**Meta**
Journal des traducteurs
Translators' Journal

# La lexicographie et l'analyse de corpus : nouvelles perspectives

Ann Bertels et Serge Verlinde

Volume 56, numéro 2, juin 2011

Les corpus et la recherche en terminologie et en traductologie
Corpora and Research in Terminology and Translation Studies

URI : https://id.erudit.org/iderudit/1006175ar
DOI : https://doi.org/10.7202/1006175ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Bertels, A. & Verlinde, S. (2011). La lexicographie et l'analyse de corpus : nouvelles perspectives. *Meta*, *56*(2), 247–265. https://doi.org/10.7202/1006175ar

Résumé de l'article

L'objectif du présent article est de montrer comment les descriptions lexicographiques traditionnelles peuvent être enrichies à partir des nouvelles techniques d'analyse et d'exploitation de corpus. Nous étudions des verbes dénotant la notion de hausse, en anglais, en français et en néerlandais, et à cet effet, nous procédons à des analyses de corpus parallèles et de corpus monolingues ciblés. Les corpus parallèles fournissent des indications sur la fréquence d'emploi et sur l'équivalence des traductions. Ces données quantitatives sont soumises à des analyses MDS (*MultiDimensional Scaling* ou positionnement multidimensionnel) afin d'établir les profils de traduction des verbes. Les corpus monolingues ciblés permettent d'affiner ces informations et de relever les collocatifs pertinents, afin de montrer les propriétés combinatoires des verbes. Les résultats des différentes analyses de corpus, en termes de profils de traduction et de profils combinatoires, contiennent des indications précieuses pour enrichir les descriptions lexicographiques traditionnelles des dictionnaires de traduction. La méthodologie et les résultats des analyses de corpus, ainsi que les défis pour la lexicographie, seront exposés.

# La lexicographie et l'analyse de corpus : nouvelles perspectives

**ANN BERTELS**
*Katholieke Universiteit Leuven, Leuven, Belgique*
ann.bertels@ilt.kuleuven.be

**SERGE VERLINDE**
*Katholieke Universiteit Leuven, Leuven, Belgique*
serge.verlinde@ilt.kuleuven.be

**RÉSUMÉ**

L'objectif du présent article est de montrer comment les descriptions lexicographiques traditionnelles peuvent être enrichies à partir des nouvelles techniques d'analyse et d'exploitation de corpus. Nous étudions des verbes dénotant la notion de hausse, en anglais, en français et en néerlandais, et à cet effet, nous procédons à des analyses de corpus parallèles et de corpus monolingues ciblés. Les corpus parallèles fournissent des indications sur la fréquence d'emploi et sur l'équivalence des traductions. Ces données quantitatives sont soumises à des analyses MDS (*MultiDimensional Scaling* ou positionnement multidimensionnel) afin d'établir les profils de traduction des verbes. Les corpus monolingues ciblés permettent d'affiner ces informations et de relever les collocatifs pertinents, afin de montrer les propriétés combinatoires des verbes. Les résultats des différentes analyses de corpus, en termes de profils de traduction et de profils combinatoires, contiennent des indications précieuses pour enrichir les descriptions lexicographiques traditionnelles des dictionnaires de traduction. La méthodologie et les résultats des analyses de corpus, ainsi que les défis pour la lexicographie, seront exposés.

**ABSTRACT**

This article shows how new approaches in corpus analysis could enrich traditional lexicographic descriptions. We examine a set of trends verbs (i.e, denoting an increase), in English, French and Dutch, building on several analyses of parallel corpora and well-targeted monolingual corpora. Parallel corpora give information about the frequency and equivalence of translations. MDS (*MultiDimensional Scaling*) analyses on these quantitative data yield interesting results, in terms of translation profiles. Corpora in the target language allow us to refine these results and to extract salient collocates, and they show the combinatorial properties of trends verbs. The results of all these corpus analyses, by means of translation profiles and lexical profiles, can be used to enrich traditional lexicographic descriptions in translation dictionaries. The methodology and results of the corpus analyses, as well as some challenges for lexicography, will be presented.

**MOTS-CLÉS/KEYWORDS**

dictionnaires de traduction, corpus parallèles, corpus monolingues ciblés, profils de traduction, profils lexicaux combinatoires
translation dictionaries, parallel corpora, target monolingual corpora, translation profiles, lexical combinatorial profiles.

## 1. Introduction

Au début des années 1980, le projet COBUILD a lancé des descriptions lexicographiques fondées sur l'analyse de corpus textuels de plusieurs millions de mots (Sinclair 1987). Depuis lors, on a vu se généraliser une approche lexicographique basée sur des analyses de corpus (*corpus-driven lexicography* [lexicographie guidée par corpus] et *corpus-based lexicography* [lexicographie fondée sur corpus]). Les approches fondées sur corpus fournissent des exemples authentiques, alors que les approches guidées par corpus permettent de découvrir des régularités ou patrons linguistiques.

Ces dernières années, les techniques de collecte et d'analyse de données ont beaucoup évolué. En effet, de nos jours, un nombre toujours grandissant de documents numériques sont disponibles sur Internet. Pensons au *British National Corpus* (*BNC*)[1], à *Frantext*[2], à *Webcorp*[3] ou aux recours au Web utilisé comme corpus (« *Web as corpus* ») (Kilgarriff et Grefenstette 2003 ; Hundt, Nesselhauf, *et al.* 2007). L'exploitation de ces corpus électroniques est considérablement facilitée par la mise à disposition d'outils d'analyse très performants, tels que *Lexico3*[4], *Hyperbase*[5], *WordSmith*[6] ou même le logiciel d'analyse statistique *R*[7].

