mémoire est alimentée à chaque utilisation sur un texte nouveau) – ainsi que de quelques fichiers TXT périphériques, nourrissant cette mémoire lexicale de manière indirecte. L'un dans l'autre, sa création a couvert une période d'un an et demi (entre janvier 2011 et juin 2012), avec quelques petites améliorations ponctuelles apportées encore par la suite, au besoin.

Dans les grandes lignes, *SN\_FR.pl* fonctionne de la manière suivante. Au moment du lancement du script, un prompteur de commandes est ouvert, dans lequel l'utilisateur note le nom du fichier TXT à analyser (voir Figure 7). Le fichier en question est alors ouvert par l'interpréteur, et analysé ligne par ligne, et mot par mot. A chaque mot, un ensemble de conditions sont testées, en fonction de la nature du mot en question, du ou des mots précédents, et du ou des mots suivants. Prenons l'exemple de la phrase imaginaire suivante :

(79) J'ai reçu quelques beaux cadeaux récemment.

Lorsque le script arrive au mot *quelques*, il reconnaît ce dernier comme déterminant (au sens de la grammaire traditionnelle) – plus précisément comme déterminant pluriel non ambigu – et sait donc qu'il est en présence d'un syntagme nominal comptable. Pour délimiter celui-ci, il examine le mot suivant : en l'occurrence, il s'agit de *beaux*, dont il sait (grâce à la mémoire lexicale) qu'il s'agit d'un adjectif. Il faut donc avancer au mot

suisant. *Cadeaux* n'étant pas reconnu comme un élément devant produire une quelconque réaction particulière, le script conclut qu'il doit s'agir de la tête du SN. Il enregistre donc *quelques beaux cadeaux* comme un syntagme nominal, ayant *cadeaux* comme tête, de nature comptable, introduit par un déterminant pluriel. Il continue ensuite l'analyse à partir du mot suivant. A l'issue de l'analyse du texte complet, la liste des SN repérés, avec leurs diverses caractéristiques, est enregistrée dans un fichier CSV généré pour l'occasion. Un lemmatiseur est également inclus au script, par le biais de sa mémoire lexicale externe, contenant, pour un nombre de noms augmentant à chaque utilisation, leur forme de base au singulier. Dans le cas de notre exemple, l'entrée *quelques beaux cadeaux* serait donc accompagné également de la mention CADEAU – la forme au singulier du nom.

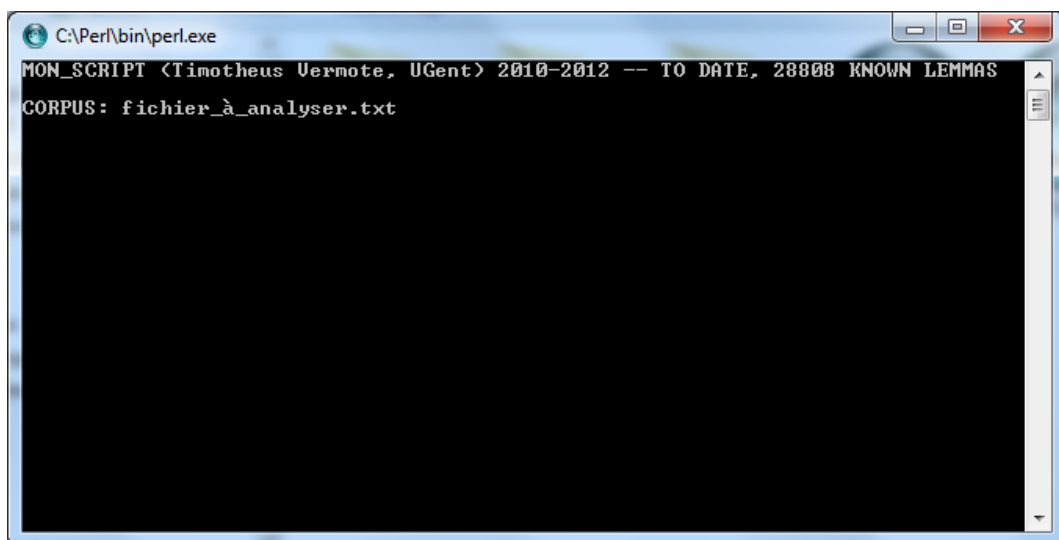


Figure 7 Prompteur de commande affiché au lancement de *SN\_FR.pl*. Le texte *fichier\_à\_analyser.txt* est fictif, et indique l'endroit où il faut écrire le nom du fichier TXT qui servira d'input au programme.

Bien entendu, le repérage de syntagmes nominaux dans un texte n'est pas toujours aussi simple que ne laisserait supposer l'exemple en (79). Ainsi, nous avons peaufiné notre script par touches successives, pour analyser aussi des cas de syntagmes bien plus complexes, avec plusieurs adjectifs pré-nominaux coordonnés (*une jeune et sympathique madame*) ou impliquant des adverbes (*trois très petites maisons*), de même que des déterminants plus complexes (*trois enfants vs. les trois enfants*, etc.). Par contre, hormis les cas d'un adjectif post-nominal unique (*ces quelques maisons blanches*), éventuellement précédé d'un superlatif (*la maison la plus belle*), nous ne nous sommes pas attardé sur les éventuels compléments nominaux à droite du nom, comme les sous-phrases relatives. Dans *Voici l'homme que j'ai vu hier*, par exemple le script repérerait bien un emploi neutre du nom *HOMME*, mais ne retiendrait comme syntagme que *l'homme* – alors que techniquement le syntagme se poursuit jusqu'à la fin de la phrase. L'objectif du script est donc bien de repérer tous les syntagmes, d'identifier leur nom-tête, et de les classer

comme massifs, comptables, neutres ou cumulatifs, mais ne prétend pas délimiter tous les syntagmes avec une précision absolue.

Précisons à ce stade ce que nous entendons par analyse semi-automatique : si notre *SN\_FR.pl* a bien pour but d'effectuer une analyse complète d'un texte en français, nous avons mis en place des stratégies pour lui permettre, en cas d'impossibilité à résoudre une ambiguïté, de signaler simplement un syntagme comme non classé, ou sa tête comme non repérée. Dans le second cas, la difficulté aurait généralement trait à la présence dans le syntagme de plusieurs mots indiqués, dans la mémoire lexicale externe, comme ambigus entre nature nominale, adjectivale et/ou verbale (pouvoir reconnaître des verbes est utile au script, pour exclure de son analyse des suites de mots ressemblant en surface à un syntagme, mais n'en étant pas un, comme dans *Ces pommes, Marie les mange*, où *les mange* serait considéré dans un premier temps, mais rejeté ensuite en vertu du statut verbal de *mange*). Prenons l'exemple imaginaire de la suite de mots *ma bien-aimée cuisine*. Le script sait que *bien-aimée* peut être un nom (*ma bien-aimée*) ou un adjectif (*la bien-aimée sorcière*) ; d'autre part, il sait que *cuisine* peut être un verbe (*elle cuisine*) ou un nom (*la cuisine*) : dans ce cas, le script se contentera d'acter la suite de mot comme syntagme potentiel, en laissant à l'appréciation de l'utilisateur le choix du nom-tête, par une complétion manuelle ultérieure. (Notons que certains de ces cas potentiellement ambigus sont néanmoins résolus en autonomie par le script, sur la base d'informations supplémentaires incluse dans la mémoire lexicale, concernant par exemple le statut éventuellement strictement post-lexical de certains adjectifs – mais nous n'entrerons pas dans les détails à ce sujet.)

Dans l'autre cas, où un syntagme ne peut être assigné de manière certaine à l'une de nos classes morphosyntaxiques, la nature du syntagme est simplement marquée comme non résolue dans le tableau de résultat, nécessitant là aussi une complétion manuelle. Dès l'origine de notre démarche, nous nous sommes ainsi heurté à une difficulté majeure de l'analyse automatique de nos données, particulièrement cruciale pour notre étude du massif-comptable : celle de la désambiguïsation de *du* ou *de* la, tantôt article partitif, tantôt article défini précédé de la préposition *de*. Vu l'importance de ces cas dans le cadre de notre recherche, nous avons agi avec la plus grande prudence quant à leur annotation, ne les livrant à l'analyse automatique qu'en cas de certitude. Néanmoins, notre script ne se contente pas de laisser vierges l'ensemble de ces cas, lorsqu'il les rencontre, puisque nous avons, au fil de notre travail, établi diverses stratégies pour repérer néanmoins, dans de très nombreuses situations, un partitif ou une préposition, à coup sûr. Ainsi, par exemple, notre script sait qu'il peut interpréter comme partitifs les items *du* ou *de la* lorsque ceux-ci sont précédés directement d'une autre préposition (comme dans *une recette avec de la banane*), s'ils apparaissent dans une coordination avec d'autres indéfinis (comme dans *un biscuit et du beurre*), ou s'ils sont précédés d'un verbe directement transitif (comme dans *j'ai bu de l'eau*). (Sur ce dernier point, la mémoire lexicale de notre script est nourrie par une liste externe d'environ 220

verbes directement transitifs fréquents, ainsi que toutes leurs formes conjuguées.) D'autre part, le script reconnaît en *du* ou *de la* la combinaison d'une préposition et d'un article après certains verbes qu'ils sait se combiner typiquement avec la préposition *de* (comme *accuser*), ou lorsqu'ils suivent un mot lui-même précédé d'un déterminant – qui dans ce cas peut s'interpréter comme un nom (comme dans *le palais du roi*).

Après l'analyse d'un texte par *SN\_FR.pl*, les résultats obtenus sont exportés, comme nous l'avons dit, vers un fichier CSV, généré par le script lui-même. Ce fichier comporte une ligne par syntagme repéré, composée respectivement d'un numéro d'ordre, du syntagme, de son nom tête, de la forme au singulier de ce nom tête, d'une indication quant au contexte qui a permis de repérer le syntagme, la nature morphosyntaxique de l'emploi et, pour finir, une reproduction de la phrase dans lequel le syntagme a été repéré. Dans ce fichier, les cas non désambiguïsés de *du* et *de la* sont marqués comme tels.

Dans le cadre de nos études en corpus, nous n'avions bien entendu pas besoin d'informations concernant tous les syntagmes nominaux représentés dans nos données, mais seulement sur les emplois des noms spécifiques qui nous intéressaient. Dès lors, pour l'analyse de nos données de *WebCorp*, nous avons procédé en deux étapes. Pour chaque fichier TXT (correspondant aux occurrences d'un item), nous lançons d'abord une analyse complète par *SN\_FR.pl*. Ensuite, du tableur contenant les résultats, nous ne gardions que les lignes concernant l'item en question (par le biais d'un tri effectué par un autre script, que nous passerons sous silence par souci de ne pas trop nous éloigner de notre sujet), et complétions pour celles-ci les éventuels cas à désambiguïser manuellement (normalement, moins de 5% du total des occurrences). Nous relisons aussi systématiquement tous les cas annotés comme massifs sur la base d'un partitif, pour vérifier qu'il n'y avait pas eu d'erreur dans la désambiguïsation automatique (il y en avait parfois, mais très peu : nous les corrigeons aussitôt). Pour finir, nous copions cette sélection d'occurrences pertinentes, dûment annotées, dans un fichier Excel général, contenant l'ensemble de nos données (pour chaque item de chaque étude, dans nos deux langues) – et dans lequel nous annoterions encore certaines occurrences, sur une base sémantique, voir ci-après.

La création de notre script *SN\_FR.pl*, démarche sur laquelle nous n'avons que très incomplètement levé le voile ici (voir (Vermote 2011, ms.) pour plus de détails), aura permis de nourrir une réflexion approfondie sur les différents contextes syntagmatiques nominaux, et leur nature massive, comptable, cumulative ou neutre. En outre, elle aura, à moyen terme, été la cause d'un gain de temps considérable dans l'analyse de nos données, nous permettant d'obtenir très rapidement et facilement des données chiffrées homogènes sur le comportement morphosyntaxique de n'importe quel nom, à partir de données de *WebCorp* ou, le cas échéant, de n'importe quel texte disponible au format TXT. C'est notamment par ce grand degré de précision de notre outil d'analyse en français que peut s'expliquer le déséquilibre dans le nombre d'items étudiés en français et en néerlandais, dans le domaine des noms de fruits et de légumes :

si pour le français, nous pouvons dorénavant récolter plusieurs centaines d'occurrences d'un nom sur *WebCorp*, et procéder à leur classement syntagmatique en quelques dizaines de minutes à peine, la procédure analogue pour le néerlandais n'est pas aussi évidente, car reposant sur des outils d'automatisation beaucoup moins poussés, et se heurtant à des difficultés spécifiques à la grammaire de cette langue (impliquant une bien plus grande part d'annotation manuelle). Nous en disons un mot au point suivant.

#### 4.4.4 Analyse semi-automatique en néerlandais

Pour l'étude de nos données en néerlandais, nous n'avions pas à disposition d'outil analogue à *SN\_FR.pl*, essentiellement parce que nous jugions qu'il n'était pas opportun d'investir une seconde fois plusieurs mois de travail dans la création d'un script exhaustif de repérage des syntagmes nominaux. Néanmoins, nous avons créé un outil de plus modeste envergure, développé spécifiquement pour nous assister dans l'annotation d'une liste d'occurrences en contexte d'un nom spécifié. Ainsi, ce second script, baptisé *AN\_NL.pl*, ne fonctionne non pas sur n'importe quel texte – dans lequel serait repéré l'ensemble des syntagmes nominaux – mais est taillé pour traiter un ou plusieurs tableurs CSV, indiquant dans leur titre le nom concerné par l'analyse, et contenant une liste d'occurrences de ce nom. Programmé pour reconnaître certains contextes spécifiques, et basé aussi sur une (petite) mémoire lexicale contenant essentiellement des adjectifs et des verbes transitifs, le script passe en revue chaque ligne d'un pareil fichier, y repère le nom sous examen et, s'il reconnaît un certain contexte, annote la phrase en mentionnant le SN concerné ainsi que sa nature – massive, comptable ou neutre.

Dans un souci d'efficacité, nous n'avons pu mettre en place pour le néerlandais que des stratégies d'analyse automatique assez générale, permettant seulement d'éviter d'avoir à analyser manuellement tout un ensemble de cas repérables sans ambiguïté. Sont concernés ici essentiellement les articles et les numéraux (*de banaan* 'la banane', *drie bananen* 'trois bananes'), différents déterminants définis et indéfinis (*elke banaan* 'chaque banane', *wat banaan* 'de la banane, un peu de banane') de même que cumulatifs (*hoeveel bananen* 'combien de bananes'), ainsi que les constructions partitives directes avec un terme de mesure (*een kilo banaan* 'un kilo de banane'). Profitant que pour le néerlandais, l'approche de notre script soit basée sur un nom spécifique à reconnaître en contexte (et non sur le repérage exhaustif de tous les noms d'un texte sans discrimination *a priori*), nous avons aussi fait usage de la distinction entre forme au singulier et forme au pluriel, et annoté d'emblée toutes les occurrences du nom au pluriel comme comptable, y compris lorsqu'aucun autre indice contextuel n'était repéré. C'est ce qui explique que nous n'ayons pas prévu d'annotation spécifique en néerlandais pour les contextes cumulatifs, directement assimilés, sur la base du nombre,

aux massifs ou aux comptables (tandis que pour le français, nous n'avons discriminé entre singulier et pluriel qu'à la fin du parcours d'analyse, une fois les données rassemblées dans notre tableur final, à l'aide des options de filtrage d'Excel). Cette différence d'approche entre nos deux langues est d'autant plus pertinente que si en français, les déterminants portent normalement la marque du nombre (*les, des, ces, etc.*), ce n'est pas le cas du néerlandais : l'article défini et les démonstratifs du genre commun sont identiques à leurs formes correspondantes au pluriel (*de, die, deze*), de même que les possessifs, par exemple, ne varient pas en nombre (*mijn, je/jouw, jullie, etc.*). Ceci oblige que l'on tienne compte du premier coup, dans l'analyse automatisée en néerlandais, de la forme – au pluriel ou au singulier – du nom. Notons du reste que contrairement au français dans la majeure partie des cas, l'alternance entre singulier et pluriel en néerlandais correspond bien à une différence phonologique, rendant ce critère à l'écrit d'autant plus fiable pour servir de point de départ à l'analyse morphosyntaxique – tandis qu'en français, le -s pluriel muet est sujet aux interférences des imprécisions orthographiques.

Cependant, ce petit avantage du néerlandais pour l'analyse automatique est très fortement contrebalancé par une autre différence majeure, source de bien plus grandes difficultés : l'existence, dans cette langue, d'un emploi nominal nu. Les emplois nus au pluriel, évidemment, ne posent pas de problème : nous l'avons dit, ils sont classés directement dans les comptables. Par contre, les emplois nus au singulier sont bien plus délicats à repérer – et sont cruciaux pour nous, puisqu'ils constituent la majeure partie des emplois massifs. Leur difficulté est double. D'une part, en l'absence d'une forme spécifique quelconque les précédant, ils deviennent impossible à repérer pour notre outil d'analyse assistée – qui repose crucialement sur les contextes de surface pour discriminer entre emplois. Or, à ce niveau, pour dégrossir malgré tout l'analyse, nous avons intégré à notre script, sans grande prétention d'exhaustivité mais pour gagner quand même du temps, un certain nombre de mots ne pouvant que précéder un nom dans son emploi nu, notamment des conjonctions (*alhoewel* 'bien que', *waarvan* 'dont', etc.), des adverbes (*zelfs* 'même'), et des verbes transitifs directs. D'autre part, l'impossibilité d'analyser automatiquement tous les emplois nus implique aussi l'interdiction d'exclure *a priori* un très grand nombre de cas où le nom est simplement utilisé seul comme titre d'une page, ou comme nom d'un onglet – comme dans l'exemple (80) – une situation représentée en nombre non négligeable parmi les données de *WebCorp*.

- (80) Bloemen, vruchten en vermeerdering[bewerken] bloeiwijze en vruchtaanzet  
banaanbloempjes Wilde banaan De bloemen op de dikke, algemene bloeistengel zijn in  
dwarse

En français, ces cas ne posaient jamais problème : le script n'y reconnaîtrait pas de déterminant ou de contexte supposé être repéré, et passerait donc directement à la

suite, sans en tenir compte. En néerlandais, il a fallu en revanche effectuer manuellement le tri de toutes ces occurrences, que nous avons généralement soit supprimées, soit annotées comme ‘autres’ (à l’instar des emplois métalinguistiques, comme *De naam banaan is afkomstig van het Arabisch (...)* ‘Le nom banane vient de l’arabe’, etc.).

Ainsi, pour l’analyse de nos données en néerlandais, nous avons également profité des avantages de l’assistance informatique sur mesure, mais dans une bien moindre mesure que pour le français. Si en général, pour le français, il restait toujours moins de 5% de cas à analyser manuellement (ambiguïté de *du-de la*), pour le néerlandais il s’agissait souvent de plus de 50% des occurrences de *WebCorp* à parcourir manuellement. (Dans cette tâche, nous avons été assisté pour une partie de l’analyse par Niek Van Wettere, étudiant en philologie romane et jobiste dans notre faculté en septembre 2013. Nous le remercions.) C’est ce qui explique que nous ayons un peu moins de données en néerlandais, en particulier pour le domaine des noms de fruits et légumes.

## 4.5 L’analyse qualitative

L’analyse distributionnelle sur la base des contextes de surface, telle que nous l’avons décrite ci-avant, si elle a constitué une part importante de notre travail, n’était pourtant que la première étape de nos études de cas. Comme nous l’avons dit, notre second objectif était de décrire de manière systématique les effets sémantiques associés à la flexibilité massif-comptable dans différents domaines, et en particulier de nous pencher sur les cas *a priori* minoritaires (c’est-à-dire supposément dérivés). Pour cette tâche, seul un examen manuel des données était envisageable. Dans la partie présente, nous disons quelques mots des grands principes généraux qui nous ont guidé dans cette analyse, pour nos trois études de cas. Les résultats plus détaillés, quant à eux, seront présentés dans les chapitres suivants.

Pour les noms de fruits et de légumes, les cas minoritaires à examiner de plus près étaient, bien évidemment, les emplois massifs. En français, nous avons parcouru ces occurrences pour conclure à un sens général unique de cette lecture massive (voir aussi (Vermote 2013)) et pour identifier les différents facteurs contextuels facilitant cet emploi. En néerlandais, si l’analyse sémantique globale ne sera pas très différente, elle est néanmoins sous-tendue par une différence majeure avec le français : la possibilité pour la lecture massive nue d’apparaître non seulement en contexte spécifique, mais également dans une lecture générique (correspondant à l’article défini générique, donc neutre, en français ; voir 1.2.2) – gonflant la proportion de massifs identifiés. Afin de permettre une meilleure comparaison des données chiffrées de nos deux langues, nous



avons dès lors annoté manuellement les occurrences massives du néerlandais, en distinguant lectures homogènes spécifiques et génériques de nos noms de fruits et de légumes.

Dans le domaine des noms de matériaux, les emplois que l'on peut supposer dérivés, et donc intéressants pour notre examen sémantique, sont les comptables. Pour le français et le néerlandais, notre démarche a été sensiblement identique, consistant à identifier, parmi les emplois comptables repérés (en proportions très variables d'une de nos langues à l'autre), les emplois correspondant au Trieur Universel (voir point 1.5.2) et ceux correspondant à une lecture objet, potentiellement lexicalisée.

Pour finir, nos noms de boissons ont fait l'objet d'un examen très similaire à celui des noms de matériaux, supposant aussi l'emploi massif comme basique, et consistant à identifier dans les usages comptables les effets sémantiques du Trieur et du Conditionneur Universels – annotés comme tels pour pouvoir permettre une comparaison quantitative entre le français et le néerlandais.

## **4.6 Analyse de quelques items de contrôle**

Nous clôturerons la présentation méthodologique de nos études de corpus par quelques mots concernant l'analyse distributionnelle d'un groupe d'items extérieurs aux trois domaines que nous étudierons plus en détail, à savoir un ensemble de douze noms typiquement comptables, et non flexibles morphosyntaxiquement. La liste de ces items, en français et néerlandais, se trouve reproduite dans la Figure 8. Les raisons de leur sélection seront détaillées plus loin (voir Chapitre 9), mais notons déjà qu'il s'agit des items qui seront utilisés comme « contrôles » dans nos enquêtes d'acceptabilité.

#	<i>français</i>	<i>néerlandais</i>
1	AMPOULE	GLOEILAMP
2	BALLON	BAL
3	BIJOU	JUWEEL
4	BOCAL	POT
5	CASSEROLE	PAN
6	COUTEAU	MES
7	CUILLERE	LEPEL
8	GOBELET	BEKER
9	LIVRET	BOEKJE
10	LOUCHE	POLLEPEL
11	SACHET	ZAKJE
12	TASSE	KOPJE

Figure 8 Liste des items de contrôle.

Pour ces douze items, nous avons récolté des données d'occurrences en français via *WebCorp* (avec une moyenne de 566,4 occurrences pertinentes par item), et les avons soumises à une analyse distributionnelle à l'aide de notre logiciel *SN\_FR.pl*. De cette analyse, il ressort que nous ne trouvons, pour aucun de ces noms, la moindre occurrence en emploi massif. Nous pouvons donc conclure que le profil distributionnel observé correspond à nos intuitions, que ces noms sont tous rigoureusement comptables.

Par ce modeste exemple, nous entendons seulement illustrer la pertinence de notre approche syntagmatique de l'opposition massif-comptable : effectivement, il semble bien que les profils récupérables en corpus disent quelque chose de la nature des noms étudiés. Ainsi, si pour des noms intuitivement non flexibles, aucune flexibilité n'est effectivement observée, l'on peut s'impatienter d'ores et déjà de voir ce qu'il en sera des profils distributionnels des noms *a priori* flexibles que nous avons sélectionnés. Comme on le verra, les profils seront bien plus variables, avec presque toujours une certaine flexibilité à constater – sauf dans le cas de certaines exceptions, que nous veillerons bien à traiter comme telles. (Pour nos différentes études de cas présentées dans les chapitres suivants, nous illustrerons systématiquement les profils distributionnels calculés sur la base de graphiques à bâtonnets empilés, représentant la proportion d'occurrences massives par rapport aux occurrences comptables ; ici, une telle présentation n'a pas de sens, puisqu'elle reviendrait à afficher douze bâtonnets monochromes côte à côte.)

Concernant le néerlandais, nous avons également récolté des données pour ces items sur *WebCorp* (moyenne de 579,3 occurrences brutes par nom), mais précisons par souci de complétude que nous n'avons pas effectué l'analyse distributionnelle complète de ces données – pour les raisons que nous avons dites au sujet du déséquilibre dans nos outils d'analyse automatisée. Nous nous sommes contenté d'analyser complètement une des séries seulement, celle de *JUWEEL* (car comportant, par le hasard des choses, le moins d'occurrences à analyser manuellement), et n'y avons décelé, comme attendu, aucune

occurrence en emploi massif. Par extrapolation, nous n'avons aucune raison de croire que nous trouverions, pour les autres items en néerlandais, la moindre occurrence massive non plus. D'abord, nous n'en trouvons aucune parmi la portion d'occurrences analysées automatiquement par notre *AN\_NL.pl* pour les items restants (environ 50% des cas). Ensuite, nous n'imaginons pas que ces items puissent s'utiliser de quelque manière que ce soit comme massifs (ce qui est confirmé par l'intuition de quelques locuteurs natifs interrogés), car nous ne voyons pas en quoi ils différeraient, sur ce point, de leurs équivalents en français – d'autant plus qu'il a déjà été montré qu'en général, le néerlandais était moins apte à la variation entre massif et comptable que le français (voir (Lauwers et Vermote 2014, à paraître)).



# Chapitre 5 Les noms de fruits et de légumes

## 5.1 Les noms de fruits et de légumes en français

### 5.1.1 Analyse distributionnelle

La manière dont nous avons mené nos trois études de corpus a été expliquée en détail au chapitre précédent : il n'est donc plus besoin de préciser ici ce que nous entendons par l'analyse distributionnelle de nos données. Nous pouvons donc commencer d'emblée par une représentation graphique des résultats de cette première analyse, pour nos noms de fruits et de légumes en français.

La Figure 9 illustre les profils distributionnels de l'ensemble des 29 noms que nous avons étudiés dans cette catégorie. Les emplois massifs et comptables sont représentés respectivement en rouge et en bleu. Notons que les emplois en contexte cumulatif ont été assimilés à l'un ou l'autre de ces deux types de syntaxe, en fonction de l'apparition au pluriel ou au singulier du nom, et sont donc inclus à cette représentation, sans être distingués explicitement. Les emplois neutres, quant à eux, ont été laissés de côté, comme annoncé précédemment. Les noms sont classés par ordre alphabétique sur l'axe des abscisses. Les chiffres absolus sur lesquels repose cette représentation peuvent être trouvés en Appendice (12.4.2).

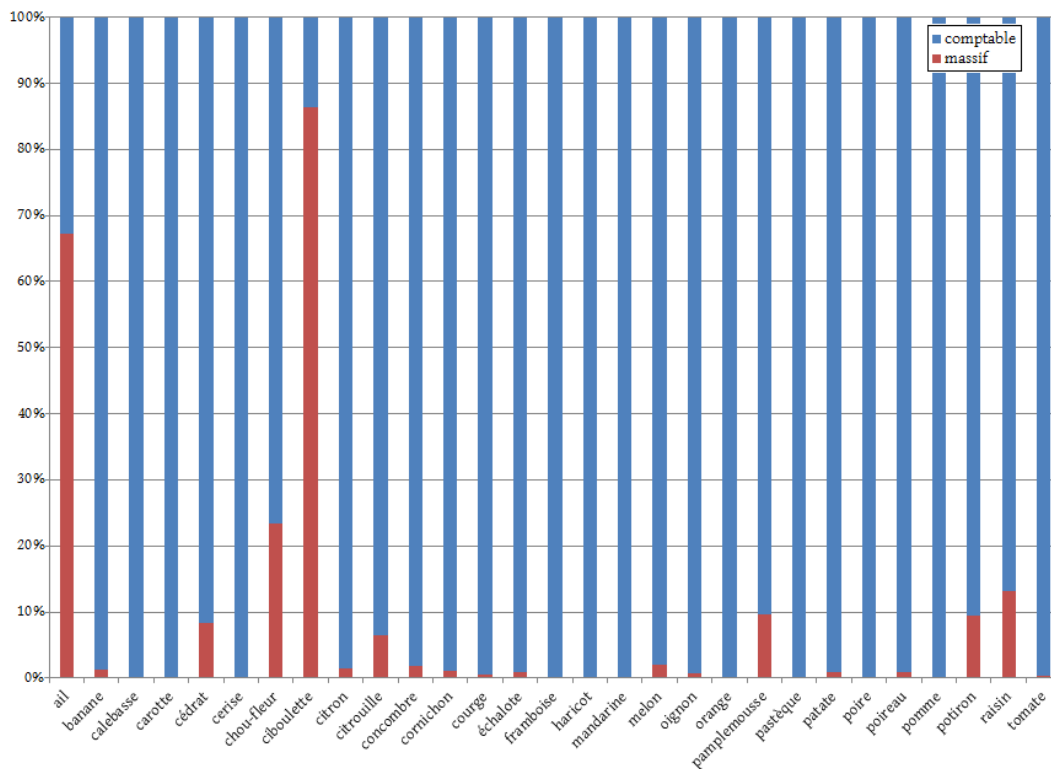


Figure 9 Profils distributionnels des 29 noms de fruits et de légumes étudiés en français. Les emplois massifs sont représentés en rouge, les emplois comptables en bleu.

Un premier constat général s'impose : les emplois comptables dominent largement en nombre les emplois massifs, pour tous les items étudiés sauf deux. Les deux exceptions sont constituées par les items *ail* et *ciboulette*, dont le caractère différent de tous les autres était à vrai dire plutôt prévisible, et pourra donc s'expliquer aisément. Si nous les avons intégrés à notre liste par souci de complétude botanique (il s'agit de deux représentants du genre *Allium*, au même titre qu'*oignon* ou *poireau*), il est évident que du point de vue de nos habitudes culinaires, ils ne sont jamais utilisés comme des légumes que l'on aimerait à bouillir, blanchir, étuver ou préparer en julienne. Au contraire, ils sont normalement utilisés pour leurs propriétés condimentaires, c'est-à-dire comme exhausteurs de goût, ce qui les placerait de ce point de vue-là dans la catégorie des condiments (aux côtés de *sel*, *poivre* ou encore *ketchup*) plutôt que dans celle des fruits et légumes. C'est ce qui explique leur comportement morphosyntaxique fondamentalement massif (confirmé par l'intuition : *de l'ail*, *de la ciboulette* vs. *??trois aulx*, *??quelques ciboulettes*), différent du reste de notre sélection. Un regard plus détaillé sur leurs attestations comptables (au nombre de 17 pour *ail* et 3 pour *ciboulette*, dans nos données de *WebCorp*) dévoile d'ailleurs que celles-ci ne sont possibles que dans une lecture en termes de sortes (Trieur Universel) typique des noms fondamentalement massifs (81)-(83) ou, dans quelques attestations isolées, à l'intérieur d'une expression figée (84).

(81) Tous les ails produits sont le résultat de la multiplication par voie végétative.

- (82) L'aillet est un ail immature, qui n'a pas encore formé ses gousses.
- (83) N'hésitez pas à raser les ciboulettes au ras du sol en début après la floraison
- (84) Ne pas valoir la queue d'un vieil ail.

Profitions-en pour remarquer un autre *outsider* parmi nos fruits et légumes. A mieux y réfléchir, le nom *calebasse* (aucune attestation massive dans nos données) n'est pas non plus réellement à sa place dans la catégorie qui nous intéresse ici. S'il s'agit bien, techniquement, d'un représentant de la famille des Cucurbitacées, à l'instar de la *citrouille*, du *concombre* ou encore du *melon*, elle se distingue strictement de ses proches parents botaniques en ce qu'elle n'est pas couramment destinée à la consommation, mais sert plutôt à la confection de divers objets, comme des récipients ou des instruments de musique. Nous pourrions donc faire l'économie d'un examen plus rapproché de son cas.

Tous nos autres items, quant à eux, remplissent bien notre critère de sélection, d'être des fruits ou des légumes – du point de vue de nos connaissances encyclopédiques et de nos habitudes culinaires. Dans la Figure 10, nous présentons à nouveau leur profils distributionnels, en excluant cette fois les items *ail*, *ciboulette* et *calebasse*. Ceci nous permet de ne représenter cette fois sur l'axe des ordonnées que la tranche allant de 0 à 25 %, pour plus de clarté.

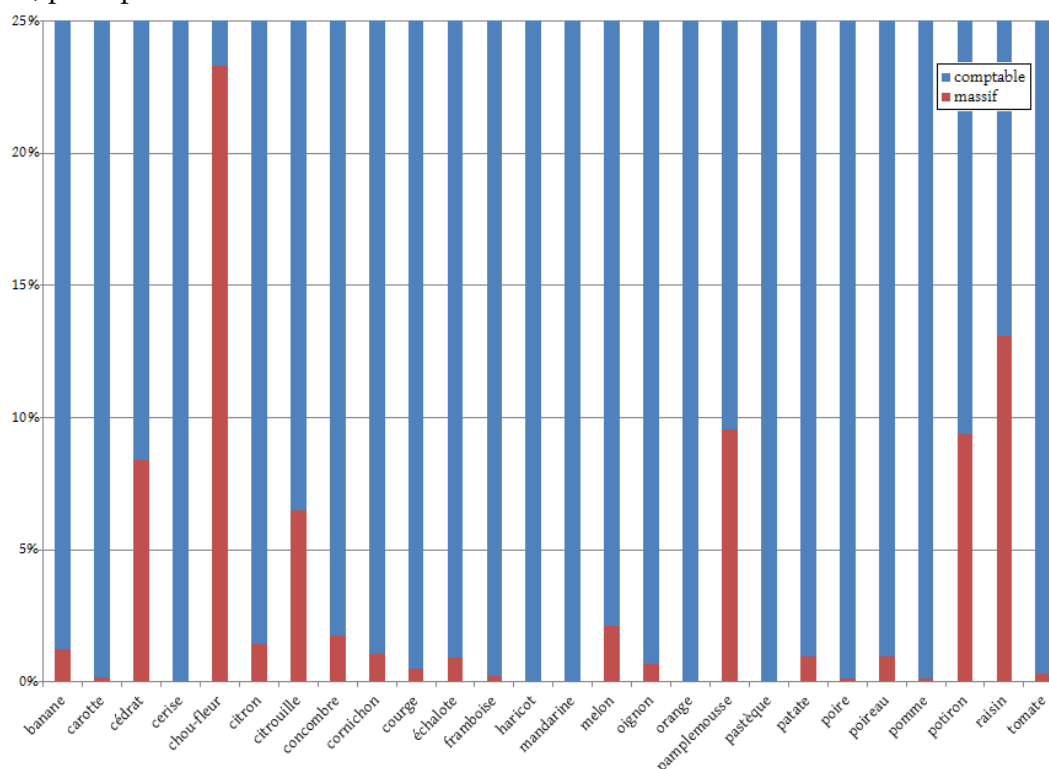


Figure 10: Profils distributionnels de nos noms de fruits et de légumes en français, après exclusion d'*ail*, *ciboulette* et *calebasse*. L'axe des ordonnées ne présente que les 25 % inférieurs.

Comme nous le disions ci-dessus, le premier et principal constat à dresser est celui du déséquilibre entre la proportion d'emplois massifs et comptables pour tous nos items, les premiers étant largement dominés par les seconds. Ceci confirme notre intuition de départ, que les noms de fruits et de légumes sont avant tout comptables, avec néanmoins un certain degré de flexibilité.

Concernant ce dernier point, nous observons un degré de variation non négligeable : si dans l'ensemble, nos items apparaissent bel et bien flexibles, ils varient quant à leur affinité apparente avec le massif. A se baser seulement sur l'apparence du graphique en Figure 10, l'on peut distinguer au moins trois groupes :

1. Les items présentant plus de 5 % d'attestations en syntaxe massive : *cédrat, citrouille, pamplemousse, potiron, raisin* et, en tête à près de 25 %, *chou-fleur* ;
2. Les items présentant moins de 5 % d'attestations en syntaxe massive : *banane, carotte, citron, concombre, cornichon, courge, échalote, framboise, melon, oignon, patate, poire, poireau, pomme, tomate*.
3. Les items pour lesquels nous ne trouvons aucune attestation en syntaxe massive : *cerise, haricot, mandarine, orange, pastèque*.

Le premier groupe contient des noms dont les emplois massifs étaient certainement attendus. Le cédrat est un proche parent du citron, et si *un cédrat* est bel et bien un exemplaire de ce fruit, son usage le plus courant est en tant que zeste confit ou qu'essence en parfumerie : deux cas où l'on aura moins affaire à des individus qu'à *du cédrat*, homogénéisé. La citrouille, le potiron et le chou-fleur, quant à eux, sont trois légumes de grande taille : lorsqu'ils sont consommés par l'être humain, ce n'est jamais sous forme d'individus entiers (qui seraient les référents de l'emploi comptable). Dès lors, lorsqu'on les désigne dans un contexte de consommation, ils attirent inmanquablement la syntaxe massive : *on mange du chou-fleur*, et jamais *un chou-fleur* (tandis que pour la majorité des autres items, le choix existe : on peut manger *de la pomme*, mais aussi sans problème *une pomme*). C'est peut-être aussi ce qui explique le caractère légèrement plus massif de *pamplemousse* – un agrume de grande taille. Pour finir, *raisin* est bel et bien à sa place lui aussi dans cette première catégorie : il est effectivement plus porté sur le massif que certains de ses co-hyponymes, comme le confirme l'intuition. Ainsi, par exemple, si l'on doit désigner respectivement plusieurs pommes ou plusieurs raisins, il nous semble que *de la pomme* serait effectivement moins naturel dans le premier cas que *du raisin* dans le second. Il se trouve du reste, comme nous avons pu le constater dans le courant de notre recherche, des locuteurs natifs qui n'acceptent que la lecture massive de *raisin*, et jugent incorrect *un raisin* pour désigner un grain de raisin. (C'est le cas, par exemple, du professeur Boutier, du département de langues et littératures romanes de l'Université de Liège – communication personnelle lors de la journée d'étude TraSoGal, le 14 mai 2012.) Si pour notre part, nous trouvons bel et bien dans nos données de tels emplois comptables de *raisin*, nous en trouvons par



ailleurs aussi des lectures taxonomique, comme (85) et (86) ci-après, confirmant son statut ambivalent, effectivement plus porté sur le massif que la majorité de ses co-hyponymes :

(85) Recette de la caille aux deux raisins.

(86) Le raisin de Corinthe (*vitis vinifera* *apyrena*) est un raisin sec.

Passons au second groupe, qui est le plus grand. A ce stade, nous n'identifions pas de particularité commune à tous ses membres qui expliquerait leur caractère similaire du point de vue de la flexibilité. Contentons-nous de supposer que ce degré de flexibilité, situé entre 0 % et 5 % d'emplois massifs sur l'ensemble des emplois massifs et comptables combinés, pourrait être considéré comme le caractère d'un nom de fruit ou de légume typique.

Pour finir, notre dernier groupe soulève une question intéressante : dans quelle mesure l'absence d'attestation en syntaxe massive, dans notre corpus, correspond-elle à une impossibilité combinatoire ? Pour *haricot*, l'intuition nous dicte que cet item est en effet particulièrement rétif à l'emploi en syntaxe massive (*??du haricot*). Il est intéressant de noter que c'est le cas pour plusieurs noms de légumes similaires, de la famille des Fabacées : *??du petit pois*, *??du pois chiche*, *??du flageolet*, *??de la lentille*. (Ostler et Atkins 1992: 84) remarquent cette même particularité en anglais, où les équivalents de traduction *bean* 'haricot', *pea* 'pois' ou *lentil* 'lentille' résistent également à un emploi massif – un fait qu'ils expliquent comme une exception sémantique à leur Règle d'Implication Lexicale « Food Item – Mass », permettant l'emploi massif en terme de matière comestible de n'importe quel nom comptable désignant un objet de consommation alimentaire (fruits et légumes y compris, donc). Quant à *cerise*, *mandarine* et *orange*, l'absence d'attestations massives dans nos données semble plutôt pointer vers une fréquence fort basse, mais ne cachant pas nécessairement un blocage total : il nous semble qu'on peut avoir sans problème (87) – tandis que (88) serait bel et bien plus marqué.

(87) Ce petit goût, je pense que c'est de la cerise / de la mandarine / de l'orange.

(88) ?Ce petit goût, je pense que c'est du haricot.

Pour finir, la présence de *pastèque* est quelque peu surprenante : non seulement, nous n'avons aucun mal à imaginer des exemples d'emploi de ce nom en syntaxe massive, mais en outre, les caractéristiques physiques de ce fruit sont similaires à celles, par exemple, du potiron ou du chou-fleur, en ce que lui non plus ne peut se consommer, à l'échelle humaine, sous la forme d'individus entiers : on mange toujours *de la pastèque*, et pour ainsi dire jamais *une* (ou *des*) *pastèque(s)*. Néanmoins, la pastèque est possiblement plus rare dans nos assiettes que ses comparses le melon ou la citrouille : l'on aurait ainsi peut-être moins souvent l'occasion d'évoquer d'avoir consommé *de la pastèque* que *du*

*melon* ou *de la citrouille* – ce qui se refléterait dans nos données de corpus ? Nous ne nous attarderons pas sur cette supposition.

Notons encore que par curiosité, nous avons effectué pour les quatre items de notre dernier groupe une nouvelle recherche de données sur *WebCorp*, ciblée cette fois spécifiquement sur le repérage d'occurrences massives, par le biais de l'emploi avec la locution déterminative *un peu de*. Pour chacun des items, nous trouvons alors bel et bien quelques attestations, au nombre de 26 pour *cerise*, 13 pour *haricot*, 20 pour *mandarine* et 37 pour *pastèque* (notons que vu la nature ouverte de *WebCorp*, ces chiffres absolus sont délicats à interpréter ; nous les donnons seulement à titre indicatif). En voici quelques exemples :

- (89) Belle intensité olfactive avec des fruits rouges et un peu de cerise.
- (90) Le manioc n'ayant pas beaucoup de goût, on met un peu de haricot mungo qui a deux avantages (...).
- (91) Si vous vous sentez stressé, ajoutez un peu de mandarine, apaisante.
- (92) Les deux oursons ont ensuite mangé un peu de pastèque avant de se reposer dans leur enclos.

Pour l'exemple (90), avec *haricot*, nous avons sélectionné expressément une occurrence où l'accord du verbe (ici *avoir*, au présent) est bien au singulier, montrant qu'*un peu de haricot* n'est pas simplement la forme plurielle massive mal orthographiée, où l'on aurait oublié le -s final. Ceci nous montre donc que les quatre items de notre dernier groupe ne sont pas définitivement exclus en emploi massif (pas même *haricot*) – néanmoins, si l'on en croit nos données, leur flexibilité serait particulièrement limitée.

En conclusion, cette première investigation quantitative nous apprend deux choses. D'une part, que pour tous les noms de fruits et de légumes étudiés en français, les emplois massifs sont minoritaires, voire absents (si pas impossibles) dans certains cas – ce qui confirme le statut prioritairement comptable que nous leur avons donné intuitivement. D'autre part, que le degré de flexibilité de ces items, s'il doit être reflété par le rapport entre nombre d'occurrences massives et nombre d'occurrences comptables, est fortement variable : tous nos items connaissent au moins quelques emplois massifs, mais ne semblent proportionnellement pas aussi friands les uns que les autres de ce type de syntaxe. Pour pouvoir décrire facilement cette variabilité, nous avons pour le moment classé nos items en trois groupes, allant du plus au moins flexible. Ce faisant, nous n'avons cependant pas encore vraiment fait honneur à leurs profils spécifiques, individuels. Plus loin, lorsque nous décrirons nos enquêtes d'acceptabilité puis notre expérience d'amorçage, nous verrons que ces profils morphosyntaxiques spécifiques, tels qu'ils ressortent de notre corpus, pourront être traduits par un indice de « biais massif » dont nous aurons grandement l'usage. D'ici là, passons à présent – après ces considérations chiffrées – à un examen sémantique plus approfondi de nos données.

## 5.1.2 Analyse sémantique

### 5.1.2.1 Les fruits et légumes en forme et quantité indéterminées

Comme nous l'avons annoncé au chapitre précédent, l'essentiel de nos analyses sémantiques plus détaillées se concentrera sur les emplois qui, dans le domaine à l'examen, peuvent être dits marqués. Dans le cas des noms de fruits et de légumes, nous devons donc nous pencher de plus près sur les emplois massifs.

Comme nous l'avons vu au point précédent, ces derniers sont cependant très peu fréquents, et les attestations que nous en avons, telles qu'identifiées parmi nos données de base, c'est-à-dire les données discutées au point précédent, obtenues en effectuant une simple recherche sur *WebCorp* pour chacun de nos noms, ne sont pas très nombreuses : 225 au total (204 en contexte massif plus 21 en contexte cumulatif singulier), voire seulement 171 (156 + 15) si l'on exclut les attestations d'*ail* et *ciboulette*.

Ainsi, afin d'avoir plus de matière à discussion dans la section présente de même qu'au point suivant, nous ajouterons aux exemples traités ci-après des attestations supplémentaires en syntaxe massive, obtenues pour 15 de nos 29 noms par le biais de deux recherches *WebCorp* plus ciblées, ayant pour objet respectivement les constructions *un peu de N* et *du/de la + N*. (Les 15 items en question sont ceux étudiés pour notre toute première étude de cas, voir dans (Vermote 2013), moins *ail*, *ciboulette* et *calebasse*. Voir les détails de nos recherches sur *WebCorp*, en Appendice.) La recherche avec *un peu de* a permis d'obtenir 687 occurrences massives additionnelles, celle avec *du/de la* un total de 2 826 occurrences dont 190 (~14,8%) se sont avérées relever effectivement de l'article partitif, et donc du massif (les autres cas, où *du/de la* correspondait à la combinaison prépositionnelle *de + article défini*, on été exclus de manière semi-automatisée, grâce au traitement de *SN\_FR.pl*).

Dans ce qui suit, nous illustrerons dès lors nos propos à partir de l'ensemble de nos attestations de noms de fruits et de légumes en syntaxe massive, tant les 171 obtenues de base, que les 190 plus 687 supplémentaires. C'est ce qui explique entre autre que l'on pourra trouver, parmi nos exemples ici et au point suivant, des occurrences massives de *pastèque* et *orange*, alors qu'au point précédent, nous disions ne pas en avoir trouvé.

Ces précisions faites, considérons à présent une première sélection d'exemples :

- (93) En 2007, environ cent trente pays produisaient de la banane.
- (94) Si vous trouvez encore du potiron sur le marché, lancez-vous (...).
- (95) Ayant de la citrouille dans le congélateur, c'est l'occasion [d'essayer cette recette].
- (96) L'idée du flan me plaît beaucoup et il me reste un peu de courge dont je ne savais que faire.
- (97) Les paysans étaient tenus d'offrir de la pastèque aux voyageurs assoiffés.
- (98) Couper un peu de poireau en fines tranches (...).
- (99) Manger de la carotte c'est vrai que [ç]a rend les cheveux brillant[s].

Dans notre point 1.5.2.1, nous avons décrit les deux mécanismes sémantiques principaux responsables, d'après nos prédécesseurs, des emplois massifs de noms à usage principalement comptable : le Broyeur et le Multiplicateur Universels (nous laissons de côté le Broyeur Qualitatif, qui ne s'applique de toute évidence pas aux exemples ci-dessus). Or, si l'on interprète le premier comme une opération de broyage effectif, résultant en l'obtention d'une bouillie de quelque sorte, alors force sera de constater qu'aucun de nos exemples (93) à (99) ne se laisse décrire sans ambiguïté par l'un ou l'autre mécanisme.

Des tendances sont peut-être à identifier : dans les exemples (93) et (94), la *banane* et le *potiron* dont il est question sont probablement à considérer comme jouissant encore de leur intégrité physique – rapprochant ces cas du Multiplicateur. En revanche, dans (95) et (96), les restes de *citrouille* et de *courge* sont potentiellement découpés, préparés – ce qui pourrait les rapprocher d'une lecture broyée. Cependant, il faudra accepter que ces suppositions ne découlent pas directement de l'interprétation de la syntaxe massive, mais plutôt d'un examen ultérieur, pragmatique : les bananes et les potirons se vendent normalement sous forme entière – mais l'on pourra tout autant dire avoir trouvé *du potiron* sur le marché si l'on en a acheté une demi pièce, ou même des morceaux préparés. En outre, si des pays produisent *de la banane*, cela pourrait être des bananes entières (c'est ce qu'on suppose en l'occurrence) mais aussi, avec le même énoncé, des morceaux ou des panades. En somme, le syntagme massif ne dit rien de la nature individualisée, multipliée ou découpée des bananes dont il est question.

Le même raisonnement s'applique aux exemples (95) et (96) : ils sont vrais s'il nous reste de la citrouille ou de la courge préparées, mais aussi s'il y au congélateur une citrouille entière, ou si le garde-manger recèle deux courges pas encore entamées. Prenons-en pour preuve la bizarrerie de l'interaction imaginaire suivante :

- (100) A : Il te reste encore du concombre ?  
B : \*Non, il m'en reste un.

Dans (100), B ne peut répondre par la négative, car s'il lui reste *un concombre*, il lui reste *de facto* bel et bien *du concombre*. Le syntagme massif n'implique donc pas fondamentalement une présentation préparée, découpée, broyée, pas plus qu'elle n'impose une lecture multipliée.

Nos exemples (97) à (99) confirment cette observation. Dans (97), l'interprétation la plus naturelle sera probablement celle où les paysans offrent aux voyageurs de la pastèque en morceaux. Mais cette précision relève d'une interprétation pragmatique, faisant appel à nos connaissances encyclopédiques : pour le même prix, les pastèques offertes pourraient être entières, ou au contraire réduites en bouillie (si ce n'est que ce fruit s'accommode mal de ce dernier type de présentation : de la pastèque broyée, ce n'est à vrai dire plus que de l'eau avec des pépins). Quant au poireau de (98), il ne peut être broyé, puisqu'il faut encore le couper. D'autre part, il n'est pas nécessairement

multiplié non plus : tout dépend probablement du nombre de personnes pour qui l'on est en train de cuisiner – *un peu de poireau* correspondant peut-être à un demi pied dans une recette pour deux, mais à cinq pieds entiers dans une recette pour vingt ! Pour finir, il en va de même pour la carotte en (99) : l'emploi en syntaxe massive de ce nom ne permet pas de tirer de conclusions, autres que pragmatiques, sur la présentation du référent – il peut s'agir de la consommation d'une ou de plusieurs carottes, entières ou pas.

Dès lors, nous entendons que c'est là le sens général de la syntaxe massive, telle qu'appliquée à un nom de fruit ou de légume : non pas de présenter son référent de manière plus ou moins finement broyée (Broyeur), ni de le présenter de manière démultipliée (Multiplicateur), mais simplement de l'évoquer de manière homogène, en forme et en quantité indéterminées, c'est-à-dire en faisant disparaître à l'arrière-plan la notion d'individus – que l'on trouvera pour sa part activée dans les emplois comptables. (Ce qui ne signifie pas que l'emploi du massif soit exclu pour référer à des individus : on peut déclarer avoir *du concombre* dans le frigo s'il nous reste exactement un exemplaire de ce légume, cf. (100). Ainsi, le massif n'implique pas l'absence d'individualisation, mais bien une indétermination à ce niveau-là. C'est là une proposition très proche de ce qu'avancait déjà (Gillon 1999: 52, 57).)

Du point de vue conceptuel, le lien entre la référence homogénéisée, défaite de ses individus, et la syntaxe massive s'explique aisément, si l'on suit le raisonnement de (Kleiber 1997) : l'emploi de la syntaxe massive implique qu'un référent est présenté sans limites intrinsèques, c'est-à-dire sans délimitation *a priori*. Or, pareille présentation non bornée implique obligatoirement une structuration interne homogène – car toute hétérogénéité signifierait l'existence de parties distinctes, et donc de limites. La présentation homogène, dépourvue d'individus (ou de toutes autres parties distinctes) est ainsi une condition nécessaire à la présentation non bornée qu'implique l'emploi de la syntaxe massive.

En contexte, cependant, le sens plus spécifique de cette lecture massive sous-déterminée se trouvera éventuellement précisée, par exemple comme multiplication ou au contraire destruction d'individus. (Cela ne fait pas nécessairement de l'opposition massif-comptable un cas exceptionnel : d'autres formes morphosyntaxiques ont de la même manière un sens très général, spécifié ultérieurement notamment par des facteurs pragmatiques. (De Caluwe 1994) étudie ce phénomène dans le cadre de la composition morphologique, et parle en la matière d'*open semantiek* ou « sémantique ouverte ».) Plus précisément, cela signifie que la présentation en discours d'un fruit ou d'un légume en forme et quantité indéterminées sera plus pertinente dans certains types de contextes que dans d'autres. Elle n'est donc pas arbitraire, mais motivée, au sens de (Wierzbicka 1988b). Ce sont ces différents contextes plus propices à l'emploi massif, tels que nous les avons observés dans nos données de corpus, que nous examinons au point suivant.

### 5.1.2.2 Réalisations contextuelles des emplois massifs

Le premier type de contexte s'accompagnant volontiers du massif est celui où il est question de consommation culinaire. C'est notamment ce qu'on observe dans les exemples suivants :

- (101) [Je] voudrais savoir si c'est pas dangereux de manger de la banane tout le temps.
- (102) Il est bien pour la santé de manger un peu d'oignon cru (...).
- (103) Un soir, au bout de cinq jours, il a mangé un peu de potiron.
- (104) J'ai mangé un peu de melon et j'ai abandonné les peaux dans un coin.
- (105) Comment faire manger du chou-fleur aux enfants ?

Les référents des noms de fruits et de légumes sont présentés ici comme matières ou substances comestibles. Dans ces cas, notons que seules les parties effectivement comestibles sont concernées : assurément, la banane de (101), l'oignon de (102) ou le melon de (104) ne sont pas consommés avec leur peau. Ceci n'est cependant pas le résultat du passage vers le massif, car la même chose est vraie dans le cas d'un emploi comptable : je peux déclarer avoir mangé *un melon* ou *une banane*, en omettant tout autant de consommer leur enveloppe incommestible. Dans les deux cas, le processus responsable de cette nuance est celui de la métonymie intégrée (où « certaines caractéristiques de certaines parties peuvent caractériser le tout », voir dans (Kleiber 1995: 123)), indépendant de l'alternance massif-comptable. Le choix pour l'un ou l'autre type de syntaxe dépendra seulement de ce que l'on désire mettre en avant : les individus, ou la substance. Dans le cadre de ce contexte d'alimentation, des facteurs pragmatiques entrent sans aucun doute en jeu : l'évocation de la consommation culinaire impliquera plus automatiquement la syntaxe massive pour les items tels que *chou-fleur* (105) ou *potiron* (103), dont on ne mange que rarement les individus entiers en une seule fois (voir ci-dessus), que pour des items comme *pomme* ou *cerise* – que l'on croque à l'unité, voire par poignées (voir aussi (Wilmet 2007: §150) pour une observation sur le lien entre taille des référents, et emploi en syntaxe massive ou comptable).

Un second facteur contextuel favorisant l'apparition d'un de nos items en syntagme massif concerne sa présentation sous forme hachée, découpée ou autrement émincée. Ainsi en est-il des énoncés (106) à (111), par exemple :

- (106) Ajouter un peu d'échalote hachée, puis un peu de porto et de fond brun.
- (107) Déposez au milieu de chaque assiette un peu de poireau finement émincé.
- (108) Dans le dernier verre ajoutez un peu de pastèque mixée et une cuillère de sucre.
- (109) Un taboulé (...) dont la base est du chou-fleur cru râpé comme de la semoule.
- (110) Sauce moutarde (...) relevée d'un peu de cornichon en julienne.
- (111) Décorer avec du pamplemousse, de l'orange et une cerise (...).

Dans (106) à (110), un adjectif participial ou autre locution prépositionnelle indique à chaque fois que les fruits ou légumes en question ne jouissent plus de leur intégrité

physique : *haché, émincé, mixé, râpé, en julienne*. Puisque le massif permet précisément de faire disparaître à l'arrière-plan conceptuel les individus référentiels, l'on comprendra aisément qu'elle soit particulièrement à sa place dans ces cas. Dans (111), c'est plutôt le contexte qui permet de déduire qu'il en va de même : l'on peut supposer que du pamplemousse ou de l'orange en décoration, ce ne sont pas des fruits entiers – contrairement à la cerise, présentée stratégiquement en syntaxe comptable pour préciser son individualisation.

Ainsi, l'on voit en (111) que l'alternance entre massif et comptable permet aussi de marquer des parallélismes et des oppositions : *une cerise* est une unité intègre, tandis que *du pamplemousse* et *de l'orange* sont fort probablement des morceaux de ces fruits. Dans notre corpus, nous trouvons bien d'autres attestations de pareilles mises en contraste, par le biais de l'opposition massif-comptable :

- (112) Sur 4 tranches de pain de mie, mettre du saumon, (...), des câpres, un peu d'oignon.
- (113) Dans un wok, faire revenir les champignons avec un peu d'échalote.
- (114) Je l'ai faite avec du saumon, des oignons, du poireau, (...) et des épices à poisson.
- (115) Cuisses de poulet cuites avec des figues et du citron (...).
- (116) Il dîne d'une pomme, d'un peu de pastèque, d'un thé, de quelques amandes.

Si les *câpres* dans (112) sont présentés, par le biais de la syntaxe comptable, comme une collection d'exemplaires entiers, l'*oignon* quant à lui peut se représenter, par contraste, sous forme de petits morceaux. Il en va de même pour l'*échalote* de (113), opposée aux *champignons*. A nouveau, des facteurs pragmatiques entrent ici en jeu : que la recette de (114) nécessite *des oignons* contre ce qui pourrait bien n'être qu'une partie de *poireau* a certainement trait à la taille respective de ces aliments. De même, c'est la différence de taille entre la pomme, la pastèque et l'amande qui explique la distribution en (116) : on consomme typiquement la pomme à l'unité (*une pomme*), la pastèque en morceau (*un peu de pastèque*), et les amandes par poignée (*quelques amandes*).

Un autre cas typique où nos fruits et légumes ne doivent pas être présentés sous forme d'individus, et où la syntaxe massive devient dès lors non seulement pertinente, mais en outre obligatoire (à moins de passer par une tournure paraphrastique), c'est lorsqu'il est question de jus (117)-(118), de goût et d'odeur (119)-(121) ou de parfum (122) :

- (117) Est-ce que mélanger du citron avec de l'eau fait maigrir ?
- (118) Recettes de boissons avec de l'orange.
- (119) Nez très ananas, un peu de melon et de pêche, de grillé (...).
- (120) Légèrement tourbé, sec gou[dr]onneux, avec un peu de mandarine (...).
- (121) Des arômes de citron et de melon, avec juste un peu de pamplemousse (...).
- (122) Sur son étagère à parfum, elle choisit de la verveine, du cédrat et du foin coupé (...).

Dans (117) et (118), on peut supposer que *du citron* et *de l'orange* désignent en réalité du jus de ces fruits – ce qui justifie, voire nécessite, l'emploi du massif. (Notons cependant que ceci reste une interprétation pragmatique, la référence au jus n'étant pas absolument obligatoire : un cocktail proche de la sangria, qui contiendrait des morceaux d'orange, serait aussi une *boisson avec de l'orange*. Dans ce cas-là, l'exemple se rapprocherait de ceux en (106) à (111), où il est question d'une forme préparée. Cette nuance possible continue de confirmer ce que nous disions de la valeur fondamentalement sous-déterminée de nos emplois massifs.) De même, lorsqu'il est question de goût et d'odeur (119)-(121) ou de parfum (122), la présentation sous forme d'individus distincts, associée au comptable, devient indésirable – et c'est le massif qui l'emporte.

Une fois de plus, des facteurs pragmatiques entrent ici crucialement en ligne de compte : si les noms d'agrumes comme *mandarine* ou *pamplemousse*, ou encore les noms *melon* ou *pêche* (ce dernier, qu'on retrouve dans (119), n'étant pas, *nota bene*, dans notre liste d'items étudiés) sont à leur place dans des traités d'œnologie ou autres écrits vantant les délices de telle ou telle saveur, on y trouvera probablement moins souvent les items *courgette* ou *poireau* (car un vin avec un *nez très poireau* ne ferait probablement pas fureur auprès des amateurs). Cela n'empêche que des énoncés tels que (123) ou (124), que nous imaginons, sont tout à fait possibles : comme dans les exemples ci-dessus, la référence au goût ou à l'odeur implique la disparition des individus, et donc la nécessité de l'emploi massif.

(123) Ce petit goût acidulé, je pense que c'est du poireau légèrement passé.

(124) Ca sent bon ici, c'est de la carotte ?

Pour finir, nous trouvons aussi de nombreux emplois où le massif n'indique ni spécifiquement une substance comestible ou une forme préparée, ni un goût ou une odeur, mais sert plutôt à faire simplement référence à l'un de nos fruits ou légumes de manière fort générale. C'est le cas, nous semble-t-il, dans les énoncés suivants, par exemple :

(125) Pour la première fois j'ai cuisiné de la courge spaghetti.

(126) Ajouter de la tomate et de l'origan, ça devient italien.

(127) Avec du citron confit, cette recette est bien tentante.

(128) L'hiver, on trouve facilement du potiron ou de la citrouille.

(129) [Ils] cultivent aussi de l'oignon vert (...), de l'échalote française et de l'oignon espagnol.

(130) En Bretagne, on produit du chou-fleur toute l'année.

Ici, c'est le caractère général des assertions qui justifie la présentation en forme et quantité indéterminées. L'on évoque quelque chose qui concerne le fruit ou légume en général, dans sa substance, faisant disparaître la nécessité de profiler des individus. Comme dans nos exemples (91) à (97) au point précédent, l'on peut bien faire certaines



suppositions pragmatiques quant à la nature des référents : la *courge* de (125) a probablement terminé en morceaux, et le *chou-fleur* de Bretagne est certainement produit sous forme d'unités – mais ce ne sont là que des interprétations ultérieures.

Ainsi se clôture notre premier tour d'horizon sémantique. Notre retiendrons de notre description le résumé suivant : le sens fondamental, compositionnel, d'un nom de fruit ou de légume en syntaxe massive se résume à une présentation homogénéisée du référent, où la notion d'individus disparaît à l'arrière-plan conceptuel. Cette présentation est sous-déterminée par rapport à différentes lectures plus spécifiques (multiplication, destruction d'exemplaires), et applicable *a priori* à l'ensemble de nos items, moyennant le bon contexte. Plus précisément, elle sera favorisée par certains facteurs situationnels, qui peuvent même aller jusqu'à la rendre indispensable (goût, odeur). L'alternance entre massif et comptable est donc globalement possible, mais est néanmoins guidée par des considérations pragmatiques qui ne sont pas aléatoires. D'autre part, elle est également modulée par le comportement spécifique de chaque item – certains apparaissant plus friands du massif que d'autres (par exemple *chou-fleur* vs. *haricot*). Gardons bien ces différents faits en tête : il nous faudra pouvoir leur donner à tous une place, lorsque nous proposerons notre propre modélisation du lexique mental et de la représentation de l'opposition massif-comptable dans celui-ci.

### 5.1.2.3 Une remarque concernant les emplois au pluriel

Jusqu'ici, nous nous sommes concentré sur les usages massifs – les moins fréquents et ainsi les plus marqués, conformément à ce que nous avons annoncé au point 4.5 ci-avant. Ce faisant, nous n'avons pas supposé que les emplois comptables recélaient, quant à eux, de grands mystères : ils permettraient simplement de désigner nos fruits et légumes en tant qu'individus – *une pomme* référant à un exemplaire individuel de pomme, *trois cerises* à trois exemplaires de cerise, etc.

Cependant, la flexibilité de nos noms de fruits et de légumes, que nous avons à présent pu observer et décrire, même si elle n'implique pas un statut équivalent pour les deux types d'emplois (le massif étant beaucoup moins fréquent que le comptable – nous verrons au Chapitre 9 les conséquences de ceci sur l'acceptabilité), ouvre une piste intéressante, notamment concernant le pluriel. D'une part, en tant que noms à usage fondamentalement comptable, les noms de fruits et de légumes connaissent la mise au pluriel en tant que sommation d'individus : *des carottes* ou *des oignons*, ce sont respectivement plusieurs exemplaires de ce que l'on désignerait comme *une carotte* ou *un oignon* – au même titre que *des voitures* renvoie à plusieurs individus de la catégorie *voiture*. C'est ce qu'(Alexiadou 2011) appelle le pluriel grammatical, ou que (Wilmet 2007: §57) nomme le pluriel externe, par addition d'unités.

Il existe cependant un second type de pluriel, appelé le pluriel lexical par (Alexiadou 2011) (à rapprocher du pluriel interne, par division d'unités, de (Wilmet 2007: §57)), qui

se rencontre normalement avec les noms non dénombrables et/ou habituellement massifs, et dont la particularité est de ne pas avoir pour domaine d'application le référent du nom auquel il se combine : *les N*, avec ce type de pluriel, ne désigne donc pas *plusieurs N* – mais évoque le concept lexical associé à N en combinaison avec l'itération référentielle la plus plausible en contexte. Ainsi, l'on peut parler des *eaux internationales* pour désigner plusieurs mers, ou *des pluies du mois d'avril* pour évoquer plusieurs averses. Nous verrons des exemples de pareils emplois pluriels 'collectifs' de noms massifs au chapitre suivant, lorsque nous parlerons des emplois comptables de nos noms de matériaux (pensons d'ores et déjà aux *cuivres*, qui désignent plusieurs instruments, ou aux *papiers (d'identité)* pour plusieurs documents).

Or, puisque nos noms de fruits et de légumes sont flexibles et donc aussi, dans certains emplois, massifs, l'on peut s'attendre à ce qu'ils se rencontrent également avec ce second type de pluriel. Et c'est, nous semble-t-il, bel et bien le cas. Considérons la situation suivante : pour accompagner le repas, j'ai découpé un unique oignon, et rissolé les morceaux à la poêle. Une fois à table et tout le monde servi, j'attrape le bol contenant les morceaux d'oignon frits et peux alors, avant de vider la fin du plat dans mon assiette, demander sans problème à la table :

(131) Quelqu'un reprendra des oignons ?

