

L'opposition massif/comptable au niveau lexical et supra-lexical

Vermote, Timotheus

Universiteit Gent
timotheus.vermote@ugent.be

1 La localisation de l'opposition massif/comptable

Si l'on compare les noms *eau*, *sable* ou *vent* aux noms *voiture*, *chaise* ou *chat*, un constat spontané (et peut-être naïf) est qu'en vertu de leurs propriétés distributionnelles, les premiers sont des noms massifs (1), et les seconds des comptables (2).

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

En effet, les premiers se combinent sans peine avec l'article partitif ou la locution déterminative *un peu de*, tandis que les seconds prennent volontiers l'article indéfini singulier *un(e)*, se mettent au pluriel, et se laissent dénombrer. Cependant, d'un autre côté, ces critères distributionnels syntaxiques ne peuvent être le fin mot de l'histoire, puisque l'on a aussi (avec plus ou moins de hardiesse, selon les cas) les énoncés en (3) et (4).

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

En outre, si l'on accepte que les énoncés en (1) et (2) illustrent, d'une façon ou d'une autre, des emplois « de base », tandis que ceux en (3) et (4) seraient « dérivés », que dire alors des exemples (5), (6) et (7), où il semble moins évident d'établir lequel des deux usages serait dérivé de l'autre, et pourquoi ?

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

(Dans les exemples ci-dessus, nous ne prétendons pas que les phrases contrastées sont équivalentes d'un point de vue sémantique – à notre sens, elles ne le sont pas – mais seulement que les deux emplois, massif et comptable, semblent aussi naturels l'un que l'autre.)

Les théories de la localisation de l'opposition massif/comptable dans le système linguistique doivent donc tenir compte de ces deux réalités préliminaires : 1) la plupart des noms ont une préférence tranchée, identifiable intuitivement, pour l'un des deux types d'usage et 2) outre ces préférences, l'opposition est particulièrement flexible, puisqu'elle permet à bon nombre de noms d'alterner (plus ou moins librement) entre les deux lectures.

Parallèlement à l'existence de ces deux faits cruciaux, l'on retrouve dans la littérature sur le sujet deux grandes familles de théories, qui se distinguent selon qu'elles mettent l'accent principal sur la première,

ou sur la seconde des deux observations. C'est ainsi que l'on peut distinguer d'une part les théories lexicalistes de l'opposition massif/comptable (qui mettent en avant les restrictions lexicales) et, de l'autre, les approches grammaticales, ou purement syntaxiques (qui insistent sur la flexibilité). Dans ce qui reste de cette introduction, nous rappelons brièvement les idées centrales de ces deux types d'approches, identifions leurs principales (et souvent célèbres) imperfections respectives, et annonçons ce qui, d'après nous, reste à résoudre sur la question.

Pour les tenants d'une approche lexicaliste, les noms sont marqués au niveau du lexique par un trait (ex. Galmiche, 1989 ; Gillon, 1999 ; Nicolas, 2002 ; Cheng *et al.* 2008) ou une combinaison de traits (ex. Landman, 2011 ; Zhang, 2013), qui opèrent des distinctions lexicales entre les noms. En essence, il y a donc certains noms qui sont massifs, et d'autres qui sont comptables. Les arguments en faveur d'une telle vision reposent généralement sur deux constats : 1) les locuteurs ont des intuitions fortes et convergentes concernant le statut massif ou comptable de bon nombre de noms (cf. Nicolas, 2002), et 2) les effets de l'opposition massif/comptable sont visibles même lorsque les noms ne sont pas marqués explicitement (Galmiche, 1989 – *Le sang avait coulé, il y en avait par terre* vs. **Le livre était tombé, il y en avait par terre* ; Le Bruyn *et al.*, 2011 ms.).¹ Quant à la flexibilité des noms, qui peuvent parfois connaître les deux emplois moyennant un contexte adéquat, elle s'explique soit parce que certains noms sont polysémiques et ainsi lexicalisés dans leurs deux usages (ex. *agneau* : animal/viande, cf. Kleiber, 1999 ; Nicolas, 2002), soit par des mécanismes de transfert sémantiques, permettant d'obtenir pour certains noms des emplois moins habituels, mais néanmoins autorisés car associés à un glissement de sens spécifique. Ces possibilités de glissements sont généralement considérées comme systématiques, et leurs effets sémantiques ont été décrits traditionnellement à l'aide de machines métaphoriques (réputées parfois Universelles) : le Broyeur (en s'inspirant de Pelletier, 1975²) (8), le Trieur (Bunt, 1985) (9), le Conditionneur (Bunt, 1985 ; Jackendoff, 1991) (10), ou encore le Multiplicateur (Galmiche, 1989) (11).

(8) Après le passage du camion, il y avait du chat partout.

(9) Voici un sable de très grande qualité. Nous vendons des terreaux naturels et enrichis.

(10) Nous prendrons une bière et deux limonades.

(11) Ce weekend, nous avons fait du kilomètre. Il y a du touriste à Paris cette année.

A cette approche lexicaliste, qui situe la distinction entre massif et comptable dans le lexique, s'oppose naturellement la vision grammaticale, qui n'attribue pas le trait aux noms, mais aux unités syntagmatiques, par le biais de la syntaxe. Ainsi, d'après une vision grammaticale de l'opposition, tous les noms ont le même statut (et sont généralement considérés comme massifs, par défaut), tandis que le trait comptable ne peut être obtenu qu'au travers d'une marque syntaxique spécifique, par exemple un déterminant numéral ou un pluriel. L'opposition massif/comptable est ainsi non pas un trait des N, mais bien des SN. A la tête des tenants d'une telle approche, l'on trouve dans la littérature récente Borer (2005), et ses partisans (ex. Bale et Barner, 2009 ; De Belder, 2011), bien que l'idée générale la précède, et soit bien plus ancienne (ex. Taylor, 2002 ; Allan, 1980 ; Weinreich, 1966, et même Damourette et Pichon, 1927).³ Dans pareille vision, les noms peuvent bien avoir des préférences réelles et mesurables à partir de critères distributionnels (cf. Allan, 1980), mais elles découlent de critères conceptuels, et non syntaxiques. Pour De Belder (2011), ceux-ci relèvent de notre connaissance encyclopédique (qui ne fait pas partie du système linguistique formel, qui lui ne connaît pas d'exceptions, et donc pas de flexibilité) : 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 rien de problématique⁴.