Dans le présent article, nous présentons les résultats d'une étude exploratoire dans le domaine de la lexicographie bilingue (Bertels, Fairon, *et al.* 2009). Nous expliquons comment utiliser les nouvelles techniques d'analyse et d'exploitation de corpus avec pour objectif l'enrichissement des descriptions lexicographiques traditionnelles dans les dictionnaires de traduction. Dans le cadre de l'expérimentation que nous présentons ici, nous avons étudié les traductions (section 2.2) et les contextes d'emploi (section 3.2.) de trois séries de verbes dénotant la notion de hausse (*trends verbs*) (Verlinde 1995) : en anglais (A) **increase**, **raise**, **rise**, **boost**, **recover**, en français (F) *augmenter, progresser, gagner, croître, accroître, agrandir, grandir* et en néerlandais (N) stijgen, toenemen, verhogen et vergroten. Les verbes de hausse sont certes plus nombreux, mais nous nous sommes limités aux verbes les plus fréquents qui figurent aussi bien dans les corpus parallèles et que dans les corpus monolingues ciblés.

## 2. Corpus parallèles

### *2.1. Corpus parallèles disponibles*

David Lee recense sur son site toute une série de corpus disponibles, parmi lesquels quelques corpus parallèles, qui se sont développés sur le modèle du corpus parallèle canadien Hansard[8]. Les corpus parallèles qui nous ont permis d'établir les profils de traduction appartiennent au projet OPUS[9] (Tiedemann et Nygaard 2004). Ce projet vise à rassembler des corpus parallèles librement accessibles et à les enrichir d'outils de recherche. Le projet OPUS regroupe plusieurs textes multilingues en accès libre, notamment des textes informatiques, juridiques et biomédicaux et des sous-titres. Dans notre expérimentation consacrée aux verbes de hausse, nous avons analysé des comptes rendus du Parlement européen (Europarl). OPUS se caractérise par une approche novatrice qui réside dans un alignement réalisé au niveau du mot. Les liens de mot à mot ont été obtenus à l'aide de Giza++[10], un logiciel utilisé dans la traduction automatique statistique.

Tous les verbes de hausse (en français, en anglais et en néerlandais) ont été rassemblés dans une base de données, avec les traductions dans les deux autres langues.

Cette base de données est interrogeable via une interface web et des scripts en php. Les requêtes génèrent des listes de paires de mots (verbe-traduction), classées par fréquence décroissante de cooccurrence.

### *2.2. Analyse MDS des corpus parallèles du projet OPUS*

Dans le but d'enrichir les descriptions lexicographiques traditionnelles, nous proposons d'exploiter de plusieurs façons les informations de traduction véhiculées dans les corpus parallèles. Pour chaque paire de langue, nous disposons ainsi d'une liste de traductions par verbe étudié, munies de leur fréquence de traduction. Ces informations permettent d'aller au-delà des descriptions lexicographiques traditionnelles, qui se limitent généralement à une liste de traductions, le plus souvent relevant de la même catégorie grammaticale que le verbe traduit. Dans un premier temps, les fréquences de traduction nous renseignent sur la ou les traductions privilégiées (p. ex., *augmenter* se traduit le plus souvent par **increase**). Elles nous donnent aussi une première indication des tendances globales en termes de ressemblances et différences de traduction entre les verbes étudiés. Dans un deuxième temps, les fréquences de traduction nous permettent de procéder à des analyses plus poussées, notamment des analyses de positionnement multidimensionnel (dites MDS, pour *MultiDimensional Scaling*).

Les analyses MDS sont des analyses exploratoires qui facilitent l'interprétation d'une quantité importante de données complexes. Si les informations de traduction d'un seul verbe sont assez faciles à interpréter, il n'en est pas de même lorsque l'on compare, pour plusieurs verbes, de nombreuses traductions et des fréquences de traduction très variées. Comment ces verbes se comportent-ils les uns par rapport aux autres ? Dans quelle mesure présentent-ils des profils de traduction similaires ?

Les analyses MDS prennent comme point de départ des relevés de données (par exemple les fréquences de traduction) pour plusieurs variables (par exemple les verbes de hausse). Le but des analyses MDS est de déterminer dans quelle mesure les variables ont des comportements analogues ou non, c'est-à-dire dans quelle mesure les verbes de hausse ont des profils de traduction similaire ou non. L'analyse peut être exécutée par un certain nombre de logiciels d'analyse statistique, tels la suite R, offerte en accès libre, ou le logiciel commercial XLSTAT[11]. L'analyse MDS appliquée à nos données produit une série de similarités et de dissimilarités entre les verbes à partir de leurs fréquences de traduction. Ces ressemblances et différences sont visualisées au moyen de petites et grandes distances, qui sont projetées dans un espace à deux dimensions. Les verbes qui se traduisent par des mots similaires à des fréquences similaires se regroupent en nuage de points. Par contre, les verbes qui présentent un profil de traduction très spécifique auront tendance à occuper une position isolée par rapport aux autres verbes. Ces verbes isolés se caractérisent généralement par des traductions très particulières et/ou par des fréquences de traduction très élevées.

Une analyse MDS est effectuée à partir d'une table de contingence qui contient les différents verbes de hausse (dans les colonnes), leurs traductions (dans les rangées) et les fréquences de traduction (dans les cases). Les informations de traduction des différents verbes de hausse sont donc croisées et intégrées dans une grande table de contingence. Pour des raisons de représentativité, nous ne retenons que les traductions dont la fréquence est supérieure à 10. Pour permettre l'analyse MDS dans le

logiciel R, nous assignons la fréquence minimale 1 aux cases vides (tableau 1[12]). À partir de cette table de contingence, le logiciel génère une matrice de dissimilarité, en calculant la différence entre chaque paire de cases. Ceci consiste à comparer toutes les fréquences de traduction de tous les verbes et de caractériser ou de situer les verbes les uns par rapport aux autres en fonction des traductions partagées (ou non) ayant des fréquences de traduction similaires (ou non).