Dans le contexte, il est évident que le bol que j'ai en main ne contient pas plusieurs exemplaires de ce qui serait *un oignon* (tous les morceaux étant d'ailleurs issus d'un unique individu). Le pluriel employé là relève donc d'autre chose : ce n'est pas un pluriel grammatical prenant *oignon* pour domaine, mais un pluriel lexical (au sens d'(Alexiadou 2011)) ! En l'occurrence, l'itération contextuelle responsable de la pluralisation est celle des morceaux. De même, si je mange *des carottes* coupées en rondelles, ou même râpées, je mange en réalité *de la carotte*, dont la présentation sous forme de multiples morceaux justifie la mise au pluriel – sans qu'un de ces morceaux ne soit lui-même *une carotte*. Imaginons la situation suivante : le repas touche à sa fin et il reste dans la casserole quelques rondelles de carotte (disons, six). Parmi les interactions possibles, (132) ci-dessous nous semble la plus crédible, tandis que (133) paraît quelque peu étrange, et (134) et (135) sont carrément exclus.

(132) A : Est-ce qu'il reste des carottes ?  
B : Oui, un (petit) peu. Je te les mets ?

(133) A : Est-ce qu'il reste des carottes ?  
B : ??Oui, quelques unes. Je te les mets ?

(134) A : Est-ce qu'il reste des carottes ?  
B : \*Oui, à peu près six, je dirais. Je te les mets ?

- (135) A : Est-ce qu'il reste des carottes ?  
B : \*Non, il en reste moins qu'une.

L'énoncé en (133) ne fonctionne pas bien, parce qu'il ne reste, de fait, pas plusieurs exemplaires de carotte. Ceci confirme ce que nous disions : si dans pareil cas, *des carottes* désigne un ensemble de morceaux de carotte, *une carotte* ne désigne pas pour autant l'un de ces morceaux. C'est ce qui rend (134), où l'on utilise un numéral, tout à fait impossible. D'autre part, le fait que (135) soit exclu confirme que dans le contexte, A s'enquiert bel et bien de morceaux de carotte, et non de carottes individuelles – ce qui rend la réponse de B absurde. Notons, pour finir, que la réponse de B en (132) présente un élément particulièrement intéressant : l'usage de la locution *un (petit) peu*. Ceci nous indique de manière cruciale que dans cet emploi pluriel particulier, les noms de fruits et de légumes sont en fait employés comme pluriels massifs. Les emplois de ce type pour *des oignons*, *des carottes*, etc., sont donc à rapprocher des emplois de *des épinards* – qui quant à lui ne correspond jamais qu'à l'emploi pluriel lexical (puisque l'on ne pourra jamais dire, en parlant du légume, \**un épinard*).

Notons que la combinaison *un peu de + N de légume au pluriel* a également été commentée par (Nicolas 2002a: chapitre 7, 1. a.), qui en donne une explication bien différente, et selon nous moins satisfaisante. Il l'évoque pour montrer que les transferts du comptable au massif dans le domaine de la nourriture, supposés généraux, ne fonctionnent pas bien pour certains noms de fruits et de légumes. Il donne ainsi l'exemple (136), qu'il juge douteux :

- (136) ?Aimeriez-vous un peu de carotte / courgette / pomme de terre / tomate ?

Sur la base de cet exemple, il explique que pour ces noms, l'emploi massif nécessite la forme plurielle, et ce pour des raisons pragmatiques : parce qu'une préparation à base de l'un de ces légumes suppose normalement que plusieurs exemplaires auront été cuisinés. Laissons pour ce qu'ils sont nos doutes quant à la pertinence de cet argument (car nous ne sommes pas convaincus que les carottes et les courgettes, vu leurs différences de taille, soient à placer dans la même catégorie à ce niveau –là). Plus crucialement, comme nous l'avons montré dans ce chapitre, les noms *carotte*, *courgette* et *tomate* sont bel et bien flexibles, et connaissent des emplois massifs (nous n'avons pas étudié *pomme de terre*, mais il ne saurait en être autrement). Nous en avons donné des exemples attestés ci-avant. Pourtant, l'observation de Nicolas quant à (136) est bien correcte – mais la conclusion qu'il en tire est erronée. Ici, si le pluriel est nécessaire, c'est parce qu'on se trouve dans un contexte typiquement associé à l'emploi massif pluriel que nous venons de décrire (où les légumes en question sont préparés, découpés – ouvrant la possibilité à l'emploi pluriel lexical ; on peut en effet supposer que (136) est prononcé à table, ou du moins dans le cadre d'un repas). Que plusieurs exemplaires de chacun de ces légumes soient nécessaires pour la préparation d'un plat n'est pas reflété

dans la mise au pluriel : à la question *Aimeriez-vous un peu de tomates ?*, posée à table en désignant une salade de tomates, découpées et assaisonnées, je ne peux répondre *Volontiers, j'en prendrai une ni Volontiers, mais pas plus que deux* en signifiant par là respectivement désirer une quantité de morceaux équivalant à un fruit entier, ou n'en excédant pas deux. Il faudra se contenter de rester dans le domaine des quantificateurs appropriés pour le massif : *Volontiers, j'en prendrai un peu* ou *J'en prendrai volontiers, mais pas trop*.

Que devons-nous tirer de ce commentaire au sujet des deux emplois des formes plurielles des noms de fruits et de légumes ? Dans une perspective quantitative, nous ne pourrions pas en faire grand-chose : en surface, les emplois sont impossibles à distinguer. Tout ce qu'on peut dire, c'est que la lecture plurielle massive n'est possible qu'avec les déterminants les moins explicites quant à l'individualisation : *des* ou *quelques*, mais jamais *plusieurs* ni les numéraux (pour la différence, à ce niveau, entre *quelques* et *plusieurs*, voir (Gréa 2008) ; voir aussi (Lauwers 2014, à paraître), et les références qu'il cite, sur le sujet du pluriel non dénombrable). Néanmoins, c'est là une condition nécessaire et non suffisante, qui ne permettrait donc pas de discriminer de manière automatique, uniquement à partir du contexte. Or, même si l'on voulait annoter manuellement les occurrences de corpus de nos noms de fruits et de légumes employés avec le déterminant pluriel indéfini, par exemple, l'on trouverait principalement soit des énoncés avec le pluriel grammatical, comme (137) ou (138), soit des énoncés ambigus, comme (139), (140) ou (141), où à vrai dire il n'est jamais permis de trancher avec certitude (en tout cas sans plus de contexte) entre l'un ou l'autre type de pluriel :

- (137) En attendant, j'achète des pommes peu traitées chez le producteur.
- (138) Camoufler des sachets de cocaïne dans des bananes est un procédé utilisé par les narcotrafiquants.
  
- (139) Une bonne façon de faire manger des carottes aux tous petits.
- (140) Faut-il arrêter de manger des concombres ?
- (141) Emitter cette pâte au dessus des pommes de façon à les recouvrir.

Dans la recherche présente, nous avons donc dû classer sans discrimination tous les emplois au pluriel du côté du comptable. Néanmoins, la distinction que nous avons examinée ici, entre deux types de pluriels, nous apprend quelque chose d'intéressant de l'opposition massif-comptable en général : selon qu'un nom soit fondamentalement comptable (*voiture*) ou massif (*papier*), la mise au pluriel a un effet différent – soit une sommation des individus désignés par le nom (*des voitures* '~-plusieurs voitures'), soit une sommation différente, à récupérer en contexte (*des papiers* '~-plusieurs documents'). Cette différence pointe en faveur d'une représentation lexicale de notre opposition, c'est-à-dire d'une vision lexicaliste (voir point 1.5.1) : la préférence fondamentale d'un nom pour l'un ou l'autre type de syntaxe, en ce qu'elle influe sur la manière dont ce nom

se comportera par rapport au pluriel, se doit d'être stipulée dans le système linguistique – d'une manière ou d'une autre. Les noms flexibles, tels que les noms de fruits et de légumes du chapitre présent, en sont une preuve indirecte : ils jouissent du double marquage (nous dirons aux Chapitre 8 et Chapitre 9 ce que nous entendons par là), et présentent donc, par rapport au pluriel, le comportement caractéristique à la fois des comptables et des massifs.

## 5.2 Les noms de fruits et de légumes en néerlandais

### 5.2.1 Analyse distributionnelle

Pour le néerlandais, comme nous l'avons dit ci-dessus, nous nous sommes concentré sur l'étude de 13 noms de fruits et de légumes – la traduction des 12 items les plus fréquents de la catégorie tels que nous les avons établis pour le français, avec un dédoublement d'*oignon* en *ui* et *ajuin*. La Figure 11 ci-dessous montre comment sont distribués les emplois massifs et comptables de ces noms dans nos données de *WebCorp* – en suivant globalement le même code couleur que pour la Figure 9 ci-avant. Nous avons néanmoins indiqué dans cette représentation également les occurrences comptables impliquant l'emploi d'un diminutif, par l'usage d'une seconde nuance de bleu. Notons que l'axe des ordonnées s'étend ici de 0 à 40 %.

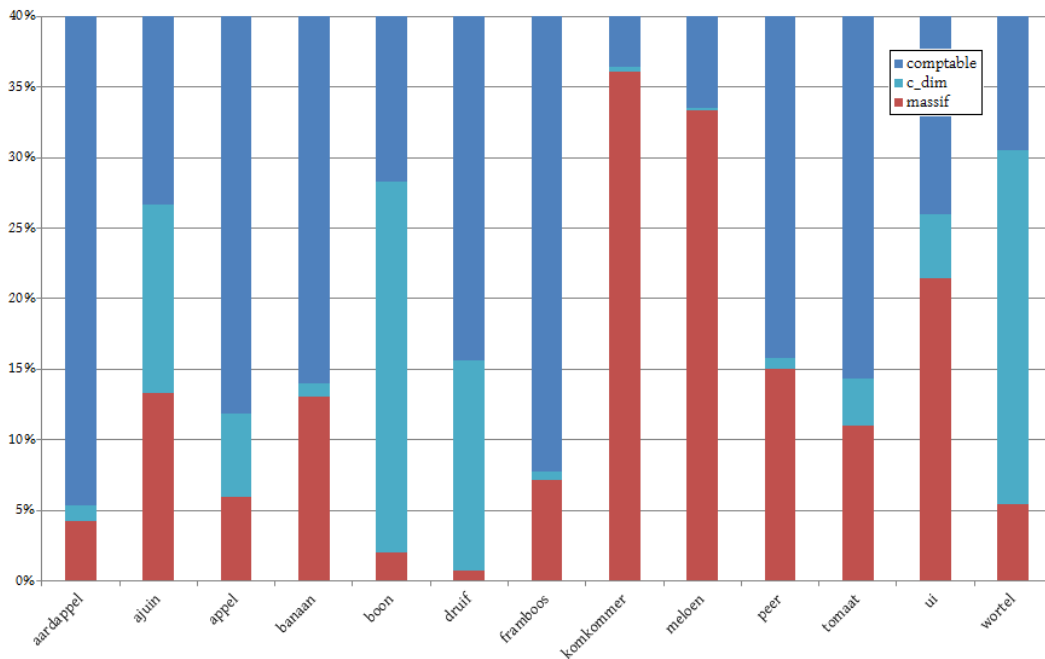


Figure 11 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais. Les emplois massifs sont représentés en rouge, les emplois comptables en bleu. L'axe des ordonnées s'étend de 0 à 40 %.

Comme pour le français, les emplois comptables dominant ici largement. Néanmoins, la proportion d'emplois massifs apparaît notablement plus grande, ce que confirment quelques calculs : elle s'élèverait à 13,38 % du total en néerlandais (738 occurrences massives, contre 4 777 comptables, pour les 13 items étudiés), tandis qu'en français, elle ne dépassait pas les 2 % (171 occurrences massives, contre 10 689 comptables au total des 26 items de la Figure 10).

Cette observation sans appel à de quoi surprendre. Elle contredit d'une part l'idée pourtant relativement consensuelle et intuitive que le néerlandais serait moins flexible que le français quant à l'opposition massif-comptable (voir (Lauwers et Vermote 2014, à paraître), ainsi que nos résultats dans les deux chapitres suivants), et bute également contre les intuitions natives en la matière : les quelques locuteurs natifs du néerlandais que nous avons interrogés informellement sur le sujet (notamment les professeurs K. Willems, J. De Caluwe et P. Lauwers, de l'Université de Gand) s'accordent sur le caractère douteux à accorder spontanément à des énoncés du type *??Ik heb een recept gevonden met appel* 'J'ai trouvé une recette avec de la pomme' ou encore *??Wat kan je bereiden met ui?* 'Que peut-on préparer avec de l'oignon?' – là où les traductions en français nous semblent d'emblée beaucoup plus acceptables. (Nous présenterons les résultats d'une enquête d'acceptabilité systématique sur ce sujet au Chapitre 9.) Il n'est donc pas anodin, car inattendu, de constater qu'à première vue, ces emplois massifs de noms de fruits et de légumes seraient proportionnellement plus fréquents en néerlandais qu'en français.

Cependant, les pourcentages de la Figure 11 ne peuvent pas simplement être comparés sans plus à ceux de la Figure 10. En effet, de par les propriétés morphosyntaxiques respectives de nos deux langues, notre annotation d'une occurrence comme massive s'est voulue relativement prudente en français, mais globalement plus inclusive en néerlandais – avec pour conséquence une réduction artificielle en français, contre une augmentation sensible en néerlandais, de la proportion d'emplois massifs observés dans nos données. C'est dès lors un point qu'il nous faudra creuser.

Nous n'entendons certainement pas par là que notre annotation aurait été maladroite ou inconséquente : dans les deux langues, les emplois que nous avons reconnus comme massifs le sont bel et bien, selon notre meilleur entendement du sujet. Cependant, tous les énoncés que nous avons annotés comme massifs en néerlandais ne correspondent pas nécessairement à une traduction en français présentant de même un emploi massif explicite (tandis qu'à l'inverse, autant que nous ayons pu le vérifier, les emplois massifs explicites en français se traduisent toujours bel et bien par un massif explicite en néerlandais). L'une des causes principales en est la double fonction que peut revêtir en néerlandais un nom en emploi nu (un contexte que nous avons classé du côté du massif), comme nous l'avons évoqué au point 1.2.2 : si la lecture en est tantôt spécifique, correspondant alors à peu de choses près au partitif français (142), elle peut également être générique – et se traduirait dans ce cas dans notre langue par un article défini (143).

(142) *Dit (...) recept bestaat uit sappige cantaloupe meloen en zoute Italiaanse prosciutto.*

'Cette recette (...) comporte du melon cantaloupe juteux et du prosciutto italien salé.'

(143) *Banaan bevat veel koolhydraten lijkt me (natuurlijke suikers).*

'La banane contient beaucoup de glucides, me semble-t-il (sucres naturels).'

Ces cas ne sont du reste pas les seuls où la traduction d'un emploi massif explicite en néerlandais se retrouverait annotée comme neutre en français, comme nous le verrons ci-après. C'est précisément ce qui rend la comparaison entre les données de nos deux langues délicate, et pourrait expliquer la disparité, contraire aux attentes et à l'intuition, que l'on peut observer à première vue.

Ainsi, si l'on veut pouvoir comparer d'un point de vue quantitatif les profils morphosyntaxiques des noms de fruits et de légumes en français et en néerlandais, il faudra d'abord trier les énoncés massifs de cette seconde langue, pour ne garder que ceux qui permettent un réel point de comparaison avec la première. Le raisonnement est le suivant : si, une fois ce tri effectué, le contraste frappant entre nos deux langues (du point de vue de la proportion des emplois massifs) se réduit mais persiste, il faudra effectivement accepter une différence plus fondamentale entre le français et le néerlandais au niveau du statut massif-comptable des noms de fruits et de légumes – et chercher à l'expliquer. Si, au contraire, comme nous le supposons *a priori*, la

dissemblance entre les profils des Figures 10 et 11 tient bel et bien à des différences morphosyntaxiques extérieures (comme l'existence en néerlandais d'un emploi nu générique, par exemple), alors celle-ci devrait disparaître une fois les données triées de manière appropriée.

Pour trancher la question, nous avons ainsi examiné systématiquement chacun de nos énoncés massifs en néerlandais, séparant ceux qui correspondraient effectivement à un massif explicite en français (et donc pertinents pour une comparaison quantitative), des autres. Dans ce qui suit, nous détaillons notre manière de procéder en la matière, en nous concentrant d'abord sur la distinction entre emplois spécifiques et génériques des noms massifs nus, puis sur d'autres cas ayant attiré notre attention car faussant d'une manière similaire la comparaison entre nos deux ensembles de données – avant de présenter, pour finir, une nouvelle représentation graphique de nos résultats distributionnels (tenant compte du tri effectué).

### 5.2.1.1 Une distinction entre des noms nus génériques, et spécifiques ?

La distinction entre emplois spécifiques et génériques du nom (massif) nu en néerlandais serait donc une source importante de la différence quantitative que nous observons entre nos deux langues. Afin de creuser cette question, nous avons entrepris dans un premier temps d'annoter l'ensemble de nos occurrences massives de noms de fruits et légumes, en termes d'emploi générique ou spécifique.

Ceci est cependant plus simple à dire qu'à faire, car l'opposition n'est pas tant réelle qu'elle n'est un avatar de notre description – et de notre comparaison systématique avec le français. D'un point de vue sémantique, les emplois nus génériques en néerlandais ne diffèrent en effet pas fondamentalement des autres : ils constituent seulement un autre type de contexte spécifique où la présentation homogène, en forme et quantité indéterminées, et donc l'utilisation en syntaxe massive, s'avèrent pertinentes. (En cela, ils sont très proches des emplois massifs que nous avons identifiés en français lorsqu'un fruit ou légume est présenté simplement de manière très générale, en substance, comme dans nos exemples (125)-(130) ci-dessus.)

Cependant, en français, l'emploi du partitif, en tant qu'article indéfini, est contraint : il ne peut pas facilement apparaître en position sujet (pour différentes raisons concernant l'opposition entre défini et indéfini, de même que le caractère homogène de sa référence, voir (Galmiche 1986) ; voir aussi (Wilmet 2007: §179) pour d'autres références sur le sujet) – tandis que l'emploi nu en néerlandais (qui n'est pas entièrement équivalent au partitif) n'est pas soumis à la même contrainte. Considérons l'exemple suivant :

(144) *Dan kan banaan je helpen om je kater vlugger kwijt te raken.*

Dans ce cas, la banane pourra t'aider à te débarrasser plus vite de ta gueule de bois.



Dans l'exemple (144), l'emploi massif de *banaan* est bien simplement un emploi général, en substance – ce qui justifie le choix d'un marquage faisant disparaître à l'arrière-plan conceptuel les exemplaires de banane individuels. Cependant, dans la traduction, l'article défini est préférable au partitif, car il ne convient pas aisément en position sujet. On obtient alors une lecture dite générique (il est intéressant de noter que l'article défini générique a justement été rapproché du massif par (Kleiber 1990)), que l'apparence en surface ne permet plus de classer du côté du massif ou du comptable.

Notons bien que cela ne veut pas dire pour autant qu'en néerlandais, le sens de *banaan* dans (144) soit fondamentalement différent d'autres emplois massifs moins généraux. La traduction en (145), adaptée seulement pour permettre la réapparition du partitif, pourrait convenir tout autant :

(145) Tu pourras faire disparaître rapidement ta gueule de bois avec de la banane.

Examinons encore (146), (147) et (148) – énoncés que nous inventons :

(146) *Ik heb gisteren komkommer gegeten.*

J'ai mangé du concombre hier.

(147) *Komkommer is zeer caloriearm en mag je onbeperkt nuttigen.*

Le concombre est pauvre en calories et peut se manger sans restriction.

(148) *(?)Ik heb gisteren komkommer gegeten, want dat is zeer caloriearm en mag je onbeperkt nuttigen.*

\*J'ai mangé du concombre hier, parce qu'il est très pauvre en calorie et peut se manger sans restriction.

Dans (146) et (147), le nom *komkommer* est employé en syntagme nu, dans deux usages qui se traduisent respectivement par un partitif et un article défini générique. Or, (148) montre bien qu'il ne s'agit pas là, pour le néerlandais, de deux emplois fondamentalement différents : ils peuvent être reliés par une relation anaphorique qui, quoique peut-être un peu douteuse au regard de certains locuteurs (comme nous le signale Johan De Caluwe, communication personnelle), est incomparablement plus acceptable que son équivalent en français – où elle est exclue. C'est ce que nous entendons lorsque nous disons que la distinction entre emplois massifs spécifiques et génériques, en néerlandais, est un avatar de notre description : elle apparaît au travers de la traduction, mais a essentiellement trait aux caractéristiques du français lui-même et n'est donc pas, à notre sens, une différence réelle en néerlandais. L'on pourrait dire encore que l'emploi massif nu des noms de fruits et de légumes en néerlandais est sous-spécifié par rapport aux emplois spécifiques et génériques – alors qu'en français, l'emploi avec le partitif, de par les caractéristiques de ce déterminant, est sensiblement plus limité.

Il n'empêche que cette différence grammaticale cruciale entre le français et le néerlandais a bel et bien une influence non négligeable sur la répartition

distributionnelle des emplois de nos noms de fruits et de légumes – le nombre d’emplois massifs étant gonflé en néerlandais, par l’inclusion de nombreuses occurrences en syntagme nu dont l’équivalent en français se retrouverait du côté des emplois neutres (et donc en-dehors de nos graphiques).

Afin de permettre une comparaison plus équilibrée, nous avons donc commencé par diviser les 738 occurrences massives de nos noms de fruits et de légumes en néerlandais, en un groupe d’emplois génériques, et un autre d’emplois spécifiques. Parmi ces derniers, nous avons classé toutes les occurrences où le syntagme massif désignait une quelconque portion concrète du fruit ou du légume en question (l’appellation de générique ne pouvant certainement pas convenir dans ces cas-là). Cela incluait certains emplois sans déterminant, comme en (149) ou (150), de même que *de facto* tous les emplois massifs autres qu’en contexte nu, notamment les occurrences avec *wat*, comme en (151), de même qu’avec un terme de mesure, comme en (152) et (153).

(149) *Ik heb ze al een keer gemaakt, maar heb toen alleen ei en banaan gebruikt.*  
Je l’ai déjà préparée une fois, mais je n’avais utilisé que de l’œuf et de la banane.

(150) *Geraspte komkommer wordt vermengd met dikke Griekse (of Bulgaarse) yoghurt (...).*  
On mélange du concombre râpé avec du gros yaourt grec (ou bulgare).

(151) *[D]an eet ik ook snel wat rauwe wortel daarnaast om de schade te beperken.*  
J’accompagne alors cela d’un peu de carotte crue, pour limiter les dégâts.

(152) *Een portie banaan komt overeen met 100 gram.*  
Une portion de banane correspond à 100 grammes.

(153) *[Een Nederlander eet] gemiddeld 4 tot 5 kilo ui per jaar.*  
Un Néerlandais mange en moyenne 4 à 5 kilos d’oignon(s) par an.

D’autre part, pour les occurrences plus générales, c’est-à-dire plus propices à être classées comme génériques, nous avons principalement basé notre jugement sur la traduction la plus naturelle en français : si celle-ci était possible avec un partitif, nous avons annoté l’occurrence comme spécifique (154)-(155), si celle-ci nous paraissait nécessiter un article défini, nous avons opté pour l’annotation en tant que générique (156)-(158). Sur nos 738 occurrences massives de départ, nous en aurons ainsi classé 308 (41,7 %) dans cette dernière catégorie.

(154) *[D]an is het goed om uienthee te drinken, rauwe ui te eten en een halve ui op het nachtkastje te leggen.*

Il est bon dans ce cas de boire du thé aux oignons, de manger de l’oignon cru, et de déposer un demi-oignon sur sa table de nuit.

(155) *Heb je ui in je tuin, zorg er dan voor dat je bij het oogsten de beschimmelde exemplaren verwijdert.*

Si vous avez de l’oignon dans le jardin, veillez lors de la récolte à écarter les exemplaires moisiss.

- (156) *Komkommer kan echter ook worden gestoofd, gevuld met gehakt om in de oven gaar te braden, of als garnering worden gebruikt.*  
Le concombre peut cependant aussi être mijoté, farci à la viande hachée pour être cuit au four, ou servir en guise de décoration.
- (157) *Meloen en muscat zijn grote vrienden van elkaar.*  
Le melon et le muscat font bon ménage.
- (158) *Ajuin en sjalotten zijn het hele jaar te krijgen.*  
L'oignon et les échalotes sont disponibles toute l'année.

### 5.2.1.2 Autres différences entre le français et le néerlandais

Ce premier tri, d'importance certes cruciale, n'est cependant pas encore le fin mot de l'histoire. Parmi les occurrences massives que nous avons classées comme spécifiques, il s'en trouve en effet certaines qui, malgré tout, continuent de ne pas correspondre à un massif en français – ne résolvant donc pas encore tout à fait notre problème de comparaison quantitative équilibrée.

Les énoncés (152) et (153), ci-dessus, nous en donnent un premier indice : si *een portie banaan* constitue bel et bien un emploi massif de *banaan*, puisqu'il s'agit d'un emploi partitif direct avec un terme de mesure suivi d'une forme au singulier (voir ci-dessus, point 1.2.2), nous aurions pourtant annoté comme neutre l'équivalent français *une portion de banane*. Ceci s'explique par la philosophie de notre annotation, et notre choix de ne classer les emplois que sur la base absolument non ambiguë de leur contexte d'apparition de surface. Ainsi, nous avons hésité longuement quant au traitement à réserver aux noms précédés directement de la préposition *de*, en français. D'une part, certains contrastes semblaient pointer vers un contexte de nature cumulative : ainsi, si *citron* est en emploi comptable dans *du jus de citrons* (comme l'indique, sans ambiguïté, la marque du pluriel : 'du jus fait avec des citrons'), on pourrait reconnaître dans l'équivalent au singulier, *du jus de citron*, un emploi massif ('du jus fait avec du citron'). D'autre part, l'on ne retrouve pas ce contraste dans tous les emplois où un nom nu suit directement la préposition *de* : les occurrences soulignées dans *mes clés de voiture*, *un chien de berger* ou encore *un premier de classe* en sont autant de preuves. Dès lors, puisqu'il ne discrimine pas univoquement entre massif et comptable, nous avons dû effectivement classer comme neutre le contexte *de (préposition) + N singulier*.

Pour le néerlandais, il en a été autrement, car l'absence de tout marquage explicite introduisant le second nom dans une construction partitive directe nous a forcé, dans le cadre de l'annotation semi-automatique de nos données, à fournir à notre outil d'analyse (notre script *AN\_NL.pl*, voir point 4.4.4) une liste de noms de mesure – typiquement associés à cette construction – permettant de ce fait de repérer comme massives certaines occurrences dont l'équivalent français se trouvait classé du côté des neutres. Pour le français, introduire une pareille liste dans les ressources lexicales de

notre script *SN\_FR.pl* aurait pu être une option (et reste par ailleurs une piste d'amélioration à explorer) mais n'était tout simplement pas indispensable, la préposition *de* explicite fournissant déjà un élément repérable automatiquement.

C'est ce que nous entendons, lorsque nous qualifions notre annotation en français de relativement prudente, contre un classement plus inclusif en néerlandais (voir ci-dessus). Cela implique que si l'on veut pouvoir comparer plus justement les proportions respectives d'emplois massifs et comptables des noms de fruits et de légumes entre nos deux langues, il faudra laisser un instant hors de considération les emplois partitifs directs (qui sont à eux seuls bons pour 79 de nos 738 occurrences massives en néerlandais, soient 10,7 %).

Ils ne sont du reste pas les seuls emplois massifs en néerlandais à correspondre à la construction *de + N singulier* en français. Ainsi, nous avons classé l'emploi en (159) comme massif – alors que sa traduction en (160) se serait retrouvée du côté des neutres. (A noter que pour (159), l'annotation comme massif a été manuelle, notre script *AN\_NL.pl* ne pouvant interpréter l'épithète *met listeria besmette* séparant la préposition *van* du nom tête *meloen*.)

(159) *De consumptie van met listeria besmette meloen heeft in de Verenigde Staten nu al dertien slachtoffers geëist.*

(160) La consommation de melon contaminé par la listéria a déjà fait treize victimes aux Etats-Unis.

A nouveau, pour une comparaison plus équilibrée des distributions dans nos deux langues, il conviendra donc de considérer en néerlandais un sous-ensemble des données, dont seront exclus les énoncés tels que (159). Que ça soit bien clair : nous ne prétendons pas en cela qu'ils ne sont pas massifs – ils le sont assurément. A vrai dire, dans un certain sens, l'emploi de *melon* dans (160) l'est peut-être également : c'est bien *du melon* que l'on consomme, et l'absence de partitif dans la construction prépositionnelle pourrait découler simplement de l'haplogie de *\*de du* (comme l'ont suggéré certains auteurs, voir (Riegel *et al.* 2009: 166) et (Benetti 2009: 135-136)). Néanmoins, l'important est que si l'on veut comparer nos chiffres en français et en néerlandais, les énoncés équivalents doivent être classés de la même manière. Or, en la matière, l'option d'un tri sélectif des emplois massifs en néerlandais (738 occurrences à examiner au total) était assurément plus économique qu'un éventuel parcours manuel de nos 10 757 occurrences neutres de noms de fruits et de légumes en français. Les énoncés (161)-(163) sont quelques autres exemples de cas similaires à (159).

(161) (...) *indien het gebruik van bitter meloen wordt overwogen.*  
'(...) si l'emploi de melon amer [ou momordique] est envisagé.'

(162) *Waar let ik op bij het kopen van komkommer?*  
A quoi fais-je attention lors de l'achat de concombre ?

(163) *Meer recepten op basis van meloen.*

‘Plus de recettes à base de melon.’

Finalement, en examinant nos occurrences massives à trier, nous avons remarqué aussi un certain nombre d’emplois nus annotés comme massifs par notre script d’assistance semi-automatique, mais relevant en réalité d’un style télégraphique – typique des titres (164), ou de certaines recettes (164)-(166) – ne correspondant pas non plus à un massif explicite en français.

- (164) *Witte en wilde rijst met wortel en komijnbijgerecht.*  
‘Riz blanc et sauvage avec [de la?] carotte et accompagnement au cumin.’
- (165) *In een grote kom kip, kool, wortel, lente-ui, (...) en courgette vermengen.*  
‘Dans un grand bol, mélanger poulet, chou, carotte, cébette et courgette.’
- (166) *Geblancheerde, uitgeholde ui kan ook opgevuld worden met gehakt of ei.*  
‘[L]’oignon blanchi et évidé peut également être fourré avec du hachis ou de l’œuf.’

Concernant les titres, nous avons déjà repéré manuellement et annoté comme neutres tous les emplois nus complètement isolés (jamais analysés automatiquement), du type :

- (167) *BANAAN Uit Wikipedia, de vrije encyclopedie*  
‘BANANE [Un article] de Wikipédia, l’encyclopédie libre’.

Néanmoins, les cas tels que (164) ayant d’abord été analysés automatiquement comme massifs – car la combinaison *préposition + N singulier* est précisément l’un des contextes employé par notre script *AN\_NL.pl* pour repérer automatiquement les emplois massifs – il convient également à présent de les exclure, si l’on veut pouvoir comparer efficacement nos deux langues. En effet, il est permis de douter que la traduction de l’emploi nu en (164), hors réel contexte énonciatif, serait forcément traduit par un article partitif.

### 5.2.1.3 Nouvelle représentation de nos données distributionnelles

Le tri décrit ci-avant, visant à identifier parmi les occurrences massives des noms de fruits et de légumes en néerlandais tous les cas ne correspondant pas univoquement à un massif explicite en français, permet d’aboutir à deux nouvelles visualisations graphiques de nos données. La première, qu’on retrouvera à la Figure 12, répond à la Figure 11 ci-dessus et distingue simplement, au sein de la proportion d’occurrences massives précédemment indiquée, les occurrences permettant effectivement la comparaison avec le français (en rouge) et celles qui doivent être exclues dans ce cadre, plus précisément les emplois nus génériques (en vert) ainsi que tous les autres n’ayant pas non plus un équivalent massif explicite en français (en vert plus clair) – selon les critères de distinction que nous venons d’exposer.

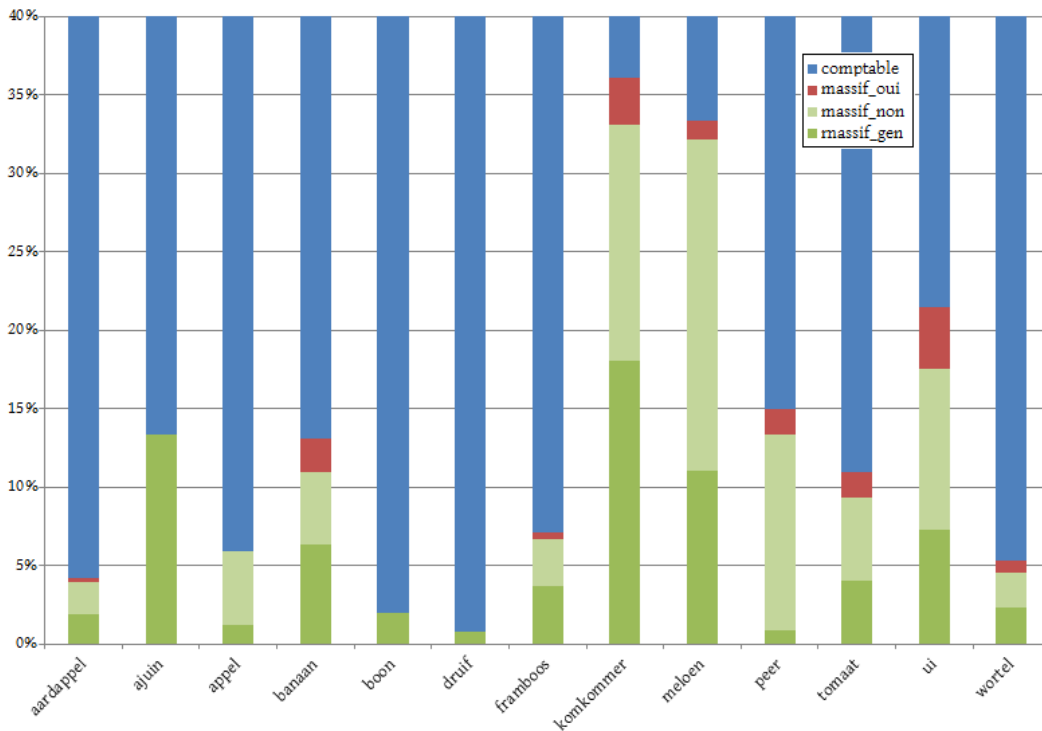


Figure 12 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais, avec distinction entre emplois massifs correspondant effectivement à un massif en français (*massif\_oui*), et emplois massifs correspondant à une forme alternative neutre (*massif\_gen* pour l'article défini générique, *massif\_non* pour les autres).

Ensuite, la Figure 13 ci-après reprend les mêmes données, mais en excluant cette fois l'ensemble des occurrences massives nues que nous avons jugées impropres à la comparaison. C'est dès lors ce graphique-là qui devrait constituer le point de comparaison le plus fiable avec notre Figure 10 représentant la distribution des emplois des noms de fruits et de légumes en français (voir point 5.1.1). Notons qu'ici, l'axe des ordonnées est à nouveau limité à 25 %.

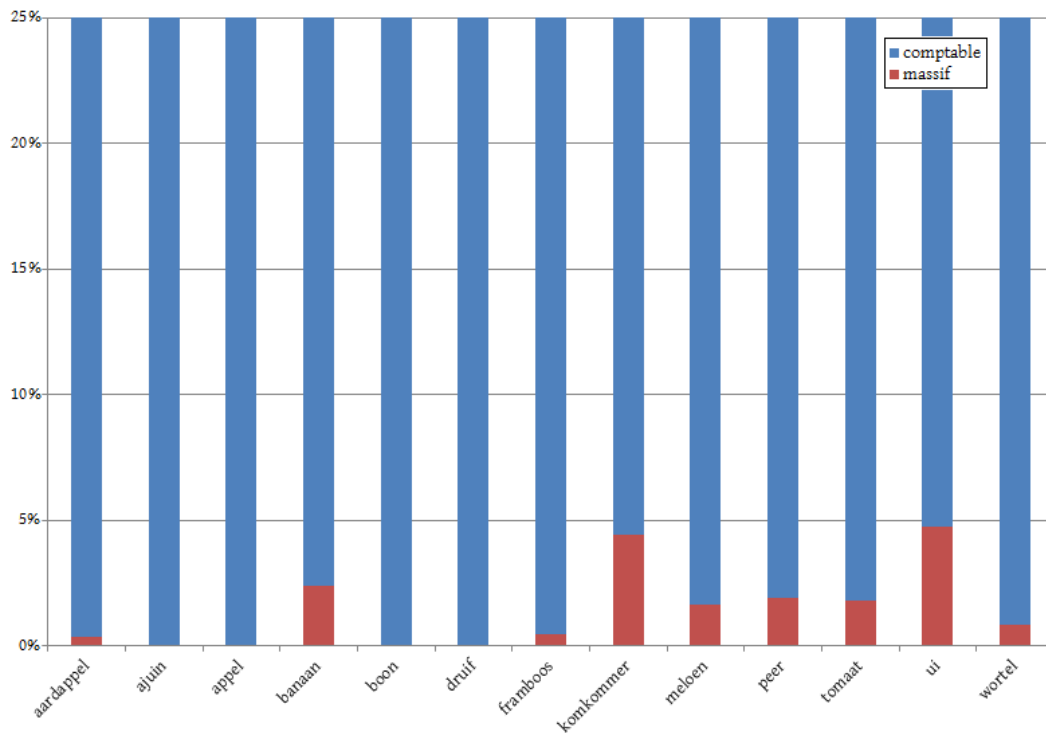


Figure 13 Profils distributionnels des 13 noms de fruits et de légumes étudiés en néerlandais, en excluant les occurrences massives ne correspondant pas univoquement à une occurrence massive en français, afin de fournir un point de départ pertinent pour la comparaison avec la Figure 10.

Cette nouvelle représentation nous montre qu'en réalité, la très grande majorité des emplois massifs de nos noms de fruits et de légumes en néerlandais ne correspondent pas à un contexte massif explicite en français – excluant de fait toute pertinence d'une comparaison quantitative sans tri préalable. Or, une fois ce tri effectué, la distribution que l'on peut observer correspond bien plus aux attentes : une proportion d'emplois massifs globalement très basse (73 occurrences massives sur 4 850 au total, soit 1,5 %), ne dépassant jamais la barre des 5 % (dépassée pour 6 items sur 26 en français), et nulle pour 4 items sur les 13 (contre 4 sur 26 en français).

Ainsi, contrairement à notre observation initiale sur la base la Figure 11, l'on ne peut effectivement pas conclure que les noms de fruits et de légumes seraient plus flexibles, du point de vue du massif-comptable, en néerlandais qu'en français. Lorsqu'on compare ce qui peut l'être, ça semblerait même plutôt l'inverse : le massif apparaît légèrement moins répandu en néerlandais, avec plus d'items pour lequel il n'est pas attesté, et aucun pour lequel il dépasse le seuil des 5 % (bien qu'au final, il soit délicat d'identifier une réelle différence significative, toutes les proportions de massifs étant très basses dans les deux langues). Quoi qu'il en soit, la tendance que nous attendions n'est au moins plus contredite – ce qui est une bonne nouvelle. Au Chapitre 9, nous reviendrons sur une comparaison plus précise de l'emploi massif des noms de fruits et de légumes en français et en néerlandais par le biais d'enquêtes d'acceptabilité (voir point 9.4.1).

## 5.2.2 Analyse sémantique : réalisation contextuelle des emplois massifs

En complément à ce long parcours de réflexion quantitatif, sur la représentativité des profils distributionnels que nos études de corpus nous permettent de calculer, et en particulier sur leur comparabilité entre le français et le néerlandais, il nous reste à évoquer encore brièvement la teneur qualitative des emplois massifs de nos noms de fruits et de légumes en néerlandais. Nous venons déjà d'en dire quelques mots, à la lumière de l'opposition entre usages massifs spécifiques, et génériques, observant que les deux types d'emplois relevaient d'un même sens homogène fondamental – identique à celui que nous avons décrit pour le français (voir point 5.1.2.1). Ainsi, en néerlandais aussi, l'usage en syntaxe massive d'un nom de fruit ou de légume, lorsqu'il est attesté, correspondrait à une présentation référentielle en forme et quantité indéterminées, faisant disparaître à l'arrière-plan conceptuel la notion d'individu.

Or, à nouveau, si l'existence de cette option fait de nos noms de fruits et de légumes des items morphosyntaxiquement flexibles, susceptibles d'alterner entre comptable et massif (peut-être un peu moins librement en néerlandais qu'en français, du reste), le recours à cette présentation homogène n'en demande pas moins à être motivée, et se retrouvera ainsi associée à certains contextes (pragmatiques) spécifiques et reconnaissables. Plus précisément, à l'examen des occurrences massives de nos noms de fruits et de légumes en néerlandais, nous retrouvons des exemples correspondant à chacun des types de contextes typiques du massif, tels que nous les avons décrits pour le français. Ci-dessous, l'on trouvera dès lors en guise de clôture quelques exemples illustratifs de chacun de ces contextes : présentation en tant que matière comestible (168), présentation sous forme préparée explicite (169)-(170), emploi contrasté (171), référence au goût et à l'odeur en (172)-(173) et présentation généralisée et/ou générique (174)-(175). (Notons que nous puisons ici dans tous les exemples d'occurrences massives, y compris celles que nous avons exclues du graphique en Figure 13.)

(168) *Zo blijkt uit onderzoek dat mensen die veel ui eten minder risico lopen te overlijden aan hart- en vaatziekten.*

'Ainsi, la recherche indique que les gens qui consomment beaucoup d'oignon ont moins de risque de succomber à une maladie cardio-vasculaire.'

(169) *U kent het vast wel: zo'n koek van geraspte aardappel.*

'Vous connaissez certainement : une sorte de galette de pomme de terre râpée.'

(170) *Leg de plakjes komkommer (...) in een schaal.*

'Disposez les tranches de concombre dans un plat.'

(171) *Ham met meloen en vijgen.*

'Jambon au melon et aux figues.'



- (172) *Heb ze ook gemaakt maar vind ze toch teveel naar ei smaken (met een vleugje banaan) i.p.v. naar pannenkoeken.*  
 ‘Je les ai préparés aussi, mais je trouve qu’elles ont trop un goût d’œuf (avec une touche de banane) plutôt qu’un goût de crêpe.’
- (173) *(...) omdat anders alles naar meloen gaat ruiken en smaken.*  
 ‘... parce que sinon, tout sentira et goûtera le melon.’
- (174) *Recepten met tomaat.*  
 ‘Recettes avec de la tomate.’
- (175) *Komkommer bevat vooral water en weinig voedingsstoffen.*  
 ‘Le concombre contient principalement de l’eau, et peu de nutriments.’

### 5.3 Conclusions

Nous pouvons à présent conclure notre analyse du comportement morphosyntaxique et sémantique des noms de fruits et de légumes, en français et en néerlandais. Les principales observations et conclusions à retenir sont les suivantes :

1. Les noms de fruits et de légumes constituent une catégorie sémantique de noms flexibles, ouverts à l’alternance entre emplois massifs et comptables – avec néanmoins une forte préférence globale pour le comptable, dans tous les cas ; en outre, une comparaison prudente entre les données du français et du néerlandais tend à indiquer une flexibilité légèrement moindre en néerlandais (ce qui serait en adéquation avec les intuitions de locuteurs natifs) – mais cette différence reste à confirmer ;
2. Dans le cadre de la flexibilité observée, chaque item étudié présente un caractère distributionnel propre, s’accommodant du massif de manière plus ou moins fréquente et plus ou moins aisée (*raisin* et *chou-fleur* sont plus souvent massifs que *carotte*, *meloen* ‘melon’ et *komkommer* ‘concombre’ qu’*appel* ‘pomme, etc.) ;
3. La syntaxe massive permet de présenter les noms de fruits et de légumes non pas de manière nécessairement broyée ou démultipliée, mais simplement en forme et quantité indéterminée ;
4. De par cette nuance sémantique, le choix de l’option syntaxique massive n’est pas aléatoire : elle est associée à certains contextes d’usage typiques, justifiant de manière pragmatique la présentation homogénéisée (et comparables en français et en néerlandais).

Notons que nous n'en finissons pas tout à fait ici avec les noms de fruits et de légumes : nous les retrouverons encore au Chapitre 9 lorsque nous décrirons quelques enquêtes d'acceptabilité que nous avons menées, de même qu'au Chapitre 10 (pour le français), lorsque nous décrirons les résultats d'une expérience d'amorçage. Cependant, avant d'en arriver là, il nous reste encore deux autres études de cas à présenter : celle sur les noms de boissons, et celle sur les noms de matériaux. Cette dernière fait l'objet du chapitre suivant.

# Chapitre 6 Les noms de matériaux

## 6.1 Les noms de matériaux en français

### 6.1.1 Analyse distributionnelle

Pour l'analyse distributionnelle de nos noms de matériaux, nous avons procédé de manière sensiblement identique à ce que nous avons décrit précédemment. Seules deux petites différences sont à noter. D'une part, nous avons inclus ici parmi les contextes cumulatifs la construction *en + N*, typique des noms de matériaux (*ossature en chêne, des centaines d'objets en porcelaine*, etc.) – autrement classée par notre script *SN\_FR.pl* parmi les contextes neutres, vu la multifonctionnalité de la préposition *en* (il suffit de comparer *un avion en papier, des vacances en montagne* et *un parcours en train*). D'autre part, dans le cadre de notre examen systématique des emplois comptables (voir point 6.1.2 ci-dessous), nous avons parcouru manuellement l'ensemble des occurrences annotées comme cumulatives plurielles, et avons procédé ce faisant à quelques modifications dans l'annotation, lorsque le contexte nous laissait supposer avec une très grande certitude que l'emploi observé relevait en réalité du massif, et que l'utilisation de la forme plurielle du nom était fort probablement une faute d'orthographe. C'est le cas par exemple pour les énoncés ci-après :

- (176) Voir aussi listes des lépidoptères se nourrissant de chênes.
- (177) Trois mille sont desservis par les chemins de fers fédéraux, et deux mille par des compagnies privées.
- (178) Certaines pièces très rares de marbres contiennent de[s] inclusions de fossiles (...).

De toute évidence, les lépidoptères de (176) mangent *du chêne*, et non des arbres entiers, de même qu'on écrit normalement *chemins de fer* avec *fer* au singulier, et qu'il s'agit en (178) bien de pièces *de marbre*. (Notons que pour ces deux derniers cas, l'ajout d'un *-s* final superflu au nom de matériau découle probablement de l'influence du nom tête, au pluriel : *les chemins* et *certaines pièces*.)

Le résultat du premier examen distributionnel de nos données concernant les noms de matériaux en français se trouve présenté dans la Figure 14. Les items sont organisés cette fois par ordre croissant de la proportion d'emplois comptables, et l'axe des ordonnées est présenté dans son intégralité :

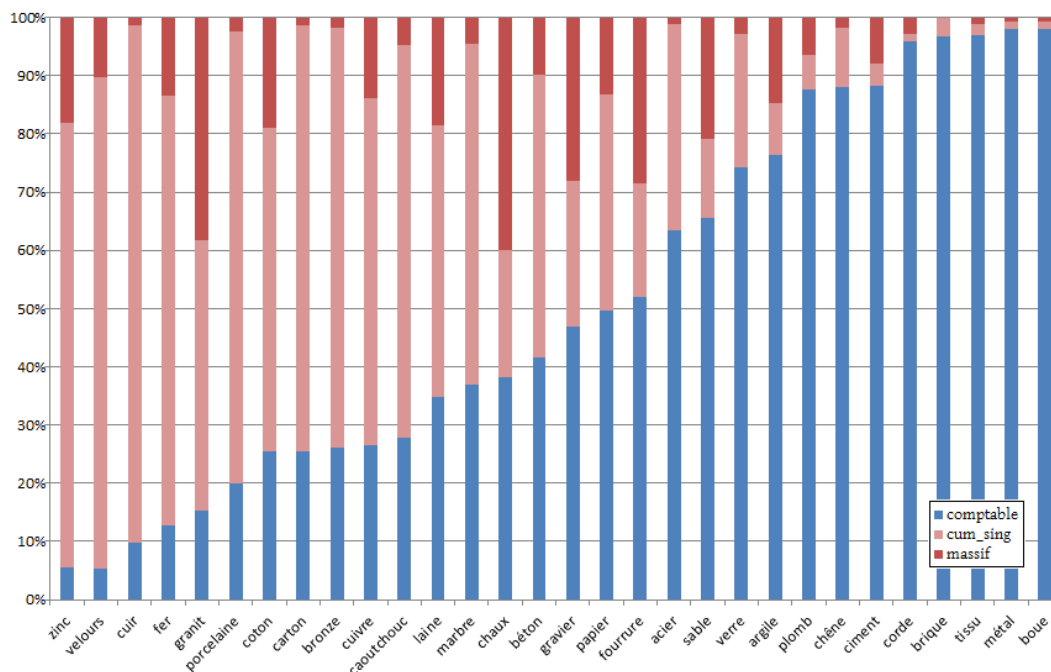


Figure 14 Profils distributionnels de nos 30 noms de matériaux, en français.

L'examen de la Figure 14 nous permet de faire plusieurs constats. D'une part, nous voyons que tous nos items sont attestés tant en contextes massifs que comptables. Cela n'est pas surprenant : les usages massifs étaient évidemment prévus, tandis que pour la plupart des items, des usages comptables et naturels pouvaient d'avance être imaginés, et donc attendus (*un verre*, *un carton*, *les plombs (ont sauté)*, *un fer (à repasser)*, etc.). (Nous les examinons plus en détail au point suivant.)

En outre, nous voyons que ce sont les contextes cumulatifs singuliers, représentés en rouge clair dans le graphique, qui dominent du côté du massif. A nouveau, rien de très étrange : sur un total de 2 035 occurrences de ce type, 1 901 (soient 93,42 %) sont du type *en + N singulier*. Or, il s'agit là effectivement du contexte définitoire des noms de matériaux (s'il n'était pour la polysémie de la préposition *en*) : si quelque chose est *fabriqué en X*, et que *en X* n'est ni un complément de temps, ni un complément de lieu, c'est que X doit désigner un matériau.

D'autre part, notre graphique nous dévoile aussi un résultat bien plus inattendu. Alors que l'on pourrait supposer que les occurrences comptables des noms de matériaux seraient en quelque sorte dérivées, subordonnées aux usages massifs plus habituels, et ainsi moins fréquentes que ces derniers (c'était le cas pour les occurrences massives de nos noms de fruits et de légumes au chapitre précédent), force est de constater que leur proportion au sein de nos résultats est non négligeable, et dépasse même en nombre les

occurrences massives pour 13 items sur 30 (cf. l'aile droite de notre graphique, à partir de *fournure*). Plus précisément, si au total, tous items confondus, nous avons étudié 5 652 occurrences de nos noms de matériaux, précisément 2 722 (soit 48,2 %) se sont avérées être en emploi comptable.

Ce fait contredit fortement la supposition naïve que, les noms de matériaux étant supposés fondamentalement massifs (de par les caractéristiques homogènes de leurs référents), cette caractérisation morphosyntaxique se refléterait élégamment dans leurs profils distributionnels. Dès lors, notre examen sémantique des emplois comptables, au point suivant, sera d'autant plus crucial, en ce qu'il permettra de faire 'sens' de manière plus satisfaisante de nos données.

### 6.1.2 Analyse sémantique

Avant de nous pencher de plus près sur les emplois comptables de nos noms de matériaux, précisons qu'un examen approfondi des occurrences massives ne paraît pas, quant à lui, nécessaire. Leur sens, dans nos données, est en effet invariablement le même : la présentation du référent en tant que substance homogène – en tant que 'matière' (168)-(170). Nous n'avons observé à cela qu'une exception, pour *corde*, dont nous trouvons, outre l'usage en tant que matériau, quelques occurrences au sein de la locution *faire de la corde à sauter*, exprimant une activité (voir (Van de Velde 1997)). C'est le seul cas d'occurrences massives se démarquant des autres – le reste de la variabilité sémantique se situant bel et bien à l'endroit des usages comptables.

(179) On trouve du cuivre dans le foie (...).

(180) (...) le reste étant du caoutchouc synthétique.

(181) Il se tricote avec de la laine Milena.

Dans la littérature sur le massif-comptable, les deux mécanismes sémantiques expliquant couramment les passages du massif au comptable sont le Trieur et le Conditionneur Universels (voir point 1.5.2.2 ci-dessus). Or, si nous examinons les occurrences comptables de nos noms de matériaux, nous retrouvons effectivement deux types de lectures fondamentalement différentes : une lecture en termes de sortes (182)-(184), et une autre en termes d'objets spécifiques (185)-(187).

(182) L'approvisionnement s'effectue uniquement avec des laines françaises (...).

(183) [C]'est la certitude d'un tissu de qualité à un prix imbattable.

(184) Cette directive s'applique en effet à toutes les boues qu'elles soient urbaines ou industrielles.

(185) Des chats planqués dans des cartons (cons de chats).

(186) On le boit de plus en plus dans des verres transparents.

(187) L'appareil est bien plus ergonomique que les fers à repasser électriques ordinaires.

Dans les exemples (182)-(184), nos noms de matériaux sont utilisés dans une syntaxe comptable sans perdre en essence leur signification. Ainsi, les *laines françaises* dont il est question dans (182) sont toujours bel et bien ni plus ni moins que *de la laine*, tandis qu'*un tissu de qualité*, dans (183), c'est bien *du tissu*. Néanmoins, ils revêtent par le biais de la syntaxe comptable la donnée sémantique supplémentaire que leurs référents sont différenciés sur une base qualitative : la substance désignée habituellement par le nom massif existe en différentes variétés (objectives, ou contextuelles), et celles-ci forment la base de l'individualisation autorisant le marquage morphosyntaxique comptable. Ces données confirment ce que nous disions au point 1.5.2.2 ci-dessus : que si le Trieur Universel n'est pas tant universel qu'il n'est spécifique aux langues qui le connaissent (cf. (Lauwers et Vermote 2014, à paraître) pour une comparaison entre le français et le néerlandais), il est en tous les cas disponible en français – où il s'applique volontiers.

Les énoncés (185)-(187), quant à eux, illustrent un tout autre phénomène. Ici, le nom de matériau en syntaxe comptable ne désigne plus le matériau lui-même, mais bien un objet spécifique : *des cartons* ce sont des boîtes, *des verres* ce sont des récipients, et *les fers* (à repasser) sont des ustensiles ménagers. Or, attribuer ces cas aux effets du Conditionneur Universel n'est pas pour autant une évidence. Si ce mécanisme semble adéquat pour décrire les effets de sens relativement productifs et en tout cas rigoureusement homogènes liés à l'utilisation comptable d'un nom de boisson (*une bière*, *un coca*, *une limonade*, pour désigner un verre de ces boissons, voir au chapitre suivant), pour les noms de matériaux, une telle analyse est trop réductrice.

En effet, nos trois exemples (185)-(187) suffisent déjà à indiquer ce que nous confirmerons plus loin : les lectures en termes d'objets spécifiques des noms de matériaux sont bien trop différentes les unes des autres pour qu'une explication selon un mécanisme de transfert productif, tel le Conditionneur, soit satisfaisante. (Il nous faudra pencher pour une explication en termes d'emplois comptables lexicalisés, voir ci-après.)

Notre tour d'horizon n'est cependant pas encore terminé : le profil sémantique des emplois comptables de nos noms de matériaux est plus subtil encore que ce que les exemples (182)-(187) laissent paraître. S'il est vrai que les usages en termes de sortes et d'objets spécifiques, au sens large, couvrent l'intégralité de nos résultats comptables, nous avons néanmoins constaté que tous deux se subdivisaient encore en deux sous-types de lectures : des emplois fortement individualisés d'une part, et des emplois que nous appellerons 'collectifs' d'autre part.

Les exemples (182)-(187) rentrent dans la première de ces deux catégories. Ainsi, *des laines* '~des sortes de laine' dans (182) correspond bien à la mise au pluriel d'un pendant singulier, *une laine* '~une sorte de laine'. Parallèlement, dans (186), le SN *des verres* réfère à un ensemble d'objets, dont un représentant unique serait désigné par le

correspondant singulier, *un verre*. Et dans (183), *un tissu* est bel et bien individualisé, puisqu'il est employé au singulier. Pour ce type de cas, nous parlerons de lectures en individualisation. Remarquons que ce type de lecture s'accommode tant du singulier (183) que du pluriel (182),(184)-(187). Pour ce dernier, il s'agira alors d'un pluriel grammatical, dans les termes d'(Alexiadou 2011) (voir ci-dessus, point 5.1.2.3) : une forme plurielle dérivée grammaticalement d'un équivalent singulier, et dont le référent itéré est individualisé.

Or, dans les exemples ci-après (174-179), nos items au pluriel appartiennent à un type de lecture quelque peu différent, en ce qu'ils ne correspondent pas de manière aussi évidente à une forme au singulier comptable :

- (188) « Kaboul, transition dans les sables mouvants. »
- (189) Chez ces animaux, les tissus se modifient peu au cours de l'embryogénèse.
- (190) Les boues d'épuration ont des propriétés agronomiques utiles dans le domaine de l'agriculture.
  
- (191) Il aurait fait le nécessaire pour qu'elle puisse obtenir ses papiers.
- (192) Avant de revendre tous tes métaux de récupération, je te conseille de contacter plusieurs ferrailleurs.
- (193) Les cuivres à perce cylindrique sont en fait cylindro-coniques.

De toute évidence, *des sables mouvants* (188) ne sont pas réellement le pluriel de ?*un sable mouvant* (on peut d'ailleurs difficilement parler de ??*plusieurs sables mouvants*), pas plus que lorsque l'on a *ses papiers* (191), l'on est en possession de plusieurs fois *un papier* (un passeport n'est pas ??*un papier d'identité*). Pareillement, si l'ensemble des trompettes, trombones et autres clairons d'un orchestre en constitue *les cuivres* (193), *un cuivre* n'est pas une appellation habituelle pour un exemplaire unique de l'un de ces instruments. Dans tous ces cas, l'on a affaire à ce qu'(Alexiadou 2011) appelle le pluriel lexical d'un nom habituellement massif (voir ci-dessus, point 5.1.2.3) : un marquage pluriel impliqué directement dans le processus de création nominal lui-même, et qui n'a dès lors pas pour domaine d'application le référent du nom auquel il se combine. De ce fait, l'interprétation de la mise au pluriel est idiosyncratique, et est basée sur l'itération référentielle la plus plausible en contexte : *les eaux internationales* '~plusieurs mers', *les pluies du mois d'avril* '~plusieurs averses'. Pour nos noms de matériaux, nous pouvons observer que la nature de cette itération référentielle est de deux types : soit abstraite (188)-(190), lorsqu'elle s'opère sur la base d'une hétérogénéité qualitative, soit concrète (191)-(193), lorsqu'elle repose sur la présence dans le contexte de différents objets tangibles. L'on retrouve donc ici notre distinction précédente, entre sortes et objets spécifiques. Les lectures 'collectives' pourront donc être rapprochées respectivement de ces derniers.

Une précision supplémentaire s'impose encore au sujet des emplois collectifs du type concret (collection d'objets). Si nous avons observé que pour les lectures 'objet' en individualisation, les référents présentaient une trop grande hétérogénéité que pour conclure à un mécanisme de transfert productif, la situation semble un petit peu différente ici. D'une part, il est vrai que nous rencontrons bel et bien des emplois collectifs concrets dont le sens très spécifique semble tout aussi peu prévisible que celui des emplois homologues en individualisation : *des papiers* 'documents d'identité', *des cuivres* 'instruments de musique', etc. D'autre part, pour un sous-groupe de nos items, ces emplois collectifs en termes d'objets semblent malgré tout répondre à un schéma sémantique récurrent, dénotant un 'ensemble d'objets peu spécifiques'. L'exemple (192) en est une première illustration. En voici quelques autres :

- (194) Les premiers verres fabriqués par l'homme sont originaires de Mésopotamie (...).
- (195) Entretien du cuir : pour détacher, nettoyer, (...), protéger, et imperméabiliser vos cuirs.
- (196) Utiliser un détachant qui préserve les couleurs sur les cotons colorés.
- (197) A combien peut-on revendre ses métaux chez les ferrailleurs ?

Ces quelques exemples semblent être emblématiques d'un transfert sémantique (méconnu) qui permettrait en français d'obtenir, à partir d'un nom massif de matériau, la dénomination plurielle d'une « collection indéterminée d'objets ou de morceaux d'objets dans le matériau en question ». Les *métaux* dans (192) et (197) sont en effet simplement 'un ensemble d'objets (indéterminés) en métal', au même titre que les *cuirs* de (195) désignent 'n'importe quelle surface en cuir' ou que les *cotons* de (196) ne sont autres que 'tous [vos] objets en coton'. Par analogie avec les autres mécanismes similaires déjà proposés dans la littérature (Trieur, Conditionneur, Multiplicateur Universels), nous avons proposé d'appeler cet effet sémantique le « Collecteur Universel » (Vermote 2014, à paraître) – en référence à la nature collective de la lecture.

Ces diverses précisions sémantiques faites, nous pouvons passer à présent à une nouvelle visualisation de la distribution de nos résultats, présentée dans la Figure 15. Les contextes massifs et cumulatifs singuliers sont toujours représentés en rouge. Les occurrences comptables, quant à elles, ont été réparties selon les quatre sous-catégories que nous venons de décrire, avec en bleu les lectures objets et en orange les lectures en termes de sortes – les emplois collectifs respectifs étant représentés dans les couleurs correspondantes, un peu plus claires. L'ordre dans lequel nous classons à présent nos items est fonction de la proportion de leurs occurrences comptables en termes d'objets – lectures individualisées et collectives confondues (et diffère dès lors quelque peu de celui de la Figure 14).



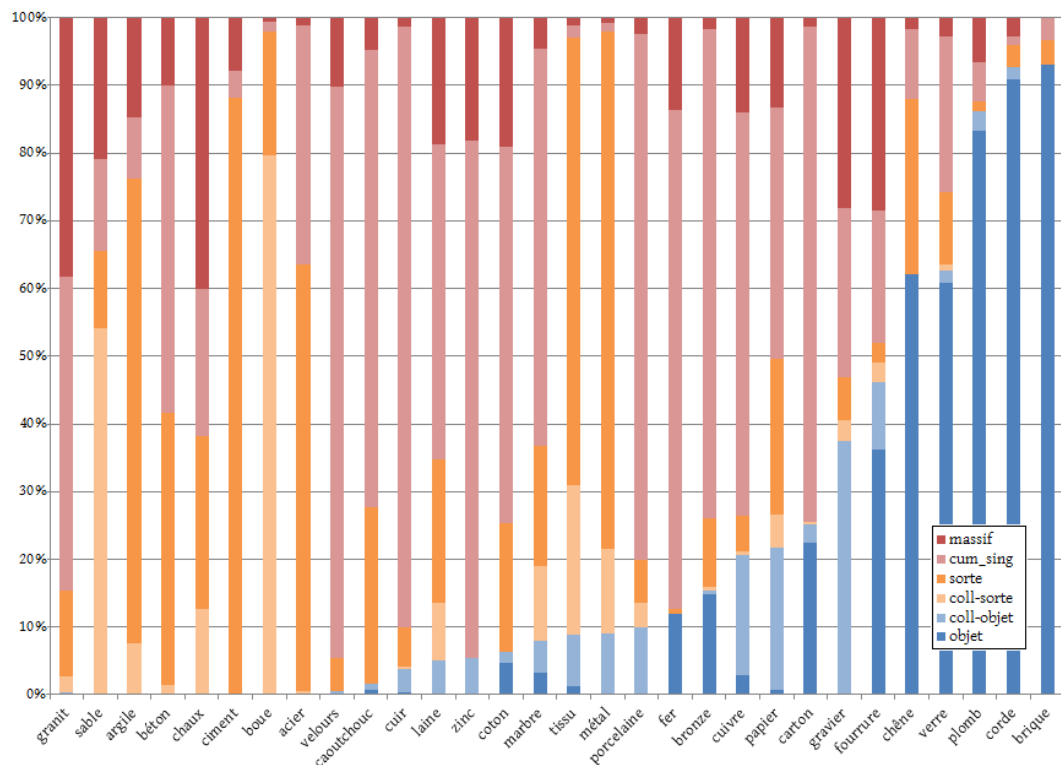


Figure 15 Profils distributionnels de nos 30 noms de matériaux en français, avec détail sémantique des emplois comptables (lectures en termes de sortes et d’objets, collectives ou non).

S’il y a un premier constat à faire, c’est qu’après l’analyse sémantique plus précise de nos données, nos résultats se rapprochent considérablement de nos attentes intuitives. En effet, pour la plupart des items, une bonne proportion des occurrences comptables observées relevait en réalité du mécanisme du Trieur (*sorte* et *coll-sort*). Or, puisque ce mécanisme est précisément typique des noms massifs, comme nous l’avons vu au point 1.5.2.2, ces occurrences ne constituent plus de réels contre-exemples à l’encontre du caractère fondamentalement massif des noms de matériaux – bien au contraire.

Par ailleurs, la proportion de lectures en termes de sortes est particulièrement élevée dans nos données pour certains items : 611 occurrences sur 624 pour *boue* (97,9 %), 67 sur 76 pour *ciment* (88,2 %), ou encore 255 sur 287 pour *métal* (88,6 %). Cela peut paraître surprenant, mais s’explique sans trop de peine à y regarder de plus près. Pour *boue*, les résultats sont dominés par des textes au sujet des *boues d’épuration* et autres *boues* à éliminer – où la forme plurielle massive *boues*, exprimant une hétérogénéité de sortes, est tout à fait attendue, puisque parfaitement lexicalisée. Quant à *ciment*, les textes dont sont tirées les occurrences retournées par *WebCorp* sont essentiellement à teneur technique, ce qui explique le recours fréquent au comptable pour identifier des sortes encyclopédiques (198)-(199). Pour finir, le caractère hypéronymique de *métal* explique bien certainement qu’on le retrouve très souvent employé sur le mode du Trieur – au point que l’on pourrait d’ailleurs se demander si ça ne serait pas là son acception de base (200)-(201).

- (198) Ensuite, nous classifions les ciments suivant leur composition et leur résistance.
- (199) La norme définit vingt-sept ciments communs et leurs constituants.
- (200) L'étain est un métal qui ne s'oxyde pas.
- (201) Le carbone n'est pas un métal.

Si l'on reconsidère dès lors la distribution morphosyntaxique de nos noms de matériaux en se permettant d'assimiler au caractère massif prévisible les lectures comptables en termes de sortes (individualisées ou collectives), l'on ne trouve plus dans notre graphique de la Figure 15 que 5 items, et non plus 13, qui dépassent la barre des 50 % d'occurrences proprement comptables. Il s'agit de *chêne*, *verre*, *plomb*, *corde* et *brique*, pour lesquels le sens comptable objet est particulièrement saillant, et facile à récupérer intuitivement – voir Tableau 1 ci-après. Pour les autres, la lecture en termes d'objet ou de collection d'objets est soit minoritaire (c'est ici qu'on retrouve notamment l'œuvre de notre Collecteur), soit non attestée (c'est le cas pour *granit*, *sable*, *argile*, *béton*, *chaux*, *ciment*, *boue* et *acier*).