Comme nous allons le montrer (ou le rappeler), ces deux types d'approche présentent chacune des lacunes notables. Premièrement, si la vision grammaticale repose grandement sur la flexibilité de l'opposition massif/comptable, force est de souligner que celle-ci n'est pas absolue, et ne s'applique pas à tous les noms de la même manière (*contra* Pelletier, 2012). Ainsi, si un nom réputé comptable peut, sans aucun problème, être combiné avec un numéral (*trois chaises, deux idées*), un nom réputé massif, quant à lui, ne jouit pas de cette possibilité avec la même aisance – même s'il peut dans certaines circonstances apparaître au pluriel. Ainsi, si *du sable* est massif, et que l'on peut avoir *des sables* dans certaines circonstances (comme dans *des sables mouvants*), l'on n'a pas pour autant **neuf sables*. Dans le même ordre d'idée, s'il est vrai que bon nombre de lexèmes présentent un certain degré de flexibilité, il est généralement possible de distinguer un emploi de base d'un emploi dérivé, ce dernier impliquant bien souvent un glissement sémantique (*une bière* n'est pas seulement 'de la bière', mais 'une portion (standard) de bière') et/ou nécessitant un contexte plus 'coercitif' pour être justifié (??*J'ai vu du chat* vs. *J'ai mangé du chat*)⁵. Ces différents faits semblent difficilement réconciliables avec une approche qui donne à tous les noms exactement le même statut dans le lexique. Or, l'idée de reléguer les intuitions des locuteurs quant au statut massif/comptable des noms à des différences encyclopédiques (concernant les référents) n'est pas convaincante non plus, car elle ne permet nullement d'expliquer des différences comme *du raisin* (massif) contre *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, l'on en revient à une vision lexicaliste, et la question se pose (à nouveau) de la représentation de cette connaissance dans le lexique.

La position lexicaliste, néanmoins, n'est pas sans ses propres manquements. D'un côté, l'idée des deux cas de figures complémentaires – polysémie lexicalisée contre mécanismes sémantiques productifs – est attrayante, mais de l'autre, la limite entre les deux n'est pas clairement définie et cela pose problème. En particulier, la flexibilité systématique au sein de certaines classes sémantiques continue d'intriguer. Sur quelle base dira-t-on, par exemple, qu'*agneau* est un terme polysémique (animal/viande, cf. Kleiber, 1999 ; Nicolas, 2002), tandis que l'emploi massif dans *En Australie, on mange du kangourou* serait dérivé d'un emploi lexical comptable, unique ? Une telle séparation est intéressante et repose probablement sur des faits de fréquence (bien qu'à notre connaissance, les théoriciens lexicalistes de l'opposition ne se soient jamais risqués à des études quantitatives) mais elle 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, concernant les emplois de base et dérivés, nous n'avons pas connaissance que de réels critères formels et systématiques (autres que l'intuition et le bon sens) aient 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 reste à le démontrer.

En somme, la question de la localisation de l'opposition massif/comptable est loin d'être résolue. Tant l'approche grammaticale que lexicaliste, en tentant d'apporter des réponses aux questions, apportent leur propre lot de problèmes. Dans cette contribution, nous entendons dès lors apporter un regard neuf sur la problématique, afin de tendre vers un modèle plus satisfaisant, qui puisse expliquer : 1) les intuitions fortes et convergentes des locuteurs concernant le statut massif/comptable des noms, 2) la flexibilité morphosyntaxique de beaucoup de noms, qui s'accommodent tantôt d'une syntaxe massive, tantôt d'une syntaxe comptable, 3) le caractère systématique de cette flexibilité au sein de certaines classes sémantiques, et 4) le lien entre statut lexical et fréquence. Pour ce faire, nous opterons pour une démarche

résolument empirique, basée sur la récolte de données (à partir de trois études de corpus et d'une enquête d'acceptabilité), dont découleront nos considérations théoriques.

2 Etudes de corpus

Les trois études de corpus menées dans le cadre de la présente recherche concernent chacune un champ sémantique spécifique, sélectionné pour présenter (au moins à l'intuition) un certain degré de flexibilité régulière et systématique, du point de vue de l'opposition massif/comptable. Les trois classes sémantiques ainsi sélectionnées sont :

- **les noms de fruits et de légumes** (27 items) : classe de noms principalement comptables (*une pomme, un potiron, une échalote*), qui autorisent un emploi massif dans une lecture en « forme et quantité indéterminée » (*de la pomme, du potiron, de l'échalote*, cf. Vermote, 2013) ;
- **les noms de matériaux** (30 items) : classe de noms rigoureusement massifs (*du sable, du verre, du cuivre*), autorisant – comme tous les massifs – un emploi comptable en termes du Trieur Universel (cf. Bunt, 1985 : *un sable* dans le sens de 'une sorte de sable'), ainsi que, pour certains items, un emploi comptable lexicalisé en termes d'un objet ou d'une collection d'objets spécifiques (*un verre* 'type de récipient', *des cuivres* 'instruments de musique') ;
- **les noms de boissons** (35 items) : classe de noms principalement massifs (*du vin, du coca, de la soupe*), qui autorisent les emplois comptables en termes de Trieur (*un vin [délicieux]*) ainsi que, pour certains items, un emploi comptable en termes de portions standardisées (Conditionneur Universel, cf. Bunt, 1985 ; Galmiche, 1989, ex. [*boire*] *une bière, une soupe*), ce dernier semblant même être l'emploi de base dans certains cas (*un espresso, un milkshake*, qui sont peut-être plus naturels que *de l'espresso, du milkshake*).

Pour chaque classe, la sélection des items s'est effectuée à partir de la fréquence – telle qu'exprimée dans la base de données LEXIQUE3 (New, Pallier & Ferrand ; <http://www.lexique.org/>) – combinée à un souci de représentativité (balayage d'un maximum des sous-catégories). La liste complète des items étudiés apparaîtra au travers de l'examen des graphiques distributionnels ci-dessous.

Pour chacun des lexèmes sélectionnés, des données textuelles ont été récoltées sur Internet, à l'aide de l'outil informatique *WebCorp* – un logiciel de concordance se servant du Web comme corpus, et disponible en libre accès⁶. A l'aide d'expressions régulières, nous avons recherché pour chaque item simultanément ses formes au singulier et au pluriel, pour des moyennes d'occurrences pertinentes par catégorie de 738,8 (fruits et légumes), 467,7 (matériaux) et 385,4 (boissons). Les occurrences ainsi récoltées ont ensuite été soumises à un examen approfondi. Tout d'abord, elles ont été triées en fonction du type de syntagme – massif, comptable, ou neutre – dans lequel l'item concerné apparaissait. Pour cette première classification, nous avons observé rigoureusement les contextes syntaxiques de surface, en nous aidant d'un programme informatique de notre propre confection (cf. Vermote, 2011, ms.), et en établissant (de manière consensuelle) les distinctions suivantes :

- **syntagmes MASSIFS** : introduits par le déterminant partitif, la locution *un peu de*, un déterminant du type *beaucoup de* suivi par une forme au singulier ou, pour les noms de matériaux, une forme au singulier dans la construction *en + N* ;
- **syntagmes COMPTABLES** : introduits par un déterminant pluriel, un déterminant du type *beaucoup de* suivi par une forme au pluriel, un déterminant singulier comptable tel que *un, aucun* ou *chaque* ou, pour les noms de matériaux, une forme au pluriel dans la construction *en + N* ;

- **syntagmes NEUTRES** : introduits essentiellement par un déterminant singulier défini (article, démonstratif, possessif), ou non introduit au niveau du SN (syntagmes nu).