Tableau 1

**Extrait de la table de contingence des verbes anglais (colonnes) et de leurs traductions en français (rangées)**

| | **increase** | **raise** | **rise** | **boost** | **recover** |
|---|---|---|---|---|---|
| **accroissement** | 411 | 1 | 13 | 1 | 1 |
| **accroître** | 1401 | 15 | 1 | 13 | 1 |
| **aggravation** | 21 | 1 | 1 | 1 | 1 |
| **aggraver** | 13 | 1 | 1 | 1 | 1 |
| **amélioration** | 36 | 1 | 1 | 1 | 1 |
| **améliorer** | 117 | 21 | 1 | 1 | 1 |
| **amplifier** | 13 | 1 | 1 | 1 | 1 |
| **attirer** | 1 | 10 | 1 | 1 | 1 |
| **augmentation** | 2931 | 11 | 197 | 1 | 1 |
| **augmenter** | 2547 | 78 | 44 | 22 | 1 |
| **avoir** | 1 | 18 | 1 | 1 | 1 |
| **creuser** | 10 | 1 | 1 | 1 | 1 |
| **croissance** | 169 | 1 | 12 | 1 | 1 |
| **croître** | 124 | 1 | 1 | 1 | 1 |

Les analyses MDS par paire de langues (langue-source → langue-cible) permettent de générer deux visualisations. La première montre la répartition en 2D des traductions des verbes analysés. Comme il s'agit d'un grand nombre de données, seules les traductions particulières très fréquentes qui occupent une position isolée sautent aux yeux. Les autres traductions, qui ont des fréquences de traduction similaires et/ou qui sont partagées par plusieurs verbes, se regroupent et forment un nuage de points représentant chacun une traduction particulière. La deuxième visualisation est celle des verbes de hausse en fonction de leurs traductions. Elle contient moins de données, à savoir uniquement les verbes étudiés, et elle est dès lors plus facilement interprétable. Elle permet de situer les verbes les uns par rapport aux autres et dès lors de vérifier, dans un ensemble de verbes que l'on peut considérer comme des parasynonymes, à quel point ces verbes se ressemblent ou se différencient.

Pour les trois langues étudiées, nous avons examiné toutes les combinaisons possibles et ainsi étendu l'analyse ébauchée dans Bertels, Fairon, *et al.* (2009), c'est-à-dire les six paires de langues : français – anglais, français – néerlandais, anglais – français, anglais – néerlandais, néerlandais – français et néerlandais – anglais. Ces nombreuses analyses MDS permettent d'étudier plusieurs comparaisons intéressantes. Dans un premier temps, nous étudions par langue-source les traductions dans les deux langues-cibles, dans le but de vérifier si les profils de traduction des verbes se ressemblent. Par exemple, pour les verbes français, nous comparons le profil de traduction en anglais à celui en néerlandais (section 2.2.1.). Une deuxième comparaison intéressante consiste à inverser le sens des traductions et donc à comparer le

profil de traduction des verbes français en anglais (F → A) avec le profil de traduction des verbes anglais en français (A → F) (section 2.2.2.). Finalement, la confrontation des résultats des analyses précédentes permettra de repérer des tendances générales en termes de profils de traduction (section 2.2.3.). Les analyses effectuées dans le cadre de cette expérimentation et discutées dans cet article sont basées sur des analyses MDS. Cependant, nous nous proposons de regarder également les listes des fréquences de traduction. En effet, comme les analyses MDS s'appuient sur les fréquences de traduction pour mettre en évidence des phénomènes saillants, un retour aux fréquences est indispensable pour faciliter l'interprétation (linguistique) des visualisations.

### *2.2.1. Les profils de traduction par langue-source*

Comme nous l'avons expliqué ci-dessus, les visualisations des analyses MDS permettent de positionner les verbes d'une langue les uns par rapport aux autres en fonction de leurs traductions dans les autres langues, plus particulièrement en fonction des fréquences de ces traductions. Considérons d'abord les verbes français (*augmenter, progresser, gagner, croître, accroître, agrandir, grandir*) traduits en néerlandais et en anglais. Comme les deux dimensions d'une matrice de traductions sont soumises à une analyse MDS, nous générons deux visualisations par paire de langues. La première montre la répartition des traductions des différents verbes analysés, alors que la deuxième montre les verbes proprement dits en fonction de leurs traductions. La visualisation des traductions en néerlandais des verbes français de hausse (figure 1) montre clairement un regroupement important à gauche au milieu, puis un axe vertical avec des traductions particulières en haut (boeken, vooruitgang, vordering) et en bas (winnen, verdienen), et finalement un axe horizontal avec des traductions particulières à droite (verhogen, toenemen, stijgen, vergroten, meer, groter).