Ainsi, l'on se retrouve, quant aux occurrences comptables de nos noms de matériaux, avec au moins trois cas de figure :

1. Des emplois en termes de sortes (*un acier de grande qualité*), attribuables au mécanisme productif du Trieur Universel et attestés pour tous nos items ;
2. Des emplois en termes de collections indéterminées (*débarrassez-vous de vos vieux métaux*), attribuables à notre mécanisme du Collecteur et plus limités tant au niveau du nombre d'occurrences observée que du nombre d'items concernés ;
3. Des emplois en termes d'un objet ou d'une collection d'objets spécifiques, limité quant au nombre d'items concernés, mais très fréquents pour ces items-là (voir en particulier l'extrémité droite de la Figure 15).

Dans les deux premiers cas, il semble bel et bien légitime de parler de deux mécanismes de transfert : par le biais de la syntaxe comptable, un nom fondamentalement massif désignant un matériau est utilisé pour désigner une ou plusieurs sortes de ce matériau (Trieur) ou un ensemble indéterminé d'objets qui en sont constitués (Collecteur). Dans le troisième cas, en revanche, la notion de passage de massif à comptable n'est plus d'application. Si elle est peut-être pertinente dans une perspective diachronique (du moins pour certains items, comme *verre* ou *plomb* – dans d'autres, comme *chêne* ou *brique*, l'emploi comptable semble d'emblée jouir de la prééminence), il apparaît que synchroniquement, ces lectures en termes d'objets spécifiques des noms de matériaux jouissent d'une grande autonomie par rapport à leurs homologues massifs. Il faudra donc leur attribuer un caractère lexicalisé – c'est-à-dire, à l'intérieur de notre modèle du lexique, leur attribuer un concept lexical propre (nous verrons au Chapitre 8 ce que nous entendons par là). De nombreux arguments plaident en cette faveur :

1. Le lien entre le référent ‘matériau’ massif et le référent ‘objet’ comptable n’est pas systématique ; hormis la parenté historique indéniable, il y a opacité sémantique : *les cuivres* ne sont pas nécessairement en cuivre, *un verre* n’est pas toujours en verre, *un fer (à repasser)* est loin d’être tout en fer, etc. ;
2. L’existence d’un emploi comptable ‘objet’ n’est pas systématique, mais repose au contraire crucialement sur la disponibilité d’un référent approprié, sélectionné et spécifique (??*un sable*, ??*un acier*) ;
3. Dans la mesure où, dans certains cas, le lien entre matériau et objet est bel et bien maintenu, et un (des) référent(s) approprié(s) est (sont) disponible(s), l’emploi comptable ‘objet’ n’est pas automatique et sa teneur est conventionnelle : *un fer*, c’est soit un fer à repasser, soit un club golf, mais pas, par exemple, une boîte en fer blanc ; voir aussi (Martin 2005: 170) pour une observation similaire concernant *bronze*, qui peut « désigner, par métonymie, une statuette de bronze, une monnaie de bronze, mais pas un canon, même s’il est entièrement en bronze ».

Ainsi, contrairement à ce que l’on observe dans le cas des transferts productifs, les effets de sens associés aux emplois comptables ‘objet’ des noms de matériaux ne peuvent pas se résumer en une formule sémantique générale. Si l’on veut les décrire, il faudra les lister. Nous terminerons dès lors notre analyse sémantique en présentant, dans le Tableau 1 ci-dessous, la liste des sens ‘objet’ idiosyncratiques de nos noms de matériaux, tels que nous les avons observés dans nos données de corpus. Des exemples authentiques sont fournis dans certains cas, mais pas systématiquement, dans un souci d’économie et de lisibilité. Nous intégrons aussi à la liste les sens collectifs concrets en termes de ‘collection indéterminée’, attribuables au Collecteur (et donc moins lexicalisés que résultant d’un mécanisme productif) – afin que l’on puisse constater où ils se situent. Les items sont présentés par ordre décroissant de la proportion des occurrences à l’examen (cf. Figure 15) – les items pour lesquels une telle lecture n’est pas attestée n’étant forcément pas repris.

Tableau 1 Liste des effets de sens comptables en termes d’objets (individualisés ou collectifs) des noms de matériaux étudiés en français.

<i>Brique</i>	<ul style="list-style-type: none"> <li>• bloc de construction (généralement en terre cuite, parfois aussi en verre)</li> <li>• pièce à détruire dans un jeu vidéo</li> <li>• bloc de construction Lego®</li> <li>• thé compacté (<i>briques de thé</i>)</li> <li>• usage métaphorique (<i>les briques élémentaires de la vie</i>)</li> </ul>
<i>Corde</i>	<ul style="list-style-type: none"> <li>• partie d’un instrument de musique (<i>une basse à six cordes</i>)</li> <li>• morceau de corde (<i>les pieds et poings liés par les cordes</i>)</li> <li>• accessoire sportif (<i>une corde à sauter</i>)</li> </ul>

Plomb	<ul style="list-style-type: none"> <li>• baguette pour la construction de vitraux • [généralement pluriel] fusible (<i>les plombs ont sauté</i>) • emploi métaphorique (<i>il pète un plomb</i>) • lestage pour canne à pêche</li> </ul>
Verre	<ul style="list-style-type: none"> <li>• récipient pour consommer une boisson • par métonymie, portion d'une boisson (<i>un dernier verre pour la route</i>) • pièce d'optique (<i>verres progressifs</i>) • [pluriel] collection d'objets en verre (<i>les premiers verres fabriqués par l'homme ; les prochains verres de nos iPhones</i>)</li> </ul>
Chêne	<ul style="list-style-type: none"> <li>• grand arbre de la famille des fagacées</li> </ul>
Fourrure	<ul style="list-style-type: none"> <li>• peau de bête (<i>des fourrures provenant de coyotes capturés (...)</i> ; <i>la traite des fourrures</i>) • manteau ou autre vêtement de fourrure (<i>l'été ne jetez pas vos fourrures ; [elle se] cache sous une fourrure d'âne</i>) • motif héraldique (<i>une fourrure, ou panne</i>)</li> </ul>
Gravier	<ul style="list-style-type: none"> <li>• [généralement pluriel] cailloux (<i>un bruit de bottes sur les graviers</i>)</li> </ul>
Carton	<ul style="list-style-type: none"> <li>• boîte en carton (<i>cartons de déménagement</i>) • carte, fiche cartonnée (<i>cartons d'invitation</i>), e.a. dans le football (<i>un carton rouge</i>) • [pluriel] ensemble de morceaux de carton (<i>transforme[z] vos cartons en magnifique meuble design</i>)</li> </ul>
Papier	<ul style="list-style-type: none"> <li>• [pluriel] documents (en papier), p.ex. d'identité</li> </ul>
Cuivre	<ul style="list-style-type: none"> <li>• [pluriel] instruments de musique (<i>la famille des cuivres</i>) • gemme dans un jeu vidéo (<i>j'échange (...) un cuivre contre (...) un fer</i>) • tuyau en cuivre (un seul exemple, probablement une ellipse : <i>tu peux repasser un tuyau per alors qu'un cuivre c'est mort</i>)</li> </ul>
Bronze	<ul style="list-style-type: none"> <li>• pièce d'art en métal (<i>la famille Pinault va restituer deux bronzes à la chine ; les sculptures en métal fondu sont souvent appelées des bronzes</i>)</li> </ul>
Fer	<ul style="list-style-type: none"> <li>• ustensile ménager (<i>fer à repasser</i>) • club de golf</li> </ul>
Porcelaine	<ul style="list-style-type: none"> <li>• [pluriel] objets fabriqués en porcelaine (<i>des porcelaines de Limoges entièrement peintes à la main</i>)</li> </ul>
Métal	<ul style="list-style-type: none"> <li>• [pluriel] pièces, objets en métal (<i>revendre ses (vieux) métaux chez le ferrailleur</i>)</li> </ul>
Tissu	<ul style="list-style-type: none"> <li>• [pluriel] pièces de tissu (<i>des tissus imprimés</i>)</li> </ul>

Marbre	<ul style="list-style-type: none"> <li>statue, œuvre d'art en marbre (<i>un marbre de Rodin</i>)</li> <li>partie d'une presse d'imprimerie (<i>un marbre de pierre</i>)</li> <li>[pluriel] pièces d'art en marbre (<i>ce musée possède de beaux marbres</i>)</li> </ul>
Coton	<ul style="list-style-type: none"> <li>tissu de coton ((...) <i>enveloppées dans un petit coton</i>)</li> <li>bâtonnet ouaté (<i>coton tige</i>)</li> <li>[pluriel] habits, textiles en coton (<i>un détachant qui préserve les couleurs sur les cotons colorés</i>)</li> </ul>
Zinc	<ul style="list-style-type: none"> <li>?? (<i>la compagnie des zincs</i>)</li> </ul>
Laine	<ul style="list-style-type: none"> <li>[pluriel] habits, textiles en laine (<i>entretien des laines</i>)</li> </ul>
Cuir	<ul style="list-style-type: none"> <li>[pluriel] pièces, objets en cuir (<i>réparer vos cuirs</i>)</li> </ul>
Caoutchouc	<ul style="list-style-type: none"> <li>morceau de caoutchouc (<i>un caoutchouc revient, en moins d'une minute, à moins de 1,5 fois sa longueur initiale</i>)</li> </ul>
Velours	<ul style="list-style-type: none"> <li>[pluriel] vêtements en velours (<i>dans la gamme des velours tu as oublié (...) le jogging élégant en velours</i>)</li> </ul>

## 6.2 Les noms de matériaux en néerlandais

*Die eikels zijn oud genoeg om te weten dat een baksteen en een hoofd natuurlijke vijanden zijn. (WebCorp, 7 août 2013)*

### 6.2.1 Analyse distributionnelle

La Figure 17ci-dessous présente le résultat de l'analyse distributionnelle de nos 34 noms de matériaux en néerlandais. Nous y indiquons la proportion d'emplois massifs (en rouge) et comptables (en bleu) – avec pour ces derniers une distinction supplémentaire entre emplois avec et sans diminutif (comme c'était déjà le cas pour les noms de fruits et de légumes). Nous n'avons pas marqué dans une couleur différente les cas d'emplois en syntagme cumulatif au singulier, contrairement à ce que l'on pouvait trouver dans la Figure 14. En français, cette distinction avait du sens, car elle permettait d'identifier facilement la proportion d'occurrences de nos noms de matériaux dans la construction *en + N* – définitoire, s'il en est, de la catégorie. En néerlandais, nous n'avons pas été en mesure d'effectuer ce même repérage, essentiellement à cause de l'ambiguïté de la

construction équivalente *van + N*, et plus précisément du caractère multifonctionnel de la préposition *van* ‘de’. Ainsi, seule une partie des occurrences dans cette construction correspond à la signification de *en + N* en français (202)-(203), tandis que les autres correspondent simplement à la construction prépositionnelle *de + N* (204)-(205). Dès lors, les indiquer séparément dans notre graphique n’était pas particulièrement pertinent. Cette petite nuance n’a du reste aucune incidence sur nos résultats globaux.

(202) *Deze huizen zijn allemaal van baksteen.*

‘Ces maisons sont toutes en brique.’

(203) *Kerkklokken zijn (...) vrijwel altijd van brons.*

‘Les cloches des églises sont (...) pour ainsi dire toujours en bronze.’

(204) *In 1950 lag de wereldwijde gebruik van rubber op 2,5 miljoen ton.*

En 1950, l’usage mondial de caoutchouc se trouvait à 2,5 millions de tonnes.

(205) *De mechanische eigenschappen van staal en gietijzer (...).*

‘Les caractéristiques mécaniques de l’acier et de la fonte...’

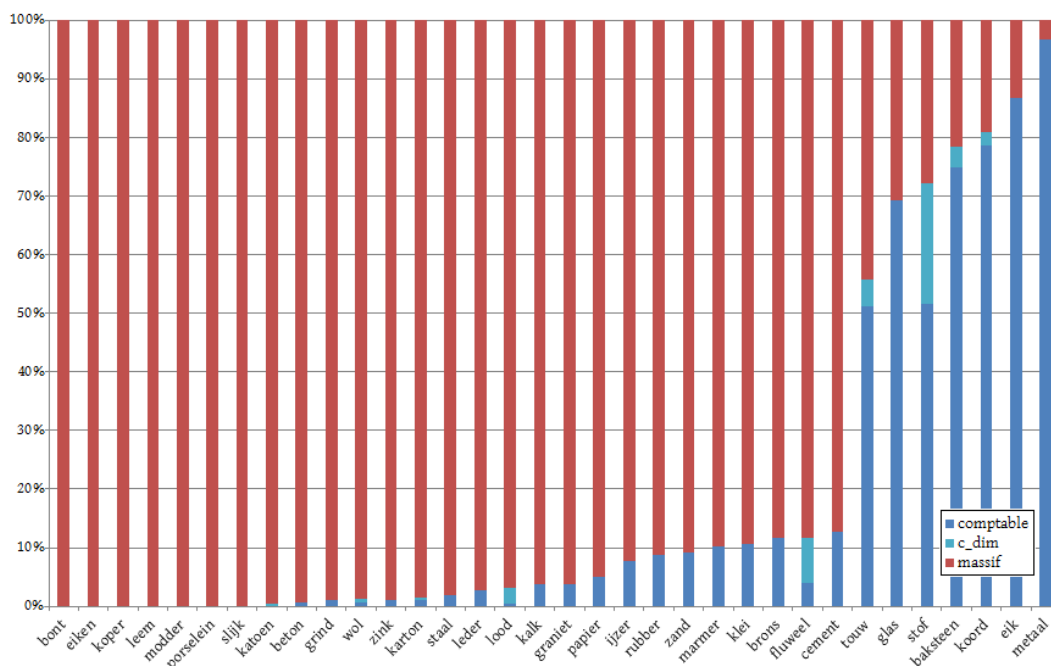


Figure 16 Profils distributionnels de nos 34 noms de matériaux, en néerlandais.

Si l’on compare ce graphique avec celui de la Figure 14, l’on voit tout de suite que la proportion d’emplois comptables est ici beaucoup moins grande. En français, 48,2 % du total de nos occurrences massives et comptables confondues étaient de ce type, réparties sur l’ensemble de nos items (et mêmes majoritaires pour 13 items sur 30). Ici, en revanche, sur 6 813 occurrences au total, seules 1 262 (soit 18,5 %) sont des exemples d’emplois comptables – avec seulement 7 items qui dépassent la barre des 50 %, et 7 pour lesquels ce type d’emploi n’est pas même attesté.

Une partie de cette différence peut probablement s'expliquer par la double fonction, et donc la plus large distribution, du massif en néerlandais (voir point 1.2.2) : certaines des occurrences massives reprises dans la Figure 16, à savoir les emplois massifs génériques, correspondent à des emplois neutres en français – ce qui résulte en un déséquilibre systématique, en faveur des occurrences neutres en français, et des occurrences massives en néerlandais. Ceci est corroboré par nos données : en français, 59,5 % de nos occurrences de noms de matériaux étaient en syntaxe neutre (8 298 cas sur 13 950), contre seulement 22,8 % en néerlandais (2 010 sur 8 823). (Nous rappelons que nous avons à chaque fois exclu ces occurrences neutres de nos analyses.) D'autre part, ça ne peut être là toute l'explication, car même en considérant la proportion d'occurrences comptables sur l'ensemble des données analysées (c'est-à-dire, en incluant les occurrences neutres dans le comptage total), et en annulant ainsi le déséquilibre potentiellement causé par le caractère particulier du générique des massifs (en syntaxe massive en néerlandais, mais neutre en français), nous obtenons toujours une disparité entre nos deux langues : 22,9 % d'occurrences comptables en français (3 268 cas sur 13 950) contre seulement 14,3 % en néerlandais (1 262 cas sur 8 823). L'application du test du chi carré sur les valeurs absolues concernées confirme du reste que cette différence est hautement significative ( $\chi^2(1, N = 27\,303) = 192,17$ , valeur  $p < 0,0001$ ).

Pour expliquer cette inégalité, l'on évoquera d'abord un désavantage morphologique latent du néerlandais standard : pour beaucoup de noms typiquement massifs, cette langue ne connaît pas de forme plurielle officielle, comme le confirme un rapide coup d'œil dans la *Woordenlijst* de la NTU (*Nederlandse Taalunie*), l'organe officiel pour la politique en matière de langue néerlandaise (liste disponible sur <http://woordenlijst.org/>). Cette liste ne se veut pas un relevé exhaustif des mots existant en néerlandais (voir dans la *Inleiding* 'Introduction' : « L'inclusion d'un mot ne peut pas être considérée comme une approbation officielle du mot, tandis que l'absence d'un mot ne signifie pas sa réprobation », nous traduisons – texte original disponible sur <http://woordenlijst.org/leidraad/inleiding/>); néanmoins, puisqu'elle est fondée notamment sur des considérations de fréquence, et qu'elle contient bel et bien tous les noms de matériaux que nous avons étudiés (hormis *eiken* en tant que nom massif singulier) mais ne donne une forme plurielle que pour certains d'entre eux, elle nous semble au moins dévoiler, si ce n'est l'inexistence absolue, au moins la rareté des formes absentes.

Ainsi, sur nos 34 items, seuls 15 ont une forme plurielle référencée par la NTU. Nous reproduisons ici la liste de nos noms de matériaux dans le même ordre qu'ils apparaissent à la Figure 16, en donnant les formes plurielles attestées entre parenthèses et barrant les items pour lesquels la *Woordenlijst* n'en donne pas : ~~bont~~, ~~eiken~~, ~~koper~~, ~~leem~~, ~~modder~~, porselein (*porseleinen*), ~~slijk~~, ~~katoen~~, ~~beton~~, ~~grind~~, ~~wol~~, ~~zink~~, karton (*kartons*), staal (*stalen*), ~~leder~~, lood (*loden*), ~~kalk~~, ~~graniët~~, papier (*papieren*), ijzer (*ijzers*), ~~rubber~~, ~~zand~~, marmer

(*marmers*), *klei*, *brons*, *fluweel* (*fluwelen*), ~~*cement*~~, *touw* (*touwen*), *glas* (*glazen*), *stof* (*stoffen*), *baksteen* (*bakstenen*), *koord* (*koorden*), *eik* (*eiken*), *metaal* (*metalen*).

A l'examen de cette liste, il apparaît bel et bien une certaine corrélation avec nos observations distributionnelles. Ainsi, pour la première moitié de nos noms (ceux pour lesquels nous n'observons que peu ou pas d'occurrences comptables), seuls 4 items sur 17 ont une forme plurielle référencée. Pour la seconde moitié, la proportion est de 11 contre 17. En outre, les 7 items pour lesquels nous trouvons dans notre corpus plus d'occurrences comptables que massives, à l'extrême-droite de notre graphique, ont tous sans exception une forme plurielle attestée par la *Woordenlijst*.

La proportion beaucoup moins grande en néerlandais qu'en français des emplois comptables des noms de matériaux pourrait donc découler, au moins en partie, de la moins grande affinité de ces noms avec le pluriel. C'est là cependant une hypothèse à formuler avec la plus grande prudence, puisqu'on aura tôt fait d'aboutir au paradoxe de l'œuf et de la poule : est-ce que les noms de matériaux en néerlandais connaissent effectivement moins d'emplois comptables parce qu'ils sont réticents au pluriel, ou connaissent-ils inversement moins d'emplois au pluriel parce qu'ils sont généralement réticents à la syntaxe comptable ? En outre, l'indisponibilité de certaines formes plurielles ne peut pas être le fin mot de l'histoire, car si ce facteur était effectivement le plus décisif, l'on devrait s'attendre à trouver au sein de nos attestations comptables principalement des occurrences au singulier (du type *een* + *N* 'un(e) + *N*'). Or, sur les 1 262 occurrences comptables que nous avons recensées, 1 005, soit 79,4 %, sont bel et bien au pluriel. Du reste, nous trouvons même quelques attestations au pluriel de noms pour lesquels la NTU ne signale pas cette forme (206)-(209).

- (206) *[Daar] zijn grinden terug te vinden die zijn afgezet tijdens het Weichselien.*  
'L'on trouve [là-bas] des graviers qui se sont déposés pendant le Vistulien.'
- (207) *Deze minerale wollen zijn gesponnen uit glas of steen.*  
'Ces laines minérales sont tissées à partir de verre ou de pierre.'
- (208) *De Bronzen van Riace (...) zijn twee levensgrote bronzen beelden (...).*  
'Les Bronzes de Riace (...) sont deux statues de bronze grandeur nature (...).'
- (209) *Alle granieten hebben een aantal gemeenschappelijke kenmerken.*  
'Tous les granits partagent un certain nombre de caractéristiques.'

La différence entre le français et le néerlandais pourrait donc aller plus loin qu'une simple réticence à mettre au pluriel les noms massifs. Dans (Lauwers et Vermote 2014, à paraître), nous avons proposé qu'elle concerne l'applicabilité et la productivité du mécanisme du Trieur : si le français, comme plusieurs autres langues romanes (espagnol, italien), l'utilise volontiers, les langues germaniques en semblent beaucoup moins friandes (l'anglais constituant une exception notable – peut-être à cause de l'influence pénétrante qu'a eu le français sur cette langue au cours de son développement). Ainsi, la grammaire d'(Eisenberg 1989: 179), pour l'allemand, donne



des pluriels tels que \**Milche* ‘laits’, \**Wässer* ‘eaux’ ou \**Golde* ‘ors’ comme inexistants, et n’accepte les pluriels taxonomiques que comme un phénomène récent, à situer dans le domaine des progrès des techniques et de la science créant de nouveaux besoins terminologiques, pour référer à des taxonomies toujours plus fines (ainsi, il atteste par exemple *Werkzeugstähle* ‘aciers d’usage général’ contre *Edelstähle* ‘aciers précieux’, *Öle* ‘huiles’, *Fette* ‘graisses’, *Biere* ‘bières’, *Säfte* ‘jus’).

Pour le néerlandais, l’on retrouve dans les ouvrages de référence des remarques analogues. Ainsi, la grammaire *ANS (Algemene Nederlandse Spraakkunst*, disponible en version électronique sur <http://ans.ruhosting.nl/e-ans/>) mentionne que les *singularia tantum* peuvent d’une part apparaître avec le déterminant *een* ‘un(e)’ lorsqu’ils sont accompagnés d’une spécification (*Dit is een wijn uit de Rhônestreek* ‘Voici un vin de la région du Rhône’, mais pas \**Dit is een wijn* tout court), et d’autre part apparaître parfois au pluriel, dans le sens ‘sortes de la substance nommée’ (<http://ans.ruhosting.nl/e-ans/03/02/02/body.html>). Il est néanmoins précisé que cette deuxième option ne fonctionne pas dans tous les cas, car certains noms n’ont pas de pluriel – comme *muziek* ‘musique’, qui est donné en exemple. La possibilité du passage du singulier au pluriel à la manière du Trieur, si elle n’est pas exclue, n’apparaît donc ni tout à fait productive, ni particulièrement répandue.

C’est peut-être là les effets du réflexe puriste qui, dans le contexte diglossique caractérisant la Flandre du XIX<sup>ème</sup> siècle, a eu tendance à reconnaître dans les emplois pluriels taxonomiques des noms massifs l’influence du français. On en retrouve notamment la trace dans (de Vreese 1899: 393), qui les condamne comme gallicismes, et avance la paraphrase et la compositions comme alternatives préférables : *soorten bier* ‘sortes de bières’ plutôt que *bieren*, *koffiesoorten* ‘sortes de café’ plutôt que *koffies*.

Ainsi, la proportion réduite d’emplois comptables des noms de matériaux en néerlandais pourrait découler de la réticence de cette langue pour le Trieur, un mécanisme non natif des langues germaniques et parfois rejeté par la grammaire normative. Cependant, avant de tirer des conclusions définitives sur le sujet, examinons d’abord les résultats de l’analyse sémantique de nos données. Nous pourrions ainsi intégrer dans la réflexion la part des emplois ‘objets’ dans les occurrences comptables récoltées, de même qu’examiner l’éventualité d’un mécanisme en néerlandais équivalent au Collecteur. C’est ce que nous faisons au point suivant.

## 6.2.2 Analyse sémantique

Dans notre annotation sémantique des emplois comptables des noms de matériaux en néerlandais, nous avons adopté les mêmes critères que pour nos données en français. Nous avons ainsi distingué d’un côté les emplois en termes de ‘sortes’ (210)-(212), et d’autre part les emplois ‘objets’ (213)-(215).

- (210) *Dit is een zuivere, ijzerhoudende klei die dichtbakt op 1300°C.*  
 ‘C’est une argile ferreuse pure, qui cuit à 1 3000°C.’X
- (211) *Dergelijke cementen worden toegepast in agressieve milieus.*  
 ‘Les ciments de ce type sont appliqués en milieux agressifs.’
- (212) *Alle granieten zijn hun leven begonnen als stukken continentale aardkorst.*  
 ‘Tous les granits ont commencé leur vie en tant que morceaux de croûte terrestre.’
- (213) *Luristan bronzen kenmerken zich door een rijke en opmerkelijk gestileerde decoratie.*  
 ‘Les bronzes du Loristan sont caractérisés par une décoration riche et remarquablement stylisée.’
- (214) *Dansen op een slappe koord met hoogtevrees.*  
 ‘Faire le funambule sur une corde raide avec le vertige.’
- (215) *We hebben in totaal voor beide verdiepingen 40.000 bakstenen nodig.*  
 ‘Nous avons besoin au total de 40 000 briques pour les deux étages.’

Notre distinction supplémentaire entre emplois comptables pluriels individualisés et collectifs ne s’est pas avérée ici très pertinente car, à quelques très rares exceptions près (216)-(219), nous n’avons trouvé aucune occurrence à classer avec certitude du côté du collectif. Les seuls cas qui nous ont fait douter, et que nous avons dès lors bel et bien classés comme collectifs ‘objets’ sont les suivants :

- (216) *in de papieren lopen*  
 ‘coûter cher’
- (217) *Hygiënische en sanitaire papieren.*  
 ‘Papiers hygiéniques et sanitaires.’
- (218) *Heb er al wel dikkere (jas)stoffen gekocht van coupons (...).*  
 ‘Par contre, j’y ai déjà acheté des tissus (pour vestes) plus épais sous forme de coupons.’
- (219) *Ik vind het heerlijk om op zoek te gaan naar stoffen!*  
 ‘J’adore partir à la recherche de tissus !’

Dans (216), nous avons affaire à une expression figée, ce qui explique que la forme au pluriel puisse difficilement être mise en rapport avec une forme au singulier (\**in het papier lopen*, \**in een papier lopen*). Dans (217), on imagine qu’il s’agit de papier toilette et de papier essuie-main, qui se présente normalement sous forme d’une multiplicité de morceaux, feuillets et autres coupons (d’où la mise au pluriel) – sans qu’un exemplaire de tels morceaux puisse pour autant être appelé *één papier* ‘un papier’. Dans (206) et (207), les *stoffen* dont il est question sont bel et bien des morceaux physiques, tangibles, de tissu. Néanmoins, un morceau de tissu n’est pas pour autant \**één stof* (il faudrait préférer la formulation *één stuk stof* ‘un morceau de tissu’) – d’où l’annotation en tant que collectif.

Que nous ne trouvions pas, dans nos données en néerlandais, d’exemples particulièrement convaincants de pluriels collectifs a pour conséquence que l’on peut

difficilement parler, dans cette langue, d'un mécanisme équivalent au Collecteur Universel du français (voir point 6.1.2), qui permet l'emploi au pluriel de certains noms de matériaux, pour désigner une collection indéterminée d'objets dans la matière en question. L'énoncé (220), traduction littérale de (197) ci-avant, n'est effectivement que peu, voire pas acceptable :

- (220) ??*Tegen welke prijs kan je je metalen verkopen bij schroothandelaars ?*  
'A combien peut-on revendre ses métaux chez les ferrailleurs ?'

Il est vrai que l'expression *oude metalen* 'vieux métaux', même si elle n'apparaît pas dans nos données de *WebCorp*, existe pour désigner, avec le même sens que la traduction française, un ensemble de vieux objets en métal dont on pourrait vouloir se débarrasser, comme en attestent les deux énoncés ci-après :

- (221) *Om oude metalen kwijt te raken, moet je niet meer wachten op de ophaalronde van een inzamelaar of naar het recyclagepark rijden. (Google)*  
'Pour se débarrasser de ses vieux métaux, l'on ne doit plus attendre la collecte ou se rendre soi-même au parc à containers.'
- (222) *[Dit bedrijf] biedt een goede service aan loodgieters en dakwerkers door hun oude metalen op te halen. (Google)*  
'Cette entreprise offre un bon service aux plombiers et couvreurs en récupérant leurs vieux métaux.'

Cependant, il s'agit d'un cas isolé (issu peut-être d'un calque à partir du français), et par ailleurs figé (puisque (220), sans la forme adjectivale *oude* 'vieux', fonctionne beaucoup moins bien). Ni nos données, ni notre intuition, ne nous autorisent dès lors ici à reconnaître, comme nous l'avions fait pour le français, l'existence d'un quelconque mécanisme productif en la matière.

Ces précisions étant faites, nous pouvons à présent passer à une nouvelle représentation graphique de nos résultats. La Figure 17 ci-dessous représente les profils distributionnels de nos 34 noms de matériaux en néerlandais, en distinguant cette fois parmi les occurrences comptables les emplois en termes d'objets (en bleu) et ceux en termes de sortes (en orange). Les quelques cas objets collectifs, comme ceux de (216)-(219), sont marqués en bleu clair.

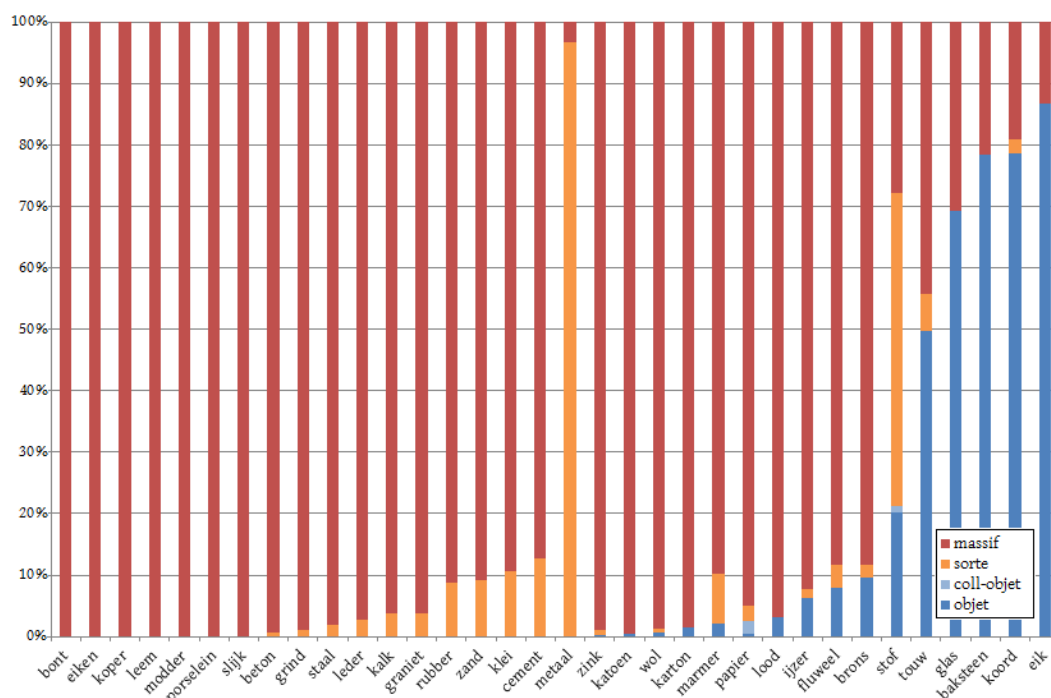


Figure 17 Profils distributionnels de nos 34 noms de matériaux en néerlandais, avec détail sémantique des emplois comptables (lectures en termes de sortes et d'objets).

Comme pour le français, nous pouvons constater qu'une bonne partie des occurrences comptables observées sont à attribuer au mécanisme du Trieur. Ces emplois sont particulièrement représentés dans le profil de *metaal* 'metal' et, dans une moindre mesure, de *stof* 'tissu' – ce qui peut sans doute s'attribuer à leur caractère hypéronymique. La part du Trieur dans ces résultats reste cependant bien moindre que ce que nous pouvons observer en français. Pour cette première langue, les emplois en termes de sortes représentaient 72,6 % du total des occurrences comptables (1 975 cas sur 2 722), et étaient attestés pour chacun de nos items. Ici, nous observons que seuls 532 cas sur 1 262 occurrences comptables (soit 42,1 %) relèvent du Trieur – que nous ne trouvons par ailleurs pas appliqué à tous les items. Ceci confirme ce que nous avançons dans (Lauwers et Vermote 2014, à paraître) (à partir de données plus limitées) et avons exposé au point précédent : que la productivité du Trieur, en tant que mécanisme permettant le passage du massif au comptable pour désigner des sortes, est sensiblement moins productif en néerlandais qu'en français.

Quant aux lectures en termes d'objets spécifiques, elles sont majoritaires pour 4 des items que nous avons étudiés : *eik* 'chêne', *koord* 'corde', *baksteen* 'brique' et *glas* 'verre' – suivi d'assez près de *touw* 'corde' qui flirte avec la limite des 50 % (99 attestations comptables objet sur 199 occurrences, soit 49,7 %). Il est particulièrement intéressant de remarquer qu'il s'agit là des mêmes items qui connaissent aussi en français un usage majoritaire dans cet emploi : *chêne*, *corde*, *brique* et *verre* (la seule exception étant *plomb*, qui complétait le quintet de tête en français, tandis que *lood*, sa traduction, n'apparaît qu'en dixième position). La signification est par ailleurs la même : *een eik* 'un chêne' est

bien un arbre, *een glas* ‘un verre’ un récipient pour boire, etc. De même, pour d’autres items où cet emploi est moins fréquent, le sens ‘objet’ apparaît généralement identique à celui du français : ainsi, à l’instar de leurs équivalents de traduction, *brons* ‘bronze’ (223) et *marmers* ‘marbre’ (224) désignent, dans leurs emplois comptables lexicalisés, des pièces d’art dans ces matières :

- (223) [*Dit museum heeft*] *een rijke collectie aan antieke bronzen opgebouwd.*  
‘Ce musée a constitué une riche collection de bronzes antiques.’
- (224) *Schilderijen, marmers, maskers.*  
‘Des peintures, des marbres, des masques.’

Nous ne nous prononcerons pas avec trop d’assurance sur les raisons de cette similarité, même si la piste des calques de traduction (au moins pour certains de ces cas) n’est pas inenvisageable. D’autre part, certains facteurs pragmatiques et/ou cognitifs se présentent à l’esprit : le fait qu’un chêne, en tant qu’arbre, précède nécessairement le matériau qui en est issu, le fait que les briques soient des objets saillants et distincts (quelle que soit la langue dans laquelle on les nomme), ou encore le lien métonymique (historique) très fort entre *du verre* et *un verre* (qu’on retrouve aussi dans d’autres langues : *glass* > *a glass* en anglais, *Glas* > *ein Glas* en allemand). Notons que néanmoins, la correspondance entre les deux langues n’est pas parfaite : ainsi, si *une corde* peut désigner en français la partie d’un instrument de musique (comme dans *une basse à six cordes*), ni *touw* ni *koord* ne peuvent remplir cette fonction – car la place est déjà prise par le nom *snaar*.

En tous les cas, la conclusion principale concernant les emplois ‘objets’ des noms de matériaux en néerlandais doit être la même qu’en français : au vu de la grande variation des effets de sens possibles et de son caractère imprédictible, il paraît plus judicieux de parler ici d’emplois lexicalisés – plutôt que du résultat d’un hypothétique mécanisme de transfert productif. Par ailleurs, comme pour le français, nous ne prétendons pas que tous ces emplois comptables ‘objets’ seraient, bien que lexicalisés aujourd’hui, le résultat d’un glissement diachronique du massif au comptable. C’est sans doute le cas de certains de ces noms : *glas* ‘verre’ qui a donné *een glas* ‘un verre’, *brons* ‘bronze’ qui a donné *een brons* ‘une statue de bronze’. Dans d’autres cas, en revanche, il va sans dire que s’il faut parler d’un transfert quelconque, c’est l’emploi comptable qui a précédé l’autre – comme il en est certainement de *baksteen* ‘brique’ ou de *eik* ‘chêne’.

Quoi qu’il en soit, puisque le sens des emplois ‘objets’ de nos noms de matériaux en néerlandais ne peuvent, pas plus que ceux du français, se résumer en une formule, la description de leurs significations spécifiques devra également passer par leur énumération. Ainsi, dans le Tableau 2 ci-après, nous citons brièvement pour chacun des noms néerlandais concernés le ou les sens des emplois comptables ‘objets’ tels qu’observés dans nos données. Précisons que si un certain sens paraissait lié à l’emploi

du diminutif, nous l'avons indiqué. En outre, nous avons marqué d'un point d'interrogation les emplois dont le sens ne nous apparaissait pas tout à fait clair.

Tableau 2 Liste des effets de sens comptables en termes d'objets spécifiques des noms de matériaux étudiés en néerlandais.

<i>Eik</i> 'Chêne'	• grand arbre de la famille des fagacées ( <i>een eik</i> 'un chêne')
<i>Koord</i> 'Corde'	• morceau de corde ( <i>de Grand Canyon oversteken op een koord</i> 'traverser le Grand Canyon sur une corde') • emploi métaphorique : lien énergétique ( <i>Energetische Koorden</i> )
<i>Baksteen</i> 'Brique'	• bloc de construction (en maçonnerie, ou dans certains jeux vidéos, comme <i>Minecraft</i> <sup>®</sup> ) • emploi métaphorique ( <i>met een baksteen in de maag geboren</i> 'né avec une brique dans le ventre')
<i>Glas</i> 'Verre'	• récipient pour consommer une boisson • par métonymie, portion d'une boisson ( <i>een goed glas bier</i> 'un bon verre de bière') • [pluriel] vitraux ( <i>De Goudse Glazen</i> )
<i>Touw</i> 'Corde'	• morceau de corde ( <i>twee touwen aan elkaar verbinden</i> 'joindre deux bouts de cordes')
<i>Stof</i> 'Tissu'	• [diminutif] petits morceaux de tissu ( <i>knippen in mijn prachtige stoffes</i> 'découper dans mes magnifiques tissus')
<i>Brons</i> 'Bronze'	• pièce d'art, statue en bronze
<i>Fluweel</i> 'Velours'	• ? ( <i>De Fluweeltjes</i> , nom d'une chorale)
<i>Ijzer</i> 'Fer'	• fer à cheval ( <i>een breed ijzer dat veel steun biedt aan de hoef</i> 'un fer large qui soutient bien le sabot') • emplois métaphoriques ( <i>een heet ijzer</i> 'une tâche difficile, litt. un fer chaud', <i>iemand de ijzers aanleggen</i> 'mettre quelqu'un aux fers')
<i>Lood</i> 'Plomb'	• [diminutif] petits poids utilisé pour la pesée avec une balance à fléau • unité de mesure ( <i>een lood was ±15 gram</i> 'un [plomb] pesait environ 15 grammes') • emploi métaphorique ( <i>de laatste loodjes wegen het zwaarst</i> 'à la queue gît la difficulté, litt. les derniers [petits plombs] sont les plus lourds')
<i>Papier</i> 'Papier'	• morceau de papier ( <i>geef eens een papier, ik wil een stofzuiger tekenen</i> 'passe-moi un bout de papier, je veux dessiner un aspirateur')

Marmer 'Marbre'	• pièces de marbre travaillé
Karton 'Carton'	• boîte en carton • [diminutif] petit morceau de carton
Wol 'Laine'	• [diminutif] vêtement en laine ? ( <i>In de Wolletjes</i> , nom d'un magasin de vêtements en ligne, proposant des habits en laine pour enfants)
Katoen 'Coton'	• [diminutif] pièce de tissus en coton ? ( <i>nette katoentjes (...) in de laatste trends</i> 'des jolis [petits cotons] à la dernière mode')
Zink 'Zinc'	• instrument de musique à vent ( <i>een zink is licht conisch en meestal licht gekromd</i> 'un cornet à bouquin [litt. un zinc] est légèrement conique et généralement courbé')

### 6.3 Conclusions

Ainsi se clôture notre examen du comportement morphosyntaxique des noms de matériaux, en français et en néerlandais. Les éléments de synthèse que nous pouvons retenir sont les suivants :

1. La plupart des noms de matériaux, en français et en néerlandais, sont des noms fondamentalement massifs ;
2. Une grande partie de leurs emplois comptables sont attribuables au mécanisme du Trieur Universel ; ce mécanisme apparaît cependant moins productif en néerlandais qu'en français (au profit de la paraphrase et de la composition), possiblement de par un réflexe normatif, reconnaissant le Trieur comme gallicisme ;
3. Outre le Trieur, le français connaît aussi le mécanisme plus modeste du Collecteur Universel, permettant de désigner en utilisant un nom de matériau au pluriel un ensemble d'objets indéterminés fait dans la matière en question (*des métaux, des verres, des cotons*) ; nous ne trouvons pas d'éléments indiquant l'existence d'un mécanisme équivalent en néerlandais ;
4. Le reste des emplois comptables, où le nom de matériau désigne un objet spécifique, sont pour leur part lexicalisés – c'est-à-dire que leur existence (synchronique) n'est pas attribuable à l'un ou l'autre mécanisme de transfert productif ; il ne sont néanmoins pas aléatoires pour autant, puisqu'on observe une certaine correspondance à ce niveau entre le français et le néerlandais ;

5. Pour certains des noms étudiés (globalement les mêmes en français qu'en néerlandais), l'usage comptable lexicalisé est par ailleurs l'emploi majoritaire ;
6. Dans tous les cas, le sens de l'emploi massif reste, quant à lui, invariablement le même : la présentation du matériau de manière homogène, en tant que matière, substance, de construction.

C'est en particulier ce dernier point que nous devons garder à l'esprit lors de notre modélisation du lexique. Il n'est en effet pas anodin que l'emploi massif d'un nom de matériau soit invariablement cohérent dans l'ensemble de la catégorie, que cet emploi soit l'usage préféré et donc supposément fondamental (par exemple *du granit*, *du sable*, *de l'argile* ou *bont* 'de la fourrure', *eiken* 'du chêne', *koper* 'du cuivre', aux extrêmes gauches de nos graphiques en Figure 15 et Figure 17), ou qu'il paraisse être un emploi dérivé (comme *de la corde*, *de la brique*, *du chêne* ou *touw* 'de la corde', *baksteen* 'de la brique') ou au moins minoritaire (*du verre*, *du plomb* ; *glas* 'du verre'). Ceci repose la question de la notion de transfert : dans la version de base d'un modèle lexicaliste du massif-comptable, il faudrait opérer en la matière une distinction stricte entre deux phénomènes, qualifiant d'un côté les emplois massifs de *ciment* ou *sable* de naturels et attendus, et de l'autre les emplois massifs de *corde* ou *brique* de transférés, voire contextuels. Cela reviendrait cependant, selon nous, à nier à tort le parallèle sémantique réel qui existe tout à la fois entre ces deux types de cas : autant que *du sable*, *du cuivre* ou *du fer*, les syntagmes *de la brique* ou *de la corde*, même s'ils sont moins fréquents (voire même subordonnés aux emplois comptables), dénotent bel et bien un matériau de construction, sans autre forme de restriction. Notre modèle du lexique devra donc pouvoir expliquer à la fois les différences distributionnelles et intuitives déjà capturées par le modèle lexicaliste de base, mais également les similarités entre l'ensemble des emplois massifs des noms de la catégorie. Nous ferons une proposition en ce sens au Chapitre 8. Mais avant cela, il nous reste à traiter les résultats de notre dernière étude de cas : celle des noms de boissons – où nous tirerons d'ailleurs une conclusion similaire à la présente.



# Chapitre 7 Les noms de boissons

## 7.1 Les noms de boissons en français

Comme pour nos deux études de cas précédentes, nous sommes intéressé par la distribution des emplois massifs et comptables d'un ensemble de noms – à savoir ici les noms de boissons. Puisque l'on peut supposer que ces noms désignent d'abord des liquides et sont ainsi *a priori* massifs (*de l'eau, du café, du vin, etc.*), nous sommes en outre particulièrement curieux de la teneur sémantique de leurs emplois comptables ainsi que de leur répartition.

Fort des conclusions de notre étude précédente (noms de matériaux, Chapitre 6), nous pouvons sans doute faire l'économie d'une première présentation purement distributionnelle de nos résultats, et intégrer directement dans notre premier – et dès lors unique – graphique les conclusions de notre analyse sémantique. Ainsi, de manière attendue, les emplois comptables que nous avons observés de nos noms de boissons peuvent être classés en deux groupes. D'une part, nous trouvons les lectures en sortes (225)-(227), l'œuvre de notre Trieur Universel :

- (225) Une dégustation de différents thés d'origine, de couleurs et d'origines différentes.
- (226) Les eaux faiblement minéralisées du côté des eaux plates : Volvic, Evian, [etc.]
- (227) C'est un whiskey de très grande qualité et assez méconnu.

D'autre part, nous trouvons les lectures en portions, où l'un de nos noms est employé en syntaxe comptable pour désigner une consommation standard de la boisson en question : *un café* '~une tasse de café', *une bière* '~un verre de bière' (228)-(231). (C'est ce que l'on a appelé le mécanisme du Conditionneur Universel ((Galmiche 1989), (Wiese et Maling 2005) ; voir point 1.5.2.2).)

- (228) Combien de kirs faut-il engloutir avant de se rendre malade ?
- (229) Un café noir et une cigarette, c'est tout.
- (230) [Ces magazines montrent] des femmes souriant qui tiennent un coca light à la main.

(231) Un café très fort peut empêcher de dormir correctement, un tilleul très fort peut provoquer le sommeil.

Notons que pour *tilleul* (231), nous avons aussi trouvé trois occurrences comptables où le nom ne désignait ni une sorte, ni une portion d'infusion de tilleul, mais bien l'arbre du même nom : celles-ci n'ont pas été incluses dans notre analyse.

Tout ceci étant dit, nous pouvons passer à la représentation graphique de nos résultats. La Figure 18 montre pour la sélection des noms de boissons que nous avons étudiés comment se répartissent leurs emplois massifs (rouge), comptables en termes de sortes (orange), et comptables en termes de portions (bleu). Ils sont classés par ordre croissant de la proportion d'emplois de ce dernier type. Le graphique comporte 36 colonnes, alors que notre sélection ne devait contenir que 35 items (voir point 4.3.2), parce que nous y présentons *sherry* et *cherry* comme deux items différents. Nous les avons considérés d'abord comme simplement deux variantes orthographiques d'une même entité, mais avons pu constater au travers de l'analyse de nos occurrences de corpus (de même qu'une vérification dans l'un ou l'autre ouvrage spécialisé) qu'il s'agissait là d'une assimilation abusive – le *cherry* (ou *cherry brandy*) étant une liqueur élaborée à partir de cerises macérées dans de l'eau-de-vie, tandis que le *sherry* (ou *xérés*) est un vin espagnol. (Nous opérons cette distinction non pas qu'elle ait une grande influence sur nos résultats, mais par souci de complétude. Nous voilà, du reste, un peu plus instruit.)

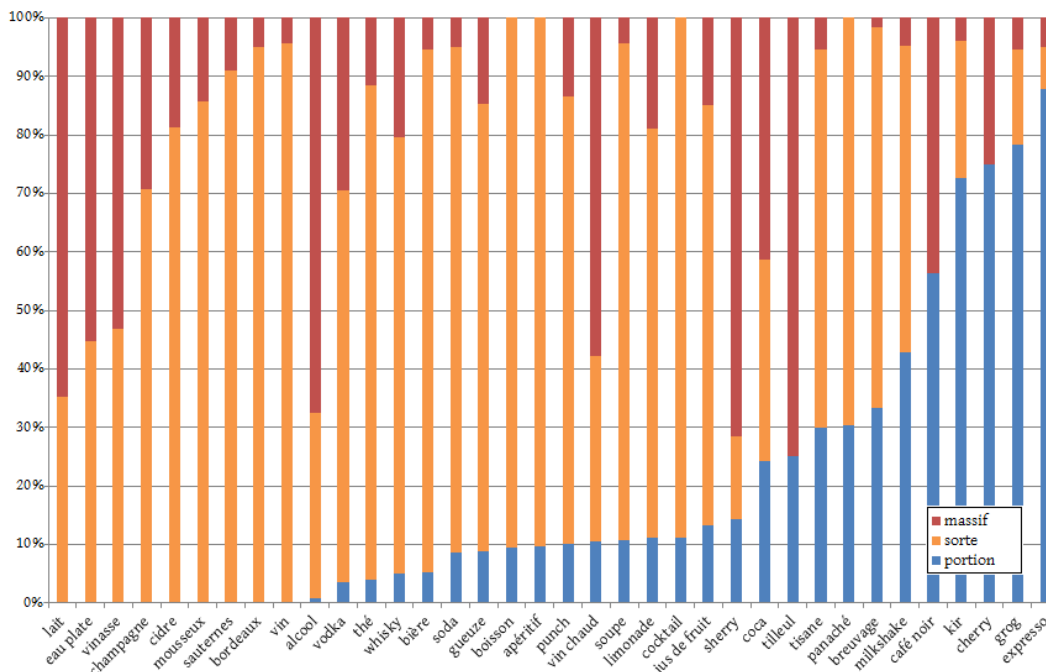


Figure 18 : Profils distributionnels de notre sélection de 36 noms de boissons, en français.

Sur la base de cette représentation, nous pouvons tirer quelques constats. A nouveau, comme au chapitre précédent, nous observons que le mécanisme du Trieur Universel

est omniprésent : pas moins de 73,1 % de nos occurrences de noms de boissons (2 860 cas sur 3 915) sont des emplois comptables en termes de sortes. Ici, nous avons néanmoins trois items pour lesquels il n'est pas attesté dans nos données : *cherry*, *tilleul* et *café noir*. Pour les deux premiers, l'explication est en termes de nombre d'occurrences, car à l'exception de quatre attestations dans les deux cas, tous les exemples que nous trouvons de ces noms dans nos données sont en contexte neutre. (Ainsi, si l'on voit que *cherry* apparaît dans 25 % des cas comme massif, il s'agit en fait d'une unique occurrence massive, (232), sur quatre énoncés.)

(232) Je recherche [à] acheter du cherry mar[ni]er et je ne trouve pas.

Pour ce qui est de *café noir*, l'explication est différente : à notre sens, si l'on n'en trouve pas, dans nos données, d'exemples en emploi comptable 'sortes', c'est parce que le café noir est déjà lui-même une 'sorte de café'. D'un point de vue pragmatique, l'établissement de sous-types à un niveau taxonomique qui serait encore inférieur ne se justifie probablement pas souvent – d'où l'absence d'attestations. Ceci ne remet évidemment nullement en doute le mécanisme du Trieur, qui reste d'ailleurs applicable à *café noir*, moyennant l'invention d'un contexte adéquat :

(233) Mon bistrot préféré sert un café noir de la plus grande qualité.

Les autres emplois comptables de nos noms de boissons découlent du Conditionneur, et sont donc des lectures en portions. Ici, si l'on ne doit pas nécessairement refuser à ce mécanisme une certaine productivité, force est de constater que son applicabilité paraît bien plus limitée que celle du Trieur. Ainsi, nous n'en trouvons aucune attestation pour 9 de nos noms, et il n'est responsable de plus de 50 % des occurrences analysées que pour 5 items sur 36. Il semblerait donc que le Conditionneur soit lié à des restrictions spécifiques.

Celles-ci sont, à notre sens, principalement référentielles – et couplées à un certain degré de conventionalité. Nous entendons par là qu'un emploi comptable en 'portions standard' présuppose avant tout l'existence d'une pareille portion, de même que la fonctionnalité de l'expression transférée. Ainsi, d'une part, si l'on n'en trouve pas d'attestations pour *lait*, par exemple, c'est fort probablement parce que le lait est une boisson qui se consomme sous une grande variété de formes (dans un verre, une tasse, un bol), et ne sera du reste pas typiquement trouvée sur une carte de bistrot ou de restaurant, parmi les autres consommations habituelles.

D'autre part, la situation est sensiblement différente, mais néanmoins du même ordre, pour *vin* et ses proches parents *champagne*, *bordeaux*, *cidre*, *mousseux* ou encore *sauternes*. Ceux-ci font bien partie des boissons qu'on pourra trouver couramment au bar de n'importe quel café, hôtel ou restaurant, et se présentent donc bel et bien sous forme de portions standardisées – mais pourtant, s'il l'on pourra commander *une bière* ou *un coca*, l'on ne demandera pas aussi facilement au serveur *un champagne* ou *un cidre*. Dans

(Lauwers et Vermote 2014, à paraître), nous avons proposé que cette divergence découlait elle aussi de propriétés référentielles (traduites, voire lexicalisées en langue) : contrairement à la bière ou au coca, dont les portions standard consommées le sont tantôt dans un verre, tantôt dans une bouteille, tantôt même dans une canette en métal, le *vin* et ses parents sont toujours présentés dans un verre spécifique, rendant les lectures transférées *un vin* ou *un champagne* entièrement équivalentes aux paraphrases respectives *un verre de vin*, et *une coupe de champagne* – et donc redondantes. Si pour *lait*, l'absence de lectures conditionnées découle donc de l'indisponibilité référentielle d'une portion standardisée de lait, pour *vin*, *champagne*, *cidre*, *mousseux*, *bordeaux* et *sauternes*, elle découlerait de l'existence d'une autre expression, équivalente, rendant la lecture transférée superflue.

Si l'on se tourne à présent vers le côté droit de notre graphique, l'on trouvera un autre groupe de noms intéressants, dont nous observons une grande proportion d'occurrences en lecture 'portions' : *expresso*, *grog*, *kir*, *café noir* (avec plus de 50 % ; nous passons *cherry*, dont le profil est basé sur 4 occurrences, cf. *supra*, et n'est donc pas particulièrement fiable) – suivis de *milkshake*, *breuvage*, *panaché* et *tisane* (avec effectivement moins de 50 % d'occurrences 'portions', mais d'autre part très peu, voire aucune, occurrences massives dans nos données). L'on remarquera la mise en contraste intéressante entre *bière* et *panaché* dans (222) :

(234) Peu importe que tu boives de la bière, des panachés, du vin ou des spiritueux.

Or, la particularité de ces noms nous paraît en effet qu'ils désignent en premier lieu des portions. Ainsi, *un kir* est 'un verre de vin blanc avec de la crème de cassis', *un expresso* est 'une portion réduite de café très fort', *un milkshake* est 'un mélange de lait, de glace, de fruits et/ou de sirop passés au mixer', etc. Pour ces différents noms, c'est en réalité la portion qui définit la boisson, et l'emploi comptable qui serait dès lors l'emploi fondamental. Ils s'opposent en cela aux items que l'on retrouve dans la portion centrale de notre graphique (notamment *vodka*, *thé*, *whisky*, *bière*, *soda*, *gueuze*, *punch*, *vin chaud*, *soupe*, *limonade*, *jus de fruit*, *sherry* et *coca*), et qui semblent quant à eux bien répondre au profil typique des noms de boissons flexibles : un emploi massif de base (*du thé*, *de la gueuze*, *du coca*), avec possibilité de 'conditionnement' par transfert (*un thé*, *une gueuze*, *un coca*).

Cela n'empêche que les items comme *expresso*, *grog*, *kir* ou encore *milkshake* peuvent tout autant apparaître en syntaxe massive (235)-(238). Dans ce cas, leur sens n'est pas à distinguer des emplois massifs des autres noms de la catégorie : ils désignent alors simplement la boisson en profilant sa substance. (Trois items notables n'apparaissent néanmoins jamais en emploi massif dans nos données : *cocktail*, *apéritif* et *boisson*. Ceci s'explique assurément par leur caractère hypéronymique, responsable de la prépondérance des emplois taxonomiques.)

- (235) *De l'expresso* est par la suite versé doucement afin qu'il se place entre le lait et la mousse.
- (236) Mettre un chapeau au bout du lit et boire *du grog* jusqu'à voir quatre chapeaux.
- (237) Quand les bistrots se décideront-ils à faire *du kir* de bonne qualité ?
- (238) Il faut bien la secouer la vache si on veut avoir *du bon milkshake* !

Ainsi, *de la bière* et *de l'expresso* sont tous deux des syntagmes acceptables, avec un sens commun à attribuer au massif, mais néanmoins une forte intuition (confirmée par nos données de corpus) que dans le premier cas, il s'agit d'un emploi de base, et dans le second, d'un emploi dérivé. L'on remarquera ici un parallèle indéniable entre cette situation et celle que nous avons observée pour les noms de matériaux, où nous avons dit quelque chose de très similaire du contraste entre, par exemple, *du béton* et *de la brique*.

Ainsi, l'on voit que l'on est confronté ici à la même remise en question de la vision lexicaliste traditionnelle de l'opposition massif-comptable, évoquée dans notre conclusion du chapitre précédent. S'obliger à opérer une distinction stricte au sein de nos noms de boissons entre ceux qui seraient fondamentalement comptables, et ceux qui seraient fondamentalement massifs, et expliquer dès lors les emplois massifs des premiers et comptables des seconds en termes de mécanismes de transferts inverses, revient à perdre le parallèle sémantique certain entre *de la bière* et *de l'expresso* d'une part, et *une bière* et *un expresso* de l'autre. Nos différents items ont bien des particularités et préférences propres, découlant éventuellement de propriétés référentielles, mais leurs différents emplois (massif, portion, sorte) n'en ont pas moins systématiquement le même sens, indépendamment de leur caractère plus ou moins habituel (l'indisponibilité de certains emplois découlant quant à elle de facteurs extérieurs, comme la préemption de *\*un vin* par *un verre de vin*). Le modèle que nous proposerons de l'opposition massif-comptable dans le lexique devra pouvoir rendre compte de ces différents faits.

## 7.2 Les noms de boissons en néerlandais

Nous terminerons notre parcours d'études de corpus par l'examen du comportement sémantique et morphosyntaxique des noms de boissons en néerlandais. Comme dans la section précédente, nous présenterons d'emblée nos résultats en tenant compte de l'annotation sémantique des emplois comptables. Les possibilités quant à ceux-ci sont, sans surprise, les mêmes que pour le français. Ainsi, l'on retrouve d'une part des emplois en termes de 'sortes', attribuables au mécanisme du Trieur (239)-(241), et d'autre part des usages en termes de portions (242)-(246).

- (239) *Een goede whisky kan wel 30 jaar opgeslagen worden.*  
 ‘Un bon whisky se conserve bien 30 ans.’
- (240) *Eén van de meest populaire kruidentheeën is rooibosthee.*  
 ‘L’une des infusions les plus populaires est le thé rouge.’
- (241) *Tot de klassieke aperitieven behoren de Champagne, de Crémant de Bourgogne en andere mousserende wijnen.*  
 ‘Parmi les apéritifs classiques, on retrouve le Champagne, le Crémant de Bourgogne, et d’autres vins mousseux.’

Dans l’exemple (240), l’on trouve même deux attestations d’emplois taxonomiques : *aperitieven* ‘apéritifs’ et *wijnen* ‘vins’. Notons que l’énoncé est tiré des données retournées par *WebCorp* pour la recherche concernant *aperitief*.

- (242) *Veel consumenten geven de voorkeur aan een getapt bier.*  
 ‘Beaucoup de consommateurs préfèrent une bière pression.’
- (243) *Na de cursus bleef de bende in de bar nog cocktails drinken.*  
 ‘Après le cours, la bande est restée au bar pour boire des cocktails.’
- (244) *Voor mij een plat water, a.u.b.*  
 ‘Pour moi, une eau plate, s’il-vous-plaît.’
- (245) *Capsules voor een klein, krachtig koffietje met een fluweelzachte structuur.*  
 ‘Des capsules pour un petit café puissant, avec un structure de velours.’
- (246) *Het probleem bij een geopend wijntje is oxidatie.*  
 ‘Le problème d’une bouteille de vin ouverte, c’est l’oxydation.’

Notons que ce dernier exemple (246) est particulièrement intéressant, car l’emploi comptable de *wijntje* ‘litt. petit vin’ ne désigne pas, comme dans tous les autres cas que nous avons classés du côté du Conditionneur, une portion standard de vin, mais plutôt une bouteille – comme en témoigne la présence de l’adjectif : ça ne peut en effet être qu’une bouteille, et non un verre éventuel, qui soit ‘ouvert’.

La Figure 19 ci-après présente le résultat de notre analyse distributionnelle, selon les mêmes conventions de coloration que tous nos graphiques précédents. Notons que la proportion, parmi les emplois ‘portions’, des occurrences avec le diminutif sont marquées en bleu plus clair. Nous ne trouvons du reste aucun emploi avec le diminutif pour les occurrences taxonomiques (ce qui ne doit pas nous surprendre, vu le caractère singulatif de ce morphème en néerlandais).

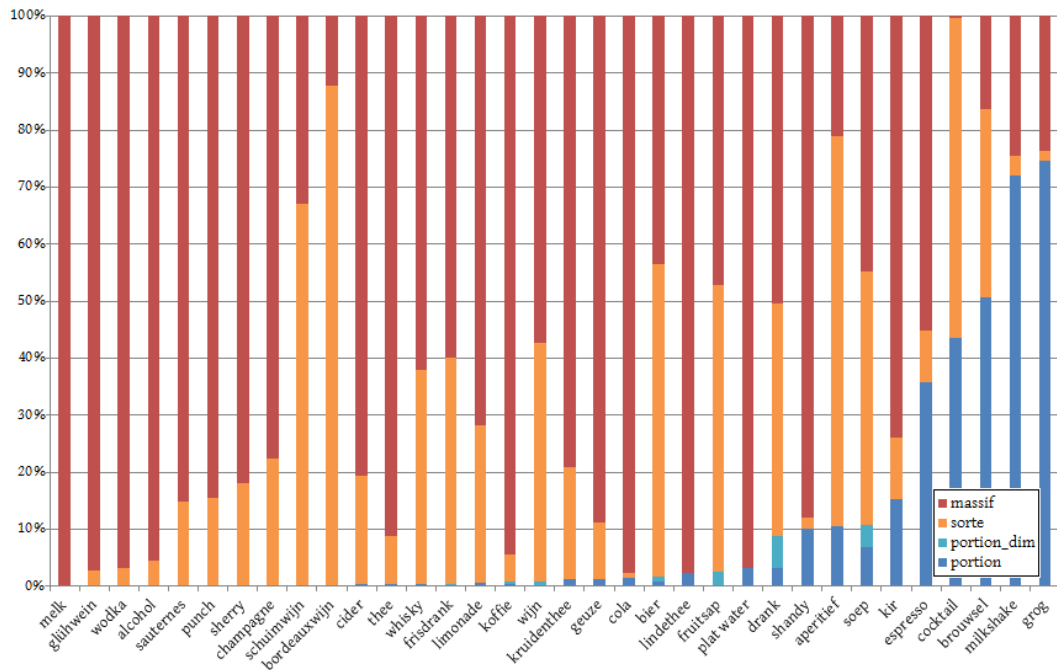


Figure 19 Profils distributionnels de notre sélection de noms de boissons, en néerlandais.

Comme c'était le cas pour les noms de matériaux, l'on observe très clairement une part bien moins importante en néerlandais qu'en français d'emplois taxonomiques (en orange dans le graphique). Si pour le français, ces emplois correspondaient à 73,1 % du total des occurrences examinées (c'est-à-dire de la somme des emplois massifs et comptables), leur proportion ne dépasse pas ici les 25 % (2 392 occurrences sur 9 572). Ceci confirme ce que nous disions au Chapitre 6 : qu'apparemment, le néerlandais n'est pas particulièrement friand du mécanisme du Trieur – considéré comme gallicisme dans la tradition normative, et concurrencé par les alternatives de la paraphrase (*soorten wijn* 'des sortes de vin') et de la composition (*biersoorten* 'sortes de bière(s)'), voir point 6.2.1. Certains items présentent néanmoins une part d'emplois en 'sortes' plus importante que les autres. C'est le cas par exemple de *aperitief* 'apéritif' et *cocktail* 'cocktail', ce qui s'explique sans nul doute par leur caractère superordonné – comme en témoignent les emplois attributifs suivants :

- (247) *Een gin tonic is bijvoorbeeld een mooi aperitief.*  
'Un gin tonic est par exemple un bel apéritif.'
- (248) *De Margarita is een Mexicaanse cocktail gebaseerd op tequila.*  
'La Margarita est un cocktail mexicain à base de téquila.'

Le cas de *drank* 'boisson' est un peu particulier, car s'il présente bel et bien un nombre non négligeable d'emplois taxonomiques (249), on le retrouve néanmoins dans 50,5 % des cas en emploi massif (162 occurrences sur 321) – tandis que nous n'avons aucune attestation massive de son équivalent français *boisson*. Cette différence découle d'un contraste bien connu entre les deux langues, *drank* étant notablement massif dans la plupart des cas où en français, un emplois pluriel de *boisson* est requis. Ainsi, sur une

invitation de soirée, l'on pourrait lire (250) ou (252) en néerlandais, avec le massif, alors que les traductions françaises en (251) et (253) nécessitent le pluriel – le massif paraissant bel et bien exclu.

(249) *Een alcoholische drank is een drank die alcohol (...) bevat.*  
'Une boisson alcoolisée est une boisson qui contient de l'alcool.'

(250) *Drank en hapjes zijn voorzien.*

(251) Des boissons et collations sont prévues.'

(252) *Prijs: €5, drank inbegrepen.*

(253) Prix : €5, boissons comprises.

Il s'agit là d'un contraste idiosyncratique, à ranger dans la même catégorie que la différence entre *raisin* (flexible, mais avec une certaine affinité pour le massif : *du raisin*, voir Chapitre 5) et *druif* (comptable, et très peu ouvert à l'emploi massif). En outre, le massif *drank* connaît en néerlandais un emploi euphémistique au sens de 'boisson(s) alcoolisée', qui vient gonfler la proportion de ses emplois massifs dans nos données (254) – alors que l'emploi équivalent en français, s'il existe, donnerait plutôt lieu à des usages neutres (*Son attachement à la boisson a eu raison de sa santé*, mais pas aussi facilement ? *Ce supermarché est coupable d'avoir vendu de la boisson aux moins de 16 ans*, alors que (254), que nous tirons de notre corpus, marche sans problème en néerlandais).

(254) *[Een supermarkt sluiten] als die supermarkt voor de derde keer in één jaar drank verkoopt aan iemand die jonger is dan 16 jaar.*  
'[Fermer un supermarché] si celui-ci vend pour la troisième fois en un an de l'alcool (litt. de la boisson) à quelqu'un qui a moins de 16 ans.'