Notons que les syntagmes neutres, qui ne discriminent pas entre lecture massive et lecture comptable, ont été laissés de côté dans la suite de la présente recherche (sans présumer de leur intérêt dans d'autres domaines).

Dans un second temps, nous avons analysé de manière qualitative les emplois supposément dérivés. Pour les noms de fruits et de légumes, cette seconde analyse a abouti principalement à une caractérisation sémantique des emplois massifs en termes de « forme et quantité indéterminées » (cf. Vermote, 2013, pour une analyse détaillée à partir d'un échantillon plus limité). Pour les deux autres catégories de noms, où ce sont supposément les usages comptables qui sont dérivés, l'examen plus approfondi a permis de distinguer parmi ceux-ci deux types de lectures, celle correspondant aux effets du Trier (12-13) et celle en termes du Conditionneur (14-15) – annotés comme tels dans notre collection d'occurrences, et traduits quantitativement.

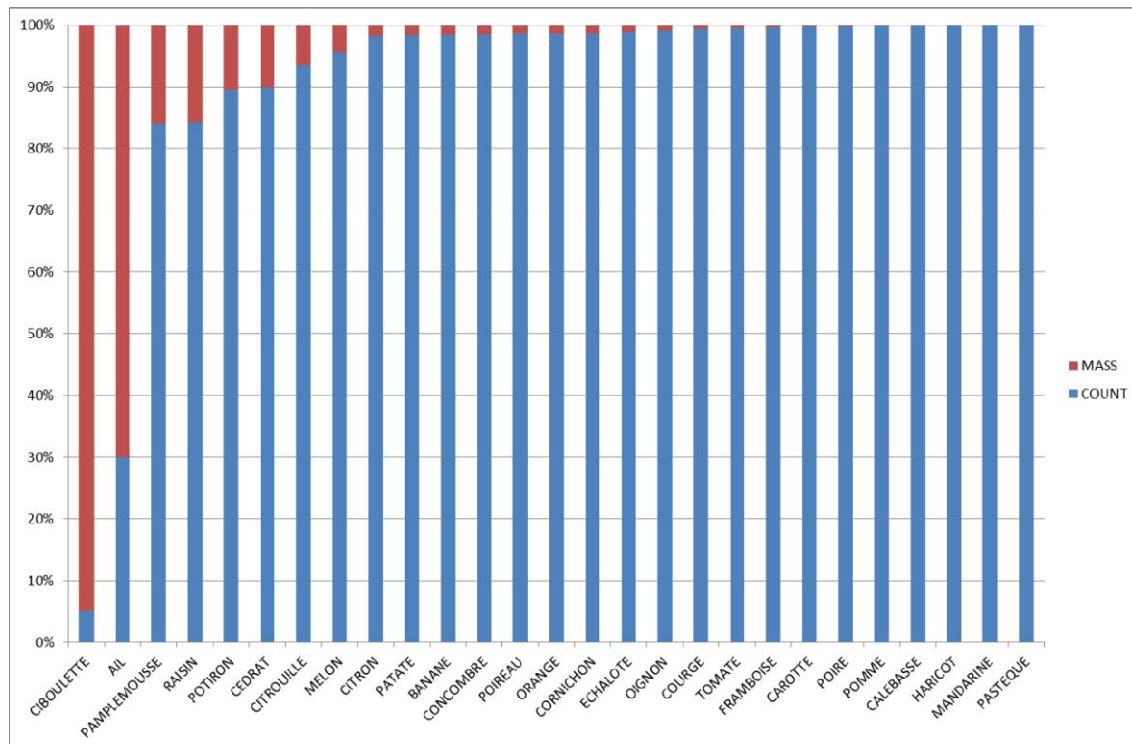
(12)Le principal défaut d'un papier mat est la grande absorption de l'encre (...).

(13)(Le) vigneron, qui n'est pas un grand amateur des vins jeunes, dits de fruit.

(14)Que valent vraiment les nouveaux fers de la marque phare du golf (...)?

(15)Il partait boire une bière.

Le résultat de ce triple examen de corpus est exprimé au travers de trois graphiques (un par classe de noms), montrant la répartition pour chaque item de ses occurrences en contexte massif et comptable ainsi que, le cas échéant, la proportion de ces dernières en termes du Trier. Dans ce qui suit, nous présentons brièvement chacun de ces graphiques, puis établissons deux constats généraux qui nous permettront de poursuivre notre réflexion empirique et théorique.



Graphique 1: noms de fruits et de légumes (27 items), profils distributionnels

Pour les fruits et les légumes (graphique 1), la principale observation est que les emplois comptables dominant fortement : cela confirme l'intuition que les noms de fruits et de légumes sont principalement attirés par la syntaxe comptable. Seuls deux items se démarquent résolument, et de manière prévisible : *ail* et *ciboulette*, qui ne sont membres de la catégorie que d'un point de vue botanique, et appartiennent autrement à la classe des condiments (rigoureusement massive). *Raisin*, que l'intuition placerait plutôt du côté des massifs, connaît bien une portion d'emplois massifs plus grande que la majorité des autres noms examinés mais ne se distingue pas en cela de certains items tels que *pamplemousse* ou (dans une moindre mesure) *potiron*. D'autre part, l'on trouve à l'extrémité droite du graphique des items dont la proportion d'emplois massifs est nulle dans notre corpus (*calebasse*, *haricot*, *mandarine*, *pastèque*). Pour ceux-ci, une recherche WebCorp additionnelle du nom dans une syntaxe explicitement massive (*un peu de + N* au singulier) permet d'établir qu'en effet, *calebasse* ne semble pas enclin à apparaître dans un syntagme massif (0 occurrences), tandis que *mandarine* (20 occurrences), *pastèque* (26 occurrences) et même *haricot* (17 occurrences)⁷ livrent bien un certain nombre d'exemples massifs apparemment naturels (16-18).

(16) Dans la coque creusée, posez un peu de mandarine gélifiée.

(17) Dans ce cas vous pouvez ajouter un peu de pastèque.

(18) J'ai donc fait revenir une saucisse, un peu de haricot vert et les raisins dans une poêle.

Pour finir, malgré les divergences (limitées) au niveau de la distribution, le sens associé à l'emploi massif d'un nom de fruit ou de légume semble cohérent d'un item à l'autre et couvrir une même palette de possibilités : goût et odeur (18), substance comestible (19), forme préparée explicite (20), homogénéisation généralisante (21) – avec pour dénominateur commun une présentation « en forme et quantité indéterminée » (cf. examen détaillé et argumentation dans Vermote, 2013). Seul *raisin* se démarquerait éventuellement des autres, avec une lecture massive plus naturelle que les autres items dans une situation où, contextuellement, des 'individus' sont bel et bien identifiés (22).

(19) Nez très ananas, un peu de melon et de pêche, de grillé (...).