FIGURE 1

**Visualisation des traductions en néerlandais des verbes français (F → N)**

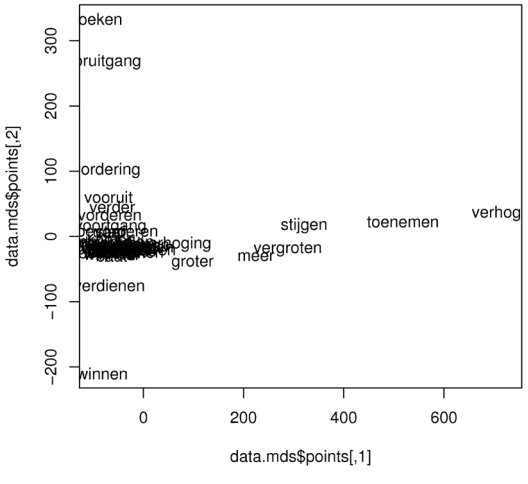
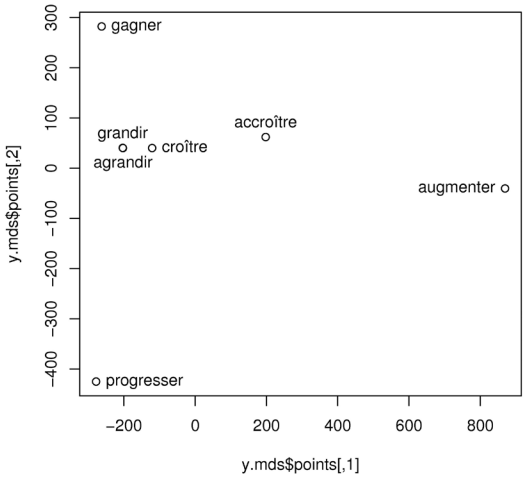


Figure 2

**Visualisation des verbes français en fonction de leurs traductions en néerlandais (F ➔ N)**




La visualisation des rapports entre les verbes français, en fonction de leurs traductions en néerlandais montre que ces traductions particulières correspondent à des verbes qui occupent une position isolée, à savoir *augmenter*, *gagner* et *progresser* (figure 2). Ces trois verbes se distinguent par un profil de traduction particulier, qui s'explique principalement par les fréquences élevées de certaines traductions privilégiées. En effet, un retour aux fréquences dans les corpus parallèles permet de mieux interpréter les visualisations. *Augmenter* (2836 occurrences dans le corpus français-néerlandais) se traduit principalement par verhogen (fréquence de traduction 775) et par toenemen (fréq. 550), *progresser* (1370 occurrences) se traduit surtout par vooruitgang boeken (fréq. 320), et *gagner* (662 occurrences) par winnen (fréq. 464) et par verdienen (fréq. 133). *Accroître* occupe une position intermédiaire, entre *augmenter* et le cluster des autres verbes, et se traduit surtout par meer (fréq. 333) et groter (fréq. 307). Si *accroître* se situe plus près du verbe *augmenter*, c'est en raison de quelques traductions privilégiées partagées, notamment vergroten (fréq. 291) et verhogen (fréq. 187). Les traductions des autres verbes sont moins exclusives et par conséquent ces verbes se situent plus près les uns des autres.

Les analyses des traductions du corpus français-anglais, et plus particulièrement la visualisation des traductions en anglais et la visualisation des verbes français traduits en anglais, confirment les observations tirées du corpus français-néerlandais. La visualisation des traductions en anglais (figure 3) montre dans la partie supérieure à gauche les traductions **progress** et **make_progress** (vooruitgang, vooruitgang boeken), ensuite à droite **increase** (verhogen, toenemen), puis à gauche un cluster avec la plupart des traductions, et finalement au-dessous du cluster **more** et **greater**. Ces deux mots correspondent parfaitement aux mots néerlandais *meer* et *groter*. Les traductions privilégiées sont **increase** (fréquence de traduction 3293) pour le verbe *augmenter* (3911 occurrences dans le corpus français-anglais), **progress** (fréq. 640) et **make_progress** (fréq. 355) pour le verbe *progresser* (1531 occurrences), et **increase** (fréq. 1721) pour *accroître* (2983 occurrences). La visualisation des verbes français en fonction de leurs traductions en anglais (figure 4, figure empruntée à Bertels, Fairon,

*et al.* 2009) confirme les conclusions de la figure 2, qui était réalisée en fonction des traductions en néerlandais. *Augmenter* et *progresser* occupent une position isolée et *accroître* reprend sa position intermédiaire. Toutefois, on constate que *gagner* rejoint le cluster des autres verbes. Bien que ce verbe se traduise par **win** (fréq. 284), **gain** (fréq. 247) et **earn** (fréq. 142), qui sont les équivalents du néerlandais winnen et verdienen, ces fréquences sont trop faibles par rapport aux fréquences de traduction des autres verbes pour faire ressortir ces traductions dans la visualisation. Par conséquent, *gagner* n'occupe pas de position isolée.

FIGURE 3

**Visualisation des traductions en anglais des verbes français (F ➔ A)**

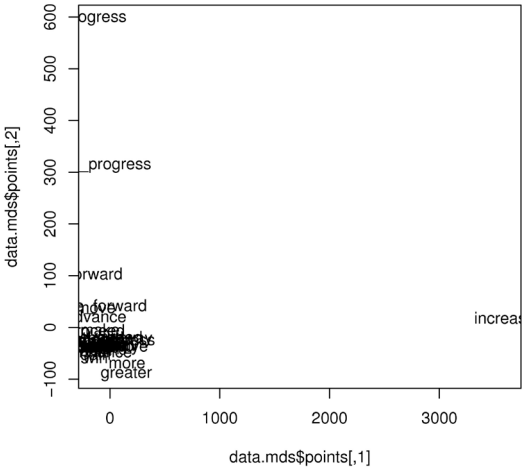
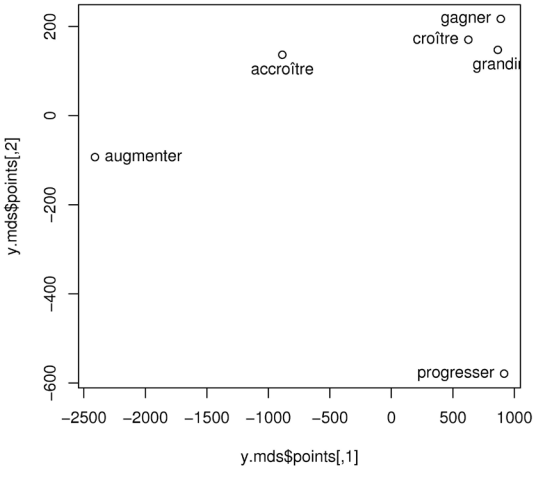