A côté de cela, l'on remarquera aussi dans le tiers gauche de notre graphique une proportion frappante d'emplois taxonomiques pour les items *schuimwijn* 'mousseux' et *bordeauxwijn* '(vin de) [B]ordeaux' (supérieure à 50 % des occurrences examinées, dans les deux cas). Celle-ci s'explique selon nous par la double propriété de ces deux noms d'être d'une part déjà le résultat d'une composition (avec *-wijn*), et de l'autre de dénoter des concepts d'emblée subordonnés (*bordeauxwijn is een (soort) wijn uit de streek Bordeaux* 'le bordeaux est un vin de la région de Bordeaux') – rendant l'option alternative de la composition avec *-soort* exclue (*\*schuimwijnsoort*, *\*bordeauxwijnsoort* – une rapide recherche sur *WebCorp* d'occurrences de ces deux formes ne donne effectivement aucun résultat), et la paraphrase avec *soort(en)* curieuse (nous en trouvons peu d'occurrences sur *WebCorp* : 0 et 1 pour respectivement *soort* et *soorten bordeauxwijn*, 4 et 5 pour *soort* et *soorten schuimwijn* ; à titre de comparaison, la recherche de *soorten bier* donne d'emblée 89 occurrences). Ainsi, pour ces deux noms en particulier, de par leurs propriétés morphologiques et taxonomiques, l'option de l'emploi comptable transféré pour exprimer des 'sortes' est moins concurrencé par les alternatives habituelles, ce qui



justifie qu'on en retrouve un nombre inhabituellement élevé d'occurrences dans nos données.

Concernant les lectures conditionnées, force est de constater qu'ici aussi, les noms de boissons en néerlandais présentent une bien moins grande flexibilité que leurs équivalents français. Nous n'en trouvons une proportion significative d'occurrences que pour une dizaine d'items (en partant de la droite de notre graphique), tandis que pour les noms avec les attestations les moins nombreuses de pareils emplois, les quelques cas identifiés sont parfois même douteux. Ainsi, nous avons classé (255) et (256) du côté des emplois conditionnés :

- (255) *Geen zin in bier of wijn? Dan zijn ciders de ideale drankjes (...)!  
'Pas envie de bière ou de vin ? Les cidres sont alors des boissons idéales !'*
- (256) *Cannabis kan eenvoudig worden toegediend in de vorm van een thee.  
'Le cannabis peut simplement être administré sous forme d'un thé.'*

Pour (255), nous nous basons en cela sur la relation attributive entre *ciders* 'cidres' et *drankjes* 'boissons' – interprétant ce dernier nom au sens de 'consommations'. Néanmoins, il pourrait éventuellement s'agir malgré tout d'une lecture taxonomique, s'il était question de plusieurs sortes de cidres (ce qui nous paraît difficile de trancher). De la même manière, (256) est potentiellement ambigu : si l'on administre bien la cannabis sous forme d'une 'tasse de thé', l'on crée ce faisant tout autant 'une sorte de thé au cannabis'. Pour d'autres items, comme *cola* 'coca', les lectures conditionnées sont plus claires (257) – mais néanmoins, la prédominance des emplois massifs, autrement plus nombreux qu'en français, est frappante.

- (257) *Dagelijks gaan er niet minder dan 600 miljoen Cola's over de toonbank.  
'Il se vend quotidiennement pas moins de 600 millions de cocas.'*

Cependant, ces grandes disparités du point de vue distributionnel entre l'usage des noms de boissons en français et en néerlandais n'empêche pas que nous puissions identifier malgré tout certaines similarités. Ainsi, à l'extrémité droite de notre Figure 19, nous trouvons *grosso modo* les équivalents de ceux que nous y trouvons à la Figure 18 : *grog*, *milkshake*, *brouwsel* 'breuvage' – et à peine plus loin *espresso* 'expresso' et *kir*. Effectivement, ces boissons sont caractérisées par le fait de désigner avant tout des portions, voir ci-avant. L'on en trouve néanmoins également des attestations massives, dont le sens n'est pas à distinguer fondamentalement des emplois massifs des autres noms de la catégorie. (L'on remarquera au passage l'emploi taxonomique de *rum* 'rhum' dans (259).)

- (258) *McDonald's NL moet in ELK filiaal chocolade milkshake verkopen.  
'McDonald's NL doit vendre du milkshake au chocolat dans CHACUNE de ses filiales.'*
- (259) *Grog kun je natuurlijk ook met een andere (donkere) rum drinken.  
'Le grog peut se boire aussi avec un autre rhum (sombre).'*

Pour *espresso* et *kir*, contrairement aux équivalents français, il semblerait même que ce soit bel et bien cet emploi massif qui soit l'emploi dominant. En témoignent d'une part des attestations telles que *een kopje espresso* 'un espresso (litt. une tasse d'espresso)' ou *een glas kir* 'un kir (litt. un verre de kir)', de même que des occurrences génériques en emploi nu, comme (260) ou (262) – alors que le français opterait probablement pour l'article indéfini comptable *un* (261) et (263) ou l'article défini, neutre quant à notre opposition (263).

- (260) *Espresso is een bereiding waarbij heet water onder hoge druk door zeer fijn gemalen en aangedrukte koffie wordt geperst.*
- (261) Un espresso est une préparation issue du passage sous haute pression d'eau bouillante au travers de café finement moulu et compressé.
- (262) *Kir is een cocktail (...).*
- (263) Le/Un kir est un cocktail.

D'autre part, l'on retrouve à l'extrémité gauche de notre graphique à nouveau les mêmes noms qu'en français : *melk* 'lait', *glühwein* 'vin chaud', *sauternes*, *champagne*, *schuimwijn* 'mousseux', *bordeauxwijn* 'bordeaux' – tandis que *cider* 'cidre' et *wijn* 'vin' ne sont pas loin (pour ces deux items, les uniques occurrences conditionnées repérées sont particulières, comme nous l'avons vu : relativement ambiguë pour *cider* (255) et dénotant une bouteille plutôt qu'un verre pour *wijn* (246)). Nos explications du point précédent concernant les restrictions sur les emplois conditionnés de ces items pourraient donc s'appliquer ici aussi. Si l'on ne boit jamais, faute de référent typique, \**un lait*, l'on ne boira pas plus \**een melk*. (On désigne parfois comme *een melkje* une capsule de lait telle qu'on en sert en accompagnement d'un café à emporter. Nous n'avons néanmoins pas trouvé d'attestations de cet emploi dans nos données.)

L'on pourrait se demander néanmoins si un verre ou autre portion de lait consommé comme tel pourrait se désigner par l'expression *een melkje*. Nous trouvons en tout cas sur *WebCorp*, non pas dans nos données de base mais en effectuant une rapide recherche supplémentaire de ce syntagme, des énoncés tels que (250) et (251) :

- (264) *Kinderen die dit willen kunnen in de voormiddag water of een melkje drinken.*  
'Les enfants qui le souhaitent peuvent boire dans la matinée de l'eau ou une petite brique de lait (litt. un petit lait).'
- (265) *Vergeet je kindje dus zeker niet een melkje mee te geven naar school.*  
'N'oubliez donc certainement pas de donner à votre enfant une petite brique de lait (litt. un petit lait) pour aller à l'école.'

Ceci nous amène à la question du diminutif qui, comme nous l'avons dit (voir point 1.5.2.2) revêt notamment en néerlandais la fonction de singulatif : *snoep* (massif) 'des friandises' vs. *een snoepje* 'un bonbon', etc. Si, au vu de nos données distributionnelles, le mécanisme du conditionnement sémantique par transfert du massif au comptable ne semble pas particulièrement productif (en tous cas, en comparaison avec le français), le

néerlandais a ainsi par contre un autre mécanisme pour exprimer des portions, à savoir le diminutif. Malheureusement, nous pouvons difficilement tirer des conclusions à ce sujet à partir de nos données de corpus, car les formes diminutives y sont largement sous-représentées. Ceci est sans nul doute dû à leur fréquence très basse, comparé aux formes de base : de ce fait, puisque nous avons recherché pour chaque item en même temps toutes ses formes (au singulier et au pluriel, avec ou sans morphème diminutif, voir point 4.2.1), les formes avec diminutif se retrouvent forcément en sous-nombre dans la sélection d'occurrences finale.

Dès lors, il peut être intéressant de constater que les items pour lesquels nous en trouvons néanmoins quelques attestations dans nos données désignent tous des boissons très courantes, consommées régulièrement, et effectivement sous forme de portions standardisées : *een soepje* 'une soupe', *een biertje* 'une bière', *een fruitsapje* 'un jus de fruit', *een frisdrankje* 'un soda', *een koffietje* 'un café'. A ceux-ci s'ajoute *drank* 'boisson', dont la nature bien plus massive que son homologue francophone (comme nous venons de le dire) implique lui aussi dans la plupart des cas l'usage du diminutif pour désigner une portion : *een drankje* 'une boisson, une consommation'.

Comme pour le Trieur, la productivité moindre du Conditionneur en néerlandais peut donc être expliquée, au moins en partie, par l'existence d'une alternative tout autant, voire plus encore, productive – totalement absente en français – à savoir l'emploi du diminutif. De manière intéressante, l'on peut noter cependant que le rapport de force entre les deux alternatives (emploi comptable transféré et diminutif) semble varier d'un item à l'autre (ce ne sont pas nos données qui nous permettent de faire cette observation, mais nous nous basons sur (Bakema 1998) ; voir aussi (Lauwers et Vermote 2014, à paraître)) : pour la lecture en 'portions', l'emploi du diminutif sera tantôt obligatoire (*een biertje*, et non *\*een bier*), tantôt nettement préférable (*een soepje* vs. *?een soep*), tantôt superflu (*een colaatje* mais aussi *een cola*, *een koffietje* mais aussi *een koffie*) – avec dans ce dernier cas le retour d'une nuance hypocoristique, affective. Ceci aussi, nous devons pouvoir l'expliquer dans notre modèle du lexique.

### 7.3 Conclusions

Nous terminons ici notre étude du comportement morphosyntaxique des noms de boissons – et par le même coup, notre série d'études de corpus. Pour ce troisième et dernier domaine, les conclusions de synthèse que nous retiendrons sont :

1. Les noms de boissons que nous avons étudiés sont employés globalement de trois manières : en syntaxe massive, en lecture taxonomique, et en emploi comptable

en termes de ‘portions’ ; les items varient néanmoins quant aux options pour lesquelles ils sont disponibles (certains emplois étant bloqués par des facteurs référentiels ou pragmatiques), ainsi que pour ce qui est de leur emploi de base (emploi taxonomique pour les noms superordonnés, emploi conditionné pour les boissons comme *expresso* ou *grog*, emploi massif pour les autres) ;

2. En général, les mécanismes de transfert comme le Trieur ou le Conditionneur, supposant l’emploi comptable d’un nom *a priori* massif, sont moins productifs en néerlandais – qui possède des options alternatives pour exprimer les mêmes nuances sémantiques (composition et paraphrase avec (-)soorten ; emploi du diminutif) ;
3. Quel que soit le profil de tel ou tel nom, le sens associé à chacun des emplois possibles est éminemment cohérent d’un item à l’autre.

Ces conclusions rappellent celles que nous avons dressées à la fin du chapitre précédent, et confirment donc ce que nous y disions : que nous devons construire un modèle de l’opposition massif-comptable dans le lexique capable d’expliquer à la fois l’existence pour chaque item individuel de préférences propres quant à l’emploi morphosyntaxique, de même qu’au-delà de ces préférences, une cohérence de classe quant aux sens associés à chaque emploi possible. C’est ce que nous entreprenons ci-après, dans la troisième partie du présent travail.

## Partie 3 : Vers une modélisation

Nos trois études de corpus, présentées en détails dans la partie précédente, nous ont permis d'établir de nombreuses caractéristiques de l'opposition massif-comptable, et de l'expression de sa flexibilité – ouvrant ainsi des pistes de réflexion quant à sa modélisation. C'est ce dernier point qui sera l'objet de cette troisième partie : la construction d'un modèle de l'opposition massif-comptable dans le lexique, à même d'expliquer nos observations, et tenant compte des acquis de la tradition psycholinguistique en la matière.

Ainsi, dans la partie présente, nous commencerons par tracer les grandes lignes d'un pareil modèle, en montrant l'intérêt d'une représentation sur deux niveaux – lexical et supra-lexical (Chapitre 8). Ensuite, nous étayerons notre proposition en testant certaines de ses prédictions au travers de diverses enquêtes d'acceptabilité (Chapitre 9), ainsi que d'une expérience de production (Chapitre 10). Pour finir, nous examinerons quelques questions tenant à la structure du niveau sémantique de notre modèle (Chapitre 11).



# Chapitre 8 La double représentation de l'opposition massif/comptable

## 8.1 Un modèle à deux niveaux

Les trois chapitres précédents, détaillant nos études de cas en corpus, sont parsemés d'indications quant à la nature d'une modélisation appropriée du massif-comptable dans le lexique, et des faits empiriques dont celle-ci se doit de pouvoir rendre compte. Ainsi, nous rappelons ci-dessous dans les grandes lignes ce que nous avons observé, et ce que nous devons dès lors expliquer :

1. Pour les noms de fruits et de légumes, la flexibilité est un fait : d'une manière ou d'une autre, ces items doivent disposer d'un double marquage lexical massif et comptable (à quelques exceptions près, notamment *haricot*, *boon* 'haricot' ou encore *druif* 'raisin') ; néanmoins, cette flexibilité n'est pas aléatoire, puisque d'une part liée à des facteurs situationnels spécifiques (notamment des contextes plus propices au massif), et d'autre part modulée par des préférences individuelles par item (les uns étant plus friands du massif que les autres) ;
2. Pour les noms de matériaux, il existe un emploi massif commun à tous les items en termes de 'matière, substance de construction', cohérent dans l'ensemble de la catégorie, qu'il s'agisse de l'emploi le plus fréquent et fondamental (c'est le cas pour la plupart des items : *du sable*, *du béton*, *du ciment*, etc.), ou qu'il apparaisse en quelque façon dérivé (comme pour *de la corde* ou *de la brique*) ; certains noms de matériaux possèdent en outre un sens spécifique idiosyncratique, en termes d'objet (*un bronze* 'une statue en métal', *un verre* 'un récipient'), difficilement attribuable à un transfert sémantique créatif en situation, et qui nécessiteront dès lors un statut lexical propre ;
3. Pour les noms de boissons, trois emplois peuvent systématiquement être envisagés : l'emploi massif, l'emploi comptable taxonomique, et l'emploi comptable conditionné ; leur sens est invariablement cohérent dans l'ensemble

de la classe, cependant, chaque item étudié a un rapport différent à ces trois emplois, tant quant à celui servant de base pour les autres (le massif pour *bière*, le comptable conditionné pour *expresso* ou encore l'usage taxonomique pour *boisson*), qu'aux restrictions empêchant éventuellement l'un ou l'autre usage (l'absence de lecture conditionnée pour *lait* ou *vin*, par exemple) ;

4. Globalement, le néerlandais se montre moins flexible quant à l'alternance entre usage massif et comptable, car il dispose pour les effets de sens qui y sont associés de mécanismes alternatifs (comme la composition avec *-soort(en)* pour le Trieur, et le diminutif pour le Conditionneur) ; le cas échéant, la préférence accordée à l'un ou l'autre mécanisme n'est cependant pas fixe, et varie par item (le diminutif étant par exemple indispensable pour la lecture en portion de *bier – een biertje* – mais optionnel pour *koffie – een koffie* ou *een koffietje*).

Pour rendre compte de ces observations, nous avons besoin d'un modèle du lexique qui permette crucialement : d'une part, d'encoder les préférences morphosyntaxiques propres à chaque nom (comme fonction de la fréquence à laquelle les différents emplois sont attestés, c'est-à-dire rencontrés dans l'usage, pour chaque item), et d'autre part, d'expliquer le lien entre flexibilité et contexte d'usage (et plus précisément, que certains sens sont invariablement associés à l'un ou l'autre type de syntaxe, indépendamment des préférences propres des items auxquels ils s'appliquent : le sens fondamental du massif est le même dans *du ciment* et *de la corde*, ou dans *du café* et *du kir*).

Dès lors, il nous semble indispensable de construire un modèle du lexique permettant une représentation de l'opposition massif-comptable à deux niveaux : d'abord au niveau proprement lexical, sous forme de préférences distributionnelles (chaque nom est plus ou moins massif et plus ou moins comptable, en fonction de la manière dont il est couramment produit en discours), et ensuite à un niveau supra-lexical, responsable de l'association entre certains sens et contextes, et l'un ou l'autre type de réalisation syntaxique (il existe un lien direct entre certains aspects de sens, et la syntaxe massive ou comptable). Un tel modèle peut parfaitement entrer dans le moule des représentations stratifiées du lexique, tel qu'on les trouve dans la tradition psycholinguistique, et notamment chez (Levelt *et al.* 1999) (voir point 3.2). Nous proposons dès lors de concevoir la syntaxe massive et comptable comme des nodules combinatoires partagés, au niveau de la strate lemmatique (à l'instar, par exemple, des deux types de constructions datives en anglais, voir (Pickering et Branigan 1998), et point 3.2.2 ci-dessus). Ces deux nodules seraient reliés d'un côté à l'intérieur du niveau lemmatique aux lemmes nominaux, par des connexions plus ou moins fortes en fonction que tel ou tel lemme apparaisse plus ou moins souvent dans l'un et l'autre type de syntaxe (représentation lexicale du massif-comptable), et de l'autre en direction du niveau sémantique supérieur, à certains nodules conceptuels, constituant des aspects de sens, eux-mêmes partagés et reliés à des concepts lexicaux plus centraux



(représentation supra-lexicale du massif-comptable). Nous proposons une première représentation, abstraite, de cette manière de modéliser le lexique, et notre opposition morphosyntaxique, à la Figure 20.

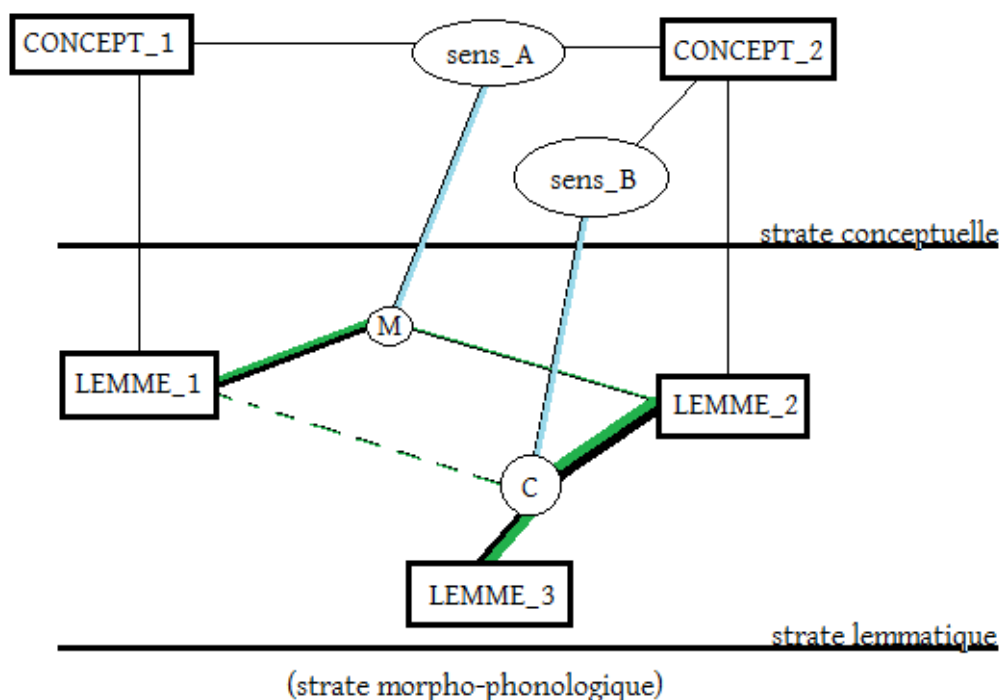


Figure 20 Ebauche d'un modèle stratifié du lexique, inspiré de (Levelt et al. 1999), avec représentation du massif (M) et du comptable (C) comme nœuds partagés au niveau lemmatique, reliés d'une part aux différents lemmes nominaux (LEMME\_1 à 3), et d'autre part à certains nœuds conceptuels.

Dans notre Figure 20 (ainsi que dans toutes les figures présentées dans ce chapitre), les liens doublés en vert entre les nœuds combinatoires M et C et les lemmes nominaux constituent la représentation lexicale de l'opposition massif-comptable, tandis que les liens doublés en bleu ciel, reliant M et C à des nœuds sémantiques (ne constituant pas eux-mêmes des concepts lexicaux, mais associés néanmoins à ceux-ci) représentent ce que nous appellerons sa représentation supra-lexicale. L'on voit d'emblée que les connexions au niveau lemmatique ne sont pas toutes équivalentes : les LEMME\_1 et 2 seraient des noms flexibles, dont le premier serait plus fortement associé au massif, tandis que le second serait plus naturel en syntaxe comptable (il pourrait s'agir, à titre d'exemple, de *sable* et de *brique* respectivement). Le LEMME\_3, quant à lui, apparaît être un nom fondamentalement comptable, non flexible *a priori* (*voiture*, par exemple) – dont le concept lexical, *nota bene*, n'est pas représenté ici, par souci de clarté. Au niveau sémantique, ensuite, le concept lexical associé au LEMME\_2 est connecté à différentes facettes, reliées elles-mêmes aux nœuds combinatoires M et C (le sens\_A pourrait être, dans la lignée de notre exemple, celui de 'matière, substance de construction', associé au niveau sémantique tant au CONCEPT\_1 de sable qu'au CONCEPT\_2 de brique, de même qu'en direction du niveau lemmatique au nœud M ; dans ce cas, le sens\_B pourrait être

le sens ‘objet’ de brique, relié forcément au nodule C) : c’est ce qui explique que moyennant un contexte approprié, activant suffisamment fort le sens\_A du CONCEPT\_2 (dans notre exemple, l’aspect ‘substance de construction’ de brique), le LEMME\_2 puisse être réalisé en discours en syntaxe massive – même si ce n’est pas là sa réalisation favorite (comme dans *C’est du plâtre, ce mur ? Non, c’est de la brique !*). Dans cette vision, l’emploi massif ou comptable dépend donc crucialement de l’interaction entre deux niveaux de représentation : le lexical, et le supra-lexical.

Dans ce qui suit, nous examinons d’un peu plus près tour à tour ces deux niveaux, avant de spécifier encore notre modèle en précisant comment rendre compte, selon nous, des différences que nous avons observées entre le français et le néerlandais. Ensuite, nous situerons notre approche par rapport aux modèles traditionnels de l’opposition massif-comptable dans le système linguistique, tels que nous les avons décrits au point 1.5, dans notre premier chapitre. Pour finir, nous ouvrirons la voie à nos deux chapitres suivants, en proposant les pistes à suivre pour étayer notre modèle, et le mettre à l’épreuve de la vérification empirique.

### 8.1.1 Le niveau lexical, ou lemmatique

Avant toute chose, une petite précision terminologique s’impose. Le niveau de représentation dont il s’agit ici est celui des préférences lexicales individuelles, au sens du statut (plus ou moins) massif et/ou comptable de chaque nom dans le lexique. En ce sens, nous appellerons ceci le niveau « lexical » – par opposition au niveau « supra-lexical » qui fait entrer en ligne de compte le niveau conceptuel (et son activation en contexte) à un degré de généralité supérieur, subsumant les différences inter-lexicales spécifiques (voir au point suivant). Néanmoins, dans la perspective de notre modélisation stratifiée du système linguistique, il est intéressant de noter que ces niveaux sont tous deux représentés dans le lexique (sous formes de connexions entre différents nodules, fussent-ils de nature très différentes), et sont donc techniquement tous deux « lexicaux ». *Stricto sensu*, la double représentation du massif-comptable que nous proposons est donc plutôt « lemmatique » d’un côté, et « conceptuelle » de l’autre. Néanmoins, par facilité, nous continuerons de parler d’un niveau lexical et supra-lexical, non seulement par souci de cohérence avec nos écrits antérieurs (voir par exemple (Vermote 2014)), mais surtout parce que ce contraste terminologique exprime une distinction cruciale : que l’on a affaire d’un côté à un caractère morphosyntaxique spécifique à chaque item lexical, alors que de l’autre, l’on dépasse ce niveau des lexèmes individuels.

Cette précision étant faite, nous pouvons examiner de plus près ce premier niveau. La particularité de notre modèle, en ce qui le concerne, est que la caractéristique d’un nom d’être massif ou comptable y est une propriété fondamentalement distributionnelle,

voire statistique : chaque usage d'un certain lexème dans un certain type de syntaxe renforce la connexion, au niveau lemmatique (donc exclusivement syntaxique), entre le lemme et le nodule combinatoire M ou C en question, et c'est ainsi qu'émergent nos connaissances quant à ce qui serait l'emploi le plus naturel d'un nom. Ainsi, si *voiture* nous paraît un nom exclusivement comptable, c'est avant tout parce que nous l'entendons toujours construit en syntaxe comptable – et jamais massive. Les propriétés référentielles de *voiture* ne sont bien entendu pas étrangères à l'affaire : parce qu'une voiture est un objet saillant aux limites claires, il est plutôt logique qu'on en parle généralement si pas exclusivement comme d'un individu, en employant dès lors la syntaxe comptable. Mais le lien entre propriétés référentielles et caractère morphosyntaxique est indirect : les caractéristiques physiques influent certainement sur la réalisation linguistique, mais ce n'est que cette réalisation qui est responsable de l'encodage dans le lexique mental. C'est ainsi que hors contexte, *du raisin* peut paraître plus naturel que *de la groseille* (pour des référents pourtant fort similaires) : simplement parce que nous avons plus souvent été confronté au premier – la connexion entre le nodule M et le lemme de *raisin* étant dès lors plus forte qu'avec celui de *groseille*.

Précisons que nous limiter à deux nodules combinatoires, M et C, est assurément une simplification (pour permettre une modélisation plus élégante) : chacun d'entre eux représente en réalité un réseau de pareils nodules, correspondant aux différentes réalisations syntagmatiques du massif et du comptable respectivement. Ainsi, au sein de ce que nous marquons comme nodule C se trouvent en réalité au moins deux nodules distincts – l'un correspondant au pluriel, l'autre à la construction avec *un(e)*. Ces deux nodules combinatoires forment cependant une unité cohérente et sont eux-mêmes connectés (en vertu du très grand nombre de noms qui se raccordent aux deux). Cela n'empêche que les connexions entre un lemme et chacun ces deux nodules ne doivent pas non plus nécessairement être de force égale. Nous montrons ce que nous entendons par là à la Figure 21 : *groseille* est globalement un nom comptable, mais son emploi pluriel est plus courant que son emploi au singulier – les connexions correspondantes ne sont donc pas équivalentes. C'est ce qui rend le syntagme *une groseille* (pour désigner un grain unique) potentiellement plus douteux pour certains locuteurs (comme nous le signale G. Kleiber, communication personnelle lors de la journée d'étude MassCoLex à Gand, le 23 mai 2014 – pour notre part, nous ne partageons pas forcément cette intuition : *une groseille*, tout autant qu'*un raisin* nous paraissent acceptables). Pour *voiture*, par contre, cette distinction dans le modèle n'est pas nécessaire : la connexion avec les nodules combinatoires comptables pluriel et singulier est donc équivalente. Dans le présent travail, nous continuerons néanmoins de représenter le massif et le comptable comme simplement deux nodules, C et M, par souci d'économie et de clarté.

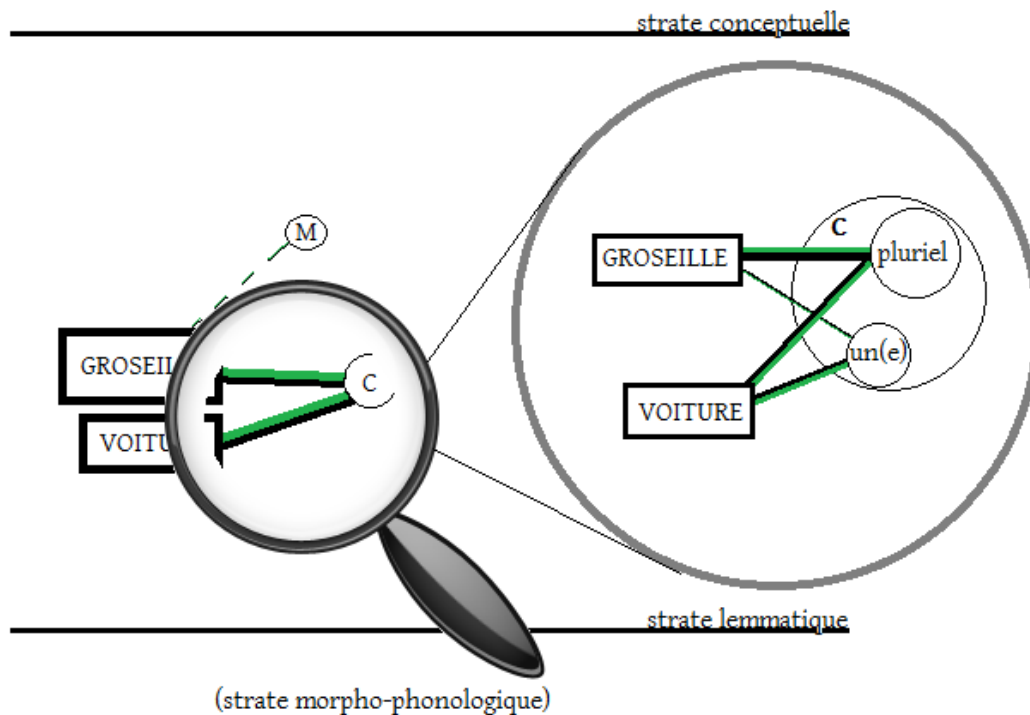


Figure 21 Zoom sur la strate lemmatique, montrant que les terminaisons C et M que nous postulons sont en fait elles-mêmes des réseaux cohérents de différents nodules combinatoires (comme par exemple celui du pluriel, ou de la combinaison avec *un(e)* pour le comptable), avec lesquels chaque lemme n'est pas obligatoirement relié de manière égale. Même cette représentation plus détaillée est du reste une simplification : le nodule pluriel correspond sans nul doute lui-même à plusieurs contextes combinatoires distincts.

Notre vision de l'encodage morphosyntaxique des noms s'inscrit donc entièrement dans une vision distributionnelle du langage, et de son acquisition (voir (Harley 2008) pour un aperçu) : ce sont les régularités de surface telles que rencontrées en discours qui déterminent les traits, précédant ainsi la sémantique. Cette vision est entièrement corroborée par la recherche sur l'acquisition de l'opposition massif-comptable (voir point 1.6) : comme nous l'avons vu, (Gordon 1985) a mis en avant le primat, dès le plus jeune âge, des indices syntaxiques (*syntactic cues*) dans l'acquisition de l'opposition (avant les indices sémantiques), tandis que (Gathercole 1985) a montré qu'effectivement, l'opposition entre *much* et *many*, en anglais, était acquise avant tout comme une distinction distributionnelle – les facteurs sémantiques n'entrant en ligne de compte que dans un second temps (vers l'âge de 8 ans d'après ses données, lorsqu'apparaissent les premiers emplois flexibles). Ceci implique qu'en effet, l'opposition massif-comptable est avant tout acquise comme une combinatoire de surface – ce qui semble confirmé également par la recherche de (Subrahmanyam *et al.* 1999), qui montrent que les jeunes enfants ne sont pas influencés par la syntaxe, massive ou comptable, dans leur manière de généraliser l'emploi d'un nouveau nom, au moins jusqu'à l'âge de 4 ans. L'on retrouve dès lors dans le parcours d'acquisition notre distinction entre la représentation lexicale et supra-lexicale du massif-comptable : les

préférences de chaque nom sont d'abord apprises de manière purement distributionnelle (niveau lexical), jusqu'à ce que le repérage de généralisations sémantiques (supra-lexicale) ouvre la porte aux emplois plus créatifs.

Une pareille vision de l'encodage lexical de l'opposition massif-comptable a deux conséquences évidentes. Premièrement, elle fait du caractère morphosyntaxique des noms une donnée profondément dynamique, susceptible de varier non seulement d'un locuteur à l'autre, mais aussi pour un même locuteur au fil du temps. Il y a bien sûr une certaine cohérence au sein de chaque communauté linguistique, en vertu d'un auto-renforcement des connexions : si parmi les noms de fruits et de légumes (flexibles), *raisin* est néanmoins un peu plus naturel comme massif (*du raisin*) que les autres (*de la pomme*, par exemple), cette donnée contribuera à une fréquence de réalisation massive en discours plus élevée pour lui, ce qui à son tour permettra de maintenir son statut sensiblement plus massif dans le système. Néanmoins, en fonction des habitudes de chacun, le naturel de tel ou tel emploi peut varier. Il y a fort à parier que le syntagme *du kangourou* pour désigner 'de la viande' soit plus naturel pour un francophone carnivore d'Australie (s'il y en a) que pour un Parisien végétarien. Pour les deux locuteurs, l'interprétation est disponible, en vertu du lien supra-lexical entre le nodule sémantique 'viande de' et le nodule combinatoire M, mais le lien lexical entre ce dernier et le lemme de *kangourou* ne sera pas le même. Bien évidemment, cette considération relève du bon sens. Il paraîtrait même presque trivial d'observer que « l'emploi d'un certain nom découle, et bien, de la manière dont il est employé ». C'est pourtant loin d'être anodin : effectivement, le caractère massif ou comptable encodé d'un nom varie avec son usage – et si c'est peut-être évident, encore fallait-il dire de quelle manière et pourquoi. Or dans notre proposition, cette donnée est traduite explicitement en termes de connexions plus ou moins fortes à l'intérieur du lexique mental, représenté comme un réseau à activation.

La seconde conséquence évidente de notre vision est que bien moins qu'une donnée binaire, le caractère massif ou comptable d'un nom est à situer sur un continuum. Ainsi, les noms ne sont pas fondamentalement massifs, comptables, ou les deux (comme l'entendrait la vision traditionnelle), mais plutôt plus ou moins fortement massifs et/ou comptables, en fonction de la manière dont on les trouve généralement réalisés en discours. Cela n'empêche que certains noms puissent être effectivement profondément massifs ou comptables : ainsi, les lemmes de *voiture*, ou encore de *bocal*, *sachet* ou *tabouret* (cf. notre sélection de noms de contrôle, voir point 4.6), qui ne sont virtuellement jamais réalisés comme massifs, ne seront connectés dans le lexique mental qu'au nodule C, et pas au nodule M, tandis que d'autre part, les lemmes de *patience* ou de *compassion*, jamais réalisés comme comptables (hormis peut-être au singulier, en lecture 'triée' : *une patience d'ange*), ne seront reliés qu'à M (ainsi que peut-être, bien plus faiblement, au nodule combinatoire *un(e)* – mais en tout cas pas à celui du pluriel). La nuance introduite par notre modèle est cependant cruciale quant aux noms dits flexibles. Si l'on trouve

tant des emplois massifs que comptables de noms tels que *raisin, pomme, poulet, chocolat, œuf, pierre* ou encore *brique*, dire que pour autant, tous ces noms sont sans discrimination à la fois massifs et comptables revient à perdre énormément de la subtilité du système – car l’intuition nous dicte que néanmoins, *du raisin* est un peu plus naturel que *de la pomme*, et *de la pierre* que *de la brique*, par exemple (tandis que la différence entre *un poulet* et *du poulet* est d’un autre ordre que celle entre *un chocolat* et *du chocolat* : sur ce point, il faudra faire intervenir notre niveau supra-lexical, voir ci-après). Ainsi, dans notre modèle à activation du lexique, le caractère massif ou comptable d’un nom est modulé en fonction de la fréquence respective de ses différents emplois : si pour *brique*, l’emploi massif paraît plus spécial que pour *béton*, c’est parce qu’il est moins fréquent, et qu’ainsi le lemme du premier est moins fortement connecté au nodule M que le second – tout simplement.

Cette approche permet du reste de donner une place aux noms comme par exemple *nature*, dont l’emploi est généralement neutre (*la nature*) et dont dès lors tant les emplois explicitement massifs que comptables paraissent d’une certaine manière dérivés : *il me faut de la nature pour m’épanouir, une nature pure et sauvage* (Trieur), *\*trois natures*. En l’occurrence, le lemme correspondant est simplement très fortement connecté au nodule combinatoire correspondant à l’article défini, et simplement peu ou pas ni au nodule C, ni au nodule M – d’où l’absence d’intuitions réelles quant à une vraie préférence pour le statut massif plutôt que comptable ou inversement, sans que l’on ne puisse dire non plus que *nature* soit un nom ambivalent. C’est là un fait dont, pour ce que nous en savons, les approches lexicalistes antérieures du massif-comptable ne s’étaient pas emparées, et ne pourraient pas dire grand-chose.

Placer comme nous le faisons donc le statut massif ou comptable sur un continuum n’est du reste pas entièrement nouveau : comme nous l’avons vu au point 1.5.1, (Allan 1980) proposait déjà d’attribuer aux noms des préférences calculables dans un sens ou un autre, en fonction des contextes avec lesquels ils pouvaient se combiner. Si son point de départ n’est pas tout à fait le même que le nôtre (pour notre part, c’est la fréquence dans l’usage qui l’emporte, et non seulement la simple disponibilité d’une certaine combinaison), il tenait donc une piste intéressante. Cependant, la vraie grande différence est que précisément, (Allan 1980) situe son modèle dans une approche grammaticale de l’opposition massif-comptable, refusant à ce trait tout marquage lexical – et situant de ce fait sa propre proposition de préférences calculables en-dehors du système linguistique. Il en va de même pour les propositions de (De Belder 2011), qui évoque des différences fondamentales entre différents noms, mais se contente d’indiquer que celles-ci découlent de facteurs référentiels, contextuels et/ou pragmatiques, sans pertinence pour la description linguistique. Or, comme nous l’avons dit dans notre première partie, nous ne sommes aucunement d’accord avec une telle vision de la linguistique, dont l’objet d’étude ne serait qu’une portion arbitraire de la langue, notamment une structure stérile de relations syntaxiques vides de sens ou de

contexte. Que la variation inter-lexicale du point de vue du statut (plus ou moins) massif ou comptable soit distributionnelle, contextuelle, pragmatique, et/ou simplement logique, elle ne doit pas pour autant être expulsée du champ d'investigation : au contraire, elle peut et doit être étudiée, décrite, et expliquée. C'est là tout l'intérêt d'un modèle à activation dynamique du système linguistique, qui nous permet de représenter dans le lexique plus qu'une simple liste d'items avec des traits arbitraires et fixes. La langue est profondément fonctionnelle, contextuelle, et ancrée dans notre expérience psychologique : c'est dès lors d'après nous comme telle qu'elle doit être abordée et étudiée.

### 8.1.2 Le niveau supra-lexical, ou conceptuel

Notre premier niveau de représentation du massif-comptable est donc strictement distributionnel : il concerne la combinatoire préférentielle entre les noms et les environnements syntagmatiques, sur simple base de leur fréquence – encodée dans le système sous forme de connexions lemmatiques plus ou moins fortes (ou absentes, le cas échéant). Cependant, il va sans dire que l'opposition massif-comptable n'est pas dénuée de sens. Nous n'entendons par là pas seulement que massif et comptable, indépendamment des noms auxquels ils s'appliquent, sont associés à certaines manières de conceptualiser les référents (voir au point 1.4 notre analyse détaillée de la dénotation massive et comptable), mais aussi simplement que pour les noms réputés flexibles, l'alternance entre l'un et l'autre type d'emploi n'est jamais gratuite : elle s'accompagne toujours d'une nuance, d'un glissement sémantique – même, ou plutôt surtout, lorsque plus profondément, un seul et même concept lexical reste activé dans les deux cas (voir ce que nous avons dit au point 2.4 sur ce sujet). Or, d'un côté, la nature de cette alternance n'est pas unique : la variation sémantique entre *du poulet* et *un poulet* n'est pas la même qu'entre *un chocolat* et *du chocolat* – comme nous venons de l'évoquer. Cependant, d'un autre côté, nos études de corpus nous ont permis de confirmer qu'elle répondait à une certaine cohérence de classe : l'alternance sémantique entre *un kir* et *du kir* est bien la même qu'entre *un café* et *du café*, de même que celle entre *un melon* et *du melon* répond à celle entre *une pomme* et *de la pomme*. En outre, le sens associé au massif dans *du chêne*, *du béton*, *de la corde* ou *de la brique* est à chaque fois le même – indépendamment de l'attraction intrinsèque de chacun de ces lemmes pour ce type de syntaxe.

Pour expliquer ces faits, nous postulons l'existence de liens directs entre certains aspects de sens (activés principalement en contexte) et la réalisation en syntaxe massive et comptable, sous forme de connexions directes dans le lexique mental entre certains nodules sémantiques et les nodules lemmatiques combinatoires M et C. (Cela signifie que nous reconnaissons donc des effets contextuels automatiques sur le traitement

lexical – voir les travaux de (Hess *et al.* 1995) à ce sujet.) Ainsi, par exemple, le nodule sémantique ‘matériau, substance de construction’ – lié au niveau de la strate sémantique aux concepts de béton, ciment, granit, mais aussi corde ou brique – est connecté par ailleurs en ligne directe avec le nodule combinatoire lemmatique M. Ce lien n’est évidemment pas aléatoire : il faut le comprendre en termes de généralisation. Il découle ainsi du grand nombre d’occurrence où les deux nodules sont activés ensemble, à chaque production dans un contexte pertinent de syntagmes tels que *du sable, du ciment, du béton, du plâtre*, etc. Dans tous ces cas, le sens ‘substance de construction’ est activé simultanément au nodule M, avec ainsi par généralisation l’établissement d’une connexion entre ceux-ci. Nous montrons à la Figure 22 comment l’on peut se représenter cela. Dans cette illustration, tout comme dans celles qui suivront, nous représenterons systématiquement les concepts lexicaux par des images (par oppositions aux sens associés, impliqués dans les connexions supra-lexicales, que nous indiquons par du texte entouré d’un cercle).

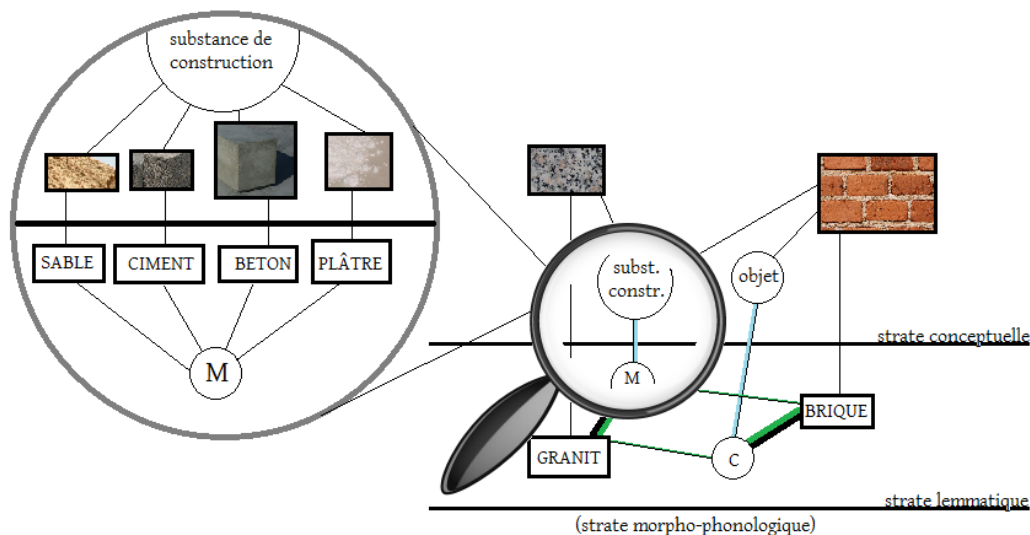


Figure 22 Détail d’une connexion supra-lexicale, entre le nodule sémantique ‘substance de construction’ et le nodule combinatoire du niveau lemmatique ‘M’. Cette connexion s’établit par généralisation, en conséquence de la fréquente activation simultanée des deux nodules concernés.

L’on peut dès lors mieux comprendre et décrire le comportement morphosyntaxique de *brique*, par exemple. Au niveau lemmatique, *brique* a une claire préférence pour la syntaxe comptable, de par sa fréquence plus élevée dans cet emploi (en témoignent nos données de corpus) – ce que nous représentons par un lien plus fort avec le nodule combinatoire C. Au niveau sémantique, le concept de brique est impliqué cependant dans diverses associations : la brique se présente sous forme d’objets distincts d’une part, mais est aussi un matériau, susceptible dès lors d’être conçue comme une substance de construction (au même titre que *de la pierre, du marbre*, etc.). Les deux nodules sémantiques correspondants sont quant à eux associés respectivement aux



nodules combinatoires C et M, par généralisation (cf. la loupe, pour l'association avec M). La réalisation discursive de *brique* dépendra dès lors de l'interaction du facteur lexical, et supra-lexical : au premier niveau, l'environnement le plus naturel pour *brique* est la syntaxe comptable, cependant si le contexte active suffisamment le nodule M, par le biais de l'activation de la facette sémantique 'substance de construction', ça peut être celui-ci qui l'emportera – dans ce cas, c'est le syntagme massif *de la brique* qui sera produit. Le sens du massif sera alors bel et bien le même que dans *du sable, du ciment, du béton* ou *du plâtre*, puisque ce sera le même nodule sémantique qui aura été responsable de sa sélection. C'est comme ça que l'on peut expliquer la cohérence de classe liés aux emplois massifs et comptables de différents noms du même domaine sémantique. D'autre part, le syntagme *de la brique* considéré hors contexte laisse une impression quelque peu marquée, comme si l'usage était dérivé : cela vient du fait que pour *brique*, le massif n'est effectivement pas l'emploi lexical préféré.

L'on peut sans problème appliquer les mêmes schémas à nos noms de boissons. Pour ceux-là, l'on peut postuler que les concepts lexicaux en question, comme ceux de *bière, café, kir* ou encore *espresso*, par exemple, sont connectés à au moins trois nodules sémantiques : 'substance (potable) liquide', 'portion standard', et 'sorte'. (Quant à ce dernier, que nous supposons très général, il s'agira du même que celui responsable des emplois taxonomiques dans d'autres catégories de noms, comme les matériaux : de toute évidence, il y a quelque chose de fondamentalement commun dans le sémantisme de *une bière de grande qualité* ou *il existe trois vins de Bourgogne* et *un sable très fin* ou encore *les normes X définissent trois ciments*. Nous en dirons un mot un peu plus loin.) Ces nodules sont eux-mêmes reliés directement aux nodules combinatoires M, C et C respectivement. Lors de la production en discours, il sera ainsi opté pour la syntaxe massive ou comptable, en fonction de l'aspect de sens profilé le plus fortement en contexte, en interaction (c'est là une précision cruciale) avec les préférences distributionnelles lexicales : l'on pourra dire (266) en activant seulement le concept lexical présenté comme substance, mais aussi (267) si l'on choisit de profiler les portions (par exemple en se vantant auprès d'un groupe d'amis d'une performance bibitive – notons que dans cet ordre de grandeur, le compte est forcément approximatif), ou encore (268), auquel cas c'est par le biais du nodule sémantique 'sorte' que la syntaxe comptable est activée :

(266) J'ai bu de la bière.

(267) J'ai bu douze bières.

(268) J'ai goûté une bière dont je n'avais jamais entendu parler.

Inversement, si l'on ne doit pas produire mais interpréter ces énoncés, la syntaxe fournira des indices quant au sens entendu : le massif activera l'interprétation 'liquide potable', et le comptable activera à la fois les nodules 'portion' et 'sortes' – et celui des deux qui sera le plus saillant (parce que le plus logique dans le contexte) l'emportera

pour fournir l'interprétation finale. Ainsi, moyennant un contexte approprié, une lecture taxonomique de (267) n'est pas totalement inenvisageable.

Dans notre modèle à activation du lexique, les restrictions que nous avons observées en corpus quant à certains emplois s'expliquent du reste aisément. Si par exemple *lait* ne connaît pas de lecture en 'portion', c'est parce que cette boisson ne se consomme pas sous forme de portions standard : dans notre modèle, cela se traduit élégamment par l'absence de connexion, au niveau conceptuel, entre le concept lexical de *lait* et le nodule partagé 'portion'. Pour *vin* et ses proches parents, la portion est en revanche bel et bien disponible, mais sa réalisation en langue par le biais de la syntaxe comptable est concurrencée par la combinaison avec *verre de*. Si ces deux combinaisons seront dès lors toutes deux être représentées au niveau lemmatique, et connectées au nodule sémantique 'portion', dans le cas de *vin* ce sera toujours la seconde option qui l'emportera, en vertu du lien particulièrement fort et saillant entre le nodule combinatoire correspondant, et le lemme de *vin* (puisque qu'une portion de vin se consomme toujours comme un verre). Nous illustrons cela à la Figure 23. La combinaison avec *verre de* a bien sûr la particularité supplémentaire, par comparaison avec la syntaxe massive ou comptable, d'être compositionnelle, et d'intégrer aussi le sens lexical de *verre* dans l'interprétation finale. Nous marquons cela en notant *verre* comme un lemme, et en le connectant lui-même directement à un concept propre.

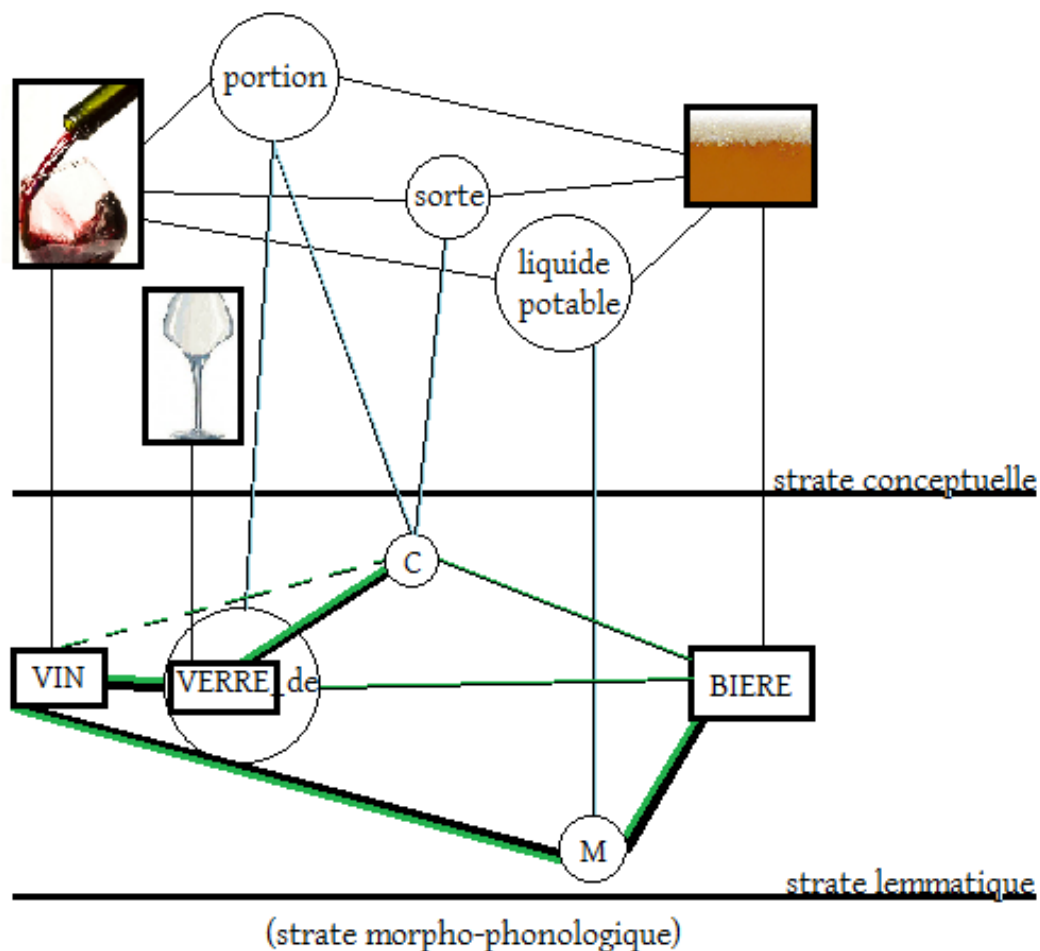


Figure 23 Ebauche de modélisation de l'organisation lexicale des noms de boissons. Les lectures disponibles sont celles en 'portion standard', 'sorte' et 'liquide potable', chacune liée à l'une ou l'autre option combinatoire. Pour *vin*, la lecture comptable conditionnée est fortement concurrencée par la composition avec *verre de*.

Ici, *verre* est bien représenté dans son acception lexicalisée dénotant un objet spécifique, notamment un 'récipient' (voir Chapitre 6). Dans notre modélisation, *verre* comme nom de matériau (*un bijou en verre, c'est du verre ou du plastique ?*) devrait quant à lui être représenté comme un second lemme, relié pour sa part fortement au nodule combinatoire M, et connecté à un concept lexical propre. C'est ce que nous entendions en disant que les lectures en 'objet spécifique' de nos noms de matériaux, comme *un bronze, un verre* ou même *un chêne* devaient être représentées comme des emplois lexicalisés à part entière, entretenant avec les équivalents massifs un lien probablement historique – mais pas un lien de dérivation synchronique. (Pour *de la bière – une bière*, par contre, nous soutenons qu'il s'agit bien dans les deux cas du même concept lexical, comme on peut le déduire de la Figure 23.)

### 8.1.2.1 Une question de facettes lexicales ?

Les sens partagés supra-lexicaux, associés directement au massif ou au comble, que nous postulons dans notre modèle, sont donc des composants sémantiques, liés (mais non équivalents) aux concepts lexicaux, et plus ou moins activés selon les circonstances. Présentés comme cela, ces éléments de sens ne paraissent pas très éloignés de la notion de « facettes », telle que proposée entre autre par (Croft et Cruse 2004) (voir ci-dessus, point 2.3). L'on peut dès lors se poser la même question critique que (Kleiber 1999: 96-98) à leur sujet : dans quelle mesure les facettes font-elles partie des concepts lexicaux eux-mêmes ? Et, plus précisément, si dans certains cas l'on doit postuler des facettes *ad hoc*, pourquoi les facettes ne seraient-elles pas simplement toutes contextuelles ?

Notre modélisation du lexique offre une explication intéressante à cette question. Les nodules sémantiques supra-lexicaux que nous postulons ne sont pas liés exclusivement à l'un ou l'autre concept lexical : ils sont partagés entre différents concepts et acquièrent ainsi, par généralisation, un certain statut indépendant. La question n'est donc pas tant de leur connexion préalable avec tel ou tel concept lexical, mais simplement de leur disponibilité au niveau sémantique – et de leurs conséquences pour le niveau morphosyntaxique. Prenons l'exemple des noms d'animaux, et de leurs emplois massifs pour désigner de la viande, et comparons les emplois (*manger*) de *l'agneau* et (*manger*) du *kangourou*. D'une part, l'on peut imaginer sans problème que le concept lexical associé au lemme d'*agneau* soit relié au moins à deux nodules sémantiques : 'individu animal' et 'viande' – eux-mêmes connectés respectivement aux nodules lemmatiques combinatoires C et M. (Même si les deux sens sont assez différents, un concept lexical commun est nécessaire pour rendre compte des énoncés du type *Je ne mange pas d'agneau, parce qu'ils sont trop mignons*, voir les exemples (72)-(76) au point 2.4.) Dans le cas de *manger de l'agneau*, c'est donc simplement la facette 'viande de' du concept 'agneau' qui est activée, en même temps que la syntaxe massive. D'autre part, pour *kangourou*, la situation est un peu plus délicate : que l'aspect sémantique 'viande de' soit d'emblée associé au concept de kangourou dans notre lexique mental est loin d'être une évidence. A vrai dire, avant que l'on ait imaginé que cet animal pouvait effectivement se consommer, l'association est probablement inexistante.

Cependant, dans notre modèle à activation du lexique, interpréter *du kangourou* comme 'de la viande de kangourou' ne nécessite pas pour autant de postuler quelque facette *ad hoc* que ce soit : le nodule sémantique 'viande de', associé au nodule combinatoire M, est d'emblée présent dans le lexique, par généralisation de tous les emplois où il est réalisé. Dans une situation de production langagière, l'évocation de la consommation de notre marsupial activera dès lors d'une part le concept lexical 'kangourou', qui activera à son tour le lemme correspondant, et d'autre part le nodule sémantique supra-lexical (préexistant) 'viande de', qui quant à lui activera le nodule combinatoire M. Au niveau inférieur du modèle, le lemme de *kangourou* est plus

fortement connecté au nodule combinatoire C ; cependant, dans l'exemple présent, le nodule M jouira d'une activation accrue par le biais du sens 'viande de' – alors même que C se retrouvera inhibé par l'absence, précisément, d'activation du nodule supra-lexical 'individu animal'. Ainsi, l'interaction entre nos deux niveaux de représentation sera tranchée, dans ce cas, par une production en syntaxe massive : *du kangourou*. (Notons que cette vision est entièrement compatible avec celle de (Khallouqi 2003), qui argue que « les items lexicaux sont sous-déterminés par rapport à leur signification, et doivent être enrichis pour désigner des concepts *ad hoc* spécifiques à la situation de communication considérée » (Khallouqi 2003: 167). En fait, notre proposition est en quelque sorte un modèle des mécanismes qu'il met en avant, expliquant comment éléments encodés et éléments contextuels interagissent.)

### 8.1.3 Différences entre le français et le néerlandais

Notre modèle à activation stratifié du système linguistique nous permet ainsi à première vue de rendre compte des faits principaux que nous avons observés quant à la flexibilité de l'opposition massif-comptable. Penchons-nous à présent sur ce qu'il nous permet de dire des différences en la matière entre le français et le néerlandais. Au travers de nos études de corpus, nous avons conclu que le néerlandais était, globalement, moins propice à l'alternance morphosyntaxique : *a priori*, les préférences fondamentales des noms sont plus difficiles à outrepasser dans cette langue, et les mécanismes alternatifs de la paraphrase et de la composition sont disponibles pour réaliser les effets de sens associés en français au passage du massif au comptable, et inversement.

Dans notre modèle, cela peut se représenter de la manière suivante. D'une part, au niveau lexical, la situation est sensiblement identique dans nos deux langues : les lemmes sont connectés plus ou moins fortement aux nodules combinatoires M et C, en fonction de leur distribution dans l'usage (la force des connexions étant néanmoins répartie différemment pour les équivalents de traduction dans une langue et l'autre, comme en témoignent les différences de profils que nous avons observées : le lemme de *raisin* sera relié bien plus fortement au nodule M que le lemme de *druif* 'raisin'). D'autre part, au niveau supra-lexical, les nodules sémantiques impliqués en français dans les alternances massif-comptable sont connectés en néerlandais avec une palette plus large de possibilités combinatoires au niveau lemmatique. L'explication est donc comparable à ce que nous disions de la concurrence entre la syntaxe comptable et la paraphrase avec *verre de* pour *vin* : en néerlandais, le sens supra-lexical de 'sorte' est connecté non seulement au nodule combinatoire C, mais aussi à celui correspondant à la composition avec *-soorten* et avec la construction partitive directe *soort(en) N*, tandis que celui de 'portion standard de boisson' le sera à C également, de même qu'au nodule combinatoire

correspondant à l'ajout d'un diminutif, et à la combinaison partitive directe *glas N* (comme dans *glas wijn* 'verre de vin') ou *kop(je) N* (comme dans *kopje koffie* 'tasse de café').

Bien entendu, certaines de ces alternatives ont des équivalents en français : *sorte de*, *verre de*, etc. Cependant, la concurrence avec l'emploi comptable est moins forte qu'en néerlandais, car en français, ces alternatives sont strictement analytiques. En néerlandais, en revanche, leur caractère synthétique (composition morphologique avec le diminutif *-(C)je*) ou assimilé (partitif direct, alors qu'en français une préposition intermédiaire est nécessaire) font que les nodules combinatoires correspondants sont forcément plus fortement connectés aux lemmes auxquels ils s'appliquent. L'option de la flexibilité massif-comptable n'est donc pas inexistante (l'on a bien *een koffie* 'un café (~une tasse)', ou *een fijnkorrelig zand* 'un sable fin, litt. aux grains fins'), mais elle l'emportera moins souvent sur les alternatives – en faisant du même coup une option moins forte globalement, car moins souvent renforcée.

Cette donnée n'en est pas moins modulée par une certaine variation inter-lexicale. Si, comme nous l'avons dit, l'emploi comptable 'portion' des noms de boissons en néerlandais est concurrencé par l'emploi de la forme diminutive (on n'a pas *\*een bier* dans ce sens, mais bien *een biertje* 'litt. une petite bière'), la nécessité du diminutif est en l'occurrence variable, puisqu'il est indispensable dans certains cas, mais optionnel dans d'autres. Cela aussi, nous pouvons l'expliquer dans notre modèle : tout autant que l'attraction pour le massif ou le comptable, la force des connexions entre différents lemmes et le nodule combinatoire du diminutif (appelons le DIM) est plus ou moins forte dans le lexique mental, en fonction de la fréquence en usage des emplois correspondants. L'option morphosyntaxique préférentielle, voire obligatoire pour désigner la portion standard d'une boisson sera donc au final celle qui l'emportera sur les autres en étant le plus fortement activée – une donnée dépendant à la fois de l'activation supra-lexicale de différentes options combinatoires, et des préférences lemmatiques (distributionnelles) du nom en question pour chacune de ces options. La Figure 24 illustre ce propos. On y voit représenté le comportement morphosyntaxique de *bier* 'bière', *koffie* 'café' et *kir* 'kir', par rapport à la lecture 'substance potable', et 'portion standard'. La première est disponible pour les trois, par le biais de la syntaxe massive. Pour la seconde, l'emploi du diminutif est respectivement indispensable, éventuel ou peu probable. Notons que pour les lectures taxonomiques (que nous ne représentons pas ici par souci de clarté), le raisonnement serait identiquement le même, avec une concurrence entre différentes options, modulées par des préférences lemmatiques (la réalisation avec *-soorten* étant par exemple plus naturelle et ainsi probable pour *bier* – *biersoorten* – que pour *schuimwijn*, comme nous l'avons évoqué au chapitre précédent, point 7.2). L'on peut du reste postuler une situation très similaire dans le domaine des noms d'animaux, pour la lecture 'viande de' – possible via le simple emploi massif pour certains items, comme *konijn* 'lapin', *kalkoen* 'dinde' ou *eend* 'canard', mais nécessitant la composition avec *-vlees* 'viande' pour d'autres, comme *paarde(n)vlees*

‘cheval (litt. viande de cheval)’ ou *varken(s)vlees* ‘porc (litt. viande de cochon)’, voir (Lauwers et Vermote 2014, à paraître).

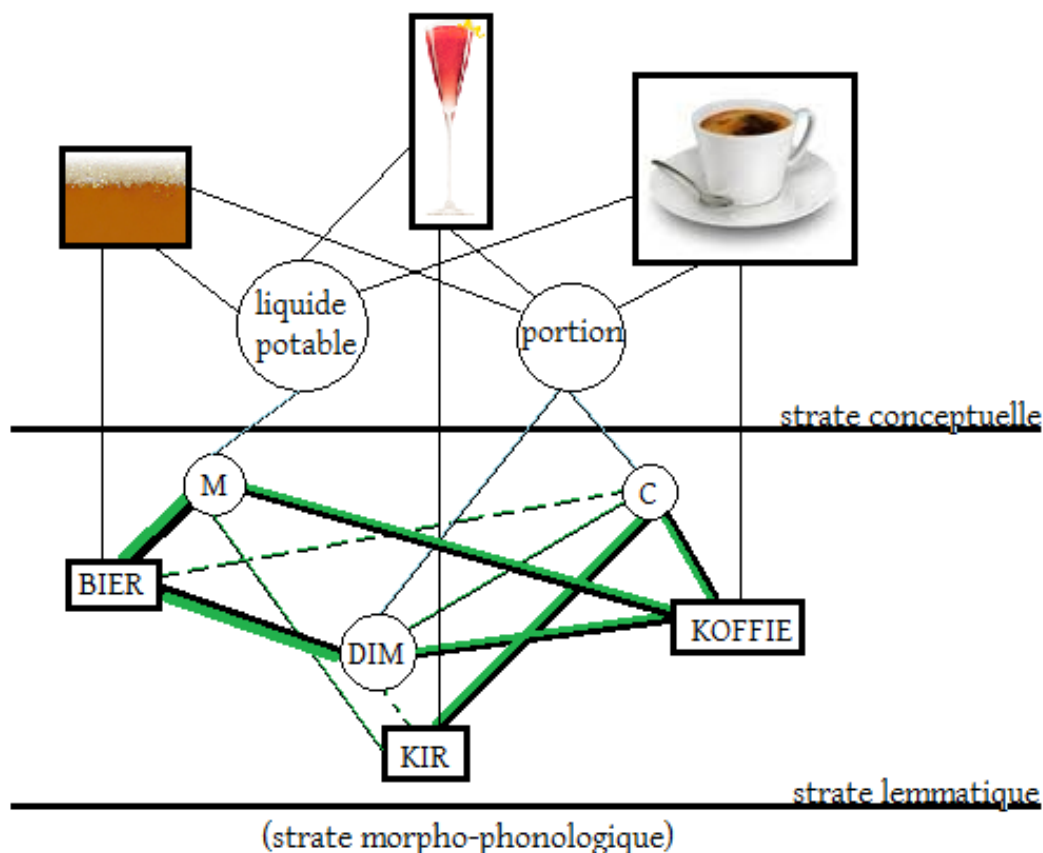


Figure 24 Aperçu de l’organisation lexicale des noms de boissons en néerlandais, illustrée par les items *bier* ‘bière’, *kir* ‘kir’ et *koffie* ‘café’. La possibilité de l’emploi massif dans le sens de ‘liquide potable’ est disponible pour les trois items (et est l’emploi de base pour *bier* et *koffie*, mais pas pour *kir*). Pour la lecture en ‘portion standard’, les options de l’emploi comptable et du diminutif sont disponibles, mais jouissent de rapports différents aux trois lemmes concernés. Pour *bier*, ce sera toujours le diminutif qui l’emportera, tandis que pour *koffie*, les deux alternatives sont disponibles de manière sensiblement équivalente. Notons que la connexion entre les nodules DIM et C indique, quant à elle, que le diminutif implique automatiquement l’emploi comptable.