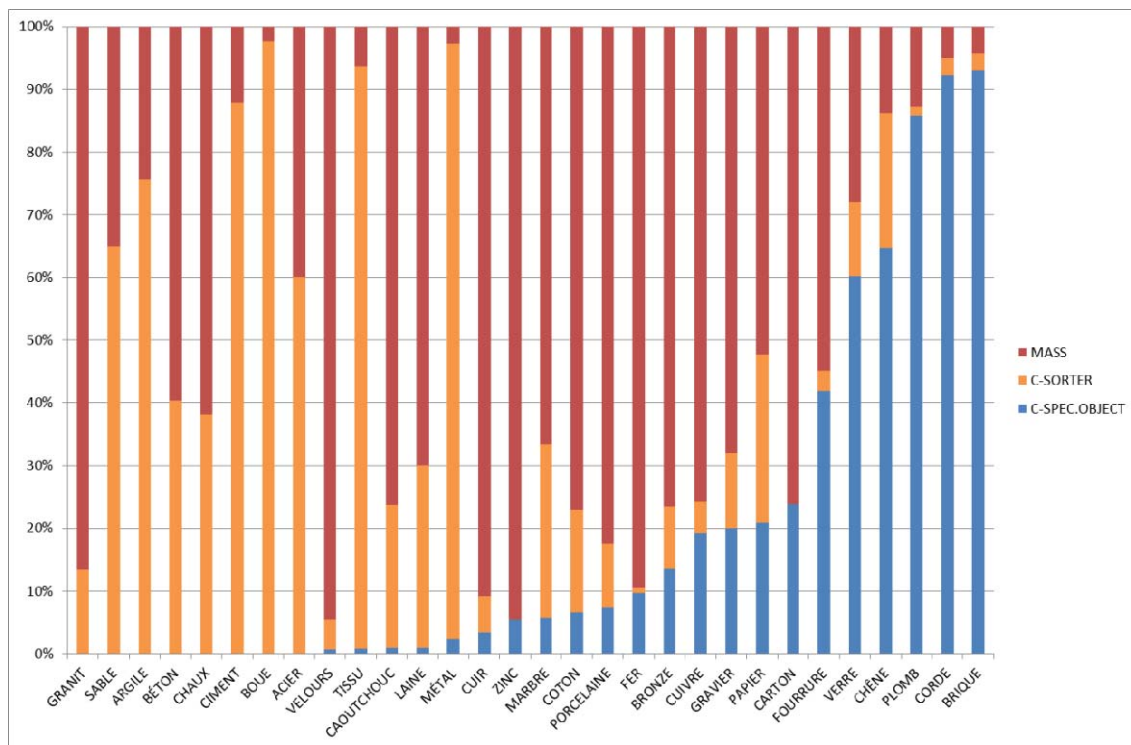
(20) Un soir, au bout de cinq jours, il a mangé un peu de potiron.

(21) Déposez au milieu de chaque assiette un peu de poireau finement émincé.

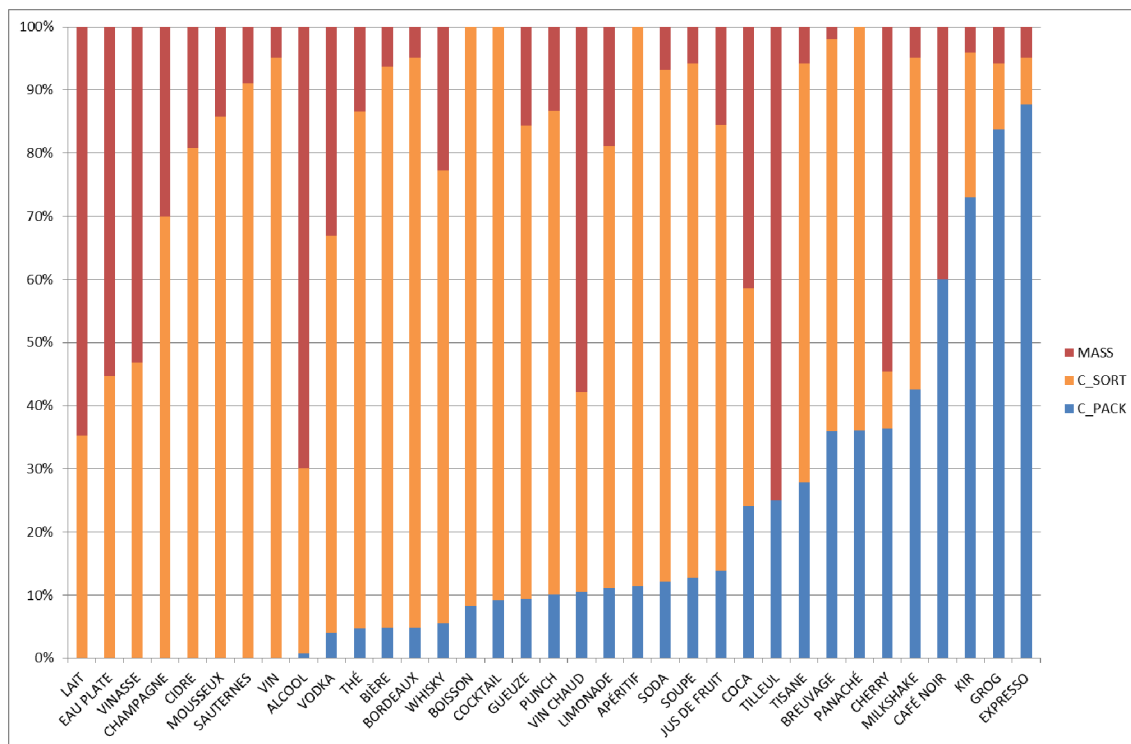
(22) Avec du citron confit, cette recette est bien tentante.

(23) Pour glacer du raisin, plongez **les grains** dans un mélange de sucre glace et de citron.

Pour les noms de matériaux (graphique 2), la situation est toute différente : ce sont bel et bien les contextes massifs (et assimilés⁸) qui dominent, confirmant – si cela était nécessaire – l'intuition que les noms de matériaux sont en premier lieu des noms massifs. Certains items connaissent néanmoins aussi un emploi proprement comptable en termes d'un objet ou d'une collection d'objets spécifiques, dont la grande variété et spécificité au niveau du sens (*un carton* 'une caisse en carton', *un fer* '[de lance, à cheval, à repasser]', *un verre* 'récipient pour boire, réalisé traditionnellement en verre', *des cuivres* 'instruments de musique en cuivre', *des papiers* 'des documents', etc.), ainsi que la nécessité de l'existence d'un référent spécifique (**des sables* ou **un granit*, en lecture objet), indiquent clairement un statut lexicalisé. Pour certains items, les données distributionnelles semblent même indiquer que cette lecture comptable serait la lecture de base, puisqu'elle domine la lecture massive par le nombre. Cela semble tout à fait en adéquation avec l'intuition, au moins pour ce qui concerne les items *brique*, *corde* et *chêne*. Néanmoins, pour tous les items (y compris ceux-là), la lecture massive est attestée, avec rigoureusement le même sens pour chaque item : celui de désigner le matériau dans sa substance (*du sable*, *du métal*, *de la corde* ; [*une construction*] *en acier*, *en cuivre*, *en brique*, etc.).