FIGURE 4

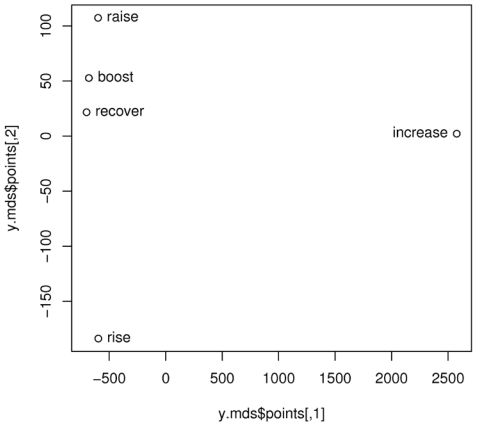
**Visualisation des verbes français en fonction de leurs traductions en anglais (F ➔ A)**




Les analyses des paires de langues anglais-néerlandais et anglais-français permettent de visualiser la répartition des verbes anglais en fonction de leurs traductions en néerlandais et en français. Il convient de noter tout d'abord que cette répartition

est largement influencée par la prédominance absolue du verbe **increase** dans les corpus parallèles : 11 221 occurrences dans le corpus anglais-néerlandais et 10 368 dans le corpus anglais-français. Les occurrences des autres verbes anglais sont nettement inférieures et se limitent à quelques centaines d'occurrences, à l'exception de **raise** (4783 occurrences dans le corpus anglais-français). Pour les deux paires de langues (A → N et A → F), nous recensons trois verbes qui occupent une position isolée (**increase**, **raise** et **rise**), non seulement en raison de leur prédominance quantitative (**increase**, **raise**), mais aussi en raison de la particularité de leurs traductions (figure 5 pour les verbes anglais en fonction de leurs traductions en néerlandais). À l'instar des verbes français, les trois pôles anglais se caractérisent par quelques traductions privilégiées. Le verbe **increase** se traduit souvent par verhogen (fréquence de traduction 1592) et meer (fréq. 1446) en néerlandais et par *augmentation* (fréq. 2931) et *augmenter* (fréq. 2547) en français. Le verbe **rise** se traduit principalement par stijgen (fréq. 168) et stijging (fréq. 134) en néerlandais et par *augmentation* (fréq. 197) et *hausse* (fréq. 117) en français. Il ressort de cette liste de traductions privilégiées que **rise** et **increase** sont souvent traduits par un substantif aussi bien en néerlandais qu'en français. Pour **increase**, cela s'explique en partie par le fait qu'il s'agit d'une forme verbale aussi bien que substantivale. Mais comme les autres verbes anglais (surtout **rise**) se traduisent également très souvent par des substantifs néerlandais et français, l'observation est tout de même pertinente. Le verbe **raise** doit sa position isolée principalement à sa polysémie (le verbe polysémique français *gagner* et ses traductions en néerlandais). En effet, il se traduit souvent par des verbes comme *soulever, poser* et *aborder* en combinaison avec un substantif comme *question*. Le phénomène de la polysémie constitue donc une source de bruit qu'il faudrait évacuer pour affiner davantage les résultats de ces analyses.

Figure 5
**Visualisation des verbes anglais en fonction de leurs traductions en néerlandais (A → N)**

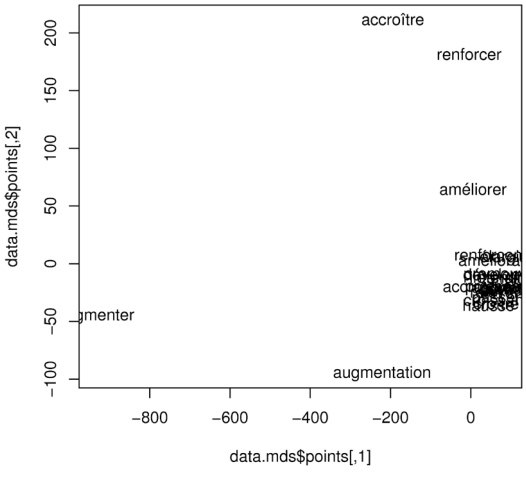



Pour les dernières paires de langues, à savoir néerlandais-français et néerlandais-anglais, nous ne retenons que quatre verbes néerlandais qui apparaissent dans les différents corpus parallèles, à savoir stijgen, toenemen, verhogen et vergroten. La notion de traduction privilégiée revêt un autre sens dans ces deux corpus parallèles, puisque les quatre verbes néerlandais semblent se traduire principalement par *aug-*

*menter* et *accroître* et par **increase**. Ces traductions sont largement prédominantes et beaucoup plus fréquentes que les autres traductions et ce, pour les quatre verbes analysés. En effet, entre 25 et 62 % des occurrences de ces quatre verbes néerlandais se traduisent par *augmenter* et entre 50 et 79 % de leurs occurrences se traduisent par **increase**. Contrairement aux verbes anglais, les verbes néerlandais sont traduits moins souvent par un substantif français, à l'exception du substantif *augmentation*. Ils se traduisent par un verbe, et de préférence *augmenter* (en français) et **increase** (en anglais). Toutefois, il est à remarquer que **increase** est aussi bien employé comme verbe que comme substantif, selon les contextes d'emploi. Malgré la prédominance des traductions *augmenter* et **increase**, la répartition des quatre verbes néerlandais en fonction de leurs traductions en français et en anglais permet tout de même de retrouver trois pôles : stijgen, vergroten, verhogen, avec au milieu toenemen (figure 6). Cette observation nous incite à étudier de plus près les fréquences de traduction des trois verbes néerlandais, en français et en anglais, dans le but de détecter des ressemblances entre les trois pôles observés pour les six paires de langues étudiées (section 2.2.3.).