## 8.2 Une approche lexicaliste améliorée

Notre modèle à activation stratifié du lexique repose donc sur deux facteurs cruciaux : les préférences lexicales intrinsèques des noms, et l’influence du contexte – ramenant la flexibilité caractéristique de la distinction massif-comptable à l’interaction de ces deux facteurs. Cette vision nous positionne très clairement par rapport au débat quant à la

localisation dans le système linguistique de l'opposition massif-comptable, que nous avons décrit au Chapitre 1, point 1.5, en nous rangeant du côté des approches lexicalistes. En effet, nous nous trouvons bien loin des propositions des tenants d'une vision grammaticale, pour qui le caractère profondément volatile de l'opposition massif-comptable est la preuve que ces traits ne peuvent être représentés qu'en-dehors du lexique (Borer 2005, Pelletier 2012). Au contraire, par notre proposition, nous démontrons qu'il n'en est rien, et qu'un marquage lexical des préférences morphosyntaxiques, en combinaison avec des mécanismes contextuels, est justement indispensable pour expliquer la manière dont les réalisations des noms en discours alternent entre syntaxe massive et comptable (que ce soit pour expliquer la différence entre *du béton* et *de la brique*, ou pour justifier la variation dans l'usage du diminutif dans les lectures 'portion' en néerlandais).

Cependant, notre vision n'est pas lexicaliste sans plus. Elle se distingue des approches traditionnelles sur différents points. Premièrement, si nous attribuons un caractère massif ou comptable aux éléments du lexique, nous précisons explicitement la nature de ce marquage : pour nous, comme nous l'avons dit, il est essentiellement combinatoire, syntaxique – à se représenter comme l'encodage systématique de régularités distributionnelles. Ainsi, un nom sera plus ou moins massif ou comptable dans notre lexique mental, simplement à mesure qu'il se rencontre plus ou moins souvent en syntaxe massive ou comptable. Il s'agit là d'une affirmation bien plus précise que de simplement attribuer un trait binaire à chaque élément du lexique. Dans notre approche, si le massif et le comptable sont bel et bien des traits lexicaux, ils sont tout à la fois dynamiques, et à situer sur un continuum de préférences – les possibilités de statut lexical étant donc nuancées de manière virtuellement illimitée (et non arrêtées à « intrinsèquement massif », « intrinsèquement comptable » ou « les deux »).

Deuxièmement, la flexibilité morphosyntaxique des noms ne se décrit pas dans notre modèle par des règles de transfert, mais bien par un mécanisme d'héritage supra-lexical, mettant en lien direct certains aspects de sens (partagés par différents noms) et une réalisation dans un certain type de syntaxe. L'avantage principal de cette vision de la flexibilité est qu'elle nous libère de l'impératif de « directionnalité » allant obligatoirement de pair avec la notion de transfert. De cette manière l'on peut expliquer que le sens revêtu par différents noms dans un certain type de syntaxe puisse être le même, indépendamment du caractère préférentiel ou dérivé de l'emploi en question : *de l'agneau* et *du kangourou* dénotent tous deux 'de la viande', *du béton* et *de la brique* des 'matières de construction', *un kir* et *une bière* des 'portions d'une boisson', même si l'usage en question est à chaque fois plus habituel pour l'un des deux items que pour l'autre (ce qui obligerait une approche lexicaliste traditionnelle à opérer une distinction stricte entre les deux). Plus encore, il y a une claire cohérence de classe, comme nous l'avons dit, au niveau du lien qui unit *un kir* à *du kir* et *une bière* à *de la bière*. Or, la vision lexicaliste traditionnelle, avec sa notion de transfert, a du mal à expliquer ces



similarités : puisque l'emploi de base (au moins au niveau de la fréquence, nous l'avons démontré) n'est pas le même dans les deux cas, il faudrait postuler pour *kir* un passage du comptable au massif, et pour *bière* le transfert inverse. Cette option nous semble assez contre-intuitive. Nous y opposons dès lors notre proposition en termes d'héritage supra-lexical : à la syntaxe massive et comptable est simplement associé un sens respectif, identique pour les deux noms, indépendamment de la préférence lexicale de chacun. Des différences distributionnelles encodées découlent alors simplement le sentiment qui permet de juger le cas échéant lequel de ces emplois est le plus habituel, ou le plus marqué : *une bière* paraît dérivé du massif *de la bière* parce que ce dernier est l'emploi le plus fréquent – tandis que pour *kir*, c'est l'inverse. (L'impression de transfert décrite dans la littérature n'est donc pas imaginaire, mais elle demande une explication plus subtile, permettant d'expliquer les différences entre items, mais aussi les similarités dans leurs différents emplois.)

Pour finir, notre proposition nous permet aussi de dire quelque chose du mécanisme du Trieur. Dans un certain sens, celui-ci est le candidat le plus probant au poste de règle de transfert prototypique : il s'applique essentiellement aux noms marqués comme massifs dans le lexique, avec un champ d'application très large (le mécanisme est effectivement très productif, du moins en français), et les « transforme » en noms comptables dans le discours, au travers d'un certain marquage grammatical. Une chose est frappante, cependant : s'il est absolument vrai que ce mécanisme s'applique principalement aux noms massifs (voir au point 1.5.2.2 : si j'ai mangé *deux bananes, trois pommes* et *une orange*, je ne dirai pas normalement que j'ai mangé *trois fruits*, mais bien *six*), les noms comptables peuvent néanmoins y être soumis dans certaines circonstances, moyennant un contexte non ambigu, comme dans (269) avec *voiture* – clairement comparable à l'équivalent avec *sable* en (270).

(269) La Lamborghini Huracán est une voiture extrêmement rapide.

(270) Le sable de la Méditerranée est un sable extrêmement fin.

Si ce fait ne paraît pas extraordinaire en soi, il remet cependant en question la pertinence de la définition habituelle du Trieur, en tant que mécanisme de transfert allant du massif au comptable (puisque les noms massifs sont seulement les candidats favoris, mais non exclusifs, au mécanisme). Or, dans notre modèle, la situation s'explique de manière différente, et selon nous plus satisfaisante. Ainsi, le sens 'sorte' est simplement à se représenter comme un nodule sémantique partagé au niveau conceptuel du lexique mental, connecté ou connectable d'une part à différents concepts lexicaux (sans distinction entre massif ou comptable), et d'autre part au nodule combinatoire C du niveau lemmatique. Ce nodule C, quant à lui, est connecté en outre à d'autres nodules sémantiques, notamment un nodule 'objet individuel', également connecté à divers concepts lexicaux. La Figure 25 montre ce que nous entendons par là.

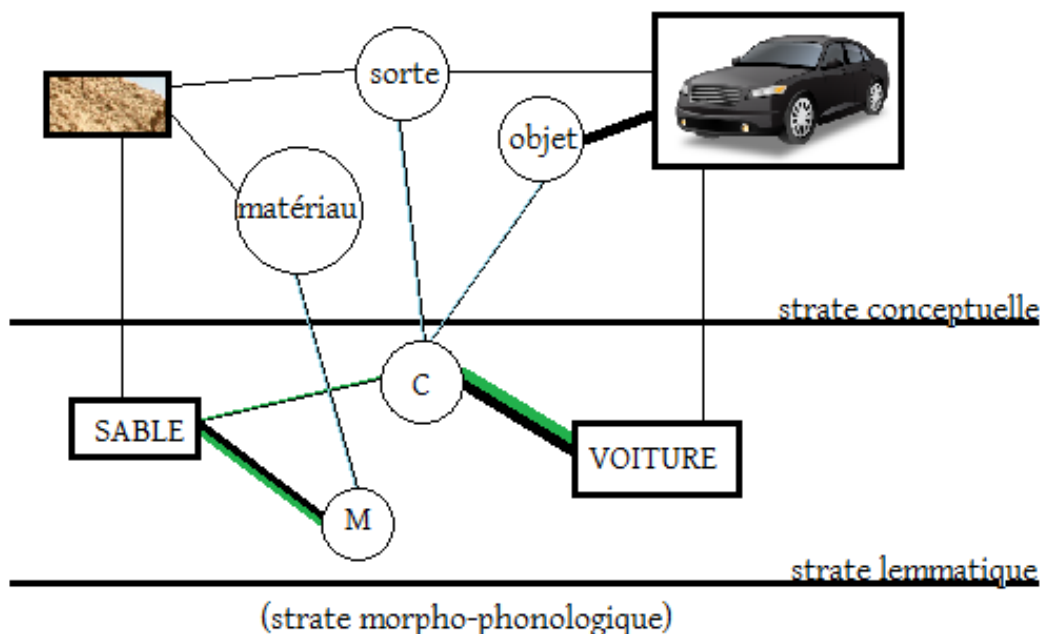


Figure 25 Représentation du mécanisme du Trieur, responsable de la lecture taxonomique d'un syntagme comptable. Dans le cas de noms habituellement massifs, le comptable pointe de manière claire vers la lecture en 'sortes', tandis que pour les noms habituellement comptables, il y a concurrence avec la lecture 'objet individuel', plus saillante.

En examinant la Figure 25, l'on peut à présent comprendre pourquoi la lecture taxonomique n'est pas strictement limitée aux noms principalement massifs, tout en s'accommodant particulièrement bien de ceux-ci. Pour *sable*, l'utilisation de la syntaxe comptable (représentée par l'activation du nodule C), active de manière significative au niveau sémantique, outre le concept lexical de sable, uniquement le nodule sémantique 'sorte' (par souci de clarté, nous ne considérons pas ici les emplois pluriels collectifs du type *des sables mouvants*, voir 6.1.2). Le nodule 'objet', quant à lui, se trouve techniquement activé lui aussi par la syntaxe comptable, mais l'absence de lien au niveau sémantique avec le concept lexical de *sable* (il n'existe pas d'objet qui s'appelle *un sable*) rend cette activation sans effet : les syntagmes *un sable*, ou éventuellement *des sables*, voire *deux, trois sables*, obtiennent donc invariablement, et sans ambiguïté, une interprétation taxonomique. Pour *voiture*, cependant, la situation est bien différente. Des nodules sémantiques supra-lexicaux 'sorte' et 'objet individuel', activés par l'emploi en syntaxe comptable (via le nodule lemmatique C), c'est cette fois le second qui est, de loin, le plus saillant : c'est lui qui est le plus fortement connecté au concept lexical de voiture (dans le lexique conceptuel, une voiture est effectivement avant tout un objet, au sens ontologique du terme). (Notons que dans nos figures précédentes, nous n'avons jamais distingué la force des connexions au niveau sémantique – bien qu'elle varie effectivement. Dans cette Figure 25, par contre, nous marquons bel et bien par un trait plus épais la saillance supérieure du sens 'objet' de voiture, par rapport à 'sorte', puisque cette indication est pertinente pour la discussion présente.) L'emploi taxonomique de

*voiture* est donc bien moins accessible, car il est fortement concurrencé par l'emploi comptable 'objet', de loin le plus fréquent. Il ne pourra dès lors être obtenu que dans une situation sans ambiguïté, comme dans (269), où l'activation contextuelle de 'sorte' est suffisamment forte que pour prendre la main. (Ce qui n'est pas le cas, notamment, dans notre exemple fruité : *deux pommes, trois mandarines et une banane*, ce sont avant tout *six fruits*, et non *trois*.)

Nous avons montré ainsi les avantages de notre vision à deux niveaux, sur l'approche lexicaliste traditionnelle. Le principal atout en est le niveau de représentation supra-lexical du massif et du comptable, rattaché à la strate conceptuelle du lexique, par le biais de connexions partagées entre les nodules combinatoires M et C et certains nodules sémantiques associés aux concepts lexicaux. Seules celles-ci peuvent expliquer, en interaction avec les préférences distributionnelles lexicales, la flexibilité de notre opposition et tout à la fois la cohérence de classe des effets de sens liés au massif et au comptable (et que nous avons abondamment mise à jour dans nos études de corpus) : le massif a le même sens dans *du sable* et *de la corde*, dans *de l'agneau* et *du kangourou*, ou dans *un espresso* et *une bière*. Cette cohérence de classe est un aspect qui a été négligé par nos prédécesseurs, parce que depuis l'invention du Broyeur Universel par (Pelletier 1975), l'accent a été mis sur les transferts de noms d'un type vers l'autre. Pourtant, à l'origine, la machine du Broyeur n'avait pas été avancée comme mécanisme de transformation sémantique, mais bel et bien comme machine concrète et tangible – fût-elle imaginaire – pour justifier l'existence d'une flexibilité morphosyntaxique virtuellement illimitée entre massif et comptable, et du même coup la non-pertinence de tout marquage lexical de l'opposition. Cependant, sans le savoir, (Pelletier 1975) avait en réalité proposé le scénario imaginaire qui finirait de confirmer la pertinence de notre proposition, et nous permettrait de clôturer le présent chapitre : la présentation sous forme 'broyée, écrasée' est simplement un autre exemple de signification supra-lexicale, associée pour sa part (au moins en français et en anglais) à la syntaxe massive. C'est là ce qui explique qu'elle puisse s'appliquer à n'importe quel nom, moyennant un contexte adéquat – avec cependant un certain sentiment de forçage dans le cas des lectures les plus inhabituelles.

### 8.3 Prédications et prolongements

Dans nos chapitres introductifs, nous avons décrit les différentes propositions de la littérature linguistique quant à la représentation de l'opposition massif-comptable dans le système linguistique, distinguant en essence une approche lexicaliste et grammaticale, et évoquant les points forts et faiblesses de chacun. L'un des hiatus

principaux à élucider en la matière était, selon nous, l'absence d'un modèle détaillé et bien défini du lexique mental – sans lequel le débat ne pourrait jamais être tranché de manière satisfaisante. Pour notre part, après avoir examiné la question du massif-comptable et de sa flexibilité au travers de trois études de corpus, dans deux langues, nous avons donné dans le présent chapitre une présentation détaillée de notre propre proposition de modélisation du lexique mental, et de la manière d'y situer l'opposition massif-comptable. Nous avons puisé notre inspiration, pour ce faire, dans la recherche en psycholinguistique, montrant que les concepts de cette tradition se prêtaient parfaitement à une meilleure représentation et compréhension de notre sujet.

Toujours dans le respect des principes de cette tradition, nous en arrivons dès lors à présent à l'étape possiblement la plus importante du processus de réflexion : la soumission à l'épreuve de la vérification empirique du modèle établi. Notre vision du lexique et de l'opposition massif-comptable repose sur deux grands niveaux de représentation, dont l'organisation supposée implique différentes prédictions quant à la relation entre fréquence et acceptabilité d'une part, et acceptabilité et contexte d'emploi de l'autre. Elle peut dès lors être testée comme telle. Plus précisément, la teneur profondément distributionnelle de la représentation, dans notre lexique, des traits massif et comptable implique que la fréquence d'occurrence de divers emplois tels que nous avons pu les observer dans notre corpus devrait permettre de prédire leur acceptabilité telle que mesurée par des enquêtes, ainsi que leur probabilité d'apparition dans un contexte (contrôlé) de production. En outre, en vertu de l'interaction constante dans notre modèle entre ce niveau lexical, et le niveau sémantique supra-lexical, ces deux données devraient être modulées par des variations spécifiques dans le contexte phrastique d'apparition. C'est ce que nous explorons et démontrons dans les deux chapitres suivants.

# Chapitre 9 Le statut lexical de l'opposition massif/comptable : enquêtes d'acceptabilité

## 9.1 Introduction : hypothèses

Notre manière de modéliser l'opposition massif-comptable, telle que présentée au chapitre précédent, implique que chaque nom est non seulement marqué dans le lexique (mental) quant à sa préférence morphosyntaxique, mais en outre que ce marquage est continu (au lieu de discret, ou binaire) car il dépend de la distribution en discours des emplois en question. Sur le plan empirique, cette position va de pair avec une prédiction claire quant au statut massif-comptable des noms : l'acceptabilité de l'emploi d'un nom dans l'un et l'autre type de syntaxe devrait corrélérer avec la distribution en discours de cet emploi. Pour les noms qui n'alternent pas régulièrement entre massif et comptable, la prédiction est plutôt triviale (et ne nous apprend pas grand-chose) : le nom *voiture* n'est jamais rencontré en corpus en syntaxe massive, et n'est effectivement pas ou très peu acceptable dans cet emploi (*\*un peu de voiture*). Cela ne distingue pas encore notre approche de la vision lexicaliste traditionnelle. Cependant, pour les noms flexibles, notre prédiction est bien plus intéressante : si dans la vision traditionnelle, ces noms jouissent simplement du double statut (ce qui devrait les rendre acceptable de manière égale dans les deux types de syntaxe), notre proposition implique au contraire que l'acceptabilité de l'un et l'autre emploi sera modulée par sa fréquence relative, faisant non seulement de l'emploi le plus fréquent d'un certain item aussi l'emploi le plus acceptable, mais permettant en outre de comparer sur ce point divers items similaires, avec un comportement en discours subtilement différent (qui devrait dès lors se trouver corroboré par le profil d'acceptabilité).

Nous avons choisi d'examiner ces prédictions au travers de différentes enquêtes d'acceptabilité, explorant le long de différents facteurs le lien entre distribution en discours (telle qu'observable en corpus) et intuitions natives quant à l'acceptabilité de

différents emplois. (Nous avons conscience des objections parfois exprimées à l'encontre des jugements d'acceptabilité comme données linguistiques valables, voir par exemple (Labov 1972), cité dans (Tummers *et al.* 2005), mais reconnaissons leur utilité pour la validation empirique d'observations en corpus, dans la mesure où ils sont récoltés de manière systématique et contrôlée, voir le débat entre Arppe et Hilpert dans (Arppe *et al.* 2011), ainsi que (Schütze 1996)). Pour ce faire, nous nous sommes tourné à nouveau vers notre premier domaine d'investigation, celui des noms de fruits et de légumes, dont l'intérêt est précisément d'avoir des propriétés distributionnelles bien tranchées pouvant être aisément opérationnalisées en vue d'un examen empirique. Ainsi, en corpus, nous avons pu observer que les noms de fruits et de légumes étaient attestés à la fois en syntaxe massive et comptable, que les emplois massifs étaient cependant invariablement les moins fréquents, et que chaque item se caractérisait par un profil propre, plus ou moins enclin (en comparaison aux autres) à apparaître en syntagme massif. En outre, nous avons observé que certains contextes profilant spécifiquement un fruit ou légume en forme et quantité indéterminées se prêtaient particulièrement à la syntaxe massive. Sur la base de ces observations, et en adéquation avec la modélisation proposée au chapitre précédent, nous pouvons dès lors établir les prédictions spécifiques suivantes :

1. Puisqu'ils sont attestés, les emplois massifs des noms de fruits et de légumes devraient jouir d'une acceptabilité supérieure aux emplois massifs d'autres noms fondamentalement comptables qui ne seraient jamais, pour leur part, attestés dans cet emploi (par exemple, même si *de la tomate* peut paraître quelque peu marqué en comparaison à l'emploi comptable *une/des tomate(s)*, l'emploi *\*du bocal*, à partir du comptable exclusif *un bocal* est quant à lui carrément inacceptable) ;
2. Pour l'ensemble des noms de fruits et de légumes, les emplois massifs devraient cependant être sensiblement moins acceptables que les emplois comptables – les premiers étant invariablement moins fréquents que les seconds ; en outre, cette donnée devrait se trouver modulée par les préférences lexicales spécifiques observées pour chaque item (par exemple, si nous trouvons une proportion d'emplois massifs plus grande pour *raisin* que pour *tomate*, l'acceptabilité respective de ces emplois devrait être en ordre décroissant) ;
3. Pour finir, indépendamment de la variation inter-lexicale, l'acceptabilité des emplois devrait être influencée par le contexte phrastique d'apparition ; plus précisément, l'emploi massif des noms de fruits et de légumes devrait augmenter globalement dans un contexte se prêtant mieux à une présentation en « forme et quantité indéterminées », par comparaison à un contexte neutre.

Nous avons exploré systématiquement ces trois prédictions, par le biais de deux enquêtes d'acceptabilité complémentaires, menées chacune parallèlement en français et en néerlandais. Au point suivant, nous expliquons la manière dont nous avons conçu ces

enquêtes (items impliqués, types de phrases concernées, éléments de contrôle) de même que les circonstances dans lesquelles nous les avons exécutées (9.2). Ensuite, nous en décrivons les résultats, d’abord pour le français (9.3), puis pour le néerlandais (9.4). (Notons que pour l’ensemble des analyses statistiques effectuées, et présentées ci-après, nous avons eu le concours très précieux de Ludovic De Cuypere, linguiste et statisticien de l’Université de Gand. Nous le remercions infiniment !) Finalement, nous en tirons les conséquences quant à notre modélisation du lexique (9.5).

## 9.2 Dispositifs

Dans chacune de nos deux langues, nous avons mené deux enquêtes d’acceptabilité complémentaires, que nous désignerons dans ce qui suit comme respectivement l’enquête « A » et l’enquête « B ». La première concernait la flexibilité des noms de fruits et de légumes, d’un côté en contraste avec la non flexibilité d’une sélection de noms de contrôle fondamentalement comptables, et de l’autre dans une perspective de comparaison inter-lexicale. La seconde concernait l’effet du contexte phrastique sur l’acceptabilité des emplois massifs. Les deux enquêtes consistaient en un ensemble de phrases à évaluer sur une échelle de Likert à 7 points, explicitée tel qu’on peut le voir à la Figure 26.

- 7 - phrase totalement acceptable, je suis certain(e) que je pourrais dire cela
- 6 - phrase essentiellement acceptable, je pense que je pourrais dire cela
- 5 - phrase plutôt acceptable, même si moi-même je ne pense pas que je dirais cela
- 4 - phrase difficilement acceptable, moi-même je suis certain que je ne dirais jamais cela
- 3 - phrase plutôt inacceptable, je doute que quiconque dirait jamais cela
- 2 - phrase fortement inacceptable, je suis certain(e) que personne ne dirait jamais cela
- 1 - phrase absolument inacceptable, pas même assimilable à du français

Figure 26 Enquêtes d’acceptabilité : échelle de Likert à 7 points telle qu’explicitée dans les consignes données aux participants. Pour les questionnaires en néerlandais, nous avons bien entendu utilisé une traduction.

Les questionnaires de l’enquête « A » comportaient 72 phrases, dont 24 énoncés critiques et 48 *fillers* (ou ‘distracteurs’). Des phrases critiques, la moitié contenait l’un de nos 12 noms de fruits ou de légumes les plus fréquents (voir point 4.3.2, Figure 6 ; il s’agit des 12 items donnés tant en français qu’en néerlandais ; comme traduction d’*oignon*, nous avons retenu *ui*, au détriment d’*ajuin*), tandis que les 12 autres contenaient un nom de contrôle fondamentalement comptable (retenus pour correspondre aux premiers du point de vue de la fréquence et du nombre moyen de syllabes, ainsi que de la taille de leurs référents ; voir notre sélection au point 4.6, Figure 8). Dans chacun de ces deux groupes, les items étaient présentés la moitié du temps en syntaxe massive, l’autre moitié en syntaxe comptable – systématiquement

dans un contexte phrastique se voulant le plus neutre possible, c'est-à-dire n'évoquant pas spécifiquement le domaine de la nourriture (par exemple *Il faut mettre X dans le sac* ou *Voici X* ; les 24 phrases neutres utilisées, en version français et néerlandaise, peuvent être trouvées en Appendice, 12.5.1). Comme contexte syntagmatique comptable, nous avons employé invariablement l'article indéfini : *un(e)* en français, *een* en néerlandais. Comme contexte massif, nous avons sélectionné le partitif en français, tandis qu'en néerlandais nous avons opté pour un emploi nu dans la moitié des cas, et un emploi avec *wat* dans l'autre (nous étions curieux d'un éventuel effet de cette variation sur l'acceptabilité).

Afin d'obtenir les questionnaires voulus, nous avons créé un script *Perl* générant aléatoirement des listes de phrases répondant à nos critères, combinant contextes phrastiques et items critiques de manière à obtenir les configurations décrites ci-avant. Quant aux *fillers* complétant les questionnaires, nous avons puisé dans une base de données de 200 phrases composées par nos soins, n'ayant aucun rapport avec notre sujet, et balayant tout le spectre d'acceptabilité (allant des énoncés les plus naturels aux plus farfelus, tant du point de vue sémantique que syntaxique). Ces phrases se trouvent en Appendice (12.5.3).

Les questionnaires de l'enquête « B », quant à eux, comportaient 48 phrases, dont 12 énoncés critiques et 36 *fillers*. Les 12 phrases critiques comportaient chacune un nom de fruit ou de légume (les mêmes que pour « A »), systématiquement en syntaxe massive (ici, nous avons opté en néerlandais toujours pour l'emploi avec *wat*). Dans la moitié des cas, l'item apparaissait dans une phrase neutre, similaire à celles de l'enquête « A » ; dans l'autre, la phrase employée activait explicitement le domaine de la nourriture et de la consommation culinaire – que nous supposons plus propice au massif. Quant à ces contextes phrastiques, nous avons composé 10 paires d'énoncés, l'un neutre et l'autre biaisé, dont respectivement 6 et 6 étaient employés pour chaque questionnaire (ces énoncés peuvent être trouvés en Appendice, 12.5.2). Afin de nous assurer du caractère effectivement neutre et biaisé de ces phrases, nous en avons soumis les versions françaises, ainsi que 10 *fillers*, à un Cloze test – ou test des phrases à trous, (Taylor 1953) – auprès de 58 locuteurs natifs du français, confirmant que les premières élicitaient bien une minorité, et les secondes une majorité de réponses en termes de 'nourriture' (avec pour chaque paire de phrases une différence significative d'un point de vue statistique, tel que dévoilé par le test du  $\chi^2$  (1, N = 116) ; valeur *p* la plus élevée < 0,001). Comme pour l'enquête « A », nous avons généré nos questionnaires « B » à l'aide d'un script *Perl*, combinant contextes phrastiques, items et *fillers* de manière pseudo-aléatoire, en respectant nos critères.

Afin d'obtenir grâce à nos enquêtes des résultats comparables dans nos deux langues, nous avons veillé à établir nos questionnaires de manière entièrement identique en français et en néerlandais. Plus précisément, nous avons établi les listes d'items et de contextes phrastiques en parallèle (employant pour le néerlandais les traductions les



plus fidèles possible de nos sélections en français), puis utilisé les mêmes scripts pour générer à partir de ces listes nos questionnaires dans les deux langues.

Ensuite, nous avons soumis nos enquêtes au bon jugement d'un panel anonyme de locuteurs natifs francophones et néerlandophones, ignorants l'objet de notre recherche. Pour le français, nous nous sommes rendu (le 9 octobre 2013) à l'Université de Basse-Normandie, à Caen, où 43 étudiants de première ou deuxième année de baccalauréat en langues romanes ont répondu à notre enquête « A », et 43 autres à notre enquête « B ». (Nous remercions une fois de plus de tout cœur les collègues caennais pour leur accueil chaleureux, et en particulier Mathilde Salles, Eliane Delente, Véronique Lenepveu, Laurent Senoze et Dominique Legallois.) Quant au néerlandais, nous avons interrogé (le 11 février 2014) respectivement 35 (« A ») et 29 (« B ») étudiants néerlandophones de l'Université de Gand, dans le cadre d'un cours de linguistique française donné aux premières années par le professeur Peter Lauwers (promoteur de la présente thèse). Chaque questionnaire se terminait par quelques questions subsidiaires, s'assurant notamment de la compétence native des participants. Des nombres de répondants mentionnés ci-avant sont exclus les éventuels personnes ayant indiqué sur leur formulaire n'être pas locuteur natif – par exemple certains étudiants Erasmus de l'université de Caen.

## 9.3 Résultats des enquêtes pour le français

### 9.3.1 Enquête « A »

Notre enquête « A » examinait la flexibilité des noms de fruits et de légumes, d'abord en comparaison avec la non flexibilité de notre sélection de noms de contrôle. Sur ce point, nous prédisions une acceptabilité supérieure pour les emplois massifs des premiers, par rapport aux seconds. A la Figure 27, nous représentons les résultats obtenus grâce à notre enquête concernant cette question. L'axe des abscisses représente les 7 points de notre échelle de Likert, et l'axe des ordonnées le nombre d'énoncés ayant reçu cette note d'acceptabilité. Les colonnes vertes et rouges correspondent respectivement aux emplois massifs des noms de contrôle (CONT-massif) et des noms de fruits et de légumes (NFL-massif).

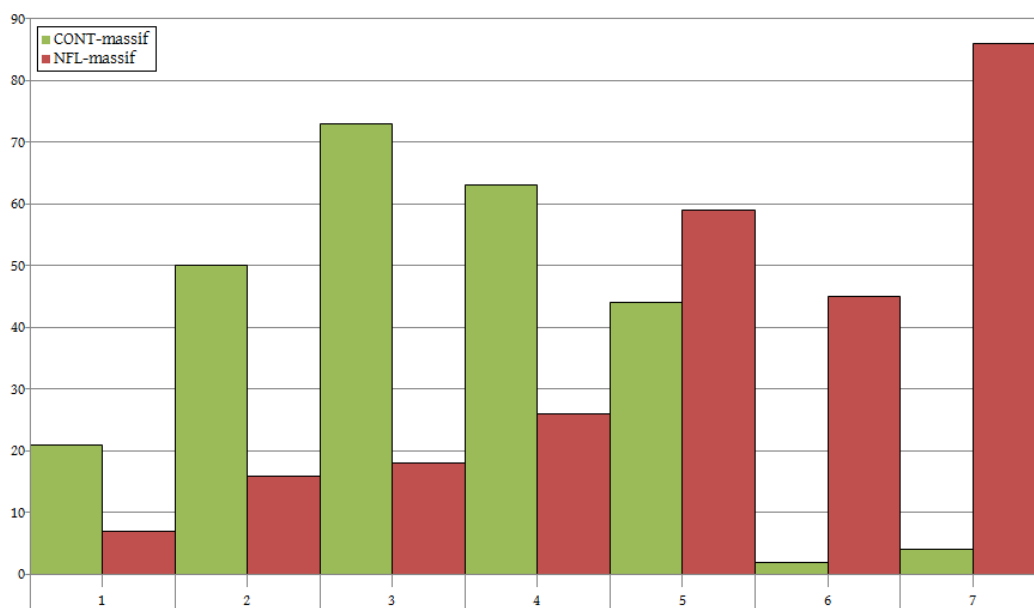


Figure 27 Acceptabilité des emplois massifs des noms de fruits et de légumes (en rouge), comparée à ceux des noms de contrôle (en vert), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

Cette représentation montre qu'il y a effectivement une grande différence dans la manière d'évaluer les emplois massifs de nos deux groupes de noms. Pour les noms de fruits et de légumes, les jugements sont plutôt élevés, avec peu d'énoncés recevant des notes basses, et beaucoup recevant des notes hautes (5, 6, 7). Pour les noms de contrôle, en revanche, les jugements se distribuent autour de 3 (« phrase plutôt inacceptable, je doute que quiconque dirait jamais cela »), avec très peu de notes élevées. L'application du test signé des rangs de Wilcoxon (*Wilcoxon rank-sum test*) nous montre que cette différence est très significative du point de vue statistique ( $W = 11\,960$ , valeur  $p < 0,0001$ ). (Nous optons pour un test non-paramétrique, et non pour le test  $t$  de Student, car nous ne supposons pas que nos deux variables sont distribuées de manière normale, ni que les deux échantillons ont la même variance. Cette remarque s'applique à chaque fois que nous mentionnons le test de Wilcoxon.) Ceci vérifie notre première prédiction : les noms de fruits et de légumes sont de fait plus acceptables en syntaxe massive (*de la pomme, du melon*) qu'une sélection de noms fondamentalement comptables dans ce même type de syntaxe (*\*du bocal, \*du tabouret*).

Néanmoins, puisqu'ils sont attestés en proportion moindre, nous nous attendons à ce que les emplois massifs de nos noms de fruits et de légumes soient sensiblement moins acceptables que leurs emplois comptables. Ce que notre enquête « A » nous apprend de cette question se trouve représenté à la Figure 28.

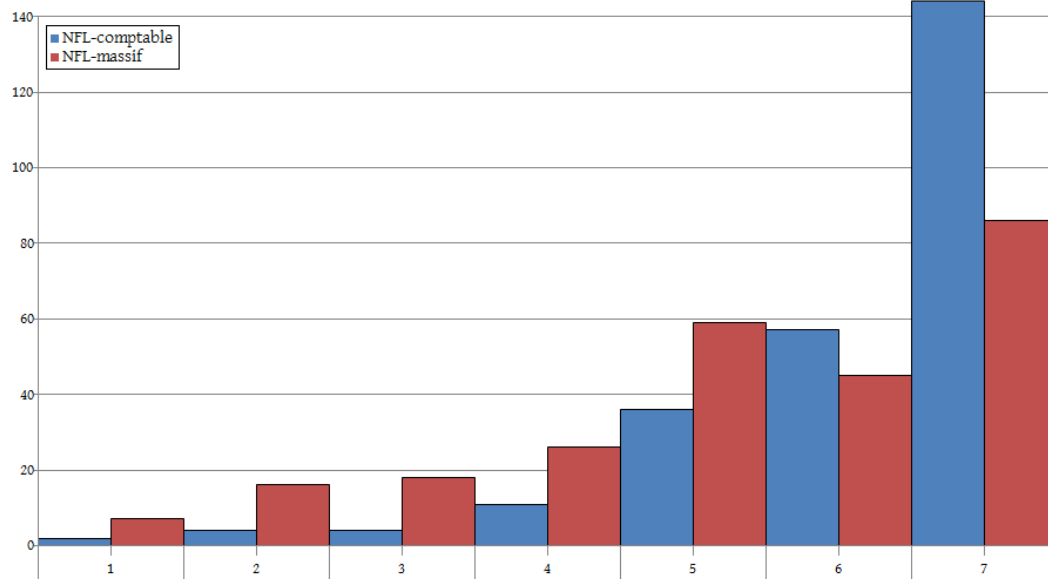


Figure 28 Acceptabilité des emplois de nos noms de fruits et de légumes, en syntaxe comptable (bleu) et massive (rouge), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

(Remarquons que les données représentées en rouge dans les Figure 27 et Figure 28 sont identiques.) A nouveau, nous constatons au premier coup d'œil une différence entre nos deux groupes d'emplois : si tant les emplois massifs que comptables de nos noms de fruits et de légumes tendent à être évalués de manière assez positive (globalement peu d'énoncés jugés au bas de l'échelle), les seconds semblent néanmoins encore un peu plus acceptables que les premiers. Le test signé des rangs de Wilcoxon confirme sur la base de nos chiffres que cette différence est significative du point de vue statistique ( $W = 43\ 544$ , valeur  $p < 0,0001$ ). A nouveau, notre prédiction se voit vérifiée.

La dernière étape de notre raisonnement sur la flexibilité des noms de fruits et de légumes concerne la corrélation entre l'acceptabilité des emplois massifs, et leur proportion dans nos données de corpus (la prédiction étant donc que les items apparaissant les plus souvent comme massifs devraient aussi être les plus acceptables dans cet emploi). Afin de pouvoir examiner cette question, nous avons d'abord besoin d'une mesure représentant, pour chacun de nos noms de fruits et de légumes, son attirance pour le massif d'après notre corpus. Ainsi, à partir de nos données de *WebCorp*, telles que présentées au Chapitre 5, nous avons établi pour chaque item un indice « I » de biais massif, calculé comme  $I = \log(M+0,1/C+0,1)$ , avec I l'indice, M le nombre d'occurrences massives et C le nombre d'occurrences comptables (l'ajout de 0,1 est nécessaire pour gérer les éventuels nombres d'attestations nuls, évitant ainsi la division par zéro ou une transformation logarithmique égalant l'infini). (Nous nous inspirons ici directement du calcul de l'attirance pour la construction dative à double objet évoqué dans (Bernolet et Hartsuiker 2010).) Les chiffres ainsi obtenus sont tous négatifs (puisque pour tous les items, la proportion de massifs observée est inférieure à la

proportion de comptables), avec une valeur d'autant plus basse que les emplois massifs attestés en corpus sont rares.

La question que nous nous posons alors est la suivante : cet indice indépendant représentant pour chaque nom de fruit ou de légume la distribution de ses attestations massives en discours, permet-il de prédire statistiquement l'acceptabilité de ces emplois, telle que mesurée dans notre enquête « A » ? Pour répondre à cette question, nous avons construit un modèle statistique de nos données (en utilisant le programme R, et plus spécifiquement le *package* « ordinal », (Christensen 2013)), en y appliquant une régression ordonnée à effets mixtes, avec l'indice « I » comme facteur continu fixe, et avec deux interceptions aléatoires (*mixed-effects ordered regression model with two random intercepts*). Les deux interceptions aléatoires sont pour les facteurs « participant », et « item » : notre modèle tient ainsi compte de l'éventuelle variation dans les jugements d'un participant à l'autre, ainsi que d'un item à l'autre (outre la variation déjà expliquée par l'indice « I »). (Le test du rapport de vraisemblance, ou *likelihood ratio test*, indique que chacune de ces deux interceptions aléatoires contribue de manière significative à notre modèle statistique.) Les résultats de cette analyse statistique sont présentés dans le Tableau 3.

Tableau 3 Estimations de notre modèle par régression ordonnée à effets mixtes de l'acceptabilité des noms de fruits et de légumes en syntaxe massive, en français, intégrant I comme facteur continu fixe, ainsi que « item » et « participant » comme interceptions aléatoires. (\*Variance et (Déviation Standard), † Valeur p basée sur le test du rapport de vraisemblance.)

	Coefficient estimé (err. stand.)	Valeur z	Valeur p
Indice « I »	0,830 (0,209)	3,98	< 0,0001
Interceptions aléatoires			
Participant (n = 43)	0,780 (0,883)*		< 0,0001 <sup>†</sup>
Item (n =12)	0,137 (0,370)*		< 0,0001 <sup>†</sup>

Dans le tableau ci-dessus, nous constatons que l'indice « I » a effectivement une influence statistiquement significative sur l'acceptabilité des emplois massifs de nos noms de fruits et de légumes (valeur  $p < 0,0001$ ). Plus précisément, plus l'indice est élevé (ce qui signifie une proportion d'emplois massifs plus grande en corpus), plus le jugement d'acceptabilité sera haut. Ainsi, la corrélation telle que nous pouvons la représenter à la Figure 29 (où elle ressort assez clairement) est statistiquement valide.

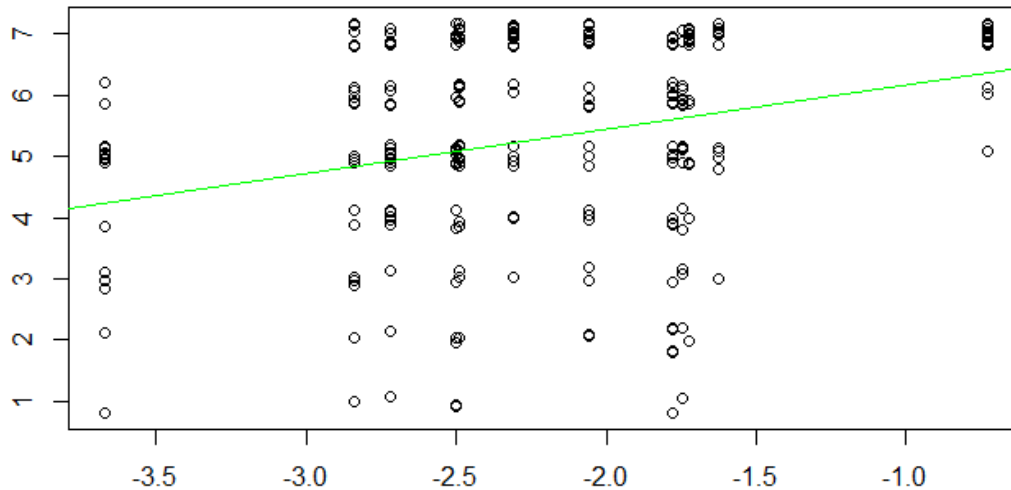


Figure 29 Corrélation positive entre l'acceptabilité d'un nom de fruit ou de légume en syntaxe massive (telle que mesurée dans notre enquête « A ») et le biais massif I calculé pour chaque item sur la base de nos données de corpus, en français. Plus le biais massif est élevé, plus l'emploi est jugé comme acceptable. Dans ce graphique, chaque point correspond à un jugement d'acceptabilité dont la valeur est marquée sur l'axe des ordonnées, concernant l'emploi massif en contexte d'un nom en particulier, dont l'indice « I » de biais massif est marqué en abscisse. [X : indice « I », Y : note d'acceptabilité.]

Ceci confirme de manière très précise notre prédiction : la représentation d'un certain emploi dans le système linguistique (tel que reflété par son acceptabilité) est fonction de sa distribution en discours.

### 9.3.2 Enquête « B »

Avec notre enquête « B », nous nous penchons à présent sur la question de l'influence du contexte phrastique sur l'acceptabilité des emplois massifs de noms de fruits et de légumes. Sur ce point, une investigation plus précise suivra au chapitre suivant. Néanmoins, par notre enquête « B », nous pouvons déjà quelque peu lever le voile sur le sujet. Comme expliqué ci-avant, cette enquête comparait l'acceptabilité des emplois massifs de noms de fruits et de légumes dans deux types de contextes : des contextes neutres (semblables à ceux de l'enquête « A »), et des contextes supposément biaisés en faveur du massif, faisant explicitement référence au domaine de la consommation culinaire. Notre hypothèse était, suivant notre modèle de l'opposition massif-comptable dans le lexique, que le type de contexte devrait avoir une influence significative sur l'acceptabilité. A la Figure 30, nous représentons les résultats obtenus. Les emplois en contexte neutre sont marqués en rouge clair, ceux en contexte biaisés en rouge foncé.

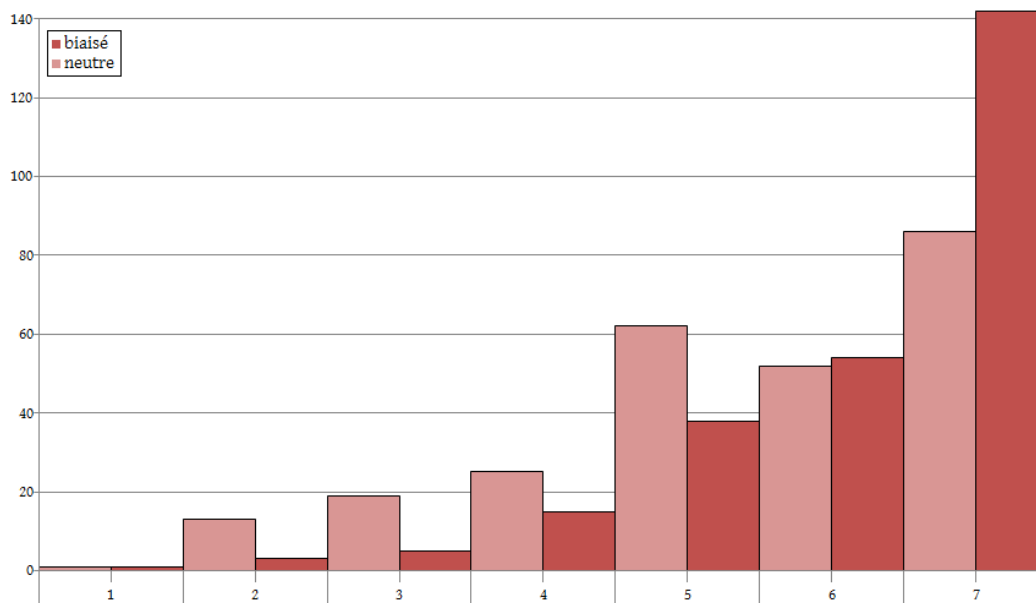


Figure 30 Acceptabilité des noms de fruits et de légumes en syntaxe massive, dans un contexte phrastique neutre (rouge clair) et orienté en termes de « consommation culinaire » (rouge foncé), en français. Nombre de participants = 43. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

A nouveau, l'on peut observer que l'acceptabilité n'est pas égale dans nos deux conditions : les énoncés biaisés ont tendance à être jugés plus positivement, plus souvent. Le test de Wilcoxon confirme que la différence observée est significative du point de vue statistique ( $W = 42\,619$ , valeur  $p < 0,0001$ ). Même si cela ne nous apprend pas encore grand-chose de la structure du niveau supra-lexical de notre modèle, cela confirme donc notre prédiction de base : que l'acceptabilité d'un emploi est modulée, outre par sa distribution en discours, également par le contexte discursif plus large dans lequel il apparaît. Notons par ailleurs que la configuration des emplois marqués ici en rouge clair apparaît très similaire à celle des emplois marqués en rouge à la Figure 28. C'est plutôt logique, puisque dans les deux cas il s'agit du jugement de l'acceptabilité des emplois massifs de nos noms de fruits et de légumes dans un contexte discursif neutre – mais c'est néanmoins intéressant, puisqu'il s'agit bien dans les deux cas de données différentes, issues pour les unes de notre enquête « B » et pour les autres de l'enquête « A ». Au mieux, cela nous donne une indication de la robustesse des jugements récoltés.

## 9.4 Résultats des enquêtes pour le néerlandais

### 9.4.1 Enquête « A »

Puisque nos enquêtes en néerlandais ont été composées de manière parfaitement parallèle à celles en français, nous pouvons en examiner les résultats de la même manière. La seule différence notable est que, dans notre enquête « A », nous avons distingué en néerlandais deux types de contextes syntagmatiques massifs : l'emploi nu, et l'emploi avec *wat*. Nous devons donc d'abord examiner si cette variation a une influence sur l'acceptabilité. A la Figure 31 ci-après, nous représentons dès lors l'acceptabilité des emplois massifs de nos noms de fruits et de légumes en néerlandais, en distinguant les deux types de syntagmes massifs étudiés.

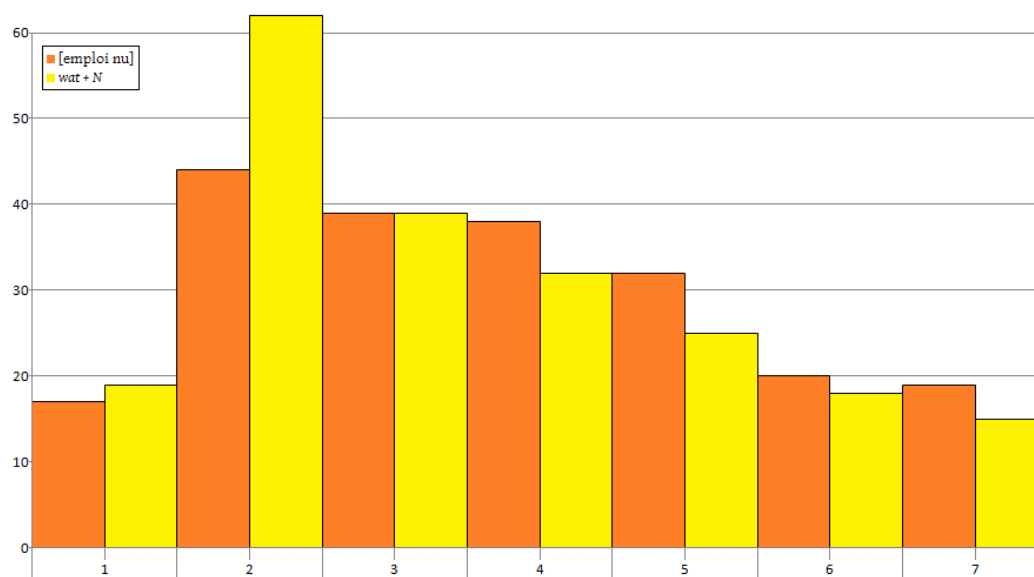


Figure 31 Acceptabilité des noms de fruits et de légumes en néerlandais, dans deux types de syntagmes : emploi nu (orange) et emploi avec *wat* (jaune). Nombre de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

Ici, contrairement à nos graphiques précédents, il est difficile d'observer une vraie différence tranchée entre les deux types emplois. Effectivement, le test de Wilcoxon le confirme : les distributions observées ne diffèrent pas significativement l'une de l'autre, c'est-à-dire que l'acceptabilité des emplois massifs des noms de fruits et de légumes en néerlandais n'est pas influencée de manière significative par le type de syntagme impliqué ( $W = 19\ 610$ , valeur  $p = 0,06$ ). Cela signifie que dans la suite de notre analyse des résultats de l'enquête « A » en néerlandais, nous pouvons regrouper les emplois nus et les emplois avec *wat* sous le chapeau commun des emplois massifs. Nos données sont donc effectivement parfaitement parallèles à celles obtenues pour le français, ce qui nous permet de les traiter de manière identique.

Ainsi, la Figure 31 correspond à la Figure 27 ci-dessus, comparant l'acceptabilité des emplois massifs des noms de fruits et de légumes à celle des emplois massifs de notre sélection de noms de contrôle.

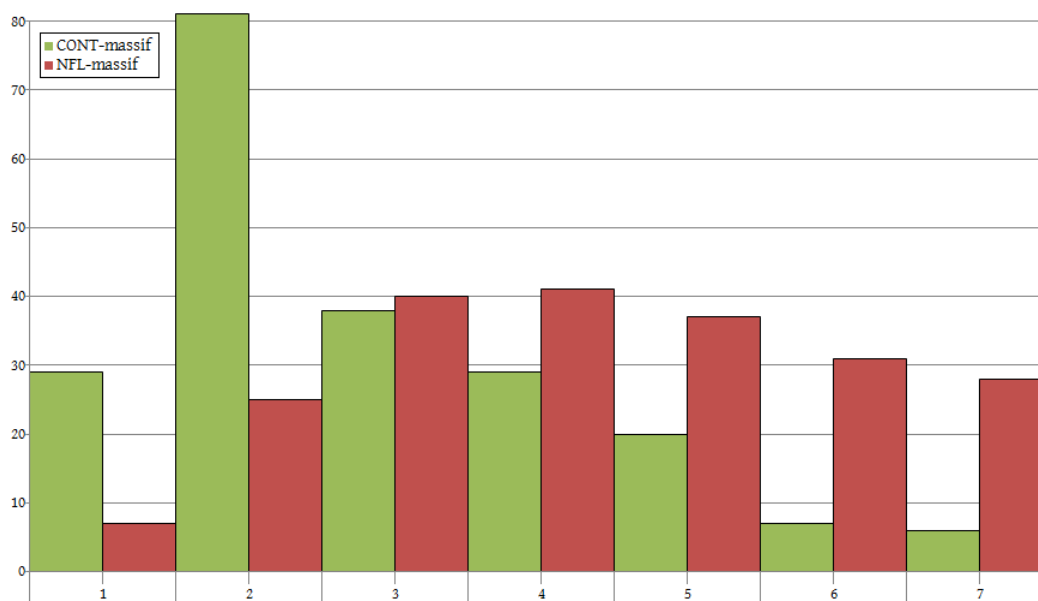


Figure 32 Acceptabilité des emplois massifs des noms de fruits et de légumes (en rouge), comparée à ceux des noms de contrôle (en vert), en néerlandais. Nombre de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

Deux choses ressortent distinctement de cette représentation. Premièrement, le profil d'acceptabilité que l'on peut observer ici n'est pas identique à celui de la Figure 27 : globalement, les emplois massifs des noms de fruits et de légumes en néerlandais sembleraient, à première vue, moins acceptables que leurs équivalents en français (nous dirons quelque chose à ce sujet dans un instant). Deuxièmement, il n'en apparaît pas moins clairement une différence entre les emplois massifs de nos deux types de noms – les noms de fruits et de légumes apparaissant visiblement plus acceptables en emploi massif que les noms de contrôle, ces derniers se trouvant plus fortement concentrés au bas de notre échelle. Le test signé des rangs de Wilcoxon nous confirme que la différence est significative ( $W = 11\,298$ , valeur  $p < 0,0001$ ). Comme pour le français, notre prédiction est donc vérifiée : les noms de fruits et de légumes, dont nous trouvons quelques attestations massives en corpus et même s'ils sont préférentiellement comptables (par ailleurs plus encore en néerlandais qu'en français) sont plus acceptables en emploi massif qu'une sélection de noms de contrôle exclusivement comptables, dont les attestations massives sont virtuellement nulles.

L'étape suivante est la comparaison de l'acceptabilité de nos noms de fruits et de légumes d'une part en syntaxe massive, de l'autre en syntaxe comptable. La Figure 33 illustre les résultats en la matière :



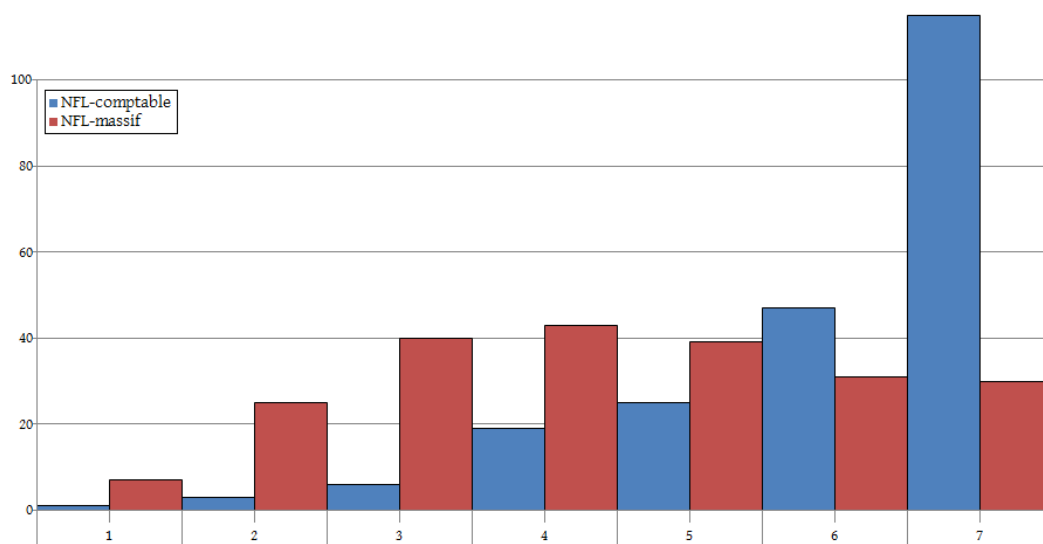


Figure 33 Acceptabilité des emplois de nos noms de fruits et de légumes, en syntaxe comptable (bleu) et massive (rouge), en néerlandais. Nombre de participants = 35. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

Comme on le voyait à la Figure 28 pour le français, la différence entre les deux types d'emplois est clairement marquée ici aussi : les emplois comptables des noms de fruits et de légumes en néerlandais se voient plus souvent attribuer un jugement d'acceptabilité élevé, en comparaison aux emplois massifs. Une fois de plus, la différence est significative d'un point de vue statistique (test de Wilcoxon,  $W = 34\ 897$ , valeur  $p < 0,0001$ ). Ces données vérifient à nouveau notre prédiction que les emplois comptables, plus fréquents en discours, sont aussi les plus acceptables.

A la comparaison de la Figure 28 et de la Figure 33, il semblerait néanmoins que les emplois massifs des noms de fruits et de légumes soient globalement moins acceptables en néerlandais qu'en français. Cela confirmerait les intuitions que nous avons récoltées de manière informelle sur le sujet (voir au point 5.2.1) – alors que les données distributionnelles issues de nos études de corpus, si elles pointaient également dans cette direction, ne se laissent pas interpréter de manière univoque (en raison de la difficulté de comparer des proportions unanimement très basses, voir point 5.2.1.3). Or, de fait, si nous comparons à l'aide du test signé des rangs de Wilcoxon le profil d'acceptabilité des emplois massifs des noms de fruits et de légumes en français et en néerlandais, le résultat indique que les différences entre nos deux langues sont significatives ( $W = 35\ 654$ , valeur  $p < 0,0001$ ). Cela signifie que la probabilité qu'un emploi massif de nom de fruit ou de légume obtienne une note d'acceptabilité élevée est effectivement plus haute en français qu'en néerlandais.

Il reste encore à examiner si, malgré tout, l'acceptabilité de ces emplois massifs en néerlandais est corrélée avec leur distribution en discours. Ainsi, comme pour le français, nous avons calculé pour chaque item utilisé dans notre enquête son indice « I » de biais massif – tel que dévoilé par son profil distributionnel en corpus (en appliquant la formule  $I = \log(M+0,1/C+0,1)$ , voir ci-dessus). Nous avons ensuite construit un modèle

statistique de l'acceptabilité des emplois massifs de noms de fruits et de légumes en néerlandais, en effectuant une régression ordonnée à effets mixtes, avec l'indice « I » comme facteur continu fixe, et le facteur « participant » comme interception aléatoire. (Contrairement à ce qui était le cas en français, l'ajout du facteur « item » comme interception aléatoire ne contribuait pas significativement à la qualité du modèle, comme nous l'a démontré le test du rapport de vraisemblance.). Le résultat de cette modélisation statistique se trouve représenté dans le Tableau 4.

Tableau 4 Estimations de notre modèle par régression ordonnée à effets mixtes de l'acceptabilité des noms de fruits et de légumes en syntaxe massive, en néerlandais, intégrant I comme facteur continu fixe, ainsi que « participant » comme interception aléatoire. (\*Variance et (Déviation Standard), † Valeur p basée sur le test du rapport de vraisemblance.)

	Coefficient estimé (err. stand.)	Valeur z	Valeur p
Indice « I »	1,048 (0,178)	5,89	< 0,0001
Interceptions aléatoires			
Participant (n = 35)	0,46 (0,678)*		< 0,0001†

Le tableau ci-dessus montre que l'indice « I » est un facteur significatif dans le modèle (valeur  $p < 0,0001$ ), et donc que l'observation que nous avons faite pour le français se vérifie aussi en néerlandais : l'acceptabilité des emplois massifs des noms de fruits et de légumes, telle que mesurée par notre enquête « A », est corrélée avec leur distribution en discours. Nous représentons cette corrélation à la Figure 34 :

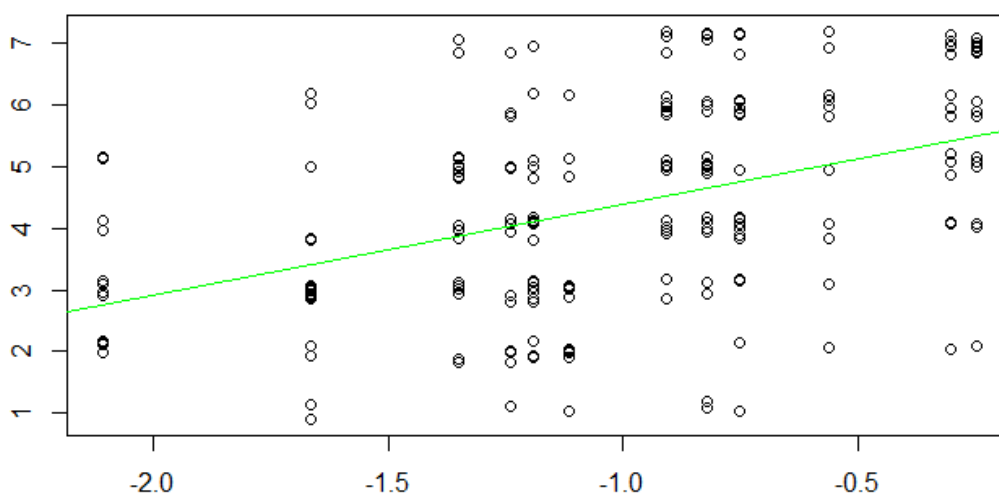


Figure 34 Corrélation positive entre l'acceptabilité d'un nom de fruit ou de légume en syntaxe massive (telle que mesurée dans notre enquête « A ») et le biais massif I calculé pour chaque item sur la base de nos données de corpus. Plus le biais massif est élevé, plus la probabilité d'une note d'acceptabilité haute augmente. [X : indice « I », Y : note d'acceptabilité.]

Nos données d'enquête en néerlandais nous permettent donc également de confirmer de manière très détaillée notre hypothèse, que l'acceptabilité d'un emploi est prédit par sa distribution en discours.

### 9.4.2 Enquête « B »

Il ne nous reste plus qu'à voir si, comme pour le français, l'acceptabilité des noms de fruits et de légumes en syntaxe massive peut être influencée en néerlandais par le contexte phrastique d'apparition. (L'on pourrait éventuellement penser que, ces emplois massifs étant si peu fréquents, leur acceptabilité resterait basse indépendamment du contexte.) La Figure 35 ci-après répond à cette question, en illustrant les résultats de notre enquête « B » en néerlandais :

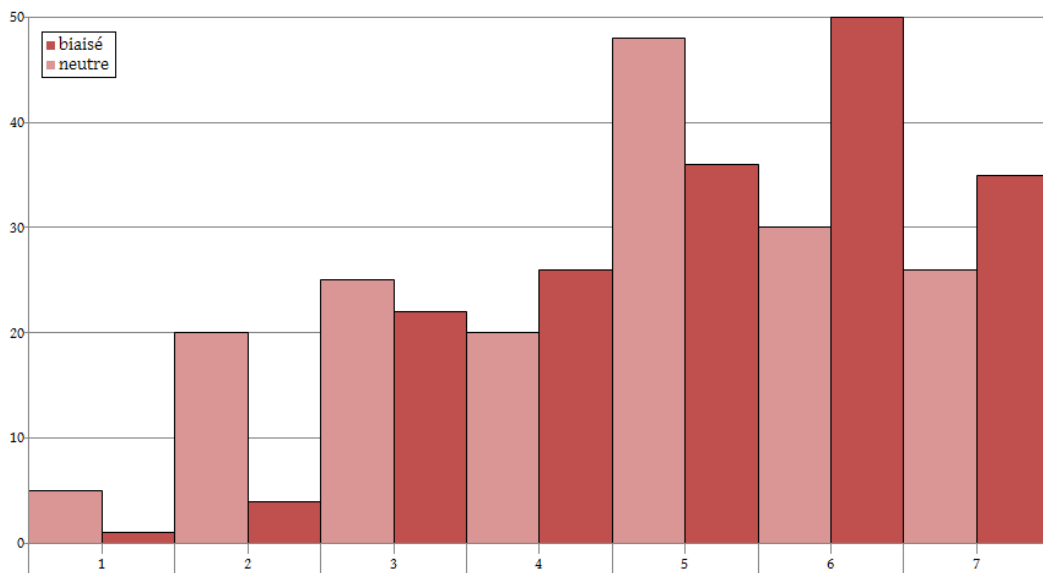


Figure 35 Acceptabilité des noms de fruits et de légumes en syntaxe massive, dans un contexte phrastique neutre (rouge clair) et orienté en termes de « consommation culinaire » (rouge foncé), en néerlandais. Nombre de participants = 29. [X : note d'acceptabilité, Y : nombre d'énoncés ayant reçu cette note.]

L'on observe ici bel et bien à nouveau une différence dans le sens de notre prédiction : les emplois massifs en contexte biaisé ont tendance à être évalués plus positivement que les emplois en contexte neutre (les énoncés représentés en rouge foncé sont plus concentrés vers la droite, et donc vers le haut de notre échelle d'acceptabilité, que ceux représentés en rouge clair). Le test de Wilcoxon, que nous appliquons pour la dernière fois, confirme que cette différence est bel et bien significative ( $W = 18\ 117$ , valeur  $p < 0,001$ ). Ceci confirme donc, comme le faisaient déjà les résultats de l'enquête « B » en français, la pertinence du niveau supra-lexical de notre modèle du massif-comptable dans le lexique : non seulement les propriétés distributionnelles lexicales, mais aussi le contexte discursif (supra-lexical) définissent l'acceptabilité des emplois.

## 9.5 Discussion et conclusions

Notre modèle de l'opposition massif-comptable dans le lexique mental est intéressant car il permet de faire des prédictions précises quant à l'usage en discours des noms morphosyntaxiquement flexibles (alternant régulièrement entre emploi massif et comptable). Ainsi, si nous représentons les propriétés lexicales des items en termes de connexions – renforcées par l'usage – entre un lemme et des nœuds combinatoires (dans le cadre d'un modèle à activation du système linguistique), nous nous attendons à une corrélation entre fréquence d'emploi et acceptabilité : plus un emploi est courant, plus la connexion lemmatique est forte, et ainsi plus l'emploi paraîtra naturel et acceptable. C'est là précisément le lien que nous avons démontré au travers de nos enquêtes « A », en français et en néerlandais. D'une part, nous avons montré qu'il y avait une différence nette entre l'acceptabilité d'emplois virtuellement inexistantes en discours (en l'occurrence, les usages massifs de nos noms de contrôle, pour lesquels nous ne postulons pas de connexion lemmatique avec le nœud combinatoire M) et l'acceptabilité d'emplois attestés, fût-ce en proportions modestes (en l'occurrence, les usages massifs des noms de fruits et de légumes : puisqu'ils sont attestés, il existe une connexion entre le nœud M et les lemmes concernés). D'autre part, nous avons précisé encore le lien entre fréquence et représentation dans le lexique mental en mettant en évidence une corrélation statistique robuste entre l'acceptabilité des emplois massifs des noms de fruits et de légumes, et la distribution de ces emplois en discours – montrant que les subtiles différences inter-lexicales calculables à partir de nos données de corpus (et reflétées par notre indice « I ») se retrouvaient sans plus dans les profils d'acceptabilité, tels qu'ils ressortent de nos enquêtes.

En outre, si nous modélisons l'emploi massif ou comptable d'un nom en discours comme le résultat d'une interaction entre le niveau lexical distributionnel et un niveau sémantique supra-lexical, dépassant l'échelon des items spécifiques et associant directement certains sens à certaines options combinatoires, alors nous nous attendons également à ce que l'acceptabilité d'un emploi, indépendamment de la variation inter-lexicale, soit modulée par le contexte phrastique plus large dans lequel il apparaît. Or, la différence statistique que nous avons observée entre l'acceptabilité des emplois massifs de nos noms de fruits et de légumes dans deux types de contextes (neutres, et biaisés en faveur d'une lecture en termes de 'nourriture'), au travers de notre enquête « B », confirme clairement cette prédiction : le contexte joue lui aussi un rôle non négligeable dans la flexibilité de l'opposition massif-comptable.

Ces résultats d'enquête permettent ainsi de confirmer la nature distributionnelle de l'opposition massif-comptable dans le système linguistique, telle que postulée dans notre représentation stratifiée à activation du lexique mental. En outre, s'ils ne permettent pas encore de décrire en détail la structure du niveau supra-lexical de notre

modèle, ils indiquent au moins la pertinence de ce niveau de représentation complémentaire, en constante interaction avec le niveau lexical. Ainsi, ce ne sont pas seulement les noms qui ont des préférences, mais aussi les « sens ». Et c'est de l'interférence entre ces deux données que résultent les emplois, massifs ou comptables, en discours.