Graphique 2: noms de matériaux (30 items), profils distributionnels



Graphique 3: noms de boissons (35 items), profils distributionnels

En ce qui concerne les noms de boissons (graphique 3), la distribution est, dans les grandes lignes, comparable à celle des noms de matériaux : majorité de cas massifs ou assimilés (*du lait – des laits [allégés], de la vodka – une vodka [de blé], etc.*), avec bon nombre d'items connaissant aussi une lecture proprement comptable spécifique, en l'occurrence en termes de 'portion standard' (*un cocktail, une gueuze, un coca, une tisane*), qui est même la lecture majoritaire dans certains cas (*un espresso, un grog, un kir*). Pour ces derniers, l'explication semble être fonction de la nature des référents : *espresso, grog, kir, milkshake*, sont autant de noms qui désignent déjà une portion – c'est-à-dire que les noms d'*espresso* ou de *kir*, par exemple, ne sont pas liés seulement à un type de liquide (*du café, du vin*), mais aussi à un mode de préparation et, crucialement, de présentation ('une portion réduite de café très fort', 'un verre de vin avec de la crème de cassis'). Cela n'empêche que ces items puissent aussi s'employer comme massifs (24-25), dans la lecture typique en termes de 'liquide' accessible, *a priori*, à tous les noms de boissons (et qui est la lecture de base pour la majorité d'entre eux). (Notons néanmoins que pour trois items – *boisson, cocktail* et *apéritif* – cette lecture massive n'est pas attestée dans notre corpus. Cette particularité distributionnelle s'explique selon nous par le statut hypéronymique de ces trois noms.)

(24) De l'*espresso* est par la suite versé doucement afin qu'il se place entre le lait et la mousse.

(25) Quand les bistrots se décideront-ils à faire du kir de bonne qualité ?

Finalement, concernant les limitations de la lecture comptable en portion (qui est, de manière célèbre, inaccessible à *vin*, par exemple, malgré l'existence de portions standard : **deux vins blancs*, dans le sens de 'deux verres de vin'), qui ne sont pas contredites par nos données distributionnelles (puisque il y a au moins 8 items pour lesquels nous ne rencontrons pas d'occurrences dans une telle lecture, cf. extrême-gauche du graphique), nous avons proposé une hypothèse explicative dans Lauwers & Vermote (2014), en termes de préemption : comme le vin (puisque c'est l'exemple le plus célèbre) est toujours consommé dans un verre, l'hypothétique syntagme comptable **un vin* pour désigner une portion standard de vin est bloqué par l'expression *un verre de vin*, qui lui est sémantiquement équivalente – ce qui n'est pas le cas pour *bière*, par exemple, *une bière* 'portion standard' se servant tantôt dans un verre, tantôt dans une bouteille, tantôt même dans une canette.

Cela clôture l'exploration de nos données de corpus. Cette investigation empirique nous permet de dresser deux constats généraux dont il nous semble essentiel de tenir compte pour une théorie satisfaisante de la localisation de l'opposition massif/comptable et qui, s'ils pourront encore être examinés, détaillés et affinés par d'autres études similaires, nous paraissent néanmoins robustement soutenus par nos données.

[1.] L'on peut dresser un parallèle entre le comportement observable d'un item lexical (profil distributionnel) et nos intuitions de locuteurs quant à ses préférences morphosyntaxiques. Ainsi, notre supposition que les noms de fruits et de légumes sont principalement comptables, et que les noms de matériaux et de boissons sont essentiellement massifs, est confirmée par les proportions d'occurrences massives et comptables observées dans notre corpus. En outre, les items qui ont un comportement observable démarqué sont aussi ceux à qui l'on attribuerait, intuitivement, un statut particulier au sein de leur catégorie : *ail, ciboulette, brique, chêne*, etc.

[2.] La flexibilité de l'opposition massif/comptable, qui est réelle et observable, est généralement régulière au sein d'une classe de noms donnée, c'est-à-dire qu'un effet de sens observé pour un item dans une certaine syntaxe (p.ex. massive) pourra être observé pour un autre item de la même classe sémantique, en utilisant la même syntaxe – indépendamment, semble-t-il, du poids distributionnel de cet usage dans le profil de l'item en question. Ainsi, les noms comptables de fruits et de légumes ne sont-ils pas tous observés aussi fréquemment les uns que les autres en syntaxe massive, mais lorsqu'ils le sont, leur sens est toujours celui d'une « indétermination en

forme et quantité » (*du concombre // de la pomme*, malgré la différence de fréquence). De même, un nom de boisson, pour désigner cette boisson en termes de substance liquide, pourra toujours être utilisé comme massif – quel que soit le statut de base, massif ou comptable, de l’item (*du vin // du kir*). Idem pour les noms de matériaux, qui pourront toujours apparaître comme massifs lorsqu’il s’agira de désigner un matériau dans sa substance (*du sable // de la brique*).

D’un point de vue théorique, les manquements des théories lexicalistes et grammaticales réapparaissent dès lors dans toute leur splendeur. D’une part, la vision grammaticale ne permet pas de dire quoi que ce soit au sujet des différences distributionnelles inter-lexicales que nous avons observées, et – surtout – du fait que celles-ci semblent refléter nos propres intuitions. D’autre part, si la vision lexicaliste prédit bel et bien des différences fondamentales entre items (à mettre en parallèle avec les jugements spontanés de locuteurs natifs), elle ne permet pas, dans sa forme de base, d’expliquer de manière satisfaisante la flexibilité morphosyntaxique de la majorité des items, et en particulier la cohérence de celle-ci à l’intérieur de certaines classes sémantiques – les mécanismes de dérivation fautant par leur aura d’universalité et leur caractère directionnel, qui sont mis à mal par nos données.