FIGURE 6

**Visualisation des traductions françaises des verbes néerlandais (N → F)**




### *2.2.2. Du français à l'anglais et inversement*

Avant de passer à une comparaison générale, nous comparons les profils de traduction des verbes français en anglais et ceux des verbes anglais en français. Comme nous l'avons mentionné ci-dessus, *augmenter* est le verbe le plus fréquent dans le corpus français-anglais (3911 occurrences) et 84 % de ses occurrences sont traduites par **increase** (substantif ou verbe). Il est suivi de *accroître* (2983 occurrences), dont 57 % des occurrences sont traduites par **increase**. Les autres verbes français, à savoir *progresser* et *gagner*, ne se traduisent presque jamais par *increase*. On observe principalement la ou les traductions littérales (**progress** et **earn**, **win**, **gain**,…). Si l'on inverse le sens de traduction, on s'aperçoit que les traductions les plus fréquentes du verbe omniprésent **increase** (10 368 occurrences) sont *augmentation* (fréq. 2931 ou 28 % des occurrences), *augmenter* (fréq. 2547 ou 24 %), *accroître* (13 %), *accroissement*

(4 %) et *hausse* (3 %). Ces traductions sont plus diversifiées, contrairement à la prédominance des 84 % (**increase**) pour *augmenter*, et font ressortir l'importance des substantifs. Les verbes *augmenter* et **increase** se caractérisent donc par une certaine réciprocité, mais non par une relation bi-univoque.

*2.2.3. Conceptualisation à travers les trois langues*

La prédominance des verbes *augmenter* et **increase** en français et en anglais et leur réciprocité partielle laissent présager d'importantes ressemblances entre les verbes qui occupent une position isolée pour les différentes paires de langues étudiées, c'est-à-dire indépendamment des langues-cibles ou du sens de traduction. Rappelons qu'une position isolée sur la visualisation de l'analyse MDS s'explique par des traductions particulières et/ou par des fréquences de traduction élevées. Le français comme langue-source montre clairement 3 verbes qui occupent une position isolée : *augmenter*, *gagner* et *progresser*. Le verbe *accroître* se situe entre *augmenter* et le cluster des autres verbes et se traduit très souvent par **increase**. Le schéma ci-dessous visualise les quatre verbes français les plus pertinents avec leurs traductions en anglais et néerlandais, complétées par les traductions en français des verbes dans les autres langues-sources. Il est à remarquer que *augmenter* et *accroître* partagent quelques verbes anglais (principalement **increase**).

F : *augmenter, augmentation, hausse*
A : *increase, raise, rise*
N : *verhogen, toenemen*

F : *accroître, accroissement*
A : *increase, raise, rise, greater, more*
N : *stijgen, vergroten, groter, meer*

F : *gagner*
A : *gain, win, earn*
N : *verdienen, winnen*

F : *progresser*
A : *progress, forward, make_progress*
N : *vooruitgang, vordering_boeken*

Ces regroupements constituent un embryon de dictionnaire de traduction pour des mots à haute fréquence. On pourrait y ajouter des indications de fréquence pour guider encore davantage l'utilisateur du dictionnaire quand il choisit une traduction.

## 3. Corpus monolingues

### *3.1. Corpus monolingues ciblés*

Les corpus monolingues ciblés utilisés dans notre expérimentation proviennent du Web et ont été rassemblés par le logiciel Corporator (Fairon 2006). Ce logiciel effectue une analyse d'un certain nombre de sources prédéfinies : des flux RSS. Nous avons ciblé des textes relevant du domaine économique à partir de flux RSS en anglais (*The Economist*), en néerlandais (*BNR*, *de Standaard*) et en français (*Le Monde*, *La Libre Belgique*, *La Tribune*). Ces textes économiques sont particulièrement intéressants, parce qu'ils se caractérisent par un emploi fréquent de verbes de hausse. Les données qui constituent les corpus ciblés se répartissent sur une période allant de mai 2007 à mai 2008, pour avoir un corpus d'une taille suffisante et disposer de plusieurs centaines d'occurrences par verbe de hausse. Ensuite, le logiciel en code source libre (*open source*) Unitex (Paumier 2003) a été utilisé pour générer des concordances, c'est-à-dire pour isoler des phrases contenant des verbes de hausse. On se reportera à Bertels, Fairon, *et al.* (2009) pour plus de détails sur la constitution des corpus monolingues ciblés.

### *3.2. Analyse MDS des corpus monolingues*

Les corpus monolingues ciblés ont également fait l'objet d'une analyse MDS, dans le but de fournir des renseignements plus détaillés sur les contextes d'emploi des verbes dénotant la notion de hausse et plus particulièrement sur leurs collocatifs (Hausmann 1989). Dans ce cas, l'analyse MDS porte sur le profil combinatoire (Blumenthal 2002 et 2006) des verbes.