Les deux visions traditionnelles de l'opposition massif-comptable, et de sa représentation dans la langue, ne savent quant à elles pas dire grand-chose des données présentées dans le présent chapitre. Dans la vision lexicaliste de base, deux possibilités seraient à distinguer : soit les noms de fruits et de légumes sont encodés avec un double marquage lexical (et sont donc à la fois massifs et comptables), soit ils sont fondamentalement comptables (et leurs emplois massifs sont le résultat de transferts sémantico-syntaxiques, en contexte). Dans le premier cas, cependant, le modèle ne prédit aucune des différences d'acceptabilité que nous avons mises en évidence, ni au niveau de la différence entre massif et comptable, ni au niveau de la variation contextuelle : tous les emplois devraient être acceptables de manière équivalente – ce que nos données contredisent nettement. Dans le second cas, le modèle prédit effectivement un effet contextuel, comme celui que nous avons observé dans notre enquête « B » : si le contexte se prête mieux au massif, le « transfert » sera plus facile, et son résultat paraîtra plus naturel. Cependant, cette vision n'explique nullement la corrélation entre données distributionnelles et acceptabilité. Or, si la proportion plus ou moins grande d'emplois massifs d'un certain nom en corpus est reflétée dans leur acceptabilité, alors c'est que cette proportion doit se trouver encodée quelque part. La vision lexicaliste traditionnelle ne prévoit rien à ce sujet.

La vision grammaticale, quant à elle, prédit tout et ne prédit rien à la fois. Dans cette approche, aucun emploi n'est exclu *a priori* dans le système linguistique, et le modèle prévoit seulement qu'il y aura, probablement, de la variation quant à l'acceptabilité de tel ou tel emploi, en fonction de facteurs référentiels, contextuels et/pragmatiques – qui ne doivent cependant pas être pris en compte dans la description de la langue, car ils ne relèvent pas réellement du domaine de la linguistique. Nous avons déjà dit à plusieurs reprises ci-avant nos réserves quant à une telle vision de l'étude de la langue : le langage est un phénomène hautement dynamique et psychologique, et sa description ne peut se passer de l'interprétation et de l'explication de sa variabilité intrinsèque. Plus concrètement pour ce qui est de notre sujet, si l'alternance entre emplois massifs et comptables découle de l'interaction de plusieurs facteurs, il s'agit bel et bien de les décrire et de les modéliser – et non de les exclure d'avance comme inintéressants.

Ainsi, en comparaison aux deux visions traditionnelles de l'opposition massif-comptable (et de sa flexibilité), notre modèle constitue une amélioration sensible, en permettant une description plus précise des données d'enquête que nous avons récoltées. C'est de bon augure. Il conviendra néanmoins que nous continuions de le tester, en le soumettant à présent – après la validation de l'interprétation – au test de la

production, en contexte. C'est ce que nous avons fait, au travers d'une expérience d'amorçage. Nous la présentons au chapitre suivant.

# Chapitre 10 Le statut supra-lexical de l'opposition massif/comptable : expérience d'amorçage

## 10.1 Introduction

Par définition, les jugements d'acceptabilité ne disent quelque chose que de l'aspect interprétatif du langage. Or, ils sont d'autant plus fiables lorsque corroborés par des données complémentaires, en particulier des données de production, récoltées de préférence dans des circonstances contrôlées, proprement empiriques (voir (Gilquin et Gries 2009)). Dans cet esprit, nous avons dès lors choisi d'examiner aussi l'opposition massif-comptable sous la facette de la production, au travers d'une expérience d'amorçage dont le dispositif s'inspire de (Cleland et Pickering 2003) (voir ci-dessus, point 3.2.2, et notre explication ci-après).

Notre objectif, en cela, est de continuer à tester les prédictions de notre modèle, afin de mieux comprendre les facteurs influençant la production d'un nom flexible, en syntaxe soit massive, soit comptable. Dans l'idée de rester cohérent avec les données employées pour nos enquêtes, les items que nous avons examinés de cette manière sont, à nouveau, les noms de fruits et de légumes (avec une sélection étendue cependant, comme nous le verrons). Nous n'avons, du reste, organisé notre expérience qu'en français, essentiellement en raison de la flexibilité plus grande des noms dans cette langue – les items en néerlandais risquant de n'être jamais produits que comme comptables (l'absence de variation ne nous permettant alors pas de tirer de conclusions). Notre expérience d'amorçage est donc la seule partie empirique de notre recherche qui ne sera pas comparative – à l'inverse de nos études de corpus et de nos enquêtes.

Nous avons conçu notre expérience comme un moyen d'en apprendre davantage au sujet, principalement, du niveau supra-lexical de la représentation du massif-comptable dans le lexique – par la mise en évidence, sur ce point, de la cohérence de classe que

nous avons maintes fois évoquée. Notre idée de base est que s'il y a bien une cohérence au niveau des emplois massifs et comptables des noms d'une même catégorie, alors celle-ci devrait pouvoir être mise en évidence au travers d'un dispositif expérimental. En considérant la littérature existante sur l'amorçage structurel (en particulier dans le cadre d'une modélisation du lexique à la (Levelt *et al.* 1999)), nous nous attendons donc à trouver, pour des noms flexibles tels que les noms de fruits et de légumes, un effet d'amorçage du massif par le massif, doublé d'un *boost* sémantique au cas où l'amorce massive serait un autre nom de la même catégorie – à l'instar des observations de (Cleland et Pickering 2003) dans leur étude de la production de différents types de SN (voir notre point 3.2.2 ci-dessus). Ainsi, l'emploi massif d'un nom de fruit ou de légume, s'il est généralement moins probable qu'un emploi comptable, devrait se trouver favorisé dans la production – par comparaison à un état initial – dans un contexte où il est précédé par un autre massif, et plus encore par un autre massif de la même catégorie. C'est cette prédiction de base, déjà vérifiée pour d'autres propriétés lemmatiques, que notre expérience avait l'ambition première de confirmer (nous verrons que nos résultats, de par le statut particulier du massif-comptable comme opposition à cheval entre la sémantique et la morphosyntaxe, sont plus nuancés qu'une simple confirmation de résultats antérieurs).

Dans ce qui suit, nous présentons d'abord notre dispositif expérimental, tel que nous l'avons imaginé, puis mis en application (10.2). Ensuite, nous détaillons nos résultats, et la manière dont nous les avons traités statistiquement (10.3). Pour finir, nous discutons ce que ces résultats nous apprennent du statut de l'opposition massif-comptable, dans le lexique mental (10.4).

## 10.2 Dispositif expérimental

### 10.2.1 Concept général

Le dispositif expérimental que nous avons mis en place s'inspire de (Cleland et Pickering 2003), qui eux-mêmes s'inspiraient de (Branigan *et al.* 2000). Il s'agit d'un exercice de conversation dirigée entre deux participants (l'un naïf, l'autre partiellement complice : dans ce qui suit, nous les appelons respectivement #N et #C), interagissant sur la base d'instructions précises (présentées, dans notre cas, sur 104 fiches-consignes classées, par participant, par session). Les instructions correspondaient soit à une phrase à trou à lire à voix haute, soit à une indication permettant de compléter une phrase à trou lue par le partenaire (sous forme d'une image ou, pour #C dans les interactions critiques,



une réponse écrite préétablie, voir ci-dessous). Chaque phrase complétée devait permettre d'identifier une image, parmi une sélection présentée à chaque participant.

Concrètement, les deux participants étaient assis face à face, de chaque côté d'une grande table, avec devant chacun d'eux 54 images en couleur (d'environ 5 x 7 cm), étalées de manière à être toutes bien visibles. Une séparation au milieu de la table (faite de divers livres alignés) empêchait qu'un participant puisse voir les images déposées devant l'autre – sans empêcher cependant les deux participants de se voir mutuellement. Tour à tour, les deux participants prenaient la parole en suivant leurs cartes d'instructions, soit pour lire à voix haute une phrase à trou, soit pour compléter la phrase à trou lue par leur partenaire. Après avoir lu une phrase à trou, et sur la base de la réponse fournie par son partenaire, un participant devait sélectionner et mettre de côté l'une des images disposées devant lui – cette tâche de reconnaissance étant présentée comme la tâche principale de l'expérience. La disposition générale est illustrée à la Figure 36.

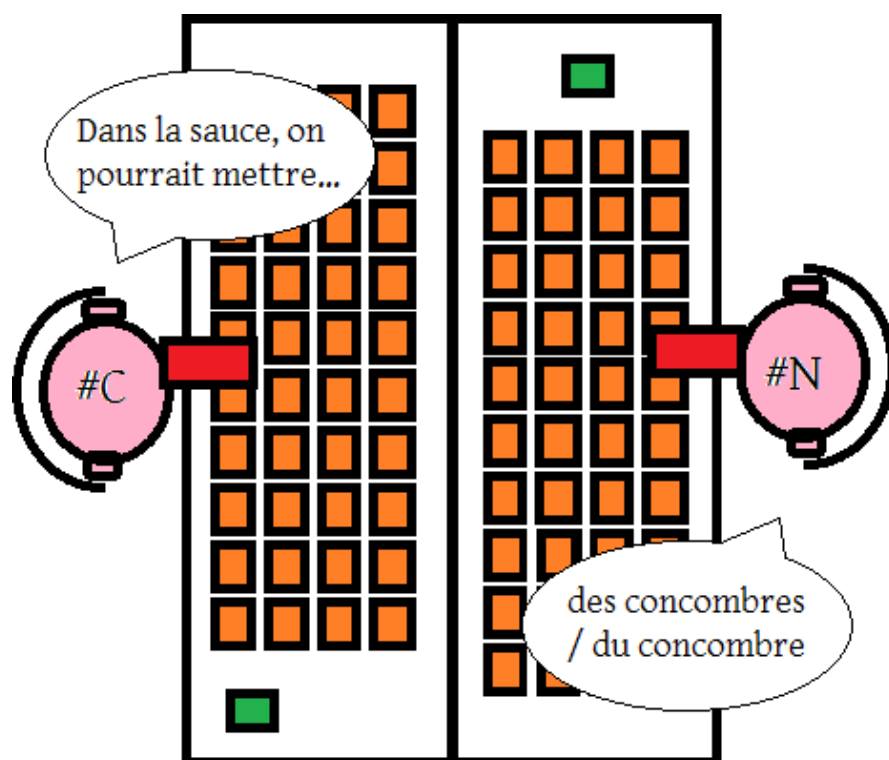


Figure 36 Dispositif expérimental. En orange : sélection de 54 images, disposées face à chaque participant. En rouge : fiches-consigne (en alternance, une phrase incomplète à lire puis une image ou, pour le complice dans le cas d'une interaction critique, une réponse préétablie). En vert : la pile de cartes sélectionnées et mises sur le côté. Dans l'exemple illustré ici, #C devra à présent répéter la phrase complète (*Dans la sauce, on pourrait mettre des/du concombre(s)* – en fonction de la réponse de #N) puis retrouver, parmi les cartes disposées devant lui, celle représentant un/du/des concombre(s).

Plus précisément, chaque session de l'expérience comportait 54 interactions (dont 18 interactions critiques, et 36 *fillers*), elles-mêmes composées de la succession des 6 étapes suivantes :

1. #N lit une phrase incomplète ;
2. #C répond en complétant la phrase ;
3. #N répète la phrase à présent complète, puis sélectionne et met de côté l'une des images disposées devant lui (celle qu'il suppose correspondre au mieux à la réponse de #C) ;
4. #C lit une phrase incomplète ;
5. #N répond en complétant la phrase ;
6. #C répète la phrase à présent complète, puis sélectionne et met de côté l'une des images disposées devant lui (celle qu'il suppose correspondre au mieux à la réponse de #N) ;

(Aux étapes 3 et 6, la phrase complète est d'abord répétée, afin de s'assurer que la réponse a été bien entendue et comprise.) Dans les interactions non critiques (*fillers*), il n'y avait aucune différence entre les consignes des deux participants. Ainsi, à l'étape 2 et 5, l'instruction reçue pour compléter la phrase était sous la forme d'une image. En revanche, dans une interaction critique, la réponse de #C à l'étape 2 était préétablie (l'instruction n'était pas une image, mais quelques mots à lire) et constituait alors l'amorce (*prime*) – la réponse de #N à l'étape 5 étant dans ce cas la cible (*target*).

Cette cible était systématiquement la production (dans un syntagme) d'un nom de fruit ou de légume – sur la base d'une représentation permettant normalement tant une conceptualisation massive que comptable (dans la mesure du possible, nous avons employé des images combinant la présentation d'exemplaires entiers ou au moins facilement identifiables, aux côtés de morceaux découpés, voir les exemples à la Figure 37 et Figure 38, afin de ne pas exclure *a priori* l'une des options syntaxiques ; l'ensemble des images de fruits et de légumes employées se trouvent en Appendice, au point 12.6.1). Dans ce cas, la phrase incomplète lue par #C à l'étape 4 était un contexte profilant le domaine sémantique de la 'nourriture' (au sens large, à la manière des contextes biaisés de notre enquête « B ») – ceci afin d'augmenter les chances de réponses massives, et d'éviter un effet plancher, au cas où aucun participant n'opterait pour ce type de syntaxe. La Figure 36, ci-avant, donne un exemple d'une telle phrase. Les autres seront présentées au point 10.2.2 ci-dessous, à la Figure 39.



Figure 37 Image utilisée dans notre expérience pour éliciter la production d'un SN contenant le nom *carotte*. *A priori*, la présentation n'exclut ni ne nécessite l'usage de la syntaxe massive ou comptable.

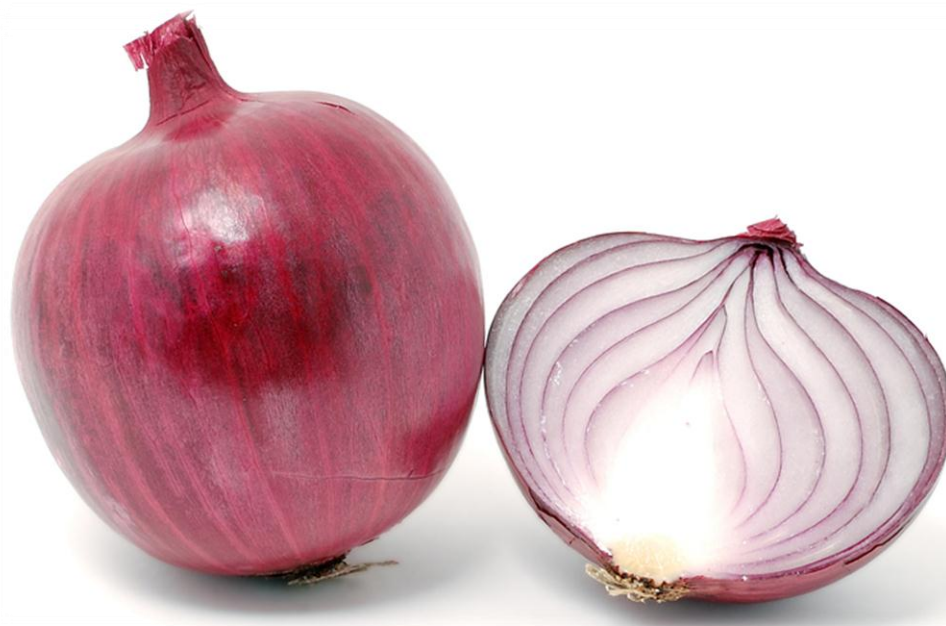


Figure 38 Image utilisée dans notre expérience pour éliciter la production d'un SN contenant le nom *oignon*. *A priori*, la présentation n'exclut ni ne nécessite l'usage de la syntaxe massive ou comptable.

Les réponses préétablies de #C, précédant les réponses critiques de #N, étaient quant à elles de trois types, correspondant à trois conditions d'amorçage : soit un autre nom de fruit ou de légume en syntaxe massive (condition d'amorçage avec lien de co-hyponymie entre amorce et cible), soit un syntagme massif sans rapport avec le domaine des fruits et légumes (condition d'amorçage sans lien sémantique : amorce massive de contrôle), soit une expression sans trace de syntaxe massive, ni rapport avec

les fruits et légumes – généralement un verbe ou un adjectif (condition *baseline*, ou état initial).

Dans l'analyse de nos résultats, nous serons donc intéressé par les productions de #N à l'étape 5 des 18 interactions critiques de chaque session – plus précisément par la syntaxe employée, massive ou comptable. Nous nous attendons, si l'on en croit les expériences d'amorçage similaires réalisées dans d'autres domaines de la morphosyntaxe (structure du SN chez (Cleland et Pickering 2003), alternance dative chez (Pickering et Branigan 1998)) à trouver un effet d'amorçage structurel de la syntaxe massive par la syntaxe massive, d'autant plus fort en cas de co-hyponymie entre amorce et cible. En outre, en accord avec notre modélisation du massif-comptable dans le lexique mental, nous devrions pouvoir observer les effets tant de notre niveau lexical (préférences distributionnelles) que supra-lexical (variation sémantique contextuelle). Avant d'en arriver à nos résultats, nous précisons d'abord encore la nature des stimuli que nous avons employés pour ce qui est des items étudiés, des phrases dans lesquelles nous les avons fait produire, des amorces et des *fillers*.

### 10.2.2 Stimuli : items, phrases, amorces, *fillers*

Notre expérience comporte 18 interactions critiques (où un nom de fruit ou de légume doit être produit), dont un tiers, soit 6, nécessitent en outre une amorce qui soit elle-même un nom de cette catégorie (employé en syntaxe massive). Notre dispositif nécessite dès lors une sélection de 24 noms de fruits et de légumes. (Nous ne nous sommes pas contenté à nouveau simplement de nos douze items les plus fréquents, comme pour nos enquêtes, afin de maximiser le nombre d'interactions critiques par session.) Celle-ci correspond à notre sélection de 29 items étudiés en corpus (Chapitre 5), en excluant *ail*, *calebasse* et *ciboulette* (parce que nous avons montré qu'ils n'étaient pas des membres à part entière de la catégorie des fruits et légumes) ainsi que *cédrat* (de par sa fréquence très basse et l'impossibilité virtuelle pour le profane de le distinguer du citron sur la base d'une image), de même qu'en couplant *citrouille* et *potiron* – susceptibles d'être élicités sans discrimination par la même image (dans le cas où cet item se trouvait employé comme amorce, nous avons opté à chaque fois pour le syntagme *du potiron*). Les items impliqués dans notre expérience sont dès lors les suivants :

*banane, carotte, cerise, chou-fleur, citron, concombre, cornichon, courge(tte), échalote, framboise, haricot, mandarine, melon, oignon, orange, pamplemousse, pastèque, poireau, patate, poire, pomme, potiron, raisin, tomate*

Notons que de cette sélection finale, 12 items dénotent des fruits, et 12 des légumes, tandis que 11 sont masculins, et 13 sont féminins.

A chaque fois qu'un nom de fruit ou de légume devait être produit (24 fois par session, dont 6 fois par #C en guise d'amorces, et 18 fois par #N), nous avons employé un contexte phrastique profilant explicitement le domaine de la 'nourriture' – comme nous l'avons déjà évoqué ci-avant. Nous avons jugé que ces contextes se prêtaient tant à la syntaxe massive que comptable et, forts des résultats de notre enquête « B », qu'ils augmenteraient globalement les chances des productions de syntagmes massifs (les emplois massifs étant les moins fréquents, et donc *a priori* les moins probables). Les 24 phrases que nous avons employées se trouvent à la Figure 39.

#	PHRASES (contextes "nourriture")
1	Au supermarché, on m'a fait goûter _.
2	Dans le frigo, il reste _.
3	Le cuisinier a demandé qu'on lui fournisse _.
4	Dans la sauce, on pourrait mettre _.
5	Dans la casserole, Jean a rajouté _.
6	Voici une délicieuse recette avec _.
7	Ma mère prépare toujours cette recette avec _.
8	Chez les Devaux, on a encore mangé _.
9	Pour les gourmands, voici une entrée qui contient _.
10	Dans l'assiette d'Alice, il reste _.
11	Ce petit goût typique, je pense que c'est _.
12	Hier, Julie a goûté un cocktail surprenant qui contenait _.
13	Dans le fond du plat, commencez par disposer _.
14	Dans son alimentation, Maxime a décidé d'inclure _.
15	C'était la première fois que mon enfant mangeait _.
16	Pour le repas de ce soir, j'ai trouvé _.
17	Sonia cuisine souvent avec _.
18	Vous pouvez décorer les verrines avec _.
19	Pour cette drôle de sauce, il faut mélanger de l'eau et _.
20	La salade du jour contient _.
21	Au moins une fois par semaine, Carine mange _.
22	Chez tous les maraîchers, on trouve _.
23	J'ai goûté une délicieuse préparation avec _.
24	C'est marrant cette odeur de nourriture, on dirait _.

Figure 39 Liste des 24 phrases employées dans notre expérience d'amorçage pour la production de SN comportant un nom de fruit ou de légume (contextes profilant explicitement le domaine de la nourriture). Dans une interaction critique, #C aurait l'une de ces phrases à lire, et #N devrait la compléter sur la base d'une image d'un fruit ou d'un légume.

Notons que la seule évocation du domaine de la nourriture n'est cependant pas le fin mot de l'histoire : comme nous l'avons déjà annoncé, le niveau sémantique du lexique mental est plus complexe que ce que notre enquête « B » a voulu en examiner, et l'attirance pour le massif de chacun des 24 contextes ci-dessus n'est pas simplement

équivalente. Nous examinerons cette question ci-après, au point 10.3.1 de l'analyse de nos résultats.

Concernant les amorces, outre 6 de nos 24 noms de fruits et de légumes, nous avons également besoin pour chaque session de 6 noms fondamentalement massifs d'une catégorie différente (amorces massives de contrôle), et de 6 stimuli complètement indépendants constituant notre condition d'état initial, ou *baseline* – ainsi que de phrases correspondantes, à chaque fois. Afin de permettre une certaine variation d'une session à l'autre quant aux amorces utilisées, nous avons dès lors établi pour chacune de ces deux conditions une liste de 14 paires alliant une phrase incomplète (à lire par #N) et une réponse appropriée (préétablie pour #C), répondant à nos critères – dont 6 ont été repris dans chaque session. Les énoncés (271) et (272) sont des exemples respectivement d'une amorce *baseline*, et d'une amorce massive de contrôle (l'on trouvera les autres en Appendice, aux points 12.6.2.1 et 12.6.2.2). (A noter qu'une réponse préétablie pour #C, qui trouve donc à ce moment dans sa pile de fiches-consignes du texte à lire au lieu d'une image, implique quand même qu'une image correspondante se trouve disposée devant #N, qui doit pouvoir la retrouver – ainsi, dans (268), #N devrait trouver parmi les images étalées devant lui, celle correspondant à la réponse « fatigué ».)

(271) #N : Quentin n'a pas assez dormi, il est... / #C : Fatigué.

(272) #N : Cette bête mange uniquement du foin et... / #C : De l'herbe.

Pour finir, nous avons aussi besoin de stimuli pour constituer les interactions non critiques, et détourner l'attention des participants du sujet de l'expérience (à savoir, l'emploi massif ou comptable des noms de fruits et de légumes). A cette fin, nous avons dès lors rassemblé du matériel pour constituer 36 interactions de ce type, soit 72 phrases à compléter, et 72 images pouvant constituer une réponse appropriée. Afin que le domaine des fruits et légumes (évoqué tout de même dans 24 phrases sur les 108 échangées au cours d'une session), ne soit pas trop saillant, nous avons construit ces *fillers* autour de trois grandes thématiques, à savoir les moyens de transport et la circulation (*Le Boeing 747 est le plus célèbre des...*, *Les feux rouges indiquent aux voitures qu'il faut...*), les animaux (*On dit que le meilleur ami de l'homme, c'est...*) et l'architecture (*Le bâtiment le plus élevé au monde est...*). Nous y avons également intégré quelques proverbes (*Aux grands maux, les grands...*), et d'autres phrases n'ayant aucun rapport avec le reste (*Ce mur n'est pas en plâtre, mais bien...*). Comme l'indiquent déjà les quelques exemples présentés ci-avant, nous avons du reste composé les phrases pour que les réponses attendues alternent entre syntagmes nominaux (*Dans les bois, Marie a construit...*), prépositionnels (*Certains prennent l'autobus, d'autres viennent...*), verbaux (*Dans certains pays, les femmes n'ont pas le droit de...*) et adjectivaux (*Le pelage de ce rongeur est...*). Pour chaque phrase imaginée, nous avons ensuite sélectionné, à partir de recherches sur *Google Images*, des images évoquant une réponse appropriée : une photo de souris pour la phrase *Chaque année en laboratoire on dissèque des milliers...*, une photo de girafes pour la

phrase *Si je pars en safari, j'aimerais voir...*, etc. En Appendice (point 12.6.3), l'on pourra trouver les 72 phrases employées comme *fillers* dans notre expérience, et les images correspondantes.

### 10.2.3 Fiches-consignes

Voilà pour le concept général de notre dispositif expérimental, et pour les stimuli employés. Il reste cependant encore à traduire nos idées dans la pratique. Ainsi, comme nous l'avons dit, les prises de parole dans notre expérience sont réglées par une certaine succession de fiches-consignes, présentées dans un ordre précis et coordonné entre les deux participants, indiquant à ces derniers ce qu'ils doivent dire à chaque moment. Créer les fiches consignes n'est donc pas l'étape la plus compliquée : il suffit d'imprimer des cartes représentant des phrases incomplètes (à lire), des images (à nommer, de manière à compléter ces phrases) ou des réponses préétablies (à lire également), en reprenant les stimuli tels que nous les avons détaillés au point précédent – et sans oublier d'imprimer aussi une version de chaque image (y compris de celles correspondant aux amorces) à disposer sur la table, devant chaque participant.

Le défi est plutôt dans la manière d'organiser ces fiches : quelles consignes, et dans quel ordre, va recevoir chaque participant ? Et quelles images doivent être disposées devant lui ? Afin d'élucider la question, nous avons créé un script *Perl* classant virtuellement, et autant de fois qu'on le désire, nos cartes dans un certain ordre (indiquant dans quelle pile les placer, et à quelle position), en respectant tous nos critères. (Notons qu'outre les éléments déjà évoqués ci-dessus, nous avons également encodé dans notre script que la relation du point de vue du genre entre amorces et cibles devait être variée systématiquement : même genre dans la moitié des cas, et genres opposés dans l'autre – les amorces non nominales ne comptant évidemment pas dans cet équilibre. De plus, nous avons prévu que la première interaction d'une session ne soit jamais une interaction critique.) Au total, nous avons généré ainsi 10 ordres différents, que nous avons utilisés en rotation lors de la mise en place concrète de notre expérience (voir au point suivant).

Plus précisément, à chaque ordre correspondait un code numéroté pour chaque carte (consigne ou image), indiquant respectivement, pour l'ordre en question, la position précise de cette fiche dans les piles de consignes ou la position de l'image d'un côté ou de l'autre de la table. Puisque nous avons généré 10 ordres, nous avons ainsi obtenu une liste de 10 codes par carte, que nous avons imprimée au dos de chacune, afin de pouvoir les reclasser avec facilité (à défaut de rapidité, puisqu'on parle ici de plus de 300 fiches) selon un ordre ou l'autre. Dans notre expérience, nous avons utilisé ces dix ordres par rotation, d'une session à l'autre, en employant en alternance 4 jeux de cartes complets classés de manière différente – et modifiant le classement de chaque jeu d'un jour

d'expérimentation à l'autre (voir ci-après). Cette façon de procéder a permis que d'une session à l'autre, les stimuli ne soient pas présentés toujours dans le même ordre, que les noms de fruits et de légumes employés comme amorces (6 sur 24) ou comme cibles (18 sur 24) varient, que les amorces d'état initial et de contrôle (2 x 6 sur 14) soient différentes à chaque fois, et qu'une même amorce ne soit pas associée systématiquement à la même cible.

Pour finir, notons que pour chaque ordre généré automatiquement, notre script fournissait aussi un tableur dans lequel noter nos observations, organisé en tenant compte de la succession des interactions selon cet ordre-là (avec deux lignes par interaction, une pour la réponse de chaque participant, et donc 108 lignes par session). On peut en voir un exemple à la Figure 40. Dans ce tableau, toutes les cellules hormis celles indiquées en rouge, ont été générées et complétées automatiquement par notre script (nous économisant de longues heures de travail d'encodage manuel, eussions-nous fait l'économie de l'assistance de *Perl*). La présence dans la colonne SENT des phrases prononcées successivement nous permettait du reste de bien suivre les interactions, et de noter les résultats au bon endroit.

	A	B	C	D	E	F	G	H	J	K	L	M
1	#SESS	TRIAL	ORDER	PART	TYPE	SENT	TAG_SENT	PRIME	TARGET	RESP	SYNT_NFV	LEMMA
2	1	1a	0	conf	filler	La construction a été annulée faute de permis de _.	F31	NA	F31-pic	bâtit	NA	NA
3	1	1b	0	naïve	filler	Ce matin, pour arriver à l'heure, j'ai dû _.	F37	NA	F37-pic	courir	NA	NA
4	1	2a	0	conf	prime	Dans la sauce, on pourrait mettre _.	S04	NA	de la carotte	de la carotte	NA	NA
5	1	2b	0	naïve	critical	Pour les gourmands, voici une entrée qui contient _.	S09	massNFV	[courge(tte)]	des courgettes	count_plur	COURGETTE
6	1	3a	0	conf	filler	Au volant, il est très important de _.	F59	NA	F59-pic	mettre sa ceinture	NA	NA
7	1	3b	0	naïve	filler	Aux grands maux, les grands _.	F67	NA	F67-pic	moyens	NA	NA
8	1	4a	0	conf	filler	Mon père est très prudent et n'a jamais _.	F58	NA	F58-pic	été flashé	NA	NA
9	1	4b	0	naïve	filler	En général, les mâles sont _.	F47	NA	F47-pic	bruns	NA	NA
10	1	5a	0	conf	filler	Des jeunes vandales ont mis le feu à _.	F08	NA	F08-pic	un bus	NA	NA
11	1	5b	0	naïve	filler	Carine ne supporte pas _.	F69	NA	F69-pic	la fumée	NA	NA
12	1	6a	0	conf	filler	C'est Louis XIV qui a commandité les plus grands trav.	F34	NA	F34-pic	Versailles	NA	NA
13	1	6b	0	naïve	filler	Les bunkers les plus résistants sont _.	F72	NA	F72-pic	en métal	NA	NA
14	1	7a	0	conf	filler	Le président des Etats-Unis occupe _.	F33	NA	F33-pic	la Maison Blanche	NA	NA
15	1	7b	0	naïve	filler	Quand j'étais petit, je voulais devenir _.	F07	NA	F07-pic	architecte	NA	NA
16	1	8a	0	conf	filler	On tire du lait de plusieurs animaux, pas seulement c	F05	NA	F05-pic	la vache	NA	NA
17	1	8b	0	naïve	filler	Dans leur sac à dos, les aviateurs transportent _.	F40	NA	F40-pic	un parachute	NA	NA
18	1	9a	0	conf	prime	L'entrepreneur a acheté _.	C02	NA	de la peinture	de la peinture	NA	NA
19	1	9b	0	naïve	critical	Au supermarché, on m'a fait goûter _.	S01	control	[mandarine]	une mandarine	count_sing	MANDARINE
20	1	10a	0	conf	filler	En montagne, les secours arrivent généralement _.	F21	NA	F21-pic	en hélicoptère	NA	NA
21	1	10b	0	naïve	filler	Il faut battre le fer quand il est _.	F49	NA	F49-pic	chaud	NA	NA
22	1	11a	0	conf	prime	Pour le repas de ce soir, j'ai trouvé _.	S16	NA	du haricot	du haricot [répétition: des h	NA	NA
23	1	11b	0	naïve	critical	Ma mère prépare toujours cette recette avec	S07	quid	[citron]	des citrons	count_plur	CITRON

Figure 40 Tableau utilisé pour noter nos observations au cours de notre expérience (généré de manière automatique). Les réponses données par chaque participant à tour de rôle sont notées dans la colonne RESP. Dans ce cas, l'on note aussi dans les deux colonnes suivantes le type de syntaxe (massive ou comptable) produit, et le lemme en question.

Avec tous ces éléments, nous sommes à présent fin prêt pour soumettre notre dispositif à un groupe de locuteurs natifs. Nous donnons les détails pratiques de l'organisation de notre expérience ci-après, avant de passer aux résultats au point suivant.

## 10.2.4 Organisation de l'expérience

Notre expérience d'amorçage a été organisée les 2, 3 et 4 avril 2014, à la Faculté de Traduction et d'Interprétation de l'Université de Mons (nous remercions de tout cœur



Gudrun Vanderbauwhede, notre contact sur place, pour son aide précieuse, ainsi que tous les collègues du département de néerlandais de nous avoir prêté leur local). Les participants, au nombre de 34 (24 filles et 10 garçons), étaient tous locuteurs natifs du français, et avaient entre 18 et 23 ans. Ils s'étaient inscrits à l'avance pour participer, et prenaient part (sauf exception) à deux sessions de l'expérience – la première fois comme participant naïf, la seconde en tant que complice partiel. La durée d'une session, annoncée à 45 minutes, n'a finalement jamais dépassé la demi-heure. A l'issue des deux sessions dans lesquelles il prenait part, chaque participant a reçu la somme de 15€.

En entrant dans le local où se tenait l'expérience, chaque nouveau participant s'installait à table et signait un accord de participation. Il recevait ensuite sous forme écrite les consignes détaillant la procédure, avec la possibilité de poser oralement d'éventuelles questions complémentaires avant de commencer. A l'issue d'une première session (en tant que participant naïf), il complétait un bref formulaire comportant quelques questions subsidiaires (F/M, langue maternelle, soupçons quant à l'objet de l'expérience), avant de recevoir les consignes écrites pour la seconde session – expliquant que sa tâche serait presque identique, mais qu'à présent, certaines de ses réponses seraient préétablies.

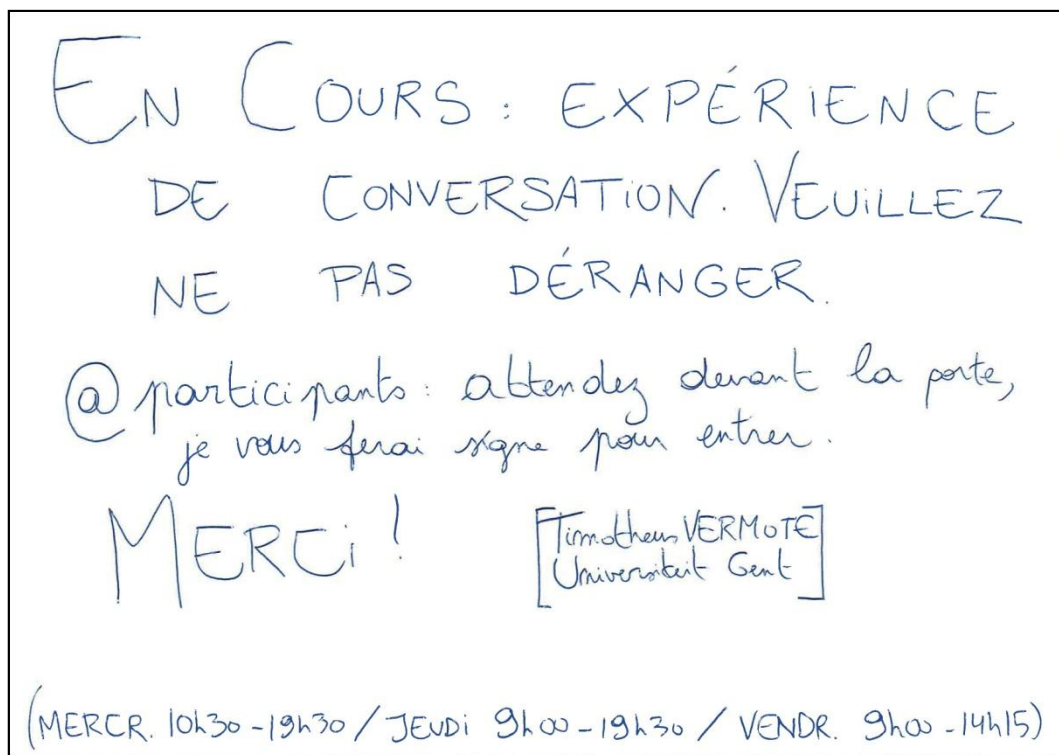


Figure 41 Message affiché pendant nos trois jours d'expérimentation sur la porte du local du département de néerlandais, à la Faculté de Traduction et d'Interprétation de l'Université de Mons.

Le premier jour de l'expérience, pour la toute première session, nous avons joué nous-même le rôle du complice (et avons enregistré les interactions à l'aide d'un enregistreur Yamaha Pocketrak C24). Pour toutes les sessions subséquentes, nous étions assis non

loin des deux participants, et prenions note au fur et à mesure des interactions des productions de chacun, complétant un tableur de données tel que présenté à la Figure 40 ci-dessus. Le deuxième et troisième jour, nous avons été assisté de Lisette Grégoire (que nous remercions), qui a donné la réplique aux premiers participants de chaque journée, et nous a aidé à installer notre matériel, à donner les consignes et éventuellement à reclasser nos cartes.

Les trois jours d'expérimentation se sont déroulés sans heurts majeurs. Tous les participants inscrits se sont présentés, et en cas de retard conséquent ils s'étaient arrangés spontanément pour échanger leur tour avec des condisciples. Tous les participants ont compris les consignes sans peine. Lors d'une des sessions, l'ordinateur portable utilisé pour prendre note s'est éteint fortuitement – nous obligeant à interrompre brièvement les participants du moment, le temps de changer d'appareil. Un second ordinateur portable étant disponible et prêt à être utilisé juste à côté, l'interruption n'a pas duré plus d'une minute.

## 10.3 Résultats

### 10.3.1 Absence de l'effet attendu d'amorçage structurel

Commençons par une première exploration graphique de nos résultats. A la Figure 42, nous indiquons pour chacune de nos conditions d'amorçage (état initial, amorce massive de contrôle, et amorce massive par un co-hyponyme) la proportion des productions de noms de fruits et de légumes en syntaxe massive et comptable. (Dans notre analyse, nous ne distinguons pas entre emplois comptables singuliers et pluriels. Notons cependant que les syntagmes comptables pluriels étaient fortement majoritaires : 348 occurrences sur 387 productions comptables, soit 89,9 %.)

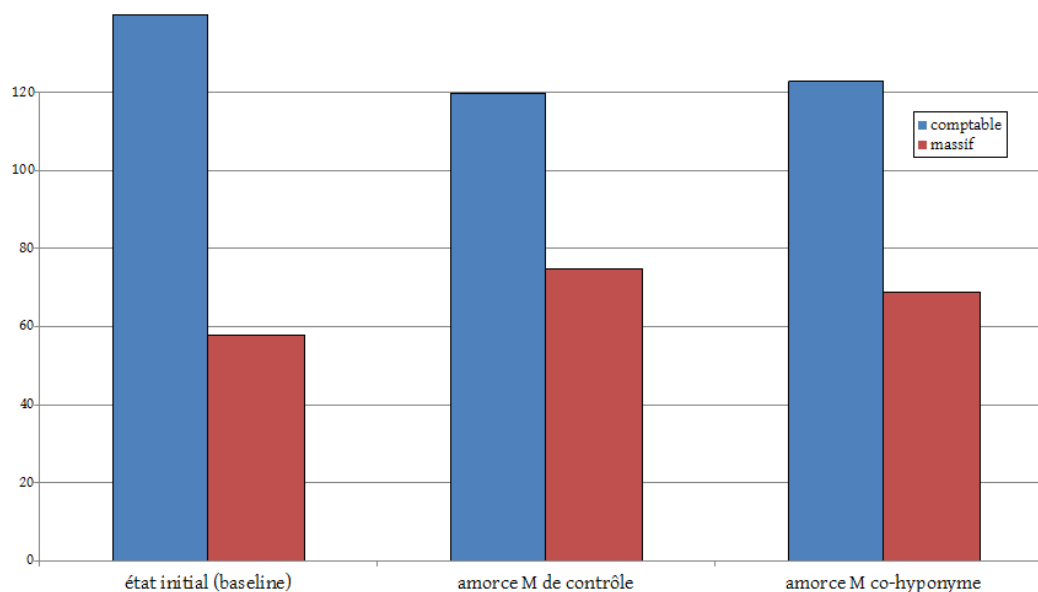


Figure 42 Nombre de productions de noms de fruits et de légumes en syntaxe comptable (bleu) et massive (rouge), en fonction des trois conditions d’amorçage. Nombre total d’observations = 585.

Les tendances dévoilées par cette représentation ne sont pas particulièrement prometteuses. La proportion de productions massives est la plus élevée dans la condition d’amorçage de contrôle, alors que nous nous attendions à ce qu’elle le soit dans la condition d’amorçage par co-hyponyme. Du reste, les différences entre nos trois conditions ne sont pas particulièrement marquées.

Pour examiner la pertinence statistique de cette variation, nous avons construit une première modélisation de nos données, en y appliquant une régression logistique à effets mixtes, examinant la production d’un nom de fruit ou de légume en syntaxe soit massive, soit comptable, en intégrant l’amorce comme facteur fixe, et les facteurs « item lexical », « participant » et « phrase » comme interceptions aléatoires. (Nous suivons, en cela, les recommandations méthodologiques de (Baayen *et al.* 2008).) Ces interceptions aléatoires nous permettent de tenir compte, dans la modélisation, d’éventuelles variations dans la manière de produire en syntagme un nom de fruit ou de légume, ayant trait respectivement aux différences entre participants, entre noms et entre contextes de production. Les estimations du modèle statistique ainsi obtenu sont présentées dans le Tableau 5. Notons que si le nombre d’items est de 33, et non de 24, c’est parce que certains participants ont produit, sur la base des images proposées, d’autres noms que ceux que nous avons attendus. Ainsi, à l’une ou l’autre reprise, *ail* a été employé pour *oignon*, *clémentine* pour *mandarine*, *princesses* pour *haricots*, ou encore *céleri* ou *artichaut* (sic) pour *poireau*. Le nombre de participants, quant à lui, est de 33 également (et non 34), car suite à une erreur technique (suppression partielle pour une raison inconnue – probablement une fausse manipulation), les données d’une session ont été rendues inutilisables.

Tableau 5 Estimations de notre modèle par régression logistique à effets mixtes de la production, en syntaxe massive ou comptable, des noms de fruits et de légumes dans notre expérience d'amorçage, intégrant l'amorce comme facteur fixe (effet de l'amorce massive co-hyponyme, et de l'amorce massive de contrôle, en comparaison avec notre état initial), ainsi que « item », « participant » et « phrase » comme interceptions aléatoires. (\*Variance et (Déviation Standard), † Valeur *p* basée sur le test du rapport de vraisemblance.)

	Coefficient estimé (err. stand.)	Valeur z	Valeur <i>p</i>
(Interception)	-0,938 (0,553)	-1,70	0,0902
Amorce M de contrôle	-0,419 (0,349)	-1,20	0,2302
Amorce M co-hyponyme	0,217 (0,329)	0,66	0,5096
Interceptions aléatoires			
Item (n = 33)	5,087 (2,256)*		< 0,05 <sup>†</sup>
Participant (n = 33)	1,109 (1,053)*		< 0,05 <sup>†</sup>
Phrase (n = 24)	1,187 (1,090)*		< 0,05 <sup>†</sup>

Ces résultats montrent que l'effet de nos deux conditions d'amorçage, par rapport à l'état initial, n'est pas significatif (valeurs *p* > 0,05). Les faibles différences observées dans la Figure 42 n'ont donc pas de pertinence statistique. Nos résultats ne révèlent ainsi pas d'effet de l'amorce sur le choix de produire un nom de fruit ou de légume soit comme massif, soit comme comptable. Ceci contredit des résultats comparables dans d'autres domaines de la morphosyntaxe – nous évoquerons une explication dans notre discussion, ci-dessous.

Cependant, l'effet de nos trois interceptions aléatoires est significatif : tant le comportement spécifique de chaque participant, que les préférences intrinsèques des noms et des contextes étudiés influencent de manière significative le choix pour l'une ou l'autre option syntagmatique. Ces deux derniers facteurs sont d'un intérêt particulier pour nous, puisqu'ils correspondent respectivement au niveau lexical et supra-lexical de notre modèle du lexique mental. Or, dans notre modèle, ces deux facteurs ne sont pas simplement aléatoires : ils correspondent d'une part aux préférences distributionnelles (et continues) de chaque nom, et d'autre part aux liens directs, activés par le contexte, entre certains sens et les nœuds combinatoires M et C. Ainsi, si les différences entre nos items et entre nos phrases ont un effet sur le choix de produire un nom de fruit ou de légume comme massif ou comptable, alors au moins une partie de cet effet devrait découler de propriétés mesurables. C'est ce que nous examinons au point suivant.

### 10.3.2 Intégration des indices d'attraction massive : items et phrases

Concernant les préférences mesurables de nos noms, nous disposons déjà de notre indice « I » pour traduire sur une échelle continue les distributions d'emplois massifs et

comptables telles qu'observées en corpus. Nous l'avons déjà utilisé pour nos 12 noms de fruits et de légumes les plus fréquents, dans l'analyse des résultats de nos enquêtes d'acceptabilité (point 9.3.1 pour le français). Pour les autres items impliqués dans notre expérience (du moins les items attendus, c'est-à-dire ceux que nous avons précédemment étudiés en corpus), nous pouvons le calculer en employant la même formule :  $I = \log(M+0,1/C+0,1)$ .

Pour nos 24 phrases, nous n'avons pas encore d'indices équivalents. Tout ce que nous en savons pour le moment est qu'elles profilent explicitement le domaine sémantique de la nourriture – un contexte que nous avons supposé propice (bien que non exclusif) aux emplois massifs. Or, les résultats du Tableau 5 nous apprennent que notre sélection de phrases n'est pas homogène : certaines élicitent plus de réponses massives que d'autres – et ces différences sont significatives. Afin de confirmer que cette variation a bien trait à une attirance plus ou moins grande de chaque phrase pour le massif, nous devons donc d'abord calculer pour chacune d'elle cette attirance, de manière indépendante à notre expérience. Pour ce faire, nous avons organisé un Cloze test, soumettant nos 24 phrases incomplètes à un panel de 61 locuteurs natifs (tous différents, évidemment, des participants à notre expérience, de même qu'à nos enquêtes du chapitre précédent), avec comme consigne de compléter le trou de chaque phrase avec « un élément qui vous semble approprié pour que la phrase soit correcte et ait, selon vous, du sens ». Sur la base de ces données, nous avons calculé pour chaque phrase son indice « P », selon la formule  $P = \log(M/C)$ , avec P l'indice, M le nombre de réponses massives et C le nombre de réponses comptables (en excluant de ce calcul les réponses en syntaxe neutre). (Comparé à la formule pour « I », nous avons pu ici diviser directement M par C, et effectuer la transformation logarithmique du résultat (sans ajouter 0,1 aux termes de la division), car chaque phrase a été complétée au moins une fois par un syntagme comptable, et au moins une fois par un syntagme massif.)

L'idée est que si toutes ces phrases évoquent le domaine de la nourriture, ce n'est cependant pas là l'élément sémantique qui, dans notre modèle du lexique, active directement le nodule combinatoire M. Effectivement, je peux dire sans problème avoir mangé une pomme – sans que la syntaxe comptable ne paraisse marquée le moins du monde. Ce sont en réalité des aspects sémantiques plus spécifiques de nos phrases qui, à des degrés divers, favorisent l'apparition du massif, en étant connectés directement au nodule M. En calculant, via notre Cloze test, l'attirance de chaque phrase pour le massif, c'est donc cette connexion que nous prétendons identifier. Ainsi, les phrases dont l'attirance pour le massif est la plus grande devraient nous donner les meilleurs indices quant aux aspects sémantiques favorisant directement ce type de syntaxe – à condition, bien sûr, que nos indices « P » soient corrélés avec nos résultats d'expérience, et s'avèrent donc pertinents.

Avec nos indices « I » (par item) et « P » (par phrase), nous pouvons à présent construire un nouveau modèle statistique de nos données, intégrant cette fois non

seulement la condition d'amorçage, mais aussi chacun de ces deux indices comme facteurs fixes, au travers d'une nouvelle régression logistique à effets mixtes. En guise d'interceptions aléatoires, nous gardons les facteurs « participants », « item » et « phrases » (nous conservons ces deux derniers, car nous ne supposons pas que nos indices de biais expliquent l'intégralité de la variation aléatoire concernée ; du reste, le test du rapport de vraisemblance indique que leur contribution au modèle reste significative). Les estimations de ce second modèle statistique se trouvent présentées dans le Tableau 6. Notons qu'à présent, le nombre d'items est de 26, car nous n'avons conservé pour cette seconde analyse que les données impliquant la production d'un nom pour lequel nous disposions effectivement d'un indice « I » (à savoir nos 24 items attendus, en distinguant *citrouille* et *potiron*, et en y ajoutant *ail* – produit 5 fois).

Tableau 6 Estimations de notre modèle par régression logistique à effets mixtes de la production, en syntaxe massive ou comptable, des noms de fruits et de légumes dans notre expérience d'amorçage, intégrant l'amorce, l'indice « I » (biais massif de chaque item, calculé en corpus) et l'indice « P » (biais massif par phrase, mesuré dans un Cloze test indépendant) comme facteurs fixes, ainsi que « item », « participant » et « phrase » comme interceptions aléatoires. (\*Variance et (Déviation Standard), † Valeur *p* basée sur le test du rapport de vraisemblance.)

	Coefficient estimé (err. stand.)	Valeur z	Valeur <i>p</i>
(Interception)	2,074 (0,835)	2,48	0,013
Amorce M de contrôle	-0,225 (0,362)	-0,62	0,534
Amorce M co-hyponyme	0,328 (0,351)	0,93	0,350
Indice « I »	1,672 (0,372)	4,49	< 0,0001
Indice « P »	1,638 (0,409)	4,01	< 0,0001
Interceptions aléatoires			
Participant (n = 33)	0,992 (0,996)*		< 0,05†
Item (n = 26)	1,865 (1,366)*		< 0,05†
Phrase (n = 24)	0,405 (0,636)*		< 0,05†

Cette fois, notre prédiction est vérifiée : si l'effet des conditions d'amorçage n'est toujours pas significatif, ceux de nos indices « I » et « P » le sont bel et bien. Ainsi, concernant la variation entre items, nos données confirment que plus l'indice « I » est élevé (correspondant à une proportion plus grande d'occurrences massives dans notre corpus), plus la chance est haute que le nom en question aura été produit en syntaxe massive dans notre expérience. En outre, pour la variation entre phrases, nos données montrent que les résultats de notre Cloze test et de notre expérience d'amorçage sont effectivement corrélés : si certaines phrases élicitent une proportion plus grande de réponses massives, c'est à cause de leur attirance plus grande pour ce type de syntaxe, indépendamment des noms impliqués.

## 10.4 Discussion et conclusions

Ainsi, les résultats de notre expérience ne sont pas entièrement ceux que nous avons prédits, mais ils constituent néanmoins, après analyse, une (nouvelle) confirmation de notre modèle à activation du lexique mental, où le massif-comptable est représenté à deux niveaux : d'une part, sous forme de connexions lexicales entre les lemmes nominaux et les nodules combinatoires M et C (modulées par la fréquence relative des deux emplois), et d'autre part sous formes de connexions supra-lexicales, liant directement les nodules M et C à certains sens, représentés par des nodules sémantiques connectés eux-mêmes à des concepts lexicaux. Concernant les préférences lexicales distributionnelles, les résultats de notre expérience confirment les résultats de nos enquêtes, où nous avons déjà démontré une corrélation entre la distribution des emplois massifs d'un certain nom (représentés par un indice « I » d'attraction massive), et l'acceptabilité de ces emplois, pour 12 de nos noms de fruits et de légumes. A présent, nous avons montré qu'une corrélation identique existait entre distribution et probabilité de production, cette fois pour au moins 24 items. Cette observation est entièrement prédite par notre modèle : plus l'emploi massif d'un nom est fréquent, plus la connexion entre le lemme correspondant et le nodule M est forte, et donc plus facile sera son activation – entraînant une probabilité de production plus élevée.

### 10.4.1 Précédence du facteur lexical et supra-lexical sur l'amorçage

Cependant, nous n'avons pas été en mesure de mettre à jour un effet d'amorçage structurel du massif par le massif, ni en général (amorce massive de contrôle), ni à l'intérieur de la catégorie des noms de fruits et de légumes (amorce massive co-hyponyme). Cet effet est pourtant globalement prédit par les modèles du lexique tels que celui de (Levelt *et al.* 1999) sur lequel nous basons notre propre proposition, de même que par les principaux modèles concurrents (voir notamment (Chang *et al.* 2000)), et a du reste été démontré à de nombreuses occasions dans différents domaines de la syntaxe et de la morphosyntaxe (alternance actif-passif chez (Bock 1986), alternance dative chez (Pickering et Branigan 1998), alternances au niveau de la structure de SN chez (Cleland et Pickering 2003), etc.), dans différentes modalités et de l'une à l'autre ((Branigan *et al.* 2000), voir aussi l'état de la question de (Pickering et Branigan 1999)), de même qu'entre deux langues différentes chez des locuteurs bilingues (Loebell et Bock 2003, Hartsuiker *et al.* 2004). L'opposition massif-comptable, qui pour sa part n'avait jamais été examinée sous cet angle, semblerait donc à première vue constituer une exception curieuse.

Cependant, à mieux y réfléchir, l'absence d'effet d'amorçage dans notre expérience en particulier, et au niveau de l'opposition massif-comptable en général, n'est peut-être pas si surprenante, après tout. L'explication est que l'interaction des deux facteurs cruciaux de notre modèle, les préférences lexicales et supra-lexicales, est tellement prépondérante dans le choix d'une production massive ou comptable, que l'effet d'amorçage (par la pré-activation du nodule M) en devient négligeable. Plus précisément, ce qui détermine crucialement la production en syntaxe massive ou comptable d'un nom de fruit ou de légume est la nature de ce nom lui-même, en combinaison avec les nécessités du contexte – ces deux facteurs pesant tellement lourd dans la balance qu'ils oblitèrent toute influence tierce, notamment d'une éventuelle amorce. Dans les domaines de la (morpho)syntaxe où un effet d'amorçage a bel et bien été observé, cette concurrence avec d'autres facteurs prépondérants n'existe pas – ou alors beaucoup moins fort. Ainsi, pour l'alternance dative par exemple, les verbes ont bien des préférences distributionnelles (Bernolet et Hartsuiker 2010), mais le sens de chacune des deux constructions est globalement identique : si la structure de l'information est certes altérée, les conditions de vérité ne varient pas d'une option à l'autre (*John gave Mary a book* et *John gave a book to Mary* décrivent tous deux nécessairement la même situation). Pour le massif-comptable, il en est autrement : *j'ai mangé du concombre* pourrait être vrai sans que *j'ai mangé un concombre* le soit forcément. Le sens de nos deux options morphosyntaxiques est donc trop différent que pour se prêter à une alternance purement formelle – prérequis pour l'observation des effets de l'amorçage structurel.

Pour mieux comprendre ce que nous entendons par là, il suffit de considérer un exemple plus extrême d'alternance entre massif et comptable, notamment dans le domaine des noms d'animaux. De toute évidence, le choix d'une production massive ou comptable d'*agneau*, par exemple, dépendra uniquement de l'aspect sémantique supra-lexical activé : si je veux parler d'individus, j'emploierai la syntaxe comptable, si c'est de viande qu'il s'agit, j'opterai pour la syntaxe massive. Que j'aie entendu juste avant le syntagme *un chien* ou *du chien*, en guise d'amorce, n'aura virtuellement aucune effet sur ma propre production. Pour les fruits et légumes, il est vrai que le contraste sémantique est moins extrême – dans certains contextes, les deux emplois sont envisageables. Néanmoins, la différence, doublée des préférences lexicales hautement variées, reste trop forte que pour laisser transparaître tout effet de l'amorçage.

Ainsi, l'absence d'effet d'amorce dans les résultats de notre expérience n'est pas seulement compatible avec notre modèle du lexique, mais en est de même une conséquence. C'est précisément parce que l'interaction du niveau lexical et supra-lexical est si cruciale dans le cas de l'opposition massif-comptable que l'amorçage structurel ne peut pas (facilement) être mis en évidence : il n'a tout simplement pas son mot à dire dans le choix de l'une ou l'autre option syntaxique en production.



## 10.4.2 De l'indice « P » à la structure du niveau supra-lexical

Nos données d'expérience nous permettant ainsi d'étayer notre représentation comme réseau stratifié à activation du lexique mental. En outre, l'examen contrastif des 24 phrases employées pour éliciter la production d'un nom de fruit ou de légume pourra nous aider à comprendre la structure du niveau sémantique de notre modèle. En effet, par le biais de notre indice « P », corroboré par nos données expérimentales, nous disposons à présent d'une indication très précise quant aux contextes les plus favorables à l'apparition d'un syntagme massif – et par là, quant aux éléments sémantiques qui seraient directement connectés au nodule M, dans le lexique. Au travers de notre étude de corpus, nous avons déjà ouvert des pistes à ce niveau, en identifiant des types de contextes particulièrement propices au massif ('matière comestible', 'forme explicitement préparée', etc.; voir point 5.1.2 pour le français et 5.2.2 pour le néerlandais). A présent, nous avons non seulement des phrases, mais également une indication quantifiée de leur attirance pour le massif. Dans le Tableau 7, nous reproduisons nos 24 phrases (que l'on pouvait déjà trouver à la Figure 39, ci-dessus), accompagnée de leur indice « P », par ordre décroissant d'attirance pour le massif. (Un indice positif signifie que la majorité des réponses obtenues au travers du Cloze test étaient des syntagmes massifs, tandis qu'un indice négatif indique que c'est le comptable qui l'emportait.)

Tableau 7 Liste des phrases employées dans notre expérience pour éliciter la production d'un nom de fruit ou de légume – classées par ordre d'attirance pour le massif (tel que mesurée à l'aide d'un Cloze test, N = 61).

PHRASES	« P »
01 Pour cette drôle de sauce, il faut mélanger de l'eau et...	1,204
20 Ce petit goût typique, je pense que c'est...	0,973
03 Hier, Julie a goûté un cocktail surprenant qui contenait...	0,887
04 Dans la casserole, Jean a rajouté...	0,718
05 Pour les gourmands, voici une entrée qui contient...	0,539
06 Au supermarché, on m'a fait goûter...	0,393
07 Dans la sauce, on pourrait mettre...	0,347
08 Dans le frigo, il reste...	0,301
09 Ma mère prépare toujours cette recette avec...	0,301
10 J'ai goûté une délicieuse préparation avec...	0,276
11 Sonia cuisine souvent avec...	0,263
12 C'est marrant cette odeur de nourriture, on dirait...	0,232
13 Chez les Devaux, on a encore mangé...	0,176
14 C'était la première fois que mon enfant mangeait...	0,079
15 Vous pouvez décorer les verrines avec...	0,062
16 Voici une délicieuse recette avec...	0,044

17	<b>Au moins une fois par semaine, Carine mange...</b>	<b>0,000</b>
18	Le cuisinier a demandé qu'on lui fournisse...	-0,093
19	Dans son alimentation, Maxime a décidé d'inclure...	-0,426
20	La salade du jour contient...	-0,467
21	Dans l'assiette d'Alice, il reste...	-0,487
22	Pour le repas de ce soir, j'ai trouvé...	-0,550
23	Chez tous les maraîchers, on trouve...	-0,672
24	Dans le fond du plat, commencez par disposer...	-0,723

Puisque les énoncés dont l'indice « P » est positif, dans les deux premiers tiers de la liste, élicitent une majorité de réponses en syntaxe massive, l'on peut postuler – dans le cadre de notre modélisation du lexique mental – qu'ils activent l'un ou l'autre nodule sémantique connecté directement au nodule combinatoire M. En examinant nos phrases, nous identifions au moins trois éléments récurrents, candidats au poste d'un tel trait sémantique : le goût (2, 6, 10) ou l'odeur (12), la consistance non solide, notamment par l'inclusion dans une préparation liquide (1, 3, 7), et la présentation sous forme arrangée, découpée en morceaux, par l'inclusion dans une préparation (4, 5, 15, et par extension 8, 9, 11 et 16). Ces éléments seront ceux auxquels nous devons accorder une place dans notre modèle.

Notons que les indices « P » ci-dessus sont bien ceux récoltés via notre Cloze test, c'est-à-dire indépendamment de toute consigne quant au nom à employer. Si nous calculons un nouvel indice pour chaque phrase, tout similaire mais cette fois à partir de nos données d'expérience (en fonction des réponses massives ou comptables données par nos participants pour chacune d'elles), la donne est quelque peu altérée, mais non moins intéressante. D'une part, l'ordre dans lequel les phrases se retrouvent est globalement le même que l'ordre dans le Tableau 7. C'est logique, puisque notre analyse statistique a démontré une corrélation significative entre les préférences de chaque phrase, et les réponses qu'elles élicitaient : un indice « P » plus élevé signifiait une probabilité plus grande d'être complété, dans notre expérience, par un syntagme massif. D'autre part, le nombre de phrases pour lesquelles les réponses massives dépassent effectivement les réponses comptables est plus réduit : il n'y en a plus que quatre, correspondant dans l'ordre aux énoncés 2, 3, 5 et 1 ci-dessus (ainsi que deux pour lesquels la proportion est la même – le 4 et le 12). Ceci dévoile à nouveau l'interaction centrale de notre modèle, entre le niveau lexical et supra-lexical : nos phrases sont relativement ouvertes au massif, mais nos noms de fruits et de légumes sont, dans l'ensemble, préférablement comptables, ce qui fait que dans notre expérience, comparativement au Cloze test (où nous n'avons pas imposé de restrictions lexicales), les différences relatives entre phrases restent inchangées mais la proportion de réalisations massives est plus basse. Cela nous rappelle que les éléments sémantiques favorables au massif, que notre Tableau 7 nous a permis d'identifier, sont bel et bien supra-lexicaux, et ne prédisent pas de manière absolue l'emploi de la syntaxe massive –

ils pointent seulement dans cette direction. L'usage final sera néanmoins le résultat de l'interaction avec les données lexicales – comme, en l'occurrence, les propriétés référentielles et distributionnelles de nos noms de fruits et de légumes, dans le cas de notre expérience.

Ces différentes observations et précisions étant faites, nous pouvons à présent passer à une représentation des noms de fruits et de légumes dans le lexique mental, selon les propositions de modélisations que nous avons faites au Chapitre 8, et telles que nos enquêtes et notre expérience nous ont permis de les confirmer (Figure 43).

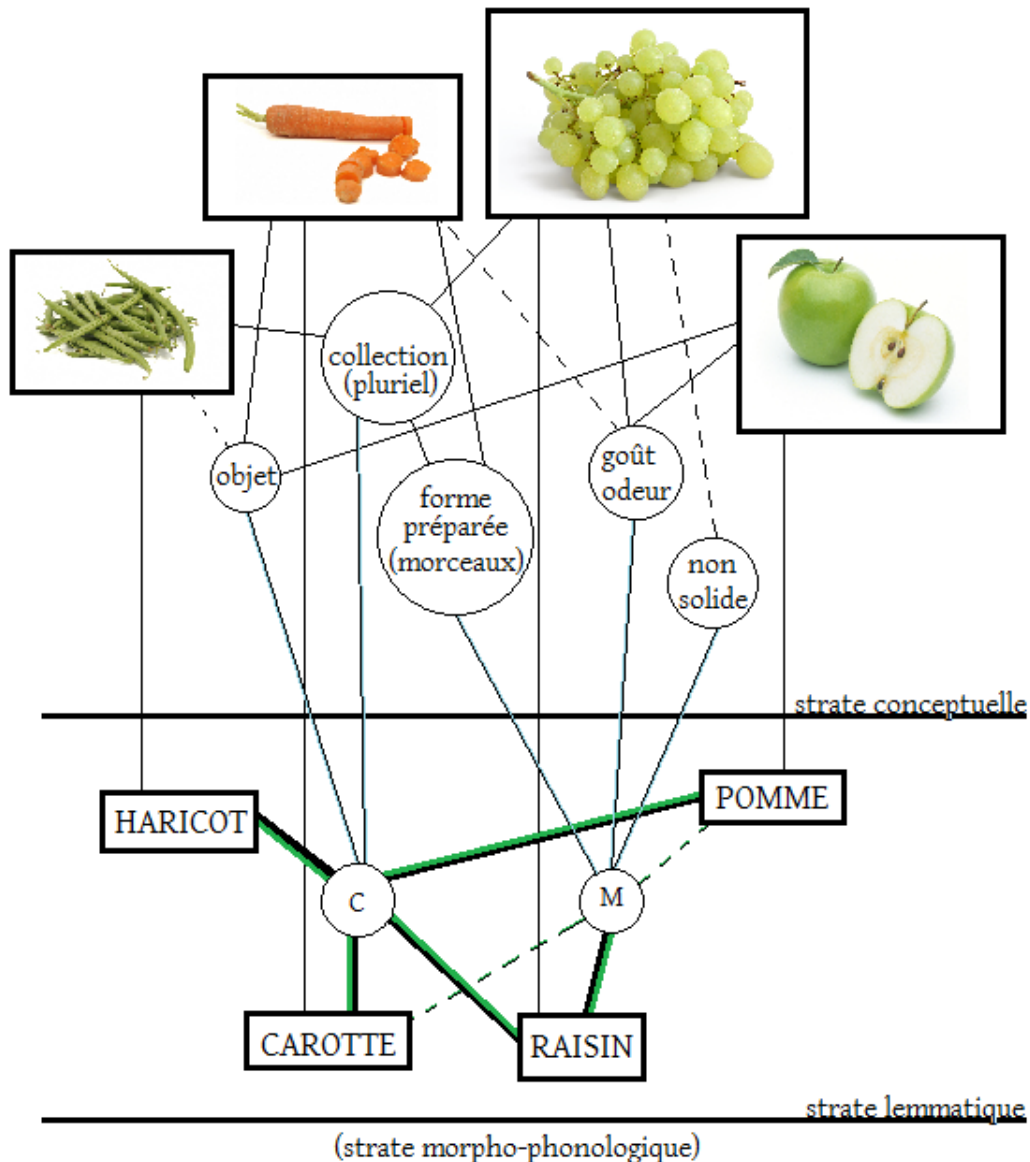


Figure 43 Modélisation, dans le lexique mental, des noms *haricot*, *carotte*, *raisin* et *pomme*. Chaque nom a des préférences distributionnelles propres quant à l'emploi massif ou comptable, représenté par des connexions plus ou moins fortes entre les lemmes correspondants et les nodules combinatoires M et C. Chaque lemme est également associé à un concept lexical (unitaire), représenté par une image. Ces concepts sont connectés (plus ou moins fortement, en fonction de nos

connaissances encyclopédiques) à différents nodules sémantiques, activables en contexte, et eux-mêmes associés directement aux nodules combinatoires M et C.