Selon nous, un modèle plus satisfaisant, et plus à même d’expliquer ces dernières, nécessite l’introduction d’un troisième niveau de représentation linguistique, à mi-chemin entre lexique et syntaxe, que nous appellerons le niveau supra-lexical – qui est celui des classes sémantiques. Notre vision se conçoit dès lors comme une approche lexicaliste améliorée, ne supposant non pas deux niveaux (lexical et grammatical), mais bien trois. D’une part, nous reconnaissons que les items lexicaux ont, généralement, des préférences très claires (et identifiables intuitivement) pour l’un ou l’autre type de syntaxe. A notre sens, ce genre d’information est bel et bien stockée dans le lexique, à partir de notre expérience en discours avec les items que nous rencontrons (perspective *usage-based* sur l’organisation lexicale, cf. p.ex. Bybee, 1988) : d’une manière ou d’une autre, nous savons qu’un certain nom est plus enclin à s’employer, par exemple, comme comptable, car c’est précisément dans ce genre de syntaxe que nous le rencontrons habituellement, en contexte non marqué. Cependant, si la frontière entre massif et comptable n’apparaît pas comme hermétique, et autorise (avec des effets de sens systématiques et relativement prévisibles) une certaine flexibilité, c’est parce qu’elle n’est pas seulement représentée à ce niveau lexical, mais que – par analogie entre items similaires, appartenant à la même classe sémantique – elle émerge aussi à un niveau supra-lexical. Ainsi, par exemple, il découle du fait qu’un grand nombre de noms de matériaux soient des massifs l’émergence d’une association supra-lexicale entre la signification ‘nom désignant un matériau’ et la caractéristique morphosyntaxique ‘massif’ : c’est ainsi que, indépendamment de leur préférence intrinsèque lexicalisée, tant *sable* (massif) que *corde* ou *brique* (comptables) peuvent s’utiliser comme massifs, lorsqu’ils désignent un référent conçu comme ‘matériau’. Un autre exemple est celui de ce qui s’est appelé le *Trieur Universel*. Dans notre modèle à trois niveaux, il s’explique aisément : une « présentation en termes de sortes » serait simplement l’une de ces catégories sémantiques supra-lexicales, à laquelle serait associée en l’occurrence la caractéristique « comptable ».

Cette façon de représenter l’opposition massif/comptable est bien différente de la vision lexicaliste traditionnelle, en ce qu’elle n’implique pas des mécanismes de transfert directionnels d’un emploi vers un autre, mais bien une relation d’héritage entre une certaine signification (supra-lexicale) et un certain emploi. Cette nuance est importante, car elle permet d’expliquer la cohérence des effets de sens associés respectivement au massif et au comptable au sein d’une classe de nom, sans que de ces deux emplois, l’un doive systématiquement se concevoir comme dérivé de l’autre : ce qui compte, pour justifier le type de syntaxe utilisé, c’est que dans le contexte donné le nom soit profilé comme appartenant à une certaine classe sémantique supra-lexicale compatible avec la syntaxe en question. Dans certains cas, le résultat correspondra à un emploi auquel nous sommes déjà habitués (ex. *du sable* ; *de l’agneau* ; *du raisin* ; *du*

champagne), dans d'autres, l'emploi paraîtra possiblement moins habituel (ex. *de la brique* ; *du kangourou* ; *de la pomme* ; *de l'expresso*) – cette différence étant simplement due aux préférences distributionnelles (enregistrées dans notre lexique par le biais de la fréquence) respectives des items en question.

3 Enquêtes d'acceptabilité

A notre connaissance, le modèle proposé ci-dessus est tout à fait novateur quant à la localisation de l'opposition massif/comptable. Dans un souci d'élaborer certaines de ses suppositions et prédictions, et pour appuyer notre investigation de corpus (comme il sied au chercheur moderne, cf. Gilquin & Gries, 2009), nous avons mené en complément une double enquête d'acceptabilité, concernant l'emploi massif et comptable des noms de fruits et de légumes, auprès de deux panels de locuteurs natifs (tous étudiants en Lettres à l'Université de Caen Basse Normandie)⁹, comptant respectivement 44 et 43 participants. Plus précisément, il s'agissait :

- d'une part, de vérifier le lien entre les différences distributionnelles constatées dans notre corpus, et les intuitions des locuteurs concernant divers emplois (l'hypothèse étant, dans la lignée de notre vision *usage-based*, qu'acceptabilité varierait avec fréquence) ;
- d'autre part, d'explorer notre hypothèse d'héritage supra-lexical, en examinant la pertinence du contexte dans l'acceptabilité de l'un ou l'autre emploi (le trait massif ou comptable associé à une certaine classe sémantique deviendrait plus acceptable à mesure que l'appartenance d'un item à cette classe serait profilée en contexte).

La première enquête concernait la flexibilité des noms de fruits et de légumes et visait à comparer l'acceptabilité d'un échantillon de 12 noms de cette catégorie (sélectionnés parmi les plus fréquents, d'après la base de données LEXIQUE3)¹⁰ dans une syntaxe massive à :

- l'acceptabilité de ces mêmes items dans une syntaxe comptable ;
- l'acceptabilité en syntaxe massive d'une sélection de 12 items de contrôle (sélectionnés pour correspondre aux premiers du point de vue de la fréquence, du nombre de syllabes, et de la taille des référents).¹¹

Pour cette première enquête, il a été demandé aux participants de juger de l'acceptabilité de 72 phrases (dont 24 contenant l'un de nos items critiques, dans une syntaxe soit massive (partitif), soit comptable (*un/une*), et 48 fillers), sur une échelle de 1 à 7 - 7 correspondant à une acceptabilité absolue. Pour tous les items critiques et les contrôles, des contextes phrastiques ont été imaginés, qui soient les plus neutres possibles, c'est-à-dire qui n'orientent pas spécifiquement l'interprétation vers un contexte de consommation culinaire (supposément plus propice à une présentation massive, cf. *infra*). Une fois les données récoltées, nous avons calculé, pour chacune de nos conditions, le score d'acceptabilité moyen. Les résultats sont les suivants :

Nom de fruit/légume		Syntaxe massive	
<i>Massif</i>	5, 29	<i>Nom de fruit/légume</i>	5, 29
<i>Comptable</i>	6, 19	<i>Nom de contrôle</i>	3, 30

Comme on peut le constater, ces résultats sont dans la droite ligne de nos prédictions. Premièrement, les emplois massifs des noms de fruits et de légumes, moins fréquents, sont jugés comme un peu moins acceptables (5,29) que leurs équivalents comptables dans le même contexte (6,19). L'application du test de Wilcoxon¹² (*Wilcoxon signed-rank test*) nous montre que la différence entre ces deux moyennes est significative d'un point de vue statistique ($p < 0,001$). Deuxièmement, ces emplois massifs de noms de

fruits et de légumes – moins fréquents, mais néanmoins attestés, cf. étude de corpus – n’en sont pas moins beaucoup plus acceptables (5,29) que les emplois massifs, dans le même contexte, de noms résolument comptables et inflexibles (3,30) – dont on peut supposer nulle la proportion d’occurrences massives en corpus. A nouveau, cette différence apparaît comme significative du point de vue statistique (test de Wilcoxon, $p < 0,001$). Ces résultats tendent à confirmer nos prédictions : 1) que les noms de fruits et de légumes sont, en effet, flexibles de manière régulière quant à l’opposition massif/comptable, et 2) qu’il existe une certaine corrélation entre la fréquence de l’un ou l’autre emploi, et son acceptabilité.