Il est vrai que les dictionnaires de traduction essaient déjà de contextualiser les traductions, mais ces contextes d'emploi sont souvent donnés en vrac, sans subdivisions ou précisions sémantiques. En plus, si le verbe français A est traduit par le verbe anglais B, il n'en va pas de même pour leurs contextes d'emploi. Les collocatifs du verbe français ne sont pas forcément les mêmes que les collocatifs du verbe anglais. Il ne suffit pas de traduire simplement les collocatifs ou contextes d'emploi, il faut vérifier la pertinence des collocatifs traduits ou en rechercher d'autres qui sont (plus) pertinents. Il faudrait en d'autres mots vérifier la pertinence statistique de l'association entre le verbe de hausse et les mots dans son contexte immédiat (ou ses cooccurrents).

Les résultats des analyses MDS permettent d'enrichir les descriptions lexicographiques traditionnelles et d'apporter des précisions supplémentaires. À l'instar du profil de traduction, nous déterminons le profil combinatoire des verbes dénotant la notion de hausse dans les trois langues. Rappelons les trois séries de verbes : **increase**, **raise**, **rise**, **boost**, **recover** en anglais (A), *augmenter, progresser, gagner, croître, accroître, agrandir, grandir* en français (F) et stijgen, toenemen, verhogen et vergroten en néerlandais (N).

#### *3.2.1. Les profils combinatoires des verbes de hausse*

##### *3.2.1.1. Comment déterminer les collocatifs pertinents ou le profil combinatoire ?*

La constitution des corpus monolingues ciblés (section 3.1.) nous a permis de constituer un sous-corpus de phrases pour chaque verbe de hausse, dans sa forme lemmatisée, et ce, dans les trois langues. Ainsi, le sous-corpus du verbe *augmenter* contient uniquement des phrases avec le verbe *augmenter*. Afin de déterminer le profil combinatoire de ces verbes, nous déterminons d'abord leurs collocatifs, c'est-à-dire leurs cooccurrents statistiquement pertinents. À cet effet, nous utilisons le logiciel *WordSmith*.

Dans un premier temps, nous dressons une liste de fréquence (option *Wordlist* dans *WordSmith*) de tous les mots par sous-corpus. Ensuite, chaque sous-corpus est analysé avec l'application *Concord*, afin de calculer les collocatifs du verbe en question. Il s'agit donc de déterminer les collocatifs (*collocates*) d'un mot de base lemmatisé (*search word*). Les collocatifs statistiquement pertinents sont calculés à l'aide de la mesure d'association du log de vraisemblance (*Log Likehood Ratio* ou LLR), dans une fenêtre d'observation (*span*) de 5 mots à gauche et à droite, pour les mots d'une fréquence minimale de 5. La mesure du LLR indique la pertinence de l'association verbe – collocatif. Dans le cadre de notre étude, le seuil de la mesure a été fixé à 50 pour l'anglais et le français et à 100 pour le néerlandais. Cela permet non seulement de limiter le nombre de collocatifs par verbe, en vue de la création de la table de contingence par langue, mais aussi de ne retenir que les collocatifs les plus pertinents, qui se caractérisent par une association (très) forte avec le verbe étudié. Les mots

fonctionnels ont été supprimés, parce qu'ils sont partagés par tous les verbes et que, dès lors, ils ne sont pas distinctifs pour visualiser les distances ou ressemblances entre les verbes analysés.

Les tableaux 2 et 3 ci-dessous montrent un extrait de la table de contingence pour les verbes anglais et français respectivement. Il est à noter que les collocatifs pertinents des verbes de hausse, dans les trois langues, sont surtout des substantifs. L'indication quantitative prise en considération n'est pas la co-fréquence d'apparition, mais la mesure statistique du LLR, qui indique la force ou la pertinence statistique de l'association. Le tableau 3 fait écho au tableau 2 dans Bertels, Fairon, *et al.* (2009).

Tableau 2

**Extrait de la table de contingence des collocatifs (rangées) des verbes anglais (colonnes)**

| | **increase** | **rise** | **raise** | **boost** | **recover** |
|---|---|---|---|---|---|
| OUTPUT | 109,2858 | 1 | 185,6028 | 137,7666 | 1 |
| PAYROLLS | 1 | 1 | 81,76154 | 1 | 1 |
| PERCENT | 1704,668 | 59,44097 | 1 | 131,3675 | 1 |
| PLAN | 1 | 1 | 55,24586 | 64,74115 | 1 |
| PLANNED | 1 | 1 | 62,57031 | 1 | 1 |
| PLANS | 1 | 1 | 138,8995 | 1 | 1 |
| POSTED | 77,70508 | 56,7392 | 1 | 1 | 1 |
| POSTS | 2 | 1 | 106,126 | 1 | 1 |
| PRESSURE | 111,5064 | 1 | 64,13657 | 1 | 1 |
| PRICE | 403,9792 | 1 | 129,8015 | 129,8007 | 1 |
| PRICES | 201,9742 | 1 | 230,2431 | 165,6286 | 1 |
| PRODUCTION | 109,4491 | 1 | 146,7692 | 76,99423 | 1 |
| PROFILE | 1 | 1 | 261,2902 | 1 | 1 |
| PROFIT | 317,8684 | 76,26538 | 415,2601 | 257,3201 | 1 |
| PROFITS | 1 | 1 | 1 | 139,9133 | 1 |
| QUARTER | 105,4518 | 1 | 1 | 1 | 1 |
| QUARTERLY | 104,9681 | 83,27461 | 104,2726 | 50,33535 | 1 |
| QUESTIONS | 1 | 1 | 214,4581 | 1 | 1 |
| QUOTE | 1 | 1 | 190,7207 | 1 | 1 |
| RATE | 62,45462 | 1 | 1 | 1 | 1 |
| RATES | 1 | 1 | 325,0611 | 70,43732 | 1 |