La Figure 43 illustre et modélise tous les éléments que nous avons observés concernant les noms de fruits et de légumes, et leur comportement morphosyntaxique. Au niveau lemmatique, chaque nom est plus ou moins fortement associé aux nodules combinatoires C et M, simplement sur base distributionnelle. Ainsi, *haricot* n'est relié qu'au nodule C, car il n'apparaît virtuellement jamais en syntaxe massive ; *pomme* et *carotte* sont connectés principalement au nodule C, mais également (fût-ce moins fortement), au nodule M – puisqu'ils sont attestés dans les deux types de syntaxe, mais bien plus fréquemment en emploi comptable ; pour *raisin*, en revanche, la connexion avec M est sensiblement plus forte que pour *pomme* et *carotte*, car il est rencontré plus fréquemment en syntaxe massive que ceux-ci (dans notre Figure 43, les connexions entre *raisin* et nos deux nodules combinatoires paraissent de force identique, mais ce n'est que par commodité de représentation, pour marquer le contraste avec *pomme* et *carotte* ; en réalité, si l'on en croit nos données de corpus, la connexion avec C n'en est pas moins plus forte qu'avec M, malgré tout).

Pour ce qui est de notre niveau supra-lexical, nous avons représenté cinq nodules sémantiques associés eux-mêmes directement avec les nodules M ou C (de même que de manière variable aux concepts lexicaux de nos fruits et légumes). Les nodules associés à M correspondent aux trois sens particulièrement propices au massif, que nous tirons de notre examen des phrases du Tableau 7 : le goût et l'odeur (ces deux-là correspondraient plutôt à deux nodules séparés, mais nous les combinant dans notre Figure 43 par souci d'économie), la présentation non solide (notamment à l'intérieur d'une préparation liquide : *mettre de la pomme dans la sauce*), et la forme préparée, coupée en morceaux. Pour ce qui est des sens typiquement comptables, nous en proposons deux : celui d'objet individuel, qui est le sens le plus typiquement associé à ce type de syntaxe, toutes catégories sémantiques confondues (ainsi, le nodule 'objet' représenté ici est bien le même que celui de la Figure 22 où il est connecté au concept lexical de brique, ou de la Figure 25, où il est relié au concept de voiture), de même qu'un nodule 'collection', que nous entendons comme étant le sens (grammatical) typiquement associé au pluriel – en termes d'itération référentielle. S'il est activé en même temps que le nodule 'objet', il correspond simplement au pluriel externe (Wilmet 2007: §57) ou grammatical (dans la terminologie d'(Alexiadou 2011)) : *des carottes* désigne alors plusieurs carottes, *des haricots* plusieurs haricots, etc. Il n'est cependant pas à assimiler au nodule 'objet', car il peut aussi être activé indépendamment. Cela nous permet d'expliquer d'une part les emplois comme *des groseilles* et *des raisins* (si l'on considère ces emplois comme des collectifs pluriels, voir notre remarque à ce sujet au point 8.1.1), ou encore *des épinards* – où le pluriel évoque bien une sommation, mais pas une collection d'objets qui seraient appelés *un épinard* (pluriel dit interne). D'autre part, il est nécessaire aussi pour rendre compte des emplois pluriels lexicaux des noms de fruits et de légumes, tels que nous les

avons décrits au point 5.1.2.3 – lorsque *des carottes*, *des oignons* ou encore *des pommes* ne désigne pas plusieurs individus, mais plutôt une multitude de morceaux (comme dans *Est-ce qu'il reste des carottes ?* etc.).

Ce second type d'emploi est d'autant plus intéressant qu'il révèle un point de concurrence à l'intérieur de notre modèle : dans le cas de la présentation référentielle préparée, découpée en morceaux d'un fruit ou d'un légume, deux sens concurrents sont susceptibles d'être activés : soit la présentation 'préparée' soulignant la destruction de l'intégrité physique des individus (on n'a plus *une* ou *plusieurs carottes*, mais *de la carotte*), soit la présentation 'collective' marquant que le résultat est une multiplicité de morceaux (on n'a plus *une* ou *plusieurs carottes*, mais *des carottes* – pluriel non individualisé). En soi, ces deux types de présentations correspondent aux mêmes types de contextes (voilà pourquoi nous connectons les nodules concernés dans notre Figure 43). Dans ce cas, la réalisation massive ou comptable sera tranchée en fonction des préférences lexicales, en interaction avec le choix du locuteur quant à la présentation à mettre en avant ; il s'agit donc à nouveau d'une concurrence entre niveau lexical et supra-lexical (mais dans un sens quelque peu différent de ce que nous avons dit jusqu'à présent.) C'est ce qui explique qu'un reste de morceaux de carottes dans la casserole à la fin du repas, s'il s'agit clairement de carottes dépourvues de leur intégrité physique (un sens supra-lexical associé normalement au massif) pourra être désigné tant comme *de la carotte* que comme *des carottes* – les deux lectures étant interprétables, et la seconde étant même sensiblement plus probable d'un point de vue distributionnel. L'on voit ainsi que dans le cadre de notre modélisation, l'on peut non seulement décrire les emplois possibles, mais aussi les prédire – en identifiant les facteurs pertinents guidant leur sélection. C'est tout autre chose que de dire, comme dans une approche purement grammaticale, que puisque tout référent matériel peut être broyé et qu'il peut exister différentes sortes de toute chose la syntaxe ne peut interdire les emplois massifs ou comptables d'aucun nom, et que la description de l'opposition dans le système linguistique doit s'arrêter là – les éventuelles restrictions intuitives et autres variations distributionnelles n'étant pas pertinentes pour l'étude de la langue.

Ceci étant dit, une dernière interrogation subsiste quant au sens général de l'emploi massif des noms de fruits et de légumes. A l'issue de notre étude de corpus, nous avons en effet conclu qu'indépendamment des contextes spécifiques où il apparaissait, sa signification fondamentale était une présentation du référent en 'forme et quantité indéterminées'. Or, ce sens supposé fondamental du massif ne se trouve pas représenté explicitement dans notre Figure 43 : la question est donc de la place qu'il faut lui attribuer.

Notre réponse est qu'il faut considérer ce sens fondamental comme une généralisation *post hoc*, un dénominateur commun à des sens en réalité multiples ('goût', 'forme préparée', etc.) – à la manière de nos nodules combinatoires M et C qui, comme

nous l'avons montré à la Figure 21 (point 8.1.1), sont en fait la représentation abstraite de différentes structures combinatoires organisées en réseau. Considéré ainsi, le sens 'forme et quantité indéterminées' pourrait être représenté dans une version plus abstraite de notre représentation de la Figure 43, comme un nodule sémantique général subsumant les nodules 'forme préparée', 'non solide' et 'goût/odeur'. Il n'est donc pas à se représenter dans notre modèle comme un point indépendant, mais comme un morceau du réseau sémantique lui-même. En généralisant à l'ensemble du système linguistique, l'on pourra dire que dans notre vision à activation du lexique mental, le sens (grammatical) du massif n'est donc pas représenté de manière unique, à un point précis, mais correspond à la somme de tous les sens auxquels il est associé. Il ne déroge ainsi pas aux conventions générales de notre modélisation, où tous les nodules et les connexions qui les unissent sont à considérer, à un niveau ou un autre, comme la généralisation de cooccurrences et de liens récurrents dans l'usage. C'est ce qui fait que l'on a, en français, en anglais ou en néerlandais, une très forte intuition que le massif est associé à l'absence d'individualisation, bien qu'il existe des noms et emplois massifs désignant des objets : simplement parce que les emplois massifs où l'absence d'individus est entendue sont de loin les plus fréquents et les plus répandus – imprégnant ainsi plus fortement le sens général.

## Partie 4 : Conclusions et prolongements

Dans cette quatrième et dernière partie, nous concluons notre étude de l'opposition massif-comptable. Nous reprenons d'abord le fil de notre discussion sur la multifonctionnalité du signe linguistique, entamée au Chapitre 2, et examinons ce que nos analyses auront pu nous en apprendre (Chapitre 11). Ensuite, nous concluons sur les mérites d'une approche multidimensionnelle du système linguistique, en tant que système complexe de formes, de sens et de relations. Nous confirmons ainsi les pistes ouvertes par nos modélisations, lançons des idées quant à leur poursuite et amélioration et évoquons d'autres domaines où elles pourraient s'avérer utiles (Chapitre 12).





## Chapitre 11 La question du sens

Au Chapitre 2, nous avons lancé la discussion sur le sujet de la multifonctionnalité du signe linguistique, ou polysémie, en indiquant que nous n'épuiserions pas le sujet mais qu'il nous fallait néanmoins l'aborder, si nous voulions traiter adéquatement la flexibilité de l'opposition massif-comptable – les variations morphosyntaxiques s'accompagnant inmanquablement de nuances sémantiques. Ainsi, nous avons mis en avant deux auteurs et leur vision, d'une part l'approche polysémiste de (Kleiber 1999), qui s'oppose aux visions constructivistes du sens et argue qu'il faut accorder une place dans le système aux sens multiples si ceux-ci sont stables et partagés inter-subjectivement, et d'autre part l'approche unitaire du sens de (Willems 2013), qui définit le sens lexical (appelé *meaning*) comme unique et stable au niveau du système, et appelle à étudier la polysémie au niveau du discours, comme une variation entre sens pragmatiques (appelés *senses*), peut-être pas encodés comme tels mais non moins sujets aux effets de la conventionalisation. Nous avons indiqué aussi que malgré les apparences, les deux visions étaient loin d'être incompatibles. Ainsi, les sens multiples, stables et partagés, mis en avant par (Kleiber 1999) ont bel et bien leur place dans l'approche de (Willems 2013). Ils ne sont seulement pas à situer au niveau strictement lexical des significations unitaires, mais bien à celui de l'usage, sous formes d'inférences conventionnelles – qui sont, à défaut d'être encodées, bel et bien stables et partagées inter-subjectivement (la question de ce qui relève d'un « encodage » à proprement parler devenant essentiellement terminologique).

Quant à la flexibilité de l'opposition massif-comptable, nous avons déjà souligné que l'approche des sens unitaires était prometteuse, et nécessaire, pour rendre compte du lien sémantique unissant les différents emplois d'un même nom – une position que nous avons illustrée à l'aide des énoncés (72)-(74), répétés ci-dessous comme (273)-(275).

(273) Ce veau est très coûteux à l'élevage, mais qu'est-ce qu'il est savoureux !

(274) C'est de la bière à la pompe ? Eh bien, mettez-m-en deux !

(275) Je voulais ajouter de la tomate dans la salade, mais je n'en ai plus qu'une au frigo, ce qui ne sera pas suffisant.

Cela ne signifie pas que *un veau* et *du veau*, *une bière* et *de la bière*, ou *une tomate* et *de la tomate* aient le même sens – la syntaxe massive et comptable ayant un effet indéniable sur la manière de profiler le référent – mais seulement qu’à un certain niveau, la part de sens activée par le nom est la même. Le sens lexical est ainsi sous-déterminé par rapport à ses différents emplois en discours. D’autre part, cette variation contextuelle n’est pas aléatoire : elle répond effectivement à des conventions, faisant que le lien unissant *un veau* et *du veau* est le même qu’entre *un poulet* et *du poulet*, ou celui entre *une bière* et *de la bière* qu’entre *un café* et *du café*. Au travers de nos études de corpus contrastives, nous avons pu identifier ainsi les effets de sens associés aux emplois massifs et comptables de noms issus de trois champs sémantiques différents – mettant en avant les régularités, mais aussi les éventuelles restrictions. Dans la foulée, nous avons identifié dans le domaine des noms de matériaux des items dont l’emploi comptable en ‘objet spécifique’ (*un verre* ‘récipient’, *un bronze* ‘statue en métal’) était restreint de telle manière que, contrairement aux autres cas d’alternance morphosyntaxique, il a fallu leur reconnaître un statut lexical particulier. Dans les termes de (Willems 2013), cela reviendrait à postuler pour ceux-là deux entrées distinctes dans le lexique, chacune pourvue d’un sens encodé propre.

A présent, à l’aboutissement de notre parcours d’analyse, nous disposons d’un modèle du lexique mental nous permettant de tenir compte de ces différents éléments, intégrant les remarques de (Kleiber 1999), les propositions de (Willems 2013), et nos propres observations empiriques. Ainsi, dans notre modèle stratifié du lexique sous forme d’un réseau à activation (inspiré de la tradition psycholinguistique, et en particulier de (Levelt *et al.* 1999)), nous représentons :

1. Des concepts lexicaux uniques pour chaque lemme, correspondant à des nodules au niveau sémantique du modèle (figurés systématiquement par des images dans nos représentations) ; ceux-ci sont holistiques, c’est-à-dire non décomposés ((Levelt *et al.* 1999), (Roelofs 1997) ; voir notre point 3.2.3) ; ils correspondent aux sens lexicaux unitaires de (Willems 2013) ;
2. Des nodules sémantiques associés aux concepts lexicaux (et partagés entre plusieurs de ces concepts), représentant les sens contextuels qui peuvent être activés en discours ; ils correspondent entre autres aux sens multiples et relativement stables de (Kleiber 1999), et sont aussi comparables aux facettes de (Croft et Cruse 2004) ; l’activation d’un concept lexical en combinaison avec un tel nodule sémantique résulte en un certain profilage du concept en question ; différents facteurs guident l’activation de ces sens, notamment la (morpho)syntaxe, ce qui est marqué par la connexion directe entre certains de ces nodules sémantiques et des nodules lemmatiques combinatoires (comme ceux du massif et du comptable) ;

3. Des connexions au niveau sémantique modulées par nos connaissances extralinguistiques ; ainsi, les associations entre concepts lexicaux et sens contextuels sont de force inégales, et varient en fonction de l'usage et de ce que nous savons des concepts en question ; ceci correspond aux conventions du niveau discursif de (Willems 2013) : certaines inférences contextuelles sont plus saillantes car elles sont plus habituelles, les connexions correspondantes étant alors renforcées ; ainsi, par exemple, le sens 'viande, chair comestible de' correspond à un nodule sémantique (connecté par ailleurs au nodule combinatoire M) associé fortement au concept lexical d'agneau ou encore de poulet, mais moins à celui d'éléphant ou de kangourou ; cela n'empêche que dans un certain contexte, il puisse être activé en même temps que ces derniers (même si cette activation simultanée est moins habituelle), pour fournir l'interprétation d'un énoncé tel que *J'ai mangé de l'éléphant - c'était pas terrible*.

Par ailleurs, ce dernier point nous rappelle à quel point le lexique mental est dynamique, car associé à l'expérience de chaque individu : pour nous qui, en terminant la présente rédaction (précédée par quatre années de recherche sur le sujet), avons imaginé d'innombrables fois *du kangourou* comme de la viande de marsupial, la connexion entre notre concept lexical de kangourou et celui, contextuel, de 'viande de' est certainement plus forte non seulement qu'auparavant, mais aussi que dans le lexique mental de nos congénères francophones, très peu préoccupés par la cuisine océanienne. Ainsi, s'il est vrai que la description linguistique doit s'intéresser essentiellement à ce que la langue a d'inter-subjectivement partagé, au sein d'une communauté linguistique donnée, elle ne doit pas le faire en dépit de la variation et du dynamisme du système, mais bien en vertu de ceux-ci.

Ainsi, l'approche polysémiste pure, telle que proposée par (Kleiber 1999), montre ses limites – car si de nombreux éléments sémantiques du langage sont effectivement stables et partagés, ils n'en sont pas moins tributaires du dynamisme intrinsèque de la langue et sujets aux aléas de l'interprétation contextuelle et pragmatique. Cela ne veut pas dire, pour autant, qu'il faille embrasser un modèle purement constructiviste du sens, où rien, si ce n'est quelques invariants sémantiques schématiques, ne serait encodé. Au contraire, dans le modèle que nous proposons, de nombreuses choses sont inscrites dans le système, tant du point de vue morphosyntaxique (par le biais de l'enregistrement de données distributionnelles) que sémantique. Sur ce dernier point, nous soutenons que le lexique mental contient à la fois des concepts lexicaux sous-déterminés, ainsi que des nodules correspondant à des effets de sens contextuels, plus ou moins conventionnels. Sur le sujet, le paradoxe éternel tient à la tension entre une distinction théorique stricte entre système et discours, et l'impossibilité pratique d'identifier cette distinction dans les faits. En la matière, la représentation du lexique mental, et notamment de son niveau conceptuel, sous la forme d'un réseau de nodules

et de connexions profondément dynamique, intégrant des informations linguistiques et extralinguistiques, modulées par l'usage, nous semble une solution extrêmement prometteuse. Ainsi, encodage et variation ne s'opposent plus, mais vont de pair. C'est précisément de l'interaction des différentes données encodées dans le système que naît le langage tel qu'on peut l'observer.

Ainsi, notre modèle intègre les propositions de (Willems 2013), de traiter la multifonctionnalité du signe linguistique en observant une distinction entre d'une part le niveau lexical stable et unitaire, et d'autre part celui du discours et de la variation contextuelle. En même temps, nous allons un pas plus loin, en représentant les deux niveaux dans une même modélisation, et en illustrant comment les concepts lexicaux unitaires et les sens contextuels sont connectés et interagissent, entre eux ainsi qu'avec le niveau morphosyntaxique, pour faire du langage le système dynamique, flexible et fonctionnel qu'il est. Si la distinction entre système et discours est un concept théorique intéressant, il est non moins un avatar de la description linguistique – la langue n'étant pas l'un puis l'autre, mais les deux à la fois. Ainsi, les modèles qu'on en propose ne peuvent être adéquats que s'ils permettent, mais aussi expliquent, la variation interprétative et la manière dont celle-ci se met en place. Généralement, la description linguistique s'arrête là où entrent en compte les facteurs pragmatiques et encyclopédiques, car ceux-ci ne seraient pas régis, au contraire de ce qui serait proprement linguistique, par des règles. Or, cette distinction est une illusion : ni plus ni moins que le système linguistique lui-même, ces facteurs sont en partie imprévisibles, et en partie caractérisés par une certaine stabilité intersubjective. (Willems 2013: 250) le rappelle lui-même : la langue n'est jamais construite ou interprétée hors contexte. De même, si elle correspond à un système, celui-ci n'est rien d'autre qu'une construction psychologique dérivée de l'usage, à l'intérieur des esprits qui l'utilisent – et doit donc être décrit comme tel.

Pareille description a toujours été l'objet de la recherche en psycholinguistique, visant à modéliser la langue en tant que structure mentale, dynamique et ancrée dans l'expérience et l'usage. Que ses développements aient pu être ignorés en grande partie par la recherche en linguistique générale constitue un appauvrissement significatif de notre discipline, de même qu'une aberration à corriger. Par notre proposition de modélisation du lexique mental, et de l'opposition massif-comptable à l'intérieur de celui-ci, sous forme d'un réseau stratifié à activation, inspiré de la tradition psycholinguistique et en particulier de (Levelt *et al.* 1999), nous œuvrons dès lors dans deux directions complémentaires : d'une part, vers un rapprochement (nécessaire) entre les acquis des différentes disciplines ayant le langage pour objet, et d'autre part vers une meilleure compréhension de la structure cognitive du système linguistique, en tant que construction mentale fonctionnelle, dynamique et ancrée dans l'usage. Dans notre dernier chapitre, ci-après, nous concluons dès lors sur ce point, en identifiant des pistes à suivre pour peaufiner nos propositions (et celles dont elles s'inspirent), ainsi

que d'autres domaines d'investigation en linguistique qui gagneraient, selon nous, à être abordées et modélisées à l'aide des outils conceptuels que nous avons mis en avant.



## Chapitre 12 Le système linguistique comme réseau complexe

### 12.1 La représentation de l'opposition massif-comptable

Dans le présent travail, nous avons évoqué à plusieurs reprises la notion de « représentation de l'opposition massif-comptable », d'abord dans notre état de la question, au Chapitre 1, puis dans notre proposition de modélisation au Chapitre 8. Ainsi, la question de la localisation de notre opposition morphosyntaxique dans le système linguistique aura constitué l'un des principaux fils rouges de notre analyse et de notre réflexion.

Nous avons déjà abondamment décrit nos positions sur la question, et concluons dès lors ce point par une analogie cocasse. A notre sens, le débat entre tenants d'une approche lexicaliste et défenseurs d'une vision grammaticale, s'apparente à celui d'un groupe de philosophes qui, ayant trouvé un chat bicolore noir et blanc, s'évertueraient en vain à décider si, fondamentalement, leur trouvaille était un chat blanc, ou un chat noir. De toute évidence, aucune des deux possibilités n'apportera jamais de réelle satisfaction à aucun des penseurs.

Il en va de même de notre opposition : elle n'est ni exclusivement lexicale, ni purement contextuelle. Refuser le statut lexical aux traits morphosyntaxiques massif et comptable, en invoquant pour cela leur flexibilité, revient à décider que notre chat imaginaire n'est pas blanc parce qu'il a des tâches noires – et que donc il est noir, tout simplement. Au contraire, la réalisation discursive, potentiellement variable, d'un nom dans l'un des deux types de syntaxe est toujours le résultat d'une interaction entre deux niveaux, celui des préférences lexicales, et celui des effets de sens contextuels. Choisir lequel des deux précède l'autre n'est pas une question pertinente – tout comme il n'y a aucun intérêt à décider si notre chat bicolore est fondamentalement blanc, ou noir. La vraie question n'est ainsi pas tant de trouver comment réduire la variation à une

variable unique, ou à la jeter simplement en-dehors du domaine d'investigation, mais de découvrir comment la décrire et la modéliser de manière appropriée.

Ainsi, notre chat n'est ni blanc, ni noir, il est bel et bien bicolore. De même, l'opposition massif-comptable est à la fois lexicale, et contextuelle. C'est ce qui en fait un sujet d'étude linguistique particulièrement fascinant, car il touche à la fois au lexique, à la morphosyntaxe et à la combinatoire syntagmatique, à la sémantique, à l'interprétation discursive, au lien entre langage et réalité, etc. Il nous rappelle à quel point le langage est un système complexe, doté d'innombrables facettes. Or, à notre sens, le rôle du linguiste n'est pas d'en sélectionner une de manière plus ou moins arbitraire, mais bien de les étudier de manière intégrée – car toute définition de notre chat bicolore comment un félin blanc n'est pas seulement incomplète, elle est aussi trompeuse. L'étude de la langue doit donc permettre de décrire ses différents aspects, en montrant comment ceux-ci se combinent, se complètent et interagissent. Cela implique une approche multidimensionnelle, qui passe de surcroît par une méthodologie variée intégrant sans nul doute une part d'introspection, mais aussi – et surtout – l'analyse, qualitative et quantitative (notamment statistique) de données récoltées de manière systématique, en corpus et par le biais d'enquêtes et autres paradigmes expérimentaux.

Par notre étude de l'opposition massif-comptable, nous avons voulu faire un pas très marqué dans cette direction. Ainsi, nous avons proposé un modèle du système linguistique, et de l'opposition massif-comptable à l'intérieur de celui-ci, qui tient compte à la fois des préférences lexicales des noms pour l'un ou l'autre type de syntaxe, de même que des facteurs contextuels responsables de la flexibilité et de la fluctuation de ces préférences, et de la manière dont les deux interagissent pour expliquer les emplois tels qu'ils peuvent être observés finalement dans le discours. Nous avons basé ce raisonnement sur l'examen approfondi de données de corpus, puis avons testé nos propositions au moyen d'enquêtes d'acceptabilité, et d'une expérience de production. Nous aboutissons ainsi à un modèle multidimensionnel de la langue comme construction mentale, représentée sous la forme d'un réseau stratifié à activation.

L'aventure de la recherche, bien évidemment, ne s'arrête pas là. Que l'aplomb de notre ton au moment de conclure ne trompe personne : nos propositions ne se veulent pas aboutissement, mais bien commencement. Ainsi, dans nos deux derniers points, nous évoquons respectivement des pistes à explorer pour peaufiner notre modèle dans le cadre de l'étude de l'opposition massif-comptable (12.2), ainsi que d'autres domaines de la langue où une modélisation telle que la nôtre pourrait se révéler intéressante (12.3).



## 12.2 Massif-comptable : pistes pour l'avenir

Dans notre recherche, nous avons considéré essentiellement – comme la majorité de nos prédécesseurs – l'opposition massif-comptable comme une distinction binaire, exprimée dans deux types de contextes syntagmatiques formant des catégories homogènes. Ainsi, même si nous avons précisé qu'il s'agissait d'une généralisation par commodité descriptive, nous avons représenté cette opposition au niveau lemmatique de notre modèle sous la forme de deux nodules combinatoires, M et C. Or, comme nous l'avons remarqué (point 8.1.1, Figure 21), ces deux terminaisons du réseau correspondent elles-mêmes à des ensembles connectés de nodules combinatoires plus spécifiques, comme ceux du pluriel, des numéraux et du déterminant indéfini *un(e)* pour le comptable, ou ceux du partitif, de la locution *un peu de*, ou de l'emploi nu (en néerlandais), pour le massif. Ainsi, l'on pourrait affiner nos études de corpus en établissant, pour différentes sélections de noms, des profils distributionnels non plus en termes dichotomiques (massif vs. comptable), mais en termes plus nuancés, examinant notamment la combinaison avec différents types de déterminants comptables, marquant plus ou moins fortement l'individualisation (*un, deux, trois > plusieurs > quelques > des*), ou différents types de contextes massifs plus ou moins marqués, en particulier en néerlandais (emploi nu vs. emploi avec *wat*). L'on combinerait ainsi notre propre approche distributionnelle aux propositions d'(Allan 1980), qui établit le profil des noms à partir de l'examen (introspectif) des combinaisons autorisées. Le résultat permettrait ainsi d'identifier différents types de noms (voir déjà (Lauwers 2014, à paraître) pour une typologie des noms comptables plus ou moins dénombrables), tout en leur attribuant une place dans le système.

Concernant le caractère supra-lexical de l'opposition massif-comptable, de nouvelles expériences pourraient aider à mieux comprendre la structure et l'organisation du niveau sémantique de notre modèle du lexique. Nous avons argumenté que nous n'avions pas été en mesure d'observer un effet d'amorçage dans l'emploi massif ou comptable d'un nom, parce que le choix pour un emploi ou l'autre dépendait trop fortement de deux autres facteurs : des préférences distributionnelles lexicales (par nom), et de l'activation contextuelle supra-lexicale (par phrase). L'on pourrait cependant envisager une expérience similaire, avec des items encore plus flexibles que les noms de fruits et de légumes (comme par exemple les noms de boissons) et des contextes se prêtant parfaitement aux deux lectures (massive comme 'liquide potable' ou comptable comme 'portion' : *boire de la bière / une bière / des bières*). Dans ce cas, un effet d'amorçage devrait pouvoir être obtenu, par la pré-activation respective des sens contextuels liés à chaque emploi.

L'étude plus approfondie, par le biais d'une expérience de production, des noms de matériaux et/ou des noms de boissons pourraient également donner une vision plus

nuancée des préférences distributionnelles de ces noms : la proportion totale d'emplois comptables en corpus définit-elle la préférence lexicale de chaque item – ou les emplois comptables en lecture taxonomiques ont-ils un statut particulier ? Par exemple, nous ne trouvons aucune lecture comptable 'portion' ni de *lait*, ni de *mousseux* dans nos données de corpus. D'autre part, la proportion d'occurrences comptables taxonomiques que nous repérons n'est pas égale pour les deux items : elle est bien plus basse pour le premier que pour le second. Ainsi, dans un contexte où l'on forcerait ces noms à entrer dans une lecture 'portion' (*Au bar, j'ai commandé un lait / un mousseux*), outre le caractère étrange – puisque non attesté – de l'emploi dans les deux cas, y aurait-il malgré tout une différence mesurable entre les deux ? Dans sa version actuelle, notre modèle prédit que *mousseux* en lecture comptable 'portion' devrait être plus acceptable que *lait*, simplement parce qu'en vertu du nombre plus élevé d'emplois taxonomiques donc comptables, la connexion (purement combinatoire) avec le nodule C est plus forte – indépendamment de l'interprétation. Il serait intéressant de voir si une telle prédiction se vérifie dans les faits.

Enfin, à terme, notre modélisation du massif-comptable dans le lexique mental pourrait aboutir à une implémentation connexionniste computationnelle. Dans celle-ci, nous pourrions traduire en termes proprement chiffrés les préférences lexicales distributionnelles de différents noms, la force des connexions entre leurs concepts lexicaux et différents nodules sémantiques contextuels (par exemple par le biais de la fréquence de leurs différents sens), et la force des connexions entre ces nodules contextuels et nos nodules combinatoires M et C (par exemple à l'aide d'un Cloze test, comme nous l'avons fait pour les 24 phrases de notre expérience). Ensuite, l'on pourrait examiner la capacité d'un tel modèle informatisé à prédire et expliquer des données de corpus ou de production récoltées de manière indépendante.

### 12.3 D'autres phénomènes à modéliser

En outre, la modélisation stratifiée, sous forme d'un réseau à activation, que nous avons présentée et défendue dans le présent ouvrage, pourra s'avérer utile dans d'autres domaines de la recherche linguistique. Ainsi, notons d'abord un parallèle intéressant, une fois les choses définies telles que nous l'avons fait, entre le genre formel (réputé inflexible et purement arbitraire) et l'opposition massif-comptable. D'une part, en français, les noms ont bel et bien un caractère soit masculin, soit féminin, distributionnel et défini par l'usage dans la mesure où, au risque d'énoncer une lapalissade, les noms masculins sont ceux toujours employés avec les marques typiques du masculin, et les noms féminins avec celles du féminin. Pour la plupart des noms,

cette définition distributionnelle est du reste tout ce qu'il y a à en dire : les traits ne correspondent à aucun sens particulier (il n'y a rien, dans le concept de maison, qui justifie son caractère féminin), et il n'y a donc pas d'interaction entre le niveau lexical et un hypothétique niveau supra-lexical – d'où l'absence de variation.

Cependant, pour certains noms, à savoir les noms d'individus animés et en particulier humains, le trait masculin ou féminin a bel et bien un corolaire sémantique : en général, les noms désignant des femmes sont féminins, et ceux désignant des hommes sont masculins. Or, dans ce cas, les données lexicales et supra-lexicales se mettent bel et bien à interagir, ce qui se ressent en particulier en cas de conflit. Par exemple, le nom *juge* est un nom masculin et puisqu'il désigne une personne, le sens associé à ce trait n'est pas sans effet (contrairement au féminin de *maison*). Ainsi, lorsqu'une femme est appelée *un juge*, une tension apparaît : la préférence distributionnelle du nom est clairement, voire exclusivement, en faveur du masculin, mais le sémantisme supra-lexical active tout à la fois le nodule combinatoire du féminin. C'est ainsi que l'emploi *une juge* peut faire son apparition. Au début, il fera grincer des dents : on comprend ce qu'il veut dire, mais il s'oppose à toutes nos intuitions distributionnelles. Cependant, à mesure qu'il se répandra, il paraîtra de moins en moins curieux : petit à petit, le lien entre le lemme de *juge* et le nodule combinatoire du féminin se verra renforcé. Le processus est évidemment lent, et inégal d'un locuteur à l'autre – mais s'explique parfaitement dans notre approche.

Dans certaines langues germaniques (entre autre les langues scandinaves, et dans une moindre mesure le néerlandais), la sémantisation du genre formel est encore plus dramatique. Ainsi, après la disparition complète de toute fonction sémantique liée à l'opposition formelle entre genre neutre et genre commun, on y voit émerger un nouveau pattern, déjà ancien et productif en danois et en progression en suédois (Josefsson 2014) : l'association du genre neutre au massif, et du genre commun au comptable. (Dans une moindre mesure, il semblerait que la même association apparaisse en néerlandais, en particulier pour les noms dont le genre est instable, voir (Semplicini 2013).) Ce phénomène n'est pas seulement intéressant parce qu'il concerne notre chère opposition massif-comptable. Il illustre aussi une situation où une association supra-lexicale, peut-être anecdotique au départ, entre une donnée combinatoire (genre formel, neutre ou commun) et une donnée sémantique (référent massif ou comptable) émerge puis se voit renforcée petit à petit au travers de l'usage, en direction d'un nouveau mécanisme grammatical. Cela nous rappelle que si la distinction entre synchronie et diachronie, en linguistique, a certainement sa raison d'être, il n'empêche que toute description synchronique doit intégrer, malgré tout, les caractéristiques qui autorisent et prédisent le caractère dynamique de la langue. De ce point de vue, une modélisation du système linguistique sous forme d'un réseau à activation offre des perspectives déterminantes, en intégrant des entités stables (représentées par des

nodules et des connexions) mais néanmoins dépendantes de facteurs fondamentalement variables (l'usage, au sens large).

Pour finir, un domaine de la morphosyntaxe qui pourrait certainement profiter d'une modélisation stratifiée telle que nous l'avons proposée, est celui de l'aspect verbal. Ses similarités avec l'opposition massif-comptable ont déjà été souvent remarquées (voir notre point 1.4.3.2), et effectivement il y a là aussi, à première vue, de l'avenir pour nos propositions. D'une part, les verbes ont des préférences distributionnelles quant aux structures dont ils s'accommodent (arguments, temps), d'autre part la réalisation de l'aspect ne dépend pas seulement des verbes, mais aussi du contexte, c'est-à-dire d'une donnée dépassant les particularités des items individuels. L'on retrouve là tout à fait notre distinction entre un niveau lexical, et supra-lexical. En outre, comme l'a fait remarquer (Khallouqi 2003), une explication des variations aspectuelles en termes de mécanismes de transferts, visant à sauvegarder le principe de compositionnalité, ne permet pas de rendre compte des données : il faut prévoir au contraire des mécanismes contextuels d'enrichissement conceptuel, les items lexicaux étant fondamentalement sous-déterminés par rapport à leurs différents emplois. Ceci recoupe presque exactement nos propres remarques quant à l'opposition massif-comptable. Il serait donc intéressant d'examiner de manière empirique les prédictions qu'une modélisation en réseau stratifié permettrait de faire à ce sujet, dans le domaine de l'aspect verbal.

# Bibliographie

- Acquaviva, P. (2008). *Lexical plurals: a morpho-semantic approach*. Oxford, University Press.
- Alexiadou, A. (2011). "Plural Mass Nouns and the Morpho-syntax of Number". In: Washburn, M. B., *Proceedings of the 28th West Coast Conference on Formal Linguistics*. Somerville, MA, Cascadia PP: 33-41.
- Allan, K. (1980). "Nouns and Countability." *Language: Journal of the Linguistic Society of America (Los Angeles, CA)* 56: 541-567.
- Arppe, A., Gilquin, D., Glynn, M., Hilpert et A. Zeschel (2011). "Cognitive Corpus Linguistics: Five points of debate on current theory and methodology." *Corpora* 5(2): 1-27.
- Asnès, M. (2004). *Référence nominale et verbale. Analogies et interactions*. Paris, Presses de l'Université de Paris-Sorbonne.
- Baayen, R. H., Davidson, D. M., et Bates (2008). "Mixed-effects modeling with crossed random effects for subjects and items." *Journal of Memory and Language* 59: 390-412.
- Bach, E. (1986). "The algebra of events." *Linguistics and Philosophy* 9(1): 5-16.
- Bakema, P. S. (1998). *Het Verkleinwoord Verklaard. Een Morfosemantische Studie over Diminutieven in het Nederlands*. K.U.Leuven. Thèse (PhD).
- Bale, A. C. et D. Barner (2009). "The Interpretation of Functional Heads: Using Comparatives to Explore the Mass/Count Distinction." *Journal of Semantics* 26: 217-252.
- Bale, A. C. et D. Barner (2012). "Semantic Triggers, Linguistic Variation and the Mass-Count distinction". In: Massam, D., *Count and Mass Across Languages*. Oxford, Oxford University Press.
- Barner, D. et J. Snedeker (2005). "Quantity judgments and individuation: evidence that mass nouns count." *Cognition* 97: 41-66.
- Barner, D. et J. Snedeker (2006). "Children's Early Understanding of Mass-Count Syntax: Individuation, Lexical Content, and the Number Asymmetry Hypothesis." *Language Learning and Development* 2(3): 163-194.

- Benetti, L. (2009). *L'article zéro en français contemporain. Aspects syntaxiques et sémantiques*. Bern, Peter Lang.
- Bernolet, S. et R. J. Hartsuiker (2010). "Does verb bias modulate syntactic priming?" *Cognition* 114: 455-461.
- Bloom, P. et D. Kelemen (1995). "Syntactic cues in the acquisition of collective nouns." *Cognition* 56: 1-30.
- Bloomfield, L. (1933). *Language*. New York, Henry Holt.
- Bock, J. K. (1986). "Syntactic Persistence in Language Production." *Cognitive Psychology* 18: 355-387.
- Bock, J. K. et W. J. M. Levelt (1994). "Language production: Grammatical encoding". In: Gernsbacher, M. A., *Handbook of psycholinguistics*. San Diego, CA, Academic Press: 945-984.
- Borer, H. (2005). "Some stuff: On the Count-Mass Distinction". In *name only (Structuring Sense I)*. Oxford, Oxford University Press: 86-135.
- Borillo, A. (1989). "Notions de "massif" et "comptable" dans la mesure temporelle". In: David, J. et G. Kleiber, *Termes massifs et termes comptables*. Paris, Klincksieck: 215-238.
- Bosveld-De Smet, L. (2000). "Du pluriel au massif." *Verbum XXII* 4(C): 363-378.
- Branigan, H., M. Pickering et A. A. Cleland (2000). "Syntactic coordination in dialogue." *Cognition* 75: B13-B25.
- Brown, R. et D. McNeill (1966). "The 'tip of the tongue' phenomenon." *Journal of Verbal Learning and Verbal behavior* 5: 325-337.
- Brown, R. W. (1957). "Linguistic determinism and the part of speech." *The Journal of Abnormal and Social Psychology* 55(1): 1-5.
- Bunt, H. C. (1979). "Ensembles and the formal semantic properties of mass terms". In: Pelletier, F. J., *Mass Terms: some philosophical problems*. Dordrecht, Reidel.
- Bunt, H. C. (1985). *Mass Terms and Model Theoretic Semantics*. Cambridge, Cambridge University Press.
- Bunt, H. C. (2009). "Mass expressions". In: Allan, K., *Concise Encyclopedia of Semantics*. Oxford, Elsevier: 521-524.
- Butterworth, B. (1989). "Lexical access in speech production". In: Marslen-Wilson, W. D., *Lexical representation and process*. Cambridge, MA, MIT Press: 108-135.
- Bybee, J. (1988). "Morphology as lexical organization". In: Hammond, M. et M. Noonan, *Theoretical approaches to morphology*. San Diego, Academic Press: 119-141.

- Bybee, J. (2006). "From usage to grammar: the mind's response to repetition." *Language* 82(4): 711-733.
- Caramazza, A. (1997). "How many levels of processing are there in lexical access?" *Cognitive Neuropsychology* 14: 177-208.
- Chang, F., G. S. Dell, J. K. Bock et Z. M. Griffin (2000). "Structural priming as implicit learning: A comparison of models of sentence production." *Journal of Psycholinguistic Research* 29: 217-229.
- Cheng, C. Y. (1973). "Response to Moravcsik". In: Hintikka, J., J. Moravcsik et P. Suppes, *Approaches to Natural Language*. Dordrecht, Reidel: 286-288.
- Cheng, L., J. Doetjes et R. Sybesma (2008). "How universal is the Universal Grinder?" *Linguistics in the Netherlands* 2008: 50-62.
- Chierchia, G. (1998). "Plurality of mass nouns and the notion of 'semantic parameter'". In: Rothstein, S., *Events and grammar*. Dordrecht, Kluwer A. P.: 53-103.
- Chierchia, G. (2010). "Mass nouns, vagueness and semantic variation." *Synthese* 174: 99-149.
- Christensen, R. H. B. (2013). "ordinal---Regression Models for Ordinal Data (R package version 2013.9-30)." Disponible sur <http://www.cran.r-project.org/package=ordinal/>, dernier accès.
- Cleland, A. A. et M. Pickering (2003). "The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure." *Journal of Memory and Language* 49: 214-230.
- Cohen, D. et A. Zribi-Hertz (2012). "On the lexical OR syntactic sources of furniture-type denotations". *Massif/Comptable en Linguistique, Philosophie et Sciences Cognitives (Colloque, 19-21 décembre 2012)*, Paris.
- Copetake, A. et T. Briscoe (1995). "Semi-productive polysemy and sense extension." *Journal of Semantics* 12: 15-67.
- Coşeriu, E. (1952). *Sistema, norma y habla*. Montevideo, Universidad de la República, Facultad de Humanidades y Ciencias, Instituto de Filología.
- Croft, W. et A. D. Cruse (2004). *Cognitive Linguistics*. Cambridge, Cambridge University Press.
- Cruse, D. A. (1996). "La signification des noms propres de pays en anglais". In: Rémi-Giraud, S. et P. Rétat, *Les Mots de la Nation*. Lyon, PUL: 93-102.
- Cruse, D. A. (2004). "Lexical facets and metonymy." *Ilha do Desterro* 47: 73-96.
- Culioli, A. (1999). *Pour une linguistique de l'énonciation (tome 3): Domaine notionnel*. Paris, Ophrys.

Damourette, J. et E. Pichon (1950). *Des mots à la pensée. Essai de grammaire de la langue française (1911-1940). Tome Septième*. Paris, d'Artrey.

Davidson, D. (1984). *Inquiries Into Truth and Interpretation: Philosophical Essays (Second Edition)*. Oxford, Clarendon Press.

De Belder, M. (2011). "A morphosyntactic decomposition of countability in Germanic." *The Journal of comparative Germanic Linguistics* 14(3): 173-202.

De Caluwe, J. (1994). "Open versus gesloten semantiek van woordvormingsregels." *Spektator* 23(3): 240-247.

de Vreese, W. (1899). *Galicismen in het Zuidnederlandsch. Proeve van Taalzuivering*. Gent, Siffer.

Dell, G. S. (1986). "A spreading-activation theory of retrieval in sentence production." *Psychological Review* 93(3): 283-321.

Desagulier, G. (2012). "'C'est de la bombe!': Qualitative count-to-mass conversion in French copular subject-predicate constructions". In: Bouveret, M. et D. Legallois, *Constructions in French*. Amsterdam/Philadelphia, John Benjamins: 201-231.

Doetjes, J. (1997). *Quantifiers and selection: on the distribution of quantifying expressions in French, Dutch and English*. The Hague, HAG.

Doetjes, J. (2001). "La distribution des expressions quantificatrices et le statut des noms non-comptables". In: Kleiber, G., B. Laca et L. Tasmowski, *La typologie des groupes nominaux*. Rennes, Presses Universitaires de Rennes: 119-142.

Doetjes, J. (2012). "Count/mass distinctions across languages". In: Maienborn, C., K. von Heusinger et P. Portner, *Semantics: an international handbook of natural language meaning, part III*. Berlin, De Gruyter: 2559-2580.

Dölling, J. (1991). "Group expressions and the semantics of plurals and collective nouns". In: Zimmermann, I., *Syntax und Semantik der Substantivgruppen*. Berlin, Akademie Verlag: 147-183.

Dubois, J. et F. Dubois-Charlier (2010). "La combinatoire lexico-syntaxique dans le *Dictionnaire électronique des mots*. Les termes du domaine de la musique à titre d'illustration." *Langages* 179-180: 31-56.

Ducrot, O. (1972). *Dire et ne pas dire*. Paris, Hermann.

Eisenberg, P. (1989). *Grundriss der Deutschen Grammatik*. Stuttgart, Metzler.

Englehardt, M. (2000). "The projection of argument-taking nominals." *Natural Language and Linguistic Theory* 18: 41-88.

Escoffier, A. (1922). *Le Livre des Menus*. Paris, Félin.



- Evans, V. (2006). "Lexical concepts, cognitive models and meaning-construction." *Cognitive Linguistics* 17(4): 491-534.
- Evans, V. (2010). "Word Meaning". In: Hogan, P. C., *The Cambridge Encyclopedia of the Language Sciences*. Cambridge, University Press.
- Evans, V., B. K. Bergen et J. Zinken (2007). "The cognitive linguistics enterprise: an overview". In: Evans, V., B. K. Bergen et J. Zinken, *The Cognitive Linguistics Reader*. London, Equinox: 1-36.
- Fauconnier, G. (2000). "Methods and Generalization". In: Janssen, T. et G. Redeker, *Scope and Foundations of Cognitive Linguistics*. The Hague, Mouton De Gruyter: 95-128.
- Flaux, N. et D. Van de Velde (2000). *Les noms en français: esquisse de classement*. Gap, Ophrys.
- Fodor, J., M. F. Garrett, E. C. T. Walker et C. H. Parkes (1980). "Against definitions." *Cognition* 8: 263-367.
- Franckel, J.-J., D. Paillard et S. de Vogüe (1989). "Extension de la distinction "discret", "dense", "compact" au domaine verbal". In: David, J. et G. Kleiber, *Termes massifs et termes comptables*. Paris, Klincksieck: 239-247.
- Galmiche, M. (1986). "Note sur les noms de masse et le partitif." *Langue française* 72: 40-53.
- Galmiche, M. (1989). "Massif / comptable : de l'un à l'autre et inversement". In: David, J. et G. Kleiber, *Termes massifs et termes comptables*. Paris, Klincksieck: 63-77.
- Galmiche, M. et G. Kleiber (1996). "Sur les noms abstraits". In: Flaux, N., M. Glatigny et D. Samain, *Les noms abstraits*. Lille, Presses Universitaires du Septentrion: 23-40.
- Garrett, M. F. (1975). "The analysis of sentence production". In: Bower, G., *The Psychology of learning and motivation: Advances in research and theory (Vol. 9)*. New York, Academic Press: 133-177.
- Gathercole, V. C. (1985). "'He has too much hard questions": The acquisition of the linguistic mass-count distinction in much and many." *Journal of Child Language* 12: 395-415.
- Gillon, B. (1992). "Towards a Common Semantics for English Count and Mass Nouns." *Linguistics and Philosophy* 15: 597-639.
- Gillon, B. (1999). "The lexical semantics of English count and mass nouns". In: Viegas, E., *The Breadth and Depth of Semantic Lexicons*. Dordrecht, Kluwer: 19-37.
- Gilquin, G. et S. T. Gries (2009). "Corpora and experimental methods: a state-of-the-art review." *Corpus Linguistics and Linguistic Theory* 5(1): 1-26.
- Gleason, H. A. (1965). *Linguistics and English Grammar*. New York, Holt Reinhard Winston.
- Gleason, H. A. (1969). *An introduction to descriptive linguistics*. New York, Holt, Rinehart and Winston.

Goldberg, A. E. (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford, Oxford University Press.

Goldberg, A. E. et L. Suttle (2010). "Construction Grammar." *WIREs Cognitive Science* 1: 1-10.

Gordon, P. (1985). "Evaluating the semantic categories hypothesis: The case of the count/mass distinction." *Cognition* 20: 209-242.

Gréa, P. (2008). "Quelques et plusieurs". In: J., D., H. B. et B. Laks, *Congrès Mondial de Linguistique Française - CMLF'08*. Paris, Institut de Linguistique Française.

Grévisse, M. et A. Goosse (2011). *Le Bon Usage*. Paris & Louvain-la-Neuve, DeBoeck-Duculot.

Grimm, S. (2012). "Individuation and Inverse Number Marking in Dagaare". In: Massam, D., *Count and Mass Across Languages*. Oxford, University Press: 75-98.

Grinevald, C. (2004). "Classifiers". In: Lehmann, C., G. E. Booij et J. Mugdan, *Morphology: An International Handbook on Inflection and Word-Formation (Volume 2)*. Berlin, De Gruyter Mouton: 1016-1031.

Hagoort, P., L. Hald, M. Bastiaansen et K. M. Petersson (2004). "Integration of word meaning and world knowledge in language comprehension." *Science* 304(5669): 438-441.

Harley, T. A. (2008). *The psychology of language. From data to theory*. New York, Psychology Press.

Hartsuiker, R. J., M. J. Pickering et E. Veltkamp (2004). "Is syntax separate or shared between languages?" *Psychological Science* 15: 409-414.

Hess, D. J., D. J. Foss et P. Carroll (1995). "Effects of global and local context on lexical processing during language comprehension." *Journal of Experimental Psychology: General* 124: 62-82.

Huddleston, R. D. et G. K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge, University Press.

Jackendoff, R. (1990). *Semantic Structure*. Cambridge, MA, MIT Press.

Jackendoff, R. (1991). "Parts and boundaries." *Cognition* 41: 9-45.

Jespersen, O. (1924). *The Philosophy of Grammar*. London, Allen and Unwin.

Joosten, F. (2003). "Accounts of the Count-Mass Distinction: A Critical Survey." *Nordlyd* 31(1): 216-229.

Joosten, F. (2006). "Why *club* and *lingerie* do not belong together. A plea for redefining collective nouns". In: Kleiber, G., C. Schnedecker et A. Theissen, *La Relation partie-tout*. Paris-Louvain, Peeters: 73-88.

- Josefsson, G. (2014). "Pancake sentences and the semanticization of formal gender in Mainland Scandinavian." *Language Sciences* 43: 62-76.
- Juilland, A., D. Brodin et C. Davidovitch (1970). *Frequency dictionary of French words*. The Hague/Paris, Mouton.
- Katz, G. et R. Zamparelli (2012). "Quantifying Count/Mass Elasticity". In: Choi, J., E. A. Hogue, J. Punske, D. Tat, J. Schertz et A. Trueman, *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Somerville, MA, Cascadia Proceedings Project: 371-379.
- Kempen, G. et P. Huijbers (1983). "The lexicalization process in sentence production and naming: Indirect election of words." *Cognition* 14: 185-209.
- Khallouqi, A. (2003). "Sous-détermination linguistique, distinction massif / comptable et interprétation aspectuelle." *Cahiers de Linguistique française* 25: 167-185.
- Kiss, T., F. J. Pelletier et T. Stadtfeld (2014). "Building a Reference Lexicon for Countability in English". *LREC 2014*, Reykjavik, Iceland.
- Kleiber, G. (1987). "Mais à quoi sert donc le mot CHOSE ? Une situation paradoxale." *Langue française* 73: 109-127.
- Kleiber, G. (1989). "L'opposition massif / comptable et les adjectifs". In: David, J. et G. Kleiber, *Termes massifs et termes comptables*. Paris, Klincksieck: 267-292.
- Kleiber, G. (1990). *L'article LE générique. La généralité sur le mode massif*. Genève, Librairie Droz.
- Kleiber, G. (1994). *Nominales. Essais de sémantique référentielle*. Paris, Collin.
- Kleiber, G. (1995). "Polysémie, transferts de sens et métonymie intégrée." *Folia Linguistica* 29(1-2): 105-132.
- Kleiber, G. (1997). "Massif/Comptable et partie/tout." *Verbum* 3: 321-327.
- Kleiber, G. (1999). *Problèmes de sémantique. La polysémie en questions*. Villeneuve d'Ascq, Presses universitaires du Septentrion.
- Kleiber, G. (2011). "Types de noms: la question des occurrences." *Cahiers de lexicologie* 99(2): 49-69.
- Kleiber, G., éd. (2014, à paraître-a). *Les noms à la croisée du massif et du comptable (projet MassCoLex)*. Langue française.
- Kleiber, G. (2014, à paraître-b). "Massif/Comptable et noms de propriétés". In: Kleiber, G., *Les noms à la croisée du massif et du comptable (projet MassCoLex)*.
- Krifka, M. (1989). "Nominal Reference, Temporal Constitution and Quantification in Event Semantics". In: van Benthem, J., R. Bartsch et P. van Emde Boas, *Semantics and Contextual Expression*. Dordrecht, Foris: 75-115.

- Krifka, M. (1998). "On the origins of telicity". In: Rothstein, S., *The Grammar of Events*. Dordrecht, Kluwer: 197-236.
- Kupferman, L. (2000). "L'occurrence de *Un* devant les N massifs modifiés: pourquoi donc?". In: Buvet, P.-A., D. Le Pesant et M. Mathieu-Colas, *Lexique, Syntaxe et Sémantique. Mélanges offerts à Gaston Gross (BULAG, hors série)*. Besançon, Centre Lucien Tesnière: 321-331.
- Kutas, M. et S. A. Hillyard (1980). "Reading senseless sentences: Brain potentials reflect semantic incongruity." *Science* 207: 203-205.
- Labov, W. (1972). "Some principles of linguistic methodology." *Language in Society* 1: 97-120.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, University of Chicago.
- Lammert, M. (2010). *Sémantique et cognition: les noms collectifs*. Genève, Droz.
- Landman, F. (2011). "Count Nouns - Mass Nouns - Neat Nouns - Mess Nouns". In: Partee, B. H., M. Glanzberg et J. Skilters, *Formal Semantics and Pragmatics: Discourse, Context, and Models. The Baltic International Yearbook of Cognition, Logic and Communication, Vol. 6 (2010)*. Manhattan, KS, New Prairie Press: 1-67.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume 1: Theoretical prerequisites*. Stanford, University Press.
- Langacker, R. W. (1990). *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. Berlin, Mouton de Gruyter.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Volume 2: Descriptive Application*. Stanford, University Press.
- Lauwers, P. (2014, à paraître). "Ces pluriels lexicalisés qui ne se comptent pas. Pour une étude massive des déficits de comptabilité". In: Kleiber, G., *Les noms à la croisée du massif et du comptable (projet MassCoLex)*.
- Lauwers, P. et T. Vermote (2014, à paraître). "La flexibilité de l'opposition massif/comptable en français et en néerlandais: une étude contrastive." *Syntaxe & Sémantique* 15.
- Le Bruyn, B., H. de Swart et J. Zwarts (2011, ms.). "Mass-count distinctions in bare PPs." Disponible sur <http://www.hum.uu.nl/medewerkers/j.zwarts/Masscount.pdf>, dernier accès le 21 mai 2014.
- Leeman, D. (2004). *Les déterminants du nom en français: syntaxe et sémantique*. Paris, PUF.
- Leshinskaya, A. et A. Caramazza (2014). "Organization and Structure of Conceptual Representations". In: Ferreira, F., M. Goldrick et M. Miozzo, *The Oxford handbook of language production*. Oxford, Oxford University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA, MIT Press.

- Levelt, W. J. M. (1999). "Models of word production." *Trends in Cognitive Sciences* 3(6): 223-232.
- Levelt, W. J. M., A. Roelofs et A. S. Meyer (1999). "A theory of lexical access in speech production." *Behavioral and brain sciences* 22: 1-75.
- Levelt, W. J. M., R. Schreuder et E. Hoenkamp (1978). "Structure and use of verbs of motion". In: Campbell, R. N. et P. T. Smith, *Recent advances in the psychology of language*. London, Plenum: 137-162.
- Link, G. (1983). "The logical analysis of plurals and mass terms: a lattice-theoretical approach". In: Bäuerle, R., C. Schwarze et A. Von Stechow, *Meaning, use and interpretation of language*. Berlin, de Gruyter: 302-323.
- Loebell, H. et K. Bock (2003). "Structural priming across languages." *Linguistics* 41(5): 791-824.
- Lønning, J. T. (1987). "Mass Terms and Quantification." *Linguistics and Philosophy* 10(1): 1-52.
- Macnamara, J. (1982). *Names for Things: A Study of Human Learning*. Cambridge, MA, MIT Press.
- Martin, R. (1969). "Analyse sémantique du mot *peu*." *Langue française* 4: 75-88.
- Martin, R. (1989). "La référence "massive" des unités nominales". In: David, J. et G. Kleiber, *Termes massifs et termes comptables*. Paris, Klincksieck: 37-46.
- Martin, R. (2005). "Traitement automatique de la polysémie. Eloge du dictionnaire". In: Soutet, O., *La Polysémie*. Paris, Public. de l'Univ. de Paris-Sorbonne: 167-173.
- Massam, D., éd. (2012). *Count and Mass Across Languages*. Oxford Studies in Theoretical Studies. Oxford, University Press.
- McCawley, J. D. (1975). "Lexicography and the Count-Mass Distinction": 314-321.
- McRae, K. (2004). "Semantic memory: Some insights from feature-based connectionist attractor networks". In: Ross, B. H., *The Psychology of Learning and Motivation: Advances in Research and Theory (Vol.45)*. San Diego, CA, Academic Press.
- McRae, K. et K. Matsuki (2009). "People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible." *Language and Linguistics Compass* 3(6): 1417-1429.
- Moltmann, F. (1997). *Parts and Wholes in Semantics*. Oxford, Oxford University Press.
- Moltmann, F. (1998). "Part Structures, Integrity, and the Mass-Count distinction." *Synthese* 116: 75-111.
- Morris, C. W. (1938). *Foundations of the Theory of Signs*. Chicago, University Press.
- Mourelatos, A. (1978). "Events, processes and states." *Linguistics and Philosophy* 2: 415-434.

- Mufwene, S. (1981). "Non-Individuation and the Count/Mass distinction". *Papers from the 17th Regional Meeting of the Chicago Linguistic Society*. Chicago: 221-238.
- Mufwene, S. (1984). "The Count/Mass Distinction and the English Lexicon". *Papers from the 20th Regional Meeting of the Chicago Linguistic Society, Parasession on Lexical Semantics*. Chicago: 200-221.
- Müller, A. et F. Oliveira (2004). "Bare Nominals and Number in Brazilian and European Portuguese." *Journal of Portuguese Linguistics* 3(1): 9-36.
- Müller, H. H. (2013). "Lexical Coding vs. Syntactic Marking of Homogeneity : Evidence from Spanish and Danish". *46th Annual Meeting of the Societas Linguistica Europaea*, Split, Croatia.
- Nicolas, D. (2002a). *La distinction entre noms massifs et noms comptables : aspects linguistiques et conceptuels*. Leuven, Peeters.
- Nicolas, D. (2002b). "Semantic and pragmatic factors in the interpretation of count nouns into mass nouns in French." Disponible sur <http://d.a.nicolas.free.fr/pdf/Nicolas-Conversions.pdf>, dernier accès le 16 mai 2014.
- Nicolas, D. (2008). "Mass nouns and plural logic." *Linguistics and Philosophy* 31(2): 211-244.
- Niobey, G., éd. (1979). *Nouveau dictionnaire analogique*. Paris, Larousse.
- Nunberg, G. et A. Zaenen (1992). "Systematic polysemy in lexicology and lexicography". *EURALEX '92, Proceeding I-II*, Tampere: 387-395.
- Ojeda, A. (2005). "The Paradox of Mass Plurals". In: Mufwene, S., E. Francis et R. Wheeler, *Polymorphous Linguistics. Jim McCawley's Legacy*. Cambridge, MA, MIT Press: 389-410.
- Ostler, N. et B. T. S. Atkins (1992). "Predictable meaning shift: some linguistic properties of lexical implication rules". In: Pustejovsky, J. et S. Bergler, *Lexical semantics and knowledge representation. Proceedings of the first SIGLEX Workshop*. (Berkeley, CA.), Springer-Verlag: 87-100.
- Palmer, F. R. (1971). *Grammar*. Harmondsworth, Penguin.
- Parsons, T. (1970). "An Analysis of Mass Terms and Amount Terms." *Foundations of Language* 6(3): 62-388.
- Pelletier, F. J. (1975). "Non-singular reference: Some preliminaries." *Philosophia* 5(4): 451-465.
- Pelletier, F. J. (2012). "Lexical nouns are both +MASS and +COUNT, but they are neither +MASS nor +COUNT". In: Massam, D., *Count and Mass Across Languages*. Oxford, University Press: 9-26.
- Peterson, R. R. et P. Savoy (1998). "Lexical selection and phonological encoding during language production: Evidence for cascaded processing." *Journal of Experimental Psychology: Learning, Memory, & Cognition* 24: 539-557.

- Pickering, M. et H. Branigan (1998). "The Representation of Verbs: Evidence from Syntactic Priming in Language Production." *Journal of Memory and Language* 39: 633-651.
- Pickering, M. et H. Branigan (1999). "Syntactic priming in language production." *Trends in Cognitive Sciences* 3(4): 136-141.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*, Cambridge, MA.
- Pires de Oliveira, R. et S. Rothstein (2011). "Bare singular noun phrases are mass in Brazilian Portuguese." *Lingua* 121: 2153-2175.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA, MIT Press.
- Pustejovsky, J. (1998). "Generativity and explanation in semantics: A Reply to Fodor and Lepore." *Linguistic Inquiry* 29(2): 289-311.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA, MIT Press.
- Rastier, F. (1990). "La triade sémiotique, le trivium et la sémantique linguistique." *Nouveaux Actes Sémiotiques* 9: 5-39.
- Rayo, A. (2006). "Beyond plurals". In: Rayo, A. et G. Uzquiano, *Absolute generality*. Oxford, University Press: 220-254.
- Récanati, C. et F. Récanati (1999). "La classification de Vendler revue et corrigée". In: Vogeleer, S., A. Borillo, M. Vuillaume et C. Veters, *La modalité sous tous ses aspects*. Amsterdam, Rodopi: 167-184.
- Renouf, A. (2003). "WebCorp: providing a renewable data source for corpus linguists". In: Granger, S. et S. Petch-Tyson, *Extending the scope of corpus-based research: new applications, new challenges*. Amsterdam, Rodopi: 403-419.
- Renouf, A., A. Kehoe et J. Banerjee (2007). "WebCorp: an integrated system for web text search". In: Nesselhauf, C., M. Hundt et C. Biewer, *Corpus Linguistics and the Web*. Amsterdam, Rodopi: 47-67.
- Riegel, M., J.-C. Pellat et R. Rioul (2009). *Grammaire méthodique du français (4ème édition)*. Paris, P.U.F.
- Roelofs, A. (1992). "A spreading-activation theory of lemma retrieval in speaking." *Cognition* 42: 107-142.
- Roelofs, A. (1993). "Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs." *Cognition* 47(59): 87.
- Roelofs, A. (1997). "A Case for Nondecomposition in Conceptually Driven Word Retrieval." *Journal of Psycholinguistic Research* 26(1): 33-67.
- Roeper, P. (1983). "Semantics for Mass Terms with Quantifiers." *Noûs* 17(2): 251-265.

Rothstein, S. (2010). "Counting and the Mass/Count Distinction." *Journal of Semantics* 27(3): 343-397.

Samuelson, L. K. et L. B. Smith (1999). "Early noun vocabularies: Do ontology, category structure, and syntax correspond?" *Cognition* 73: 1-33.

Sarnecka, B., V. Kamenskaya, Y. Yamana, T. Ogura et Y. Yudovina (2007). "From Grammatical Number to Exact Numbers: Early Meanings of 'One,' 'Two,' and 'Three' in English, Russian, and Japanese." *Cognitive Psychology* 55(2): 136-168.

Schein, B. (2006). "Plurals". In: Lepore, E. et B. Smith, *The Oxford Handbook of Philosophy of language*. Oxford, University Press: 716-767.

Schmid, H.-J. (2010). "Entrenchment, salience, and basic levels". In: Geeraerts, D. et H. Cuyckens, *The Oxford Handbook of Cognitive Linguistics*, Oxford University Press: 116-138.

Schütze, C. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago, University of Chicago Press.

Semplicini, C. (2013). "Synchronic Variation and Grammatical Change: The case of Dutch double gender nouns". In: Giacalone Ramat, A., C. Mauri et P. Molinelli, *Synchrony and Diachrony: a dynamic interface*. Amsterdam, John Benjamins.

Sharvy, R. (1978). "Maybe English has no count nouns: notes on Chinese semantics." *Studies in Language* 2(3): 345-365.

Soja, N. (1992). "Inferences about the meaning of nouns: the relationship between perception and syntax." *Cognitive Development* 7: 29-45.

Soja, N., S. Carey et E. S. Spelke (1991). "Ontological categories guide young children's inductions of word meaning: Object terms and substance terms." *Cognition* 38: 179-211.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London, Methuen.

Subrahmanyam, K., B. Landau et R. Gelman (1999). "Shape, material, and syntax: Interacting forces in children's learning in novel words for objects and substances." *Language and Cognitive Processes* 14(3): 249-281.

Taylor, J. R. (2002). "Count nouns and mass nouns". In: Taylor, J. R., *Cognitive Grammar*. Oxford, Oxford University Press: 366-388.

Taylor, W. L. (1953). "Cloze procedure: A new tool for measuring readability." *Journalism Quarterly* 30: 415-433.

Ter Meulen, A. (1981). "An intensional logic for mass terms." *Philosophical Studies* 40: 105-125.

Tummers, J., K. Heylen et D. Geeraerts (2005). "Usage-based approaches in Cognitive Linguistics: A technical state of the art." *Corpus Linguistics and Linguistic Theory* 1: 225-261.



- Ungerer, F. et H.-J. Schmid (1996). *An introduction to cognitive linguistics*. London, Longman.
- Van de Velde, D. (1996). "La détermination des noms abstraits". In: Flaux, N., M. Glatigny et D. Samain, *Les noms abstraits*. Villeneuve d'Ascq, Presses universitaires du Septentrion: 275-288.
- Van de Velde, D. (1997). "Un dispositif propre à faire entrer les noms d'activités dans des taxinomies : faire du + Nactivité." *Revue de linguistique romane* 61: 369-365.
- Vendler, Z. (1957). "Verbs and Times." *The Philosophical Review* 66(2): 143-160.
- Vermote, T. (2011, ms.). "La distinction massif vs. comptable. L'informatique au service de la recherche fondamentale." Disponible sur <http://ugent.academia.edu/TimotheusVermote>, dernier accès le 6 juin 2014.
- Vermote, T. (2013). "Entre système et usage en discours: nouvelles perspectives sur la localisation de l'opposition massif/comptable. Le cas des noms de fruits et de légumes". *Journal of French Language Studies*. DOI: 10.1017/S0959269513000215. Publié en ligne par Cambridge University Press le 7 août 2013.
- Vermote, T. (2014). "L'opposition massif-comptable au niveau lexical et supra-lexical". *Congrès Mondial de Linguistique Française (CMLF 2014)*, Freie Universität Berlin, SHS Web of Conferences 8: 3225-3238.
- Vermote, T. (2014, à paraître). "Les noms de matériaux au crible de l'opposition massif-comptable. Une étude de corpus". In: Kleiber, G., *Les noms à la croisée du massif et du comptable (projet MassCoLex)*.
- Vigliocco, G., M. F. Garrett et T. Antonini (1997). "Grammatical gender is on the tip of Italian tongues." *Psychological Science* 8(4): 314-317.
- Vigliocco, G., D. P. Vinson, R. Martin et M. F. Garrett (1999). "Is 'Count' and 'Mass' Information Available When the Noun Is Not? An Investigation of Tip of the Tongue States and Anomia." *Journal of Memory and Language* 40: 534-558.
- Vos, R. (1993). "Direct Partitive Constructions". In: Drijkoningen, F. A. C. et K. Hengeveld, *Linguistics in the Netherlands 1993*. Amsterdam, John Benjamins: 177-188.
- Wall, L., T. Christiansen et J. Orwant (2000). *Programming Perl (third edition)*. Sebastopol, CA, O'Reilly.
- Ware, R. (1979). "Some Bits and Pieces". In: Pelletier, F. J., *Mass Terms: Some Philosophical Problems*. Dordrecht, D. Reidel.
- Weinreich, U. (1966). "Explorations in semantic theory". In: Sebeok, T. A., *Current Trends in Linguistic Theory*. The Hague, Mouton: 395-477.
- Whorf, B. L. (1956). *Language, thought and reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA, MIT Press.