Notre seconde enquête, quant à elle, concernait l’influence du contexte sur l’acceptabilité d’un nom de fruit ou de légume en syntaxe massive. Le raisonnement était le suivant : si un nom de fruit et de légume (qui est *a priori*, ou du moins préférentiellement, comptable) est susceptible d’hériter du trait massif de la part d’une certaine classe sémantique supra-lexicale dans laquelle il peut s’inscrire, alors un contexte phrastique profilant d’autant plus le nom de fruit ou de légume comme membre de cette classe devrait être synonyme d’une acceptabilité accrue. Pour les fruits et légumes, nous avons postulé que cette classe supra-lexicale était celle des « produits de consommation culinaire, présentés en forme et quantité indéterminées » (par analogie avec *de l’œuf, de la pizza, du fromage, du chocolat, de la choucroute*, etc.). Nous avons dès lors conçus parallèlement 10 contextes phrastiques neutres (*En ville, j’ai acheté X*), et 10 contextes phrastiques biaisés vers une lecture en termes de « nourriture » (ex. *Au supermarché, j’ai pu goûter X*).¹³

Ainsi, pour cette seconde enquête, les participants devaient juger de l’acceptabilité de 48 phrases (16 contenant un nom de fruit ou de légume en syntaxe massive, soit dans un contexte neutre, soit dans un contexte biaisé, et 32 fillers) sur la même échelle de 1 à 7 que précédemment. Aucun participant à la seconde enquête n’avait participé à la première. Une fois les données récoltées, nous avons calculé la moyenne d’acceptabilité pour les deux valeurs de notre condition, avec les résultats suivants :

Nom de fruit ou de légume en syntaxe massive	
Contexte phrastique neutre	5,64
Contexte phrastique biaisé	6,16

L’application du test de Wilcoxon nous indique que la différence entre ces deux moyennes est statistiquement significative ($p < 0,001$). Cela signifie que notre prédiction est vérifiée : un contexte plus explicite, profilant le référent (fruit/légume) comme membre d’une classe de produits comestibles, augmente l’acceptabilité de l’apparition du nom dans une syntaxe massive.

Les résultats de cette double enquête constituent une première confirmation, timide mais robuste, de notre hypothèse de double représentation – lexicale et supra-lexicale – du trait massif ou comptable. D’un côté, les noms ont des préférences réelles, enregistrées dans le lexique comme fonction de leur distribution en discours, et reflétées au travers des jugements d’acceptabilité. De l’autre, le trait massif ou comptable pouvant être associé à certaines significations de classe (à un niveau supra-lexical), l’acceptabilité d’un syntagme pourra varier en fonction du profilage du référent dans son contexte sémantique.

4 Conclusions : un niveau de représentation supra-lexical

Dans la présente contribution, nous avons jeté un regard neuf sur la localisation de l’opposition massif/comptable dans le système linguistique. Nous avons présenté et évalué les deux approches majoritaires en la matière (lexicaliste et grammaticale), et y avons confronté de nouvelles données empiriques, issues de trois études de corpus et d’une double enquête d’acceptabilité. A partir de celles-ci,

nous avons proposé et défendu un modèle original de l'opposition massif/comptable, reposant sur trois niveaux de représentation :

- **le lexique** : la majorité des noms possèdent des préférences (identifiables intuitivement) pour l'un ou l'autre type de syntaxe, et celles-ci découlent de leur profil distributionnel, c'est-à-dire de notre expérience avec eux en discours ;
- **le niveau sémantique supra-lexical** : à certains types de significations générales, applicables à un ensemble de noms (dans une classe sémantique déterminée), est également associée une caractérisation morphosyntaxique en termes de massif/comptable, dont les noms peuvent hériter en contexte, indépendamment de leurs préférences lexicales ;
- **la grammaire / le discours** : au niveau du discours, lorsque noms et syntaxe sont combinés, l'opposition massif/comptable apparaît comme flexible et instable, précisément parce que les syntagmes observés sont le résultat de la combinaison (et donc parfois de la compétition) entre représentation lexicale et supra-lexicale.

La présente contribution suivait une méthodologie résolument empirique, et ses conclusions sont fortement teintées des accents de la linguistique cognitive et *usage-based* (voir Taylor, 2002, pour une autre évocation de l'opposition massif/comptable comme résultat de l'interaction entre facteurs formels et conceptuels). Notre modèle de la localisation de l'opposition massif/comptable demande encore à être testé plus avant, et peaufiné chemin faisant, mais il a le mérite de souffler un vent frais sur un problème ancien, en proposant une alternative originale au débat entre approches purement lexicalistes et purement grammaticales. Néanmoins, une définition précise et formalisée de ce que nous avons appelé 'classe supra-lexicale' reste à donner, de même qu'il reste à décrire par quels mécanismes cognitifs celles-ci émergent, quel est leur degré de généralité et comment est leur organisation en réseau. Nous sommes néanmoins convaincus d'avoir ouvert une piste d'investigation intéressante, et sommes bien résolu à en poursuivre l'exploration.

Bibliographie

- Allan, K. (1980). "Nouns and Countability." *Language: Journal of the Linguistic Society of America (Los Angeles, CA)* 56, 541-567.
- Bale, A.C. & Barner, D. (2009). "The Interpretation of Functional Heads: Using Comparatives to Explore the Mass/Count Distinction." *Journal of Semantics*, 26, 217-252.
- Borer, H. (2005). "Some stuff: On the Count-Mass Distinction." In: *In name only (Structuring Sense I)*. Oxford: Oxford University Press, 86-135.
- Bunt, H.C. (1985). *Mass Terms and Model Theoretic Semantics*. Cambridge: Cambridge University Press.
- Bybee, J. (1988). "Morphology as lexical organization." In: Hammond, M. and Noonan, M. (eds), *Theoretical approaches to morphology*. San Diego: Academic Press, 119-141.
- Cheng, L., Doetjes, J., & Sybesma, R. (2008). "How universal is the Universal Grinder?" *Linguistics in the Netherlands* 2008, 50-62.
- Damourette, J. et Pichon, E. (1911-1940). *Des mots à la pensée. Essai de grammaire de la langue française (7 volumes)*. Paris: d'Artray.
- De Belder, M. (2011). "A morphosyntactic decomposition of countability in Germanic." *The Journal of comparative Germanic Linguistics* 14.3, 173-202.