Tableau 3

**Extrait de la table de contingence des collocatifs (rangées) des verbes français (colonnes)**

| | **augmenter** | **progresser** | **gagner** | **grandir** | **croître** | **agrandir** | **accroître** |
|---|---|---|---|---|---|---|---|
| PARI | 1 | 1 | 124,8504 | 1 | 1 | 1 | 1 |
| PART | 105,182312 | 78,481163 | 1 | 1 | 1 | 1 | 1 |
| PARTENARIATS | 54,34153748 | 2 | 1 | 1 | 1 | 1 | 1 |
| PARTICIPATION | 59,02616882 | 2 | 1 | 1 | 1 | 1 | 1 |
| PARTS | 1 | 1 | 110,9474 | 1 | 1 | 1 | 1 |
| PC | 320,4685669 | 181,91954 | 160,4114 | 1 | 66,43745 | 1 | 1 |
| PÉAGES | 54,34153748 | 2 | 2 | 1 | 1 | 1 | 1 |
| PÉTROLE | 112,0878601 | 1 | 1 | 1 | 1 | 1 | 1 |
| PIB | 152,8404846 | 145,37674 | 1 | 1 | 56,29705 | 1 | 1 |
| POINT | 62,55287552 | 1 | 1 | 1 | 1 | 1 | 1 |
| POINTS | 80,56619263 | 1 | 87,5106 | 1 | 1 | 1 | 1 |
| POPULATION | 73,57284546 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PREMIER | 72,34870148 | 1 | 1 | 1 | 1 | 1 | 1 |
| PRESSION | 1 | 1 | 1 | 1 | 1 | 1 | 60,32048 |
| PRIX | 1001,626953 | 201,13026 | 1 | 1 | 1 | 1 | 1 |
| PRODUCTION | 900,4157715 | 128,6676 | 1 | 1 | 1 | 1 | 141,4259 |
| PRODUCTIVITÉ | 65,04171753 | 1 | 1 | 1 | 1 | 1 | 1 |
| PRODUITS | 155,7894897 | 82,955864 | 1 | 1 | 1 | 1 | 1 |
| PROMESSE | 2 | 1 | 248,5492 | 1 | 1 | 1 | 1 |
| RAPPORT | 117,4226837 | 51,848938 | 1 | 1 | 1 | 1 | 1 |

Les analyses MDS permettent d'aller plus loin que la simple liste des collocatifs avec des indications de co-fréquence ou de pertinence d'association. À partir du tableau de contingence des collocatifs d'une série de verbes, par exemple, on pourrait visualiser les ressemblances et les différences entre ces verbes, en fonction des collocatifs partagés ou non. À l'instar des visualisations des verbes de hausse en fonction de leurs traductions, on peut les situer les uns par rapport aux autres, à une distance plus ou moins grande, en fonction de leurs collocatifs. En effet, les verbes qui figurent avec les mêmes collocatifs, donc dans les mêmes contextes d'emploi, vont se retrouver ensemble. Le critère de ressemblance ou de différence n'est pas la fréquence du collocatif, mais la pertinence de l'association verbe – collocatif (voir la table de contingence).

*3.2.1.2. Différences de fréquence et nombre de collocatifs*

Avant de passer aux résultats de ces analyses MDS en fonction des collocatifs, il convient d'insister sur les grandes différences de fréquence entre les verbes de hausse dans les corpus monolingues ciblés. Par exemple, le verbe français *augmenter*, qui est très fréquent également dans les corpus parallèles (section 2.2.) est de loin le verbe le plus fréquent dans le corpus monolingue : on recense 2340 occurrences. Le sous-corpus du verbe *augmenter* contient donc 2340 phrases. Ensuite, on retrouve les deux autres verbes plutôt fréquents dans les corpus parallèles, à savoir *progresser* (1175 occurrences dans le corpus monolingue ciblé) et *gagner* (1060 occurrences dans le corpus monolingue ciblé). Pour les autres verbes, les fréquences d'occurrence dans le corpus monolingue ciblé décroissent rapidement : on ne recense plus que quelques centaines d'occurrences. Ces différences de fréquence influent aussi sur le nombre de collocatifs pertinents, puisqu'un verbe plus fréquent a d'autant plus de chances de se retrouver dans le même contexte d'emploi, ce qui fait grimper la mesure d'association ou la valeur du LLR. Ainsi, les verbes *augmenter*, *progresser* et *gagner* ont respectivement 83, 50 et 33 collocatifs pertinents. Le quatrième verbe (*accroître*) n'en a plus qu'une dizaine.

En anglais, on observe également de grandes différences de fréquence. Les verbes **increase** (1758 occurrences) et **rise** (1742 occurrences) ont des fréquences comparables, mais la fréquence de *raise* est beaucoup plus élevée (6482 occurrences). Celle de **boost** (1071 occurrences), et surtout de **recover**, est moins élevée (173 occurrences). Ces différences de fréquence entraînent également des différences quant au nombre de collocatifs pertinents : **raise** a le plus de collocatifs (122 occurrences), **increase** et **boost** ont plus ou moins autant de collocatifs (respectivement 79 et 69 occurrences). Par contre, **rise** a très peu de collocatifs (8 occurrences), en dépit de sa fréquence d'occurrence comparable à celle du verbe **increase**.

En néerlandais, les différences de fréquence dans le corpus monolingue ciblé sont même plus importantes. Le verbe stijgen dépasse de très loin les autres verbes de hausse, car il figure 13 224 fois dans le corpus néerlandais et se caractérise par 104 collocatifs

pertinents. C'est un nombre très important, en dépit du fait que le seuil de la mesure d'association (LLR) a été fixé à 100 pour le néerlandais (seuil fixé