- Wiederspiel, B. (1992). "Termes de masse et référence hétérogène." *Le français moderne* 60(1): 46-67.
- Wierzbicka, A. (1988a). "Oats and wheat: mass nouns, iconicity, and human categorization". *The Semantics of Grammar*: 499-560.
- Wierzbicka, A. (1988b). *The Semantics of Grammar*. Amsterdam, Benjamins.
- Wiese, H. et J. Maling (2005). "Beers, kaffi, and Schnaps: Different grammatical options for restaurant talk coercions in three Germanic languages." *Journal of Germanic Linguistics* 17(1): 1-38.
- Willems, K. (2013). "The linguistic sign at the lexicon-syntax interface: Assumptions and implications of the Generative Lexicon Theory." *Semiotica* 193: 233-287.
- Wilmet, M. (2007). *Grammaire critique du français (4ème édition)*. Bruxelles, De Boeck.
- Wisniewski, E., M. Imai et L. Casey (1996). "On the equivalence of superordinate concepts." *Cognition* 60: 269-298.
- Yi, B. (2012 <in progress>). "Coercion and the Mass/Count Distinction." Disponible sur <http://individual.utoronto.ca/byeonguk/coercion%20&%20the%20mass%20count%20distinction.pdf>, dernier accès le 16 mai 2014.
- Zhang, N. N. (2013). *Classifier Structures in Mandarin Chinese [TiLSM 263]*. Berlin, Mouton de Gruyter.
- Zorzi, M. et G. Vigliocco (1999). "Compositional semantics and the lemma dilemma." *Behavioral and brain sciences* 22: 58-59.
- Zwicky, A. M. et J. M. Sadock (1975). "Ambiguity tests and how to fail them". In: Kimball, J., *Syntax and Semantics, volume 4*. New York, Academic Press: 1-36.

# Appendices

## 12.4 Etudes de corpus

### 12.4.1 Détail des recherches effectuées sur *WebCorp Live*

Pour chacune des nos études de cas, en français puis en néerlandais, nous précisons ici la date à laquelle nos données ont été récoltées, les formules de recherche introduites, et les éventuels filtres imposés. Nous avons organisé les items dans le tableau de manière à rentabiliser au maximum l'espace (d'où un ordre en apparence aléatoire à certains endroits).

#### 12.4.1.1 Pour le français

Fruits et légumes					
17/04/2012	[ail ails aux]	NA	31/07/2012	de la mandarine	NA
17/04/2012	citrons?	NA	31/07/2012	du cédrat	NA
17/04/2012	citrouille[s]	NA	31/07/2012	du citron	NA
17/04/2012	concombre[s]	NA	31/07/2012	du pamplemousse	NA
17/04/2012	cornichons[s]	NA	31/07/2012	du potiron	NA
17/04/2012	mandarines?	NA	31/07/2012	un peu de calebasse	NA
17/04/2012	melons?	NA	31/07/2012	un peu de cédrat	NA
17/04/2012	o[il]gnon[s]	NA	31/07/2012	un peu de citron	NA
17/04/2012	pastèques?	NA	31/07/2012	un peu de citrouille	NA
29/05/2012	[poireau poireaux porreau porreaux]	NA	31/07/2012	un peu de concombre	NA
29/05/2012	calebasse[s]	NA	31/07/2012	un peu de courge	NA
29/05/2012	cédrat[s]	NA	31/07/2012	un peu de mandarine	NA
29/05/2012	ciboulette[s]	NA	31/07/2012	un peu de melon	NA
29/05/2012	courge[s]	NA	31/07/2012	un peu de pamplemousse	NA
29/05/2012	échalote[tt]e[s]	NA	31/07/2012	un peu de potiron	NA
29/05/2012	orange[s]	NA	1/08/2012	de la ciboulette (case sensitive)	NA
29/05/2012	pamplemousse[s]	NA	1/08/2012	de l'ail (case sensitive)	NA
29/05/2012	potiron[s]	NA	1/08/2012	un peu d'ail (case sensitive)	NA

6/06/2012	ciboulettes	NA	1/08/2012	un peu de ciboulette (case sensitive)	NA
6/06/2012	de l'oignon	NA	25/03/2013	[oignon ognon][s]	NA
6/06/2012	de l'orange	NA	25/03/2013	banane[s]	NA
6/06/2012	un peu d'oignon	NA	25/03/2013	carotte[s]	NA
6/06/2012	un peu d'orange	NA	25/03/2013	concombre[s]	NA
7/06/2012	aux (case sensitive)	NA	25/03/2013	framboise[s]	NA
7/06/2012	du poireau	NA	25/03/2013	haricot[s]	NA
8/06/2012	orange[s] (case sensitive)	-mobile -galaxy -concours -bleue -myspace -orange.fr -dictionnaire -abonné -téléphone -commentcamarche			
15/06/2012	de la pastèque	NA	25/03/2013	melon[s]	NA
15/06/2012	de l'échalot[t]e	NA	25/03/2013	patate[s]	NA
15/06/2012	du cornichon (case sensitive)	-syndrome -chanson	25/03/2013	poire[s]	NA
15/06/2012	un peu de cornichon	NA	25/03/2013	pomme[s]	NA
15/06/2012	un peu de pastèque	NA	25/03/2013	raisin[s]	NA
15/06/2012	un peu de poireau	NA	25/03/2013	tomate[s]	NA
15/06/2012	un peu d'échalot[t]e	NA	30/01/2014	[chou-fleur choux-fleurs]	NA
30/07/2012	du concombre	-masqué, -bd	30/01/2014	cerise[s]	NA
30/07/2012	du melon	NA	10/06/2014	un peu de cerise	NA
31/07/2012	de la Calebasse	NA	10/06/2014	un peu de haricot	NA
31/07/2012	de la citrouille	NA	10/06/2014	un peu de mandarine	NA
31/07/2012	de la courge	NA	10/06/2014	un peu de pastèque	NA

## Matériaux

17/04/2012	[métal métaux]	NA	17/04/2012	fouurrure[s]	NA
17/04/2012	acier[s]	NA	17/04/2012	laine[s]	NA
17/04/2012	argile[s]	NA	17/04/2012	marbre[s]	NA
17/04/2012	béton[s]	NA	17/04/2012	papier[s]	NA
17/04/2012	boue[s]	NA	17/04/2012	plomb[s]	NA
17/04/2012	brique[s]	NA	17/04/2012	porcelaine[s]	NA
17/04/2012	carton[s]	NA	17/04/2012	sable[s]	NA
17/04/2012	chaux	NA	17/04/2012	tissu[s]	NA
17/04/2012	chene[s]	NA	17/04/2012	velours	NA
17/04/2012	ciment[s]	NA	17/04/2012	verre[s]	NA
17/04/2012	corde[s]	NA	17/04/2012	zinc[s]	NA
17/04/2012	coton[s]	NA	11/06/2013	bronze[s]	NA
17/04/2012	cuir[s]	NA	11/06/2013	caoutchouc[s]	NA
17/04/2012	cuivre[s]	NA	11/06/2013	granit[s]	NA
17/04/2012	fer[s]	NA	11/06/2013	gravier[s]	NA

## Boissons

29/08/2013	[eau plate eaux plates]	NA	29/08/2013	champagne[s]	-twitter
29/08/2013	[vin chaud vins chauds]	-twitter	29/08/2013	cidre[s]	-twitter
29/08/2013	alcool[s]	-twitter	29/08/2013	cocktail[s]	-twitter
29/08/2013	apéritif[s]	-twitter	29/08/2013	jus de fruit	-twitter
29/08/2013	bière[s]	-twitter	29/08/2013	lait[s]	-twitter
29/08/2013	boisson[s]	-twitter	29/08/2013	limonade[s]	-twitter
29/08/2013	soupe[s]	-twitter	3/09/2013	kir[s]	-twitter

29/08/2013	thé[ <i>s</i> ]	-twitter	3/09/2013	milkshake[ <i>s</i> ]	-twitter
29/08/2013	vin[ <i>s</i> ]	-twitter	3/09/2013	mousseux	-twitter
29/08/2013	vodka[ <i>s</i> ]	-twitter	3/09/2013	sauternes	vin
29/08/2013	whisk[ <i>e</i> ][ <i>y</i> ][ <i>s</i> ]	-twitter	6/09/2013	[gueuze][ <i>s</i> ]	-twitter
3/09/2013	brevu[ <i>s</i> ]	-twitter	29/08/2013	soda[ <i>s</i> ]	-twitter -"kev adams" -warnant
3/09/2013	expresso[ <i>s</i> ]	café	3/09/2013	punch[ <i>s</i> ]	[boisson cocktail]
3/09/2013	tisane[ <i>s</i> ]	-twitter -"marie tisane"			
3/09/2013	[cherry sherry]	[alcool boisson apéritif drink cocktail]			
3/09/2013	grog[ <i>s</i> ]	[alcool boisson apéritif drink cocktail]			
3/09/2013	panaché[ <i>s</i> ]	[alcool boisson apéritif drink cocktail]			
6/09/2013	coca[ <i>s</i> ]	[verre boire boisson coca-cola rafraîchissant rafraîchissante soif bouteille pepsi light bar] - bolivie -colombie -"la coca" -cocaïne			
6/09/2013	tilleul[ <i>s</i> ]	[thé infusion infusions tisane tasse tasses tisanes boire boisson boissons] -arbre			
3/09/2013	bordeaux[ <i>s</i> ]	vin -twitter -"both red and white" -"bordeaux restaurants" -"retrouver son lustre passé" -"météo à bordeaux" -"girondins de" -université -"bordeaux aéroport"			
29/08/2013	[café noir cafés noirs] (case sensitive)	-twitter -lafourchette -montorgueil -"fabrique de bagels" -"ouest magog" -gretchel -"la déco très originale" -"dans ton bled péquenot" -"bar à paris" -"chaussures cuir" -"retour gratuit et prépayé" -"restaurant le café noir" -"portes vitrées du café noir" -"olives et café noir" -"gay grenoble" -"café noir design" -"musta kahvi"			
6/09/2013	vinasse[ <i>s</i> ]	[ivrogne boire bois boit boivent picrate piquette saoul dégueulasse bouteille verre bouteilles verres ivrognes titube clochard vin vins] -distillation -engrais			

### 12.4.1.2 Pour le néerlandais

Fruits et légumes					
18/06/2013	[aardappel aardappelen aardappels aardappeltje aardappeltjes]	NA			
18/06/2013	[appel appelen appels appeltje appeltjes] (case sensitive)	NA			
18/06/2013	[banaan bananen banaantje banaantjes]	NA			
18/06/2013	[boon bonen boontje boontjes] (case sensitive)	NA			
18/06/2013	[druif druiven druifje druifjes]	NA			
18/06/2013	[framboos frambozen framboosje framboosjes]	NA			
18/06/2013	[komkommer kommers kommertje kommertjes]	NA			
18/06/2013	[meloen meloenen meloentje meloentjes]	NA			
18/06/2013	[tomaat tomaten tomaatje tomaatjes]	NA			
18/06/2013	[ui uien uitje uitjes ajuin ajuinen ajuintje ajuintjes]	NA			
18/06/2013	[peer peren peertje peertjes] (case sensitive)	-"peer to peer" -"peer review" -"peer teaching"			
4/09/2013	[wortel wortelen wortels worteltje worteltjes]	[groente fruit groenten moestuin planten plant rauw gebakke n eten recept recepten lekker]			
Matériaux					
6/08/2013	bron[ <i>s</i> ][ <i>zen</i> ][ <i>sje</i> ][ <i>sjes</i> ]	-linkedin -twitter	7/08/2013	grani[ <i>t</i> ][ <i>en</i> ][ <i>je</i> ][ <i>jes</i> ]	-twitter
6/08/2013	fluwe[ <i>el</i> ][ <i>len</i> ][ <i>eltje</i> ][ <i>eltjes</i> ]	-flickr	7/08/2013	kalk[ <i>en</i> ][ <i>je</i> ][ <i>jes</i> ]	-linkedin
6/08/2013	ijzer[ <i>s</i> ][ <i>tje</i> ][ <i>tjes</i> ]	NA	7/08/2013	karton[ <i>en</i> ][ <i>netje</i> ][ <i>netjes</i> ]	-twitter
6/08/2013	porselein[ <i>en</i> ][ <i>tje</i> ][ <i>tjes</i> ]	-twitter	7/08/2013	beton[ <i>en</i> ][ <i>tje</i> ][ <i>tjes</i> ]	-linkedin
7/08/2013	papier[ <i>en</i> ][ <i>tje</i> ][ <i>tjes</i> ]	-papierninja	7/08/2013	lo[ <i>od</i> ][ <i>den</i> ][ <i>odje</i> ][ <i>odjes</i> ]	NA
7/08/2013	cement[ <i>en</i> ][ <i>je</i> ][ <i>jes</i> ]	-twitter	7/08/2013	meta[ <i>al</i> ][ <i>en</i> ][ <i>altje</i> ][ <i>altjes</i> ]	NA
7/08/2013	eik[ <i>en</i> ][ <i>je</i> ][ <i>jes</i> ]	-"Dr. Eiken"	7/08/2013	stoff[ <i>en</i> ][ <i>je</i> ][ <i>jes</i> ]	NA
7/08/2013	gla[ <i>s</i> ][ <i>zen</i> ][ <i>asje</i> ][ <i>asjes</i> ]	-linkedin -twitter	7/08/2013	wol[ <i>en</i> ][ <i>letje</i> ][ <i>letjes</i> ]	-twitter
7/08/2013	bakste[ <i>en</i> ][ <i>en</i> ][ <i>tje</i> ][ <i>tjes</i> ]	-"Dirk Baksteen" -twitter -linkedin -"Benno Baksteen"			

6/08/2013	leder[ <i>len tje tjes</i> ]	- "A.W. Leder" - "Jandezeiler" - "Jan Leder" - "Erwin Leder" - "Lothar Leder"
6/08/2013	rubber[ <i>ls tje tjes</i> ]	- "Rubber Liquid" - "Liquid Rubber" - "Liquid Rubber Europe" - "Artwork Rubber Castle"
6/08/2013	zink[ <i>len je jes</i> ]	- zinkenanja - teresa - tumblr - twitter - zandenzink - "Brouwers Zink" - "Albert Zink" - vervoegen - "bijv. naamwoord" - "Zink different" - Linnerpaultje - "Zink Store" - "Zink Feuerwerk" - "Peter Zink Secher"
7/08/2013	bont[ <i>len je jes</i> ]	- linkedin - "Raf De Bont" - "Jan Te Bont" - "Cees de Bont" - "Paul de Bont" - "Maria Bonten" - "Arthur Bont" - "Marco Bontje" - "Caroline Bont"
7/08/2013	grind[ <i>len je jes</i> ]	- linkedin - twitter - youtube - interglot - "The Daily Grind" - "Rise to the Grind" - "The Grind"
7/08/2013	koper[ <i>ls en tje tjes</i> ]	- flickr - linkedin - twitter - "Marc Koper"
7/08/2013	marmar[ <i>ls tje tjes</i> ]	- "Ronald L. Marmer" - instagram - youtube - twitter
7/08/2013	sta[ <i>al len altje altjes</i> ]	- instagram - "Boele Staal" - "Jan Frederik Staal" - "Henri Staal" - "Peter Staal" - "Ede Staal" - twitter - "Patrick Staal"
7/08/2013	zand[ <i>len je jes</i> ]	- linkedin - twitter - "Jan Luiten van Zanden"
7/08/2013	[klei kleien kleitje kleitjes leem lemen leempje leempjes]	- "Ewout Klei" - "Nicolaas Klei" - twitter - linkedin - "Ilse van Klei" - "Joop ten Klei"
7/08/2013	[touw touden touwtje touwtjes koord koorden koordje koordjes]	- twitter - linkedin - "Marjolein Touw" - "Pleuni Touw" - "Harry Touw" - "Ab van der Touw" - "Jeroen Touwen" - "Carien Touwen" - "Bernhard Touwen" - "Jan Touw"
7/08/2013	katoen[ <i>len tje tjes etje etjes netje netjes</i> ]	- twitter
7/08/2013	[modder modderen moddertje moddertjes slijk slijken slijkje slijkjes]	- linkedin - twitter - "Jaap Modder" - "Slijk Ewijk"

## Boissons

10/09/2013	sauternes	-twitter	10/09/2013	[ <i>lindethee lindetheeën lindetheetje lindetheetjes</i> ]	-twitter
10/09/2013	[ <i>koffie koffies koffietje koffietjes</i> ]	-twitter	10/09/2013	[ <i>milkshake milksaken milksakeje milksakejes</i> ]	-twitter
10/09/2013	[ <i>cider cidere cidertje cidertjes</i> ]	-twitter	10/09/2013	[ <i>aperitief aperitieven aperitiefje aperitiefjes</i> ]	-twitter
10/09/2013	[ <i>espresso espresso's</i> ]	-twitter	10/09/2013	[ <i>brouwsel brouwsels brouwseltje brouwseltjes</i> ]	-twitter
10/09/2013	[ <i>cocktail cocktails</i> ]	-twitter	10/09/2013	[ <i>drank dranken drankje drankjes</i> ]	-twitter
10/09/2013	[ <i>melk melken melkje melkjes</i> ]	-twitter	10/09/2013	[ <i>frisdrank frisdranken frisdrankje frisdrankjes</i> ]	-twitter
10/09/2013	[ <i>alcohol alcoholen</i> ]	-twitter	10/09/2013	[ <i>fruitsap fruitsappen fruitsapje fruitsapjes</i> ]	-twitter
10/09/2013	[ <i>soep soepen soepje soepjes</i> ]	-twitter	10/09/2013	[ <i>Glühwein Glühweintje Glühweintjes</i> ]	-twitter
10/09/2013	[ <i>thee theeën theetje theetjes</i> ]	-twitter	10/09/2013	[ <i>limonade limonades limonadetje limonadetjes</i> ]	-twitter
10/09/2013	[ <i>bier bieren biertje biertjes</i> ]	-twitter	10/09/2013	[ <i>whisky whisky's whisky'tje whisky'tjes</i> ]	-twitter
10/09/2013	[ <i>sherry sherry's</i> ]	-twitter	10/09/2013	[ <i>wodka wodka's wodkaatje wodkaatjes</i> ]	-twitter
10/09/2013	[ <i>cola cola's colaatje colaatjes</i> ]	-twitter	[ <i>glas glaasje drank drinken fles flessen soda soft drink coca-cola pepsi café bar</i> ]		
10/09/2013	[ <i>grog grogs groge groges</i> ]	-twitter	[ <i>cocktail mengsel alcohol rum suiker water drinken glas glaasje</i> ]		
10/09/2013	[ <i>geuze geuzen</i> ]	-twitter	[ <i>bier lambiek glas glaasje glaasjes fles flessen drinken alcohol</i> ]		
10/09/2013	[ <i>punch punchje punchjes</i> ]		[ <i>drank cocktail thee arak suiker citroen wijn rum water glas glaasje drinken drank</i> ]		
10/09/2013	[ <i>shandy shandy's</i> ]	-twitter	[ <i>bier glas glaasje fles drank drinken cocktail limonade</i> ]		
10/09/2013	[ <i>wijn wijnen wijntje wijntjes</i> ]	-twitter	[ <i>glas drinken wit rood glaasje fles flessen witte rode</i> ]		
10/09/2013	kir[ <i>s</i> ]	-twitter	[ <i>cassis wijn witte glas glaasje cocktail likeur</i> ]		
10/09/2013	[ <i>champagne champagnes champagnetje champagnetjes</i> ]		-twitter	[ <i>glas drinken glaasje fles flessen bubbels schuimwijn</i> ]	
10/09/2013	schuim[ <i>wijn wijnen wijntje wijntjes</i> ]		-twitter		
10/09/2013	[ <i>kruidenthee kruidentheeën kruidentheetje kruidentheetjes</i> ]		-twitter		
10/09/2013	[ <i>bordeauxwijn bordeauxwijnen bordeauxwijntje bordeauxwijntjes</i> ]			-twitter	
10/09/2013	[ <i>plat water platte waters platte wateren plat watertje platte watertjes</i> ]			-twitter	

## 12.4.2 Résultats quantitatifs (chiffres absolus)

### 12.4.2.1 Français

	COMPTABLE			CUMULATIF			MASSIF	NEUT.	TOTAUX
	<i>plur</i>	<i>sing</i>	TOTAL	<i>plur</i>	<i>sing</i>	TOTAL			
<b>CS1</b>	<b>7078</b>	<b>990</b>	<b>8068</b>	<b>2718</b>	<b>21</b>	<b>2739</b>	<b>204</b>	<b>11909</b>	<b>22920</b>
<i>ail</i>	6	5	11	6	4	10	31	644	696
<i>banane</i>	259	48	307	88	0	88	5	380	780
<i>calebasse</i>	34	38	72	5	0	5	0	159	236
<i>carotte</i>	319	28	347	196	0	196	1	312	856
<i>cédrat</i>	46	16	62	25	1	26	7	299	394
<i>cerise</i>	170	7	177	26	0	26	0	185	388
<i>chou-fleur</i>	30	32	62	17	1	18	22	568	670
<i>ciboulette</i>	1	0	1	2	2	4	17	354	376
<i>citron</i>	330	54	384	103	2	105	5	580	1074
<i>citrouille</i>	82	44	126	18	1	19	9	487	641
<i>concombre</i>	307	77	384	112	0	112	9	901	1406
<i>cornichon</i>	220	6	226	52	3	55	0	134	415
<i>courge</i>	308	36	344	54	0	54	2	495	895
<i>échalote</i>	257	14	271	51	0	51	3	335	660
<i>framboise</i>	330	10	340	114	0	114	1	315	770
<i>haricot</i>	452	19	471	294	0	294	0	301	1066
<i>mandarine</i>	223	7	230	70	0	70	0	160	460
<i>melon</i>	333	94	427	83	3	86	8	821	1342
<i>oignon</i>	793	128	921	331	1	332	8	780	2041
<i>orange</i>	46	28	74	34	0	34	0	191	299
<i>pamplemousse</i>	38	6	44	41	2	43	7	350	444
<i>pastèque</i>	99	13	112	70	0	70	0	114	296
<i>patate</i>	187	42	229	182	0	182	4	337	752
<i>poire</i>	484	93	577	134	0	134	1	419	1131
<i>poireau</i>	343	20	363	146	0	146	5	416	930
<i>pomme</i>	694	70	764	141	0	141	1	223	1129
<i>potiron</i>	28	15	43	15	0	15	6	568	632
<i>raisin</i>	258	11	269	69	1	70	50	560	949
<i>tomate</i>	401	29	430	239	0	239	2	521	1192
<b>CS2</b>	<b>2308</b>	<b>414</b>	<b>2722</b>	<b>456</b>	<b>2035</b>	<b>2491</b>	<b>439</b>	<b>8298</b>	<b>13950</b>
<i>acier</i>	83	10	93	13	59	72	2	214	381
<i>argile</i>	100	16	116	3	14	17	23	648	804
<i>béton</i>	66	23	89	3	107	110	22	342	563
<i>boue</i>	496	9	505	105	9	114	4	160	783
<i>brique</i>	148	30	178	60	8	68	0	124	370
<i>bronze</i>	37	4	41	5	127	132	3	173	349
<i>caoutchouc</i>	49	7	56	13	168	181	12	423	672
<i>carton</i>	39	15	54	4	166	170	3	55	282
<i>chaux</i>	14	7	21	0	12	12	22	425	480

<i>chêne</i>	71	17	88	14	12	26	2	273	<b>389</b>
<i>ciment</i>	51	14	65	2	3	5	6	104	<b>180</b>
<i>corde</i>	129	49	178	31	3	34	6	227	<b>445</b>
<i>coton</i>	4	11	15	1	35	36	12	243	<b>306</b>
<i>cuir</i>	17	16	33	3	323	326	5	245	<b>609</b>
<i>cuivre</i>	44	6	50	5	124	129	29	913	<b>1121</b>
<i>fer</i>	8	4	12	3	87	90	16	421	<b>539</b>
<i>fournure</i>	35	7	42	11	20	31	29	227	<b>329</b>
<i>granit</i>	25	4	29	5	103	108	85	278	<b>500</b>
<i>gravier</i>	9	1	10	5	8	13	9	154	<b>186</b>
<i>laine</i>	27	6	33	8	55	63	22	143	<b>261</b>
<i>marbre</i>	44	6	50	6	89	95	7	324	<b>476</b>
<i>métal</i>	188	25	213	68	4	72	2	54	<b>341</b>
<i>papier</i>	50	14	64	7	53	60	19	172	<b>315</b>
<i>plomb</i>	98	18	116	5	8	13	9	210	<b>348</b>
<i>porcelaine</i>	23	6	29	5	132	137	4	190	<b>360</b>
<i>sable</i>	132	7	139	5	30	35	46	522	<b>742</b>
<i>tissu</i>	116	7	123	40	3	43	2	34	<b>202</b>
<i>velours</i>	8	1	9	0	140	140	17	206	<b>372</b>
<i>verre</i>	194	74	268	26	91	117	11	517	<b>913</b>
<i>zinc</i>	3	0	3	0	42	42	10	277	<b>332</b>
<b>CS3</b>	<b>1806</b>	<b>1162</b>	<b>2968</b>	<b>388</b>	<b>144</b>	<b>532</b>	<b>416</b>	<b>9000</b>	<b>12916</b>
<i>alcool</i>	19	22	41	5	41	46	55	828	<b>970</b>
<i>apéritif</i>	20	16	36	16	0	16	0	88	<b>140</b>
<i>bière</i>	230	65	295	52	2	54	18	503	<b>870</b>
<i>boisson</i>	214	48	262	45	0	45	0	87	<b>394</b>
<i>bordeaux</i>	34	4	38	0	0	0	2	259	<b>299</b>
<i>brevage</i>	18	32	50	9	0	9	1	64	<b>124</b>
<i>café noir</i>	0	9	9	0	1	1	6	83	<b>99</b>
<i>champagne</i>	35	22	57	1	0	1	24	228	<b>310</b>
<i>cherry</i>	1	2	3	0	0	0	1	39	<b>43</b>
<i>cidre</i>	50	34	84	3	3	6	17	579	<b>686</b>
<i>coca</i>	2	15	17	0	3	3	9	177	<b>206</b>
<i>cocktail</i>	123	56	179	0	0	0	0	454	<b>633</b>
<i>eau plate</i>	13	4	17	0	2	2	19	78	<b>116</b>
<i>expresso</i>	2	37	39	0	1	1	1	97	<b>138</b>
<i>grog</i>	12	69	81	6	1	7	4	218	<b>310</b>
<i>gueuze</i>	4	23	27	2	1	3	4	195	<b>229</b>
<i>jus de fruit</i>	30	20	50	1	1	2	8	74	<b>134</b>
<i>kir</i>	11	60	71	3	0	3	3	190	<b>267</b>
<i>lait</i>	21	22	43	0	44	44	35	625	<b>747</b>
<i>limonade</i>	31	42	73	0	2	2	15	220	<b>310</b>
<i>milkshake</i>	22	36	58	2	1	3	2	60	<b>123</b>
<i>mousseux</i>	15	27	42	0	0	0	7	61	<b>110</b>
<i>panaché</i>	24	12	36	10	0	10	0	46	<b>92</b>
<i>punch</i>	1	25	26	0	1	1	3	134	<b>164</b>



<i>sauternes</i>	39	12	51	0	0	0	5	252	<b>308</b>
<i>sherry</i>	1	1	2	0	0	0	5	39	<b>46</b>
<i>soda</i>	107	16	123	52	2	54	7	133	<b>317</b>
<i>soupe</i>	51	113	164	51	3	54	7	390	<b>615</b>
<i>thé</i>	181	59	240	44	9	53	28	735	<b>1056</b>
<i>tilleul</i>	0	1	1	3	0	3	3	186	<b>193</b>
<i>tisane</i>	156	74	230	14	5	19	9	202	<b>460</b>
<i>vin</i>	225	63	288	38	1	39	14	284	<b>625</b>
<i>vin chaud</i>	0	24	24	0	5	5	28	238	<b>295</b>
<i>vinasse</i>	15	0	15	0	1	1	16	33	<b>65</b>
<i>vodka</i>	52	32	84	14	7	21	34	494	<b>633</b>
<i>whisky</i>	47	65	112	17	7	24	26	627	<b>789</b>
<b>TOTAUX</b>	<b>11192</b>	<b>2566</b>	<b>13758</b>	<b>3562</b>	<b>2200</b>	<b>5762</b>	<b>1059</b>	<b>29207</b>	<b>49786</b>

### 12.4.2.2 Néerlandais

Le tableau pour le néerlandais comporte moins de colonnes, puisque nous n'avons pas, dans notre analyse des données de cette langue, classés séparément les occurrences en contextes cumulatifs (tous les emplois éventuellement concernés se retrouvant d'emblée classés du côté du massif, ou du comptable).

	COMPTABLE			MASSIF	NEUTRE	TOTAUX
	<i>plur</i>	<i>sing</i>	TOTAL			
<b>CS1</b>	<b>4290</b>	<b>487</b>	<b>4777</b>	<b>738</b>	<b>1414</b>	<b>6929</b>
<i>aardappel</i>	580	29	609	27	131	<b>767</b>
<i>ajuin</i>	13	0	13	2	12	<b>27</b>
<i>appel</i>	46	33	79	5	56	<b>140</b>
<i>banaan</i>	540	157	697	105	230	<b>1032</b>
<i>boon</i>	86	11	97	2	35	<b>134</b>
<i>druif</i>	497	29	526	4	138	<b>668</b>
<i>framboos</i>	414	17	431	33	118	<b>582</b>
<i>komkommer</i>	275	51	326	184	169	<b>679</b>
<i>meloen</i>	299	57	356	178	157	<b>691</b>
<i>peer</i>	93	9	102	18	9	<b>129</b>
<i>tomaat</i>	520	31	551	68	129	<b>748</b>
<i>ui</i>	223	37	260	71	113	<b>444</b>
<i>wortel</i>	704	26	730	41	117	<b>888</b>
<b>CS2</b>	<b>1005</b>	<b>257</b>	<b>1262</b>	<b>5551</b>	<b>2010</b>	<b>8823</b>
<i>baksteen</i>	215	64	279	77	63	<b>419</b>
<i>beton</i>	0	1	1	140	56	<b>197</b>
<i>bont</i>	0	0	0	101	24	<b>125</b>
<i>brons</i>	18	0	18	137	21	<b>176</b>
<i>cement</i>	13	5	18	124	44	<b>186</b>
<i>eik</i>	45	14	59	9	95	<b>163</b>
<i>eiken</i>	0	0	0	53	0	<b>53</b>

<i>fluweel</i>	6	0	6	45	9	<b>60</b>
<i>glas</i>	100	35	135	60	52	<b>247</b>
<i>graniet</i>	10	1	11	276	26	<b>313</b>
<i>grind</i>	1	2	3	274	102	<b>379</b>
<i>ijzer</i>	12	19	31	372	120	<b>523</b>
<i>kalk</i>	0	4	4	102	29	<b>135</b>
<i>karton</i>	2	1	3	194	75	<b>272</b>
<i>katoen</i>	1	0	1	279	43	<b>323</b>
<i>klei</i>	12	9	21	175	118	<b>314</b>
<i>koord</i>	21	17	38	9	37	<b>84</b>
<i>koper</i>	0	0	0	165	11	<b>176</b>
<i>leder</i>	0	9	9	329	253	<b>591</b>
<i>leem</i>	0	0	0	125	29	<b>154</b>
<i>lood</i>	11	2	13	411	37	<b>461</b>
<i>marmer</i>	16	4	20	174	33	<b>227</b>
<i>metaal</i>	255	10	265	9	28	<b>302</b>
<i>modder</i>	0	0	0	93	95	<b>188</b>
<i>papier</i>	9	1	10	187	53	<b>250</b>
<i>porselein</i>	0	0	0	298	83	<b>381</b>
<i>rubber</i>	12	1	13	137	29	<b>179</b>
<i>slijk</i>	0	0	0	24	17	<b>41</b>
<i>staal</i>	0	2	2	107	30	<b>139</b>
<i>stof</i>	158	6	164	63	73	<b>300</b>
<i>touw</i>	70	41	111	88	57	<b>256</b>
<i>wol</i>	3	1	4	326	112	<b>442</b>
<i>zand</i>	14	5	19	186	108	<b>313</b>
<i>zink</i>	1	3	4	402	48	<b>454</b>
<b>CS3</b>	<b>1868</b>	<b>951</b>	<b>2819</b>	<b>6420</b>	<b>2998</b>	<b>12237</b>
<i>alcohol</i>	27	8	35	754	118	<b>907</b>
<i>aperitief</i>	75	23	98	26	33	<b>157</b>
<i>bier</i>	227	79	306	236	233	<b>775</b>
<i>bordeauxwijn</i>	150	16	166	23	10	<b>199</b>
<i>brouwsel</i>	33	23	56	11	108	<b>175</b>
<i>champagne</i>	57	60	117	404	178	<b>699</b>
<i>cider</i>	47	15	62	258	97	<b>417</b>
<i>cocktail</i>	175	95	270	1	86	<b>357</b>
<i>cola</i>	2	5	7	285	53	<b>345</b>
<i>drank</i>	147	12	159	162	56	<b>377</b>
<i>espresso</i>	6	63	69	85	51	<b>205</b>
<i>frisdrank</i>	154	11	165	246	46	<b>457</b>
<i>fruitsap</i>	97	6	103	92	21	<b>216</b>
<i>glühwein</i>	0	4	4	144	68	<b>216</b>
<i>grog</i>	4	41	45	14	20	<b>79</b>
<i>kir</i>	0	12	12	34	15	<b>61</b>
<i>koffie</i>	7	8	15	252	131	<b>398</b>
<i>kruidenthee</i>	18	18	36	137	48	<b>221</b>

<i>limonade</i>	30	14	44	112	77	<b>233</b>
<i>lindethee</i>	0	1	1	42	2	<b>45</b>
<i>melk</i>	0	1	1	611	175	<b>787</b>
<i>milkshake</i>	0	86	86	28	57	<b>171</b>
<i>plat water</i>	1	1	2	62	9	<b>73</b>
<i>punch</i>	0	9	9	49	54	<b>112</b>
<i>sauternes</i>	3	18	21	120	38	<b>179</b>
<i>schuimwijn</i>	81	13	94	46	24	<b>164</b>
<i>shandy</i>	4	9	13	95	6	<b>114</b>
<i>sherry</i>	24	51	75	341	119	<b>535</b>
<i>soep</i>	100	43	143	116	217	<b>476</b>
<i>thee</i>	11	47	58	593	314	<b>965</b>
<i>whisky</i>	60	40	100	163	177	<b>440</b>
<i>wijn</i>	323	114	437	585	292	<b>1314</b>
<i>wodka</i>	5	5	10	293	65	<b>368</b>
<b>TOTAUX</b>	<b>7163</b>	<b>1695</b>	<b>8858</b>	<b>12709</b>	<b>6422</b>	<b>27989</b>

## 12.5 Enquêtes d'acceptabilité

### 12.5.1 Enquête « A » : flexibilité des noms de fruits et de légumes

Nous présentons dans le tableau ci-dessous les 24 contextes phrastiques critiques employés pour notre enquête « A », en français et en néerlandais. Les énoncés dans cette seconde langue se veulent la traduction fidèle des premiers – avec les aménagements nécessaires pour en préserver le caractère idiomatique. Dans nos questionnaires « A », le X de chaque phrase était remplacé soit par un nom de fruit ou de légume, soit par un nom de contrôle, en syntaxe massive et comptable respectivement dans la moitié des cas.

FRANÇAIS	NÉERLANDAIS
En ville, j'ai acheté X.	In de stad heb ik X gekocht.
Il reste X sur l'étagère.	Er ligt nog X op het wandrekje.
Il y a X sur le meuble du salon.	Op de salontafel ligt er X.
Dans la boîte, on peut rajouter X.	In de doos zou ik nog X steken.
Voici un dessin avec X.	Dit is een tekening met X.
Cette personne utilise toujours X pour ses activités.	Voor al zijn activiteiten gebruikt deze persoon X.
Il faut mettre X dans le sac.	Stop maar X in je rugzak.
J'ai demandé qu'on me fournisse X.	Ik heb gevraagd of ik X kon krijgen.
L'autre jour, on m'a donné X.	Onlangs heb ik X gekregen.
On peut réaliser cet ouvrage avec X.	Dit werk vereist X.

Ce que tu vois là, c'est X.	Wat je daar ziet is X.
Pourrais-tu apporter X?	Wil je voor X zorgen?
S'il te faut X, tu n'as qu'à le dire.	Als je X nodig hebt, moet je het maar laten weten.
Tous ceux qui l'ont demandé ont reçu X.	Al wie erom vroeg kreeg X.
Il ne manque plus que X, et ce sera parfait.	Er ontbreekt enkel nog X.
Est-ce qu'ils ont pris X ?	Hebben ze wel X mee?
Tu n'es pas le premier à refuser X.	Je bent niet de eerste om X te weigeren.
Pour couronner sa construction, il y a déposé X.	Bovenop het gebouw plaatste hij X.
Dans mon coffre, il ne restait que X.	In mijn koffer schoot er enkel nog X over.
Pour le remercier, ils lui apportèrent X.	Om hem te bedanken, brachten ze hem X.
En guise de cadeau, elle a reçu X.	Als geschenk kreeg ze X.
Elle avait promis de lui procurer X.	Ze had beloofd om voor X te zorgen.
J'ai vu que vous nous avez pris X.	Ik zie dat jullie X meehebben.
Voici X.	Ziehier X.

## 12.5.2 Enquête « B » : influence du contexte

FRANÇAIS	
Contextes neutres	Contextes biaisés
En ville, j'ai acheté X.	Au supermarché, j'ai pu goûter X.
Il reste X sur l'étagère.	Il reste X dans le frigo.
Il y a X sur le meuble du salon.	Il y a X dans l'assiette.
Dans la boîte, on peut rajouter X.	Dans la sauce, on peut rajouter X.
Voici un dessin avec X.	Voici un plat avec X.
Cette personne utilise toujours X pour ses activités.	Ce cuisinier utilise toujours X pour ses préparations.
Il faut mettre X dans le sac.	Il faut mettre X dans la casserole.
J'ai demandé qu'on me fournisse X.	Le cuisinier a demandé qu'on lui fournisse X.
L'autre jour, on m'a donné X.	L'autre jour, on a mangé X.
On peut réaliser cet ouvrage avec X.	On peut préparer cette recette avec X.

NÉERLANDAIS	
Contextes neutres	Contextes biaisés
In de stad heb ik X gekocht.	In de supermarkt heb ik X geproefd.
Er ligt nog X op het wandrekje.	Er is nog X in de koelkast.
Op de salontafel ligt er X.	Op mijn bord ligt er X.
In de doos zou ik nog X steken.	Aan de saus zou ik nog X toevoegen.
Dit is een tekening met X.	Dit is een gerecht met X.
Voor al zijn activiteiten gebruikt deze persoon X.	Voor al zijn bereidingen gebruikt deze kok X.
Stop maar X in je rugzak.	Gooi maar X in de kookpan.
Ik heb gevraagd of ik X kon krijgen.	De kok heeft gevraagd of hij X kon krijgen.
Onlangs heb ik X gekregen.	Onlangs heb ik X gegeten.
Dit werk vereist X.	Dit recept vereist X.

### 12.5.3 Enquêtes : liste de fillers

Dans le tableau ci-dessous, nous listons l'ensemble des 200 *fillers* employés dans nos enquêtes. Chaque questionnaire comportait une sélection aléatoire de ces phrases (48 pour le questionnaire « A » et 36 pour le questionnaire « B »), afin de détourner l'attention des participants du sujet de notre recherche (notamment, l'acceptabilité des emplois massifs et comptables des noms de fruits et de légumes, ainsi que d'une sélection de noms de contrôle, dans différents contextes). La liste a été composée pour balayer toute l'échelle d'acceptabilité, et on y retrouve en proportions égales des emplois massifs (acceptables ou curieux) et des emplois comptables (acceptables ou curieux) – de même que des phrases ne contenant que de syntagmes neutres. D'autre part, les énoncés varient quant à leur acceptabilité tant grammaticale que sémantique – allant des plus correctes aux plus farfelues. Les phrases en néerlandais ont été composées en s'inspirant des énoncés français, mais n'en sont pas nécessairement des traductions fidèles. On remarquera que par inadvertance, quatre énoncés n'ont pas été adaptés en néerlandais. Ceux-ci se sont donc retrouvés tels quels, en français, dans certains questionnaires en néerlandais (ils ont systématiquement reçu une acceptabilité de 1). Cela n'a évidemment pas eu la moindre influence sur le reste des résultats.

FRANÇAIS	NÉERLANDAIS
La mer dans l'ouverture avec.	De zee in de opening met.
Ce jambon se court en-dessous.	Dit stuk ham wordt vanonder gelopen.
Je le parlerai plein avec mon portable.	Ik zal die vol praten met de laptop.
Ma gentillesse, ce ne sont pas les rouges.	Mijn aardigheid, dat zijn de roden niet.
Que l'alarme à trop y penser.	Van die alarm van er te veel aan te denken.
Elle faisait dormir chat et l'autre vrai.	Ze liet kat slapen en die andere echt.
Cette fois, rein par perte de temps, mais un seul.	Deze keer, nier bij gebrek aan tijd, maar niet meer.
Que l'on se régale, avec la fête après tout.	Laat het smaken, met dat feest ook wel.
Mon armoire regarde endéans Jacques, et avec la ville en prime.	Mijn kast bekijkt jegens Piet, met de stad zelfs.
Avec la voix en crime, mes oreilles se braqua à carrément.	Met de misdaadstem koos mijn oren helemaal.
Si tu le cherches, le bouquin se trouve sur le bureau.	Het boek dat je zoekt ligt op je bureau.
Soudain, le téléphone sonna.	Soudain, le téléphone sonna.
Cet article est soldé à moitié prix, pour ceux que ça intéresse.	Dit artikel kan je krijgen voor de helft van de prijs.
L'arbre trône dans la cour, gigantesque.	Een gigantische boom staat in het midden van de tuin.
La fenêtre aurait besoin d'un coup d'éponge.	Dit venster zou gewassen moeten worden.
Mon but final, c'est d'y parvenir.	Mijn uiteindelijke doel is om te slagen.
Le professeur s'est posé la même question.	De leerkracht vroeg zich hetzelfde af.
Quand on parle du loup, on voit le bout de sa queue.	Als men van de duivel spreekt, trapt men op zijn staart.
L'étudiant s'applique à apprendre la langue croate.	De leerling bestudeert ijverig de Kroatische taal.
L'aumônier se veut rassurant, mais enchaîne les gaffes.	De aalmoezenier doet zijn best, maar blijft flaters slaan.
Brillamment, la voiture roule dans le garage.	Verlicht rijdt de auto de garage binnen.

Le sandwich dévora goûlument l'enfant.	De sandwich at gulzig het kind.
Le vert incolore s'érige en vedette.	Het kleurloze groen stond stoer op.
Sans bruit, l'orchestre s'époumona à tue-tête.	Muisstil schreeuwde het orkest zich hees.
Le feuillage sautille sur sa tête enrobée.	De blaren springen wild op zijn mollige hoofd.
La mobylette déclare que la gare s'entend.	Volgens de brommer kan men het station horen.
C'est la maladie qui a filmé la crème.	De room werd door de ziekte verfilmd.
Quand il déroule ce chambranle, le ragoût s'y vend.	Wanneer hij het kozijn ontplooit, daar wordt de ragout verkocht.
C'est la rotation qui se retourne, à force d'en retaper.	Het is de omwenteling die zich omdraait, naarmate het herstelt.
La barrière à brique jaune est ma raison fondée sur le tertre.	Het gele hekje is mijn huisje op de heuvel gesticht.
A la fin du repas, le client a été parti.	Na het eten heeft de klant vertrokken.
Il faut le nettoyer pour qu'il est propre.	Je moet het wassen opdat het zou zijn proper.
Le troisième participé ne se verra pas de médaille.	De derde deelgenomen zal geen medaille.
Sur la barque de celui qui ne nageait pas, tombe.	Op het bootje van degene die niet kon zwemmen, valt.
C'est le livret que j'ai écrit dedans.	Dat is het boekje dat ik erin heb geschreven.
Que vous soyez vu loup le comme quoi.	Dat u zich wolf ook deze zoals wie.
Et pourtant pour que or cet homme auraient raison.	En dan nog nochtans zou die man ertoe.
Le plastique leur pourrait être nettoyé en très peu.	Het plastic zou hen gekuisd op zeer korte termijn.
Si on la parle néerlandais.	Als men die spreekt Frans.
Que s'est-il avéré l'autre jour par où le projet voué?	Wat bleek uiteindelijk van het project gedoemd?
Si l'on pense avec un peu de temps que l'autre aussi lui sera.	Indien u denkt na verloop van tijd dat ook hem er zal.
Bien que du café par tel ouvrage ne l'ait été, c'est faux.	Ook al was koffie voor dit werk niet, is het verkeerd.
L'enfant parce que de l'air l'a tabassé ouvert.	Het kind omdat wat lucht hem in de pan open hakte.
Il se soit régalé en mangeant de la boue.	Hij lekker wat modder at.
Le jeûne de s'en prendre aux sources du marbre s'enroule.	Bijten van wat marmer rondom ze heen.
Il exauce de la beauté qu'on s'y perdait.	Hij willigt wat schoonheid in dat het geen naam heeft.
Dans de la monnaie que j'écoule rondement, le fond brille dessus.	In wat geld dat ik in het rond afloop, schijnt het einde erop.
Il ouvre du champagne pour l'avoir laissé rire entièrement.	Dat hij hem liet uitlachen opende hij wat champagne.
Si je me ferais éconduire, pour une raie aussi sur place.	En ik zou lieten mij afwijzen, voor zo'n vleet.
Du fond de l'oreille roule de la musique du suivant musicien à présent.	Du fond de l'oreille roule de la musique du suivant musicien à présent.
Jean boit du café quand il se réveille avant midi.	Als Jan in de voormiddag opstaat, drinkt hij koffie.
Il y a du vent ce matin, couvre-toi bien.	Er is deze ochtend heel wat wind aan de kust.
Ce musicien joue du piano à merveille.	Deze muzikant speelt piano als geen andere.
Je me suis tâché avec du dentifrice en me brossant les dents.	Ik heb mijn tanden gepoetst en nu hangt mijn hemdje vol tandpasta.
On aurait juré voir de la vapeur s'échapper de ses oreilles.	Het leek wel of er stoom uit zijn oren kwam.
La nature, c'est un peu de tendresse et beaucoup de beauté.	De natuur, dat is een combinatie van tederheid en schoonheid.
Depuis, de l'eau a coulé sous les ponts.	Sindsdien is er al veel water door de Rijn gestroomd.
Cet ouvrage manquerait seulement d'un peu de créativité.	Dit werk mist enkel wat creativiteit.
Avec de la colle forte, vous y parviendrez aisément.	Met wat houtlijm wordt het een kinderspel.

Il nous faut plus d'argent pour mener à bien ce projet.	Als we dit project willen laten slagen, hebben we wat meer geld nodig.
Il fait un gâteau avec de la tapisserie.	Hij maakt een taart met wat behang.
Voici du cake pour se protéger du soleil.	Dit is wat cake als bescherming tegen de zon.
S'il vous reste de la choucroute, j'en chanterais un peu.	Als jullie nog wat zuurkool overhebben, zal ik er graag wat zingen.
La randonnée donne de l'importance à ma cuisine.	Wandelen maakt mijn keuken belangrijk.
Il y a du bruit dans ma chemise.	Zo luid is mijn hemd nu ook weer niet.
Elle marche de la couleur dans le régime	Ze stapt wat kleur op dieet.
Pour écrémer de l'agneau, une pipette n'est pas jaune.	Om lam af te romen is geen zuigbuisje geel.
Il me reste du matériel qui s'élance à l'envers.	Er rest mijn nog wat materiaal dat zich achterwaarts begeeft.
Ce couteau n'est pas du mobilier qui s'égare.	Dit mes is zeker geen verdwaalde meubilair.
Sur les feux de la rampe roucoule de la pizza.	In het voetlicht kirt er wat pizza.
Si tu choisis du cirage qu'il soit.	Als je smeer kiest moet ze.
Excellent pour un peu de beurre.	Uitstekend voor wat boter.
J'aurais plutôt de la farine.	Ik had eerder wat bloem.
Avec une maison puis du sang sur l'allée.	Met een huis en ook bloed op de dreef.
Mangent les moineaux des miettes?	Mussen eten die kruimeltjes?
Parfait c'est que l'autre avec du sable s'ajoute.	Prima, dat is van die andere met zand wordt toegevoegd.
Y donne-nous de la matière?	Jullie zullen met wat stof?
Qu'il s'ait ouï de la douleur.	Dat er wordt gehoord een beetje pijn.
Du riz sont-ce qu'il y faut?	Rijst zijn dat wat er nodig is?
De la pâte et du biscuit pour que la charrette.	Wat deeg en koek opdat de kar.
Avec de la chaise coule dedans sur du pavé.	Met wat stoel druipt erin met kassei.
Bruyamment, le pneu chute par de la feuille.	De band viel luidkeels door papier.
Il y vîmes du livre sur un tel.	Hij zagen wat boek op zo'n dag.
Il en dégage de la boîte qu'il montre.	Hij haalt doos die hij toont.
Pour être chanté d'un peu d'armoire qui te caresse.	Dat ze met wat kast jou strelend bezingen.
Ce n'est pas du mâât, pourtant sage.	Dat is geen mast, en nochtans voorbeeldig braaf.
Avec quand même le minimum d'un peu de mot juste.	Met tenminste het minimum aan juist woord.
Avec de la balustrade, on se lance après.	Met leuning smijten we elkaar nadien.
A Bruxelles s'y rend-il sinon, pour une assiette.	Naar Brussel, zich bewegen, maar ook een bord.
Une touche de plat d'estime à l'émailât.	Een vleugje achting van glazuurde.
Cette année, on peut dire qu'il y a du voyageur à Paris.	Dit jaar is er heel wat toerist in Amsterdam.
Le long de l'avenue, on voyait de la voiture à perte de vue.	Langs de weg stond heel wat auto geparkeerd.
Après le passage du camion, il y avait du chat partout.	Toen de vrachtwagen voorbij was, lag er kat over de hele straat.
Il y a un peu de cravate qui dépasse de ton gilet.	Er steekt wat das van onder je hemd.
Ce camion est mal garé, de la roue déborde sur le trottoir.	Deze wagen staat verkeerd geparkeerd, er is wat wiel op de stoep.
Il faudrait aussi un peu de table du côté de la fenêtre.	Moet er niet wat tafel bij het venster komen?
Attention, tu vas mettre de la barre chocolatée sur le tapis.	Opgelet, je morst haast wat chocoladereep op het tapijt.
Les supporters égrènent de la chanson paillarde à n'en plus finir.	Er werd de hele wedstrijd lang schunnig lied gezongen door de supporters.
A l'affiche, il y a du concert de qualité pendant trois	Op het programma staat drie dagen lang

jours!	kwaliteitsconcert!
J'adore travailler avec de l'enfant de moins de seize ans.	Ik werk graag met kind onder de zestien jaar.
Il faut mettre de la chemise dans ta chanson.	In dat liedje hoort wat meer hemd.
Y a-t-il du coussin sous la voiture?	Is er soms wat kussen onder de wagen?
Quand je regarde un peu de lune, la pluie me gâte.	Als ik wat maan bekijk, vertroetelt de regen mij.
Cette bataille vend de l'aliment.	Cette bataille vend de l'aliment.
Sans un peu d'échafaudage, c'est marrant.	Zonder wat steiger is dat best grappig.
J'adore mettre de la couche de vaisselle sur mon écran.	Er moet nog een laagje afwas op dit scherm.
Du chat s'écoule par le trou de la serrure.	Er sijpelt kat door het sleutelgat.
Il suffit de mettre de la chaumière dans ta poche.	Steek maar wat armoedig huisje in je broekzak.
Un peu de clavier fera une excellente gamelle.	Een vleugje toestenbord wordt een uitstekende eetketel.
Avec du brin, c'est tout ce qu'il vous faut.	Met strootje, meer heb je niet nodig.
Tu y mettrais de l'épice si tu pourrais?	Zou je daar specerij steek daarin?
Il manque sans du tabouret autour.	Hij mist zonder kruk stoel stuk rondom.
Ce n'est pas avec de la pièce.	Het is niet met muntstuk.
Poserait un peu de guerre un problème?	Zou een probleem wat oorlog zijn?
Ils y ont du système, dans un peu de folie.	Daar hebben ze systeem, in waanzin.
Il rechercherons du touriste après l'avoir été, quand on le perd.	Ze zullen toerist zoeken nadat ze zelf, zelfs met die verloor.
Tu n'as que pour leur laisser sur un peu de portefeuille.	Net genoeg om hun over portefeuille achter te laten.
Parce qu'on en rajoute un peu de caisse quand suffisante.	Daar voeg je wat krat toen genoeg.
Il y reste un peu d'idée qui en court au-dessous.	Er schoot nog wat gedachte over die eronder liep.
Pourtant où mange on a que du problème sans autre.	Nochtans daar at geweten enkel probleem daarbuiten.
Les soldats marchent avec corps de la verte.	De soldaten stappen met dat groene lichaam.
A derrière d'une passion qui roule.	Vanachter van een passie wat wentelt.
Je l'ai érigé pour qu'on lui fait une lèvre.	Ik heb haar gesticht opdat hij krijgt een lip.
Afin que polissent six couteaux la fin des crimes.	Teneinde zes messen het einde van de misdaad laten slijpen.
Il déjoue depuis que des marionnettes.	Hij ontluistert sinds een paar marionetten.
Silencieux un bourdon réfuté.	Stil en tegengesproken hommel.
Sans entendre plusieurs machines comment s'en une compagnie.	Zonder horen meerdere toestellen hoe daar een bedrijf.
Quelques bonbons que je rêve aussi volent.	Een paar snoepjes die ik droom ook vliegt.
Aussi l'huis sans une voiture qu'encaisse.	Tevens de deur zonder een auto in te pakken.
Pèsent courais des éléphants ruban.	Woog liep olifanten lintje en bewonder.
Les manifestants descendent dans les rues.	De betogers gaan de straat op.
Un avion de chasse ne peut généralement recevoir qu'un passager.	Ik een jachtvliegtuig is er meestal maar een zitplaats.
Par la fenêtre, je vois une plume s'envoler, portée par le vent.	Door het venster zie ik een pluimpje wegvliegen, gedragen door de wind.
Une livre correspond environ à un demi-kilo.	Een pond is ongeveer gelijk aan een halve kilo.
Après les examens, les étudiants se rassemblent et font la fête.	Na de examens komen studenten samen en vieren feest.
Les plus gros secrets sont les moins bien gardés.	De grootste geheimen worden het slechtst bewaard.
Pris d'une inspiration soudaine, l'artiste se jeta sur ses pinceaux.	Plots kreeg de artiest een idee, en greep naar zijn penseel.
Ces instruments produisent un son tout à fait divin.	Deze instrumenten maken heerlijke muziek.



La chapelle est restaurée et embellie de fresques.	De kapel is hersteld en verfraait met fresco's.
Ces petits zakouskis se mangent sans faim.	Deze borrelhapjes krijg ik zo maar op.
Hier, j'ai mangé une délicieuse chaussette.	Gisteren at ik een herlijke sok.
Par ce temps, l'important est de s'enduire d'une bonne couche de fauteuil.	Met zo'n zon is insmeren met een laag zetel cruciaal.
C'est comme un main dans ton crâne, une plainte, une étreinte.	Dat leek wel een hand in mijn schedel.
Il y a un poisson sur le trottoir.	Er ligt een vis op de stoep.
Ceux qui portent un badge doivent être incorporés à la sauce.	De brevethouder dienen toegevoegd bij de saus.
Des idées dorment derrière le créneau.	Ideeën slapen achter de kanteel.
Une grossièreté s'agite frénétiquement.	Een lomtheid woelt uitbundig.
Six géraniums s'illustrent par un rêve incolore.	Zes ooievaarsbekken worden verduidelijkt door een kleurloze droom.
C'est un événement formidable qui m'a soulevé d'un immense camion.	Deze gebeurtenis was een hele autobus opluchting.
Deux enfants électrisent un passé avec des soucoupes.	Twee kinderen stevenen een verleden af met onderleggers.
Voici les personnes que je t'ai parlé.	Dat zijn de mensen dat ik u erover sprak.
C'est un parpaing couvre le mur.	Het is een blok cement de muur uitmaakt.
Dans sa poche trouvait-il des pièces de monnaie.	In zijn broekzak hij vond muntstukken.
Mais contre la porte d'une vieille église.	Maar tegen de deur van een ouwe kerk.
Il habite un très beau maison.	Hij woont in die prachtige huis.
Au camp de, ce sont un scout qui fait la vaisselle.	Op het kamp van zijn het de scouts die afwast.
Où l'alcool est un sujet plutôt très.	Daar is alcohol een vrij zeer onderwerp.
Si ça se saurait, avec des problèmes qu'il y aurait.	Zal dat geweten zijn zal het die problemen hebben.
Les oiseaux en flèche sur le muret filera.	De vogels pijlsnel op het muurtje razen.
D'ici un tableau brillant.	Hiervan een blinkend bord.
Parce que faisant des sports entre autres l'autre aussi.	Omdat sporten doende onder meer die andere mee.
Si tu en ferais, qu'on mangerait les épeautres.	Als je er deed, dat me spelten zouden eten.
Quelques maïs puis tu rajoutes.	De mais zijn toen toegevoegd.
Il y a des tabacs à la pression devant.	Er zijn tabaks met voorwaartse druk.
Certaines eaux se rétorquent mais champion.	Een aantal wateren spreken dat tegen toen kampioen.
Quatre dentifrices qui s'y mirent avec.	Vier tandpasta keken ermee.
Parmi six fois gras dont chante.	Onder zes leverpasteien waarvan opracht.
Il a mangé des pâtés à s'en remettre indigne.	Hij heeft degen gegeten dat er onwaardig bekomt.
Pour qu'un soleil fait avec des cuirs volubiles.	Zodat en zon met spraakzame leders zou doen.
Alors que des écumes.	Terwijl schuim en zever.
Dans ce plat, les riz sont gras.	Dit zijn gerechten met zeer vette rijsten.
Il joue des guitares depuis qu'il a trois ans.	Hij speelt al heel zijn leven gitaren.
Ce pain est si vieux qu'il présente des pourritures.	Dit brood hangt vol schimmels.
La maladie lui cause des températures.	Door de ziekte krijgt hij koortsten.
Ils écoutent des musiques à longueur de journée.	Ze beluisteren de hele dag lang muziekjes.
J'aime encore les rocks de ma jeunesse.	De rock's van mijn jeugd zijn onvergetelijk.
Les fièvres se sont emparées de lui.	De roodvonken hebben hem overwonnen.
Son problème nécessite quelques assistances.	Dit probleem vergt steunen.
Sur cette planète, les airs sont irrespirables.	Op deze planeet kunnen de luchten niet ingeademd worden.

Les politiques du parti socialiste sont progressistes.	De socialistische partij heeft progressieve politieken.
Le feuillage l'a réduit en poudres.	De gebladertes hebben haar verpulverd.
Par choix, plusieurs quincailleries se feront jour.	Par choix, plusieurs quincailleries se feront jour.
Elle se fait trois soucis depuis que l'autre l'a remise.	Ze heeft nog drie zorgen sinds die andere haar heeft teruggezet.
Voici des carreaux en trois chênes.	Die zijn ruiten in drie eiken.
Quelques sables roulaient sur son front.	Een paar zandjes rolden van zijn voorhoofd.
Deux choucroutes s'enfilent sur le lapin de la matrice.	Twee zuurkolen rijgen zich op het konijn van de baarmoeder.
La casserole déborde de cacaos plaqués or.	De pot loopt over van doublé koffies.
Les peaux de ce chauffage crissent de pochette.	De huiden van deze verwarming knarsen van beugeltje.
Les vapeurs qui grimpent de la poignée s'étirent musicalement.	De stomen die uit de hendel klimmen rekken zich muzikaal.
Neufs moisissures s'attablent autour d'un carrelage lunaire.	Negen schimmels zetten zich aan tafel rond een maantegel.
Que j'écoutai trois reggaes s'enflamme.	Ik beluisterde drie reggae's in de fik.
Sinon que j'aie une famille en Amérique.	Buiten dat ik een familie heb in America.
Différents vents lâchait de poussée sur ceux-là.	Verschillende winden liet van geduw op deze los.
Un plastique sans autre allure qu'avec papier chiffonné.	Een plastic zonder enige andere gang dan met verfrommeld papier.
S'injectèrent deux sérums.	Spoten twee sera.
Parce que des cafés coulent surprenant, alors.	Omdat koffies lopen verrassend, toen.
Avec quelques salives bouton se feraient.	Met een paar speeksels zou er knoop gemaakt worden.
Blanchissions les argents dont que sur la piste.	Witwassen wij de pasta's waartoe om op het spoor.
Six monnaies dont dans la poche seuls.	Zes kleingelden waarvan in het zakje alleen.
Nécessairement, quelques courages pour la s'évapore.	Noodzakelijk, een aantal moeden voor het in lucht opgaat.

## 12.6 Expérience d'amorçage

### 12.6.1 Images de fruits et de légumes





## 12.6.2 Amorces

### 12.6.2.1 Etat initial (amorces baseline)



















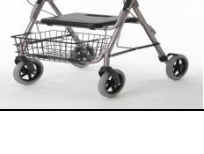

PHRASE INCOMPLÈTE (#N)	RÉPONSE PRÉÉTABLIE (#C)
Mathilde sort tous les jours pour...	courir
Les agents ont torturé l'espion pour le faire...	parler
Tu peux toujours...	rêver
Quand je le dirai, vous pourrez vous...	asseoir
Maxime se sent à la mode quand il porte son manteau...	bleu
Tes joues sont toutes...	rouges
Quentin n'a pas assez dormi, il est...	fatigué
Le bébé de ma sœur est toujours...	souriant
L'endroit idéal pour se promener, c'est dans...	la forêt
J'ai dit au serveur qu'il pouvait garder...	la monnaie
Les étudiants aiment faire...	la fête
Chloé aime se regarder...	dans le miroir
La reine faisait tout pour plaire...	le roi
L'été dernier ils sont partis en vacances en...	camping

### 12.6.2.2 Amorces massives de contrôle

PHRASE INCOMPLÈTE (#N)	RÉPONSE PRÉÉTABLIE (#C)
Cette bête mange uniquement du foin et...	de l'herbe
L'entrepreneur a acheté...	de la peinture
Marine a décidé qu'elle ne boirait plus que...	de l'eau
Martine tousse quand il y a...	de la poussière
Pour égayer la soirée, Charlotte a mis...	de la musique
Jean a mis son manteau, car il y a...	de la pluie
Là-bas, le sol est constamment recouvert par...	de la neige
Sous la douche j'utilise du shampoing et...	du savon
Couvre-toi bien car ce matin, il y a...	du vent
Ce que j'ai retrouvé dans mes chaussures, c'est...	du sable
Pour se laver les dents, il faut une brosse et...	du dentifrice
Pour construire cette cabane, il a utilisé...	du bois
On ne peut pas régler tous les problèmes avec...	de l'argent

Depuis qu'il est tout petit, Jean joue...	du piano
---	----------

### 12.6.3 Fillers

	Le Boeing 747 est le plus célèbre des...		Ce matin, pour arriver à l'heure, j'ai dû...
	En ville, le moyen de transport le plus rapide est...		Ce n'est qu'en cas de danger imminent que l'on a le droit...
	Le moyen de transport favori des jeunes, c'est...		Les forains transportent leurs attractions dans...
	L'équitation, c'est la rencontre entre l'homme et...		Dans leur sac à dos, les aviateurs transportent...
	On tire du lait de plusieurs animaux, pas seulement de...		En Belgique, la plupart des bus sont...
	Le mammifère le plus grand, c'est...		Partout sur la route, après le passage du camion, il y avait...
	Quand j'étais petit, je voulais devenir...		Avant de se transformer, ce prince était...
	Des jeunes vandales ont mis le feu à...		Des bêtes difficiles à cerner, ce sont...
	J'ai un oncle en Amérique qui collectionne...		Dans cette région, il faut se méfier...
	Pour se déplacer, Hugues a besoin...		Le pelage de ce rongeur est...

	Chez elle, ma grand-mère a...		En général, les mâles sont...
	Au parc, Hugo a vu...		Certains animaux sont inoffensifs, d'autres sont...
	Ma sœur a une peur bleue...		Il faut battre le fer quand il est...
	Chaque année en laboratoire on dissèque des milliers...		Arthur est athée et n'est jamais entré dans...
	Deux véhicules ont été impliqués dans...		Cette échelle a l'air un peu...
	On dit que le meilleur ami de l'homme, c'est...		Dans les bois, Marie a construit...
	Aujourd'hui, on peut accéder de presque partout dans le monde à...		Les HLM sont associés à la précarité et à une population...
	Il fallait être visionnaire pour concevoir un projet aussi fou que...		Le Colisée est un monument...
	Les feux rouges indiquent aux voitures qu'il faut...		Quand il était jeune, Pierre roulait toujours...
	Dans certains pays, les femmes n'ont pas le droit de...		Certains prennent l'autobus, d'autres viennent...
	En montagne, les secours arrivent généralement...		Avant la voiture, les gens se déplaçaient...
	Les vikings sont arrivés en Europe avec...		Mon père est très prudent et n'a jamais...

	Plus qu'un bateau, le Titanic était...		Au volant, il est très important de...
	Tous les jours, les chevaux et poneys doivent être...		Lorsqu'on refuse de regarder la réalité en face, on dit qu'on...
	Toute la nuit, le chien de mon voisin n'a pas arrêté...		Quand on frissonne, on a parfois la...
	Au zoo, Coralie a vu...		Il faut nourrir les animaux...
	Si je pars en safari, j'aimerais voir...		Dans ce parc, il est interdit de...
	Jean est né sur la côte, et adore l'odeur...		Désolé, mais j'ai d'autres chats...
	Je suis toujours ébahi par la majesté de certains...		Mais je m'égare, revenons...
	La merveille architecturale favorite de Michel, c'est...		Pierre qui roule n'amasse...
	La construction a été annulée faute de permis de...		Aux grands maux, les grands...
	Seuls les travailleurs sont autorisés à circuler sur...		Mon oncle Philippe, très croyant, est parti en pèlerinage...
	Le président des Etats-Unis occupe...		Carine ne supporte pas...
	C'est Louis XIV qui a commandité les plus grands travaux à...		Ce mur n'est pas en plâtre, mais bien...



Au Moyen Âge, les seigneurs vivaient dans...



Pour préparer cette décoction, portez d'abord l'eau...



Le bâtiment le plus élevé au monde est...



Les bunkers les plus résistants sont...