- Desagulier, G. (2012). "'C'est de la bombe!': Qualitative count-to-mass conversion in French copular subject-predicate constructions." In: Bouveret, M. and Legallois, D. (eds), *Constructions in French*. Amsterdam/Philadelphia: John Benjamins, 201-231.
- Galmiche, M. (1989). "Massif / comptable : de l'un à l'autre et inversement." In: David, J. and Kleiber, G. (éds), *Termes massifs et termes comptables*. Paris: Klincksieck, 63-77.
- Gillon, B. (1999). "The lexical semantics of English count and mass nouns." In: Viegas, E. (eds), *The Breadth and Depth of Semantic Lexicons*. Dordrecht: Kluwer, 19-37.
- Gilquin, G. & Gries, S.T. (2009). "Corpora and experimental methods: a state-of-the-art review." *Corpus Linguistics and Linguistic Theory* 5.1, 1-26.
- Jackendoff, R. (1991). "Parts and boundaries." *Cognition* 41, 9-45.
- 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.
- Landman, F. (2011). "Count Nouns - Mass Nouns - Neat Nouns - Mess Nouns." In: Partee, B.H., Glanzberg, M., and Skilters, J. (eds), *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.
- Lauwers, P. & Vermote, T. (2014). "La flexibilité de l'opposition massif/comptable en français et en néerlandais: une étude contrastive." *Syntaxe & Sémantique* 15. A paraître.
- Le Bruyn, B., de Swart, H. et Zwaarts, J. (2011). "Mass-Count distinctions in bare PPs." Manuscrit. Disponible en ligne sur l'adresse http://www.llf.cnrs.fr/Gens/Mardale/le.bruyn_de.swart_zwarts.pdf (dernier accès le 20 novembre 2013).
- Mondini, S., Angrilli, A., Bisiacchi, P., Spironelli, C., Marinelli, K., & Semenza, C. (2013). "Mass and Count nouns activate different brain regions: An ERP study on early components." *Neuroscience Letters* 430.1, 48-53.
- Nicolas, D. (2002). *La distinction entre noms massifs et noms comptables : aspects linguistiques et conceptuels*. Leuven: Peeters.
- 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. (eds), *Count and Mass Across Languages*. Oxford: University Press, 9-26.
- Taylor, J.R. (2002). "Count nouns and mass nouns." In: Taylor, J.R. (eds), *Cognitive Grammar*. Oxford: Oxford University Press, 366-388.
- Vermote, T. (2011, ms). "La distinction massif vs. comptable. L'informatique au service de la recherche fondamentale." Non publié. Disponible en ligne sur *Academia* (<http://ugent.academia.edu/TimotheusVermote>).
- 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.
- Weinreich, U. (1966). "Explorations in semantic theory." In: Sebeok, T.A. (eds), *Current Trends in Linguistic Theory*. The Hague: Mouton, 395-477.
- Zhang, N.N. (2013). *Classifier Structures in Mandarin Chinese [TiLSM 263]*. Berlin: Mouton de Gruyter.

¹ A cela, l'on pourrait encore ajouter une étude neurolinguistique relativement récente de Mondini *et al.* (2008), qui révèle que les noms massifs et comptables activent des réseaux corticaux différents dans le cerveau, et ce dès le stage le plus précoce de reconnaissance langagière automatique.

² Pour Pelletier (1975), qui lance le terme de *Universal Grinder*, le Broyeur est bel et bien une machine (fût-elle imaginaire), capable de broyer n'importe quel objet jusqu'à une masse homogène. Il propose de l'inventer pour justifier que n'importe quel référent, désigné habituellement par un nom en syntagme comptable, une fois broyé, pourra être désigné par ce même nom en syntaxe massive – justifiant ainsi précisément que l'opposition massif/comptable n'a pas lieu d'être située au niveau du lexique (*contra* la vision lexicaliste). Le nom de son engin a néanmoins été récupéré (dans un contexte lexicaliste) pour nommer le dispositif sémantique de transfert du comptable vers le massif, lorsqu'une présentation homogène d'un référent autrement comptable est supposée, comme dans (8), mais aussi, chez certains auteurs, *Il faudrait du concombre dans cette salade* ou encore *Je porte du lapin*, en parlant d'une fourrure (ex. Nunberg et Zaenen, 1992).

³ Desagulier (2012), quant à lui, qui travaille dans le domaine de la Grammaire des Constructions, attribue aux « constructions » (et non plus au SN) dans lesquelles les noms s'inscrivent de prédire le trait massif ou comptable – dans une approche comparable aux précédentes en ce qu'elle ne suppose pas non plus nécessairement de marquage lexical de l'opposition.

⁴ Sur ce point, Pelletier (2012) va, pour sa part, encore plus loin : s'il retire bien le trait syntaxique massif ou comptable du lexique, il constate néanmoins que « presque tous les [noms] peuvent être utilisés des deux manières » et que « dans un très, très grand nombre de cas les deux sens sont tout aussi 'naturels' (p. 24, nous traduisons), et il le remplace dès lors par un trait sémantique, bel et bien lexical, mais dont pour ainsi dire tous les noms possèdent les deux valeurs. Sa vision peut donc paraître dans un sens « hybride », mais s'apparente finalement essentiellement à la vision grammaticale – qui met en avant la flexibilité avant les préférences lexicales.

⁵ Notons encore que, si le français et l'anglais (qui servent souvent d'exemples) sont des langues relativement libérales quant à la séparation entre le massif et le comptable, il existe des langues, comme le néerlandais, où la flexibilité est strictement limitée – bon nombre de noms réputés massifs n'ayant par exemple pas même de forme plurielle attestée (voir Lauwers & Vermote 2014).

⁶ *WebCorp*, “Concordance the web in real-time”, <http://www.webcorp.org.uk/live/>.

⁷ Comme *WebCorp* n'est pas un corpus avec un nombre total de mots arrêté, les nombres absolus d'occurrences ne sont pas facilement interprétables. Nous les donnons seulement à titre indicatif.

⁸ Dans la mesure où le Trieur est un mécanisme sémantique s'appliquant *a priori* aux noms massifs (cf. le caractère inacceptable de **Aujourd'hui, j'ai mangé deux fruits : 4 pommes et 6 poires*, exemple inspiré et adapté très librement de Kleiber, 2011), nous nous permettons de compter les occurrences comptables en termes de Trieur comme appartenant indirectement au total des occurrences massives.

⁹ Nous en profitons pour réitérer l'expression de notre meilleure gratitude à nos collègues normands du laboratoire CRISCO, pour leur accueil et leur soutien, et en particulier à Eliane Delente, Véronique Lenepveu, Mathilde Salles et Laurent Senoze.

¹⁰ Noms : *banane, carotte, concombre, framboise, haricot, melon, oignon, patate, poire, pomme, raisin, tomate*.

¹¹ Contrôles : *ampoule, ballon, bijou, bocal, casserole, couteau, cuillère, gobelet, livret, louche, sachet, tasse*.

¹² Nous avons opté pour le *Wilcoxon signed rank test* (pour échantillons appariés) plutôt que le *t-test* de Student, car nos essais de résultats n'apparaissent pas être distribués normalement (cf. test de Shapiro-Wilk). Dans ce qui suit, nous avons à chaque fois opté pour le test de Wilcoxon, pour cette même raison.

¹³ Pour confirmer le biais de ce second groupe de phrases par rapport au premier, nous avons soumis les 20 contextes phrastiques (ainsi que 10 fillers) à un Cloze test (cf. Taylor, 1953), auprès d'un panel de 58 locuteurs natifs du français. De ce test s'est avéré que chaque contexte biaisé provoquait bel et bien une majorité et chaque contexte neutre une minorité de réponses en termes de « nourriture ». En outre, pour chaque paire de phrase, le contexte biaisé provoquait systématiquement une proportion de telles réponses significativement supérieure au contexte neutre (test du χ^2 , $p < 0,001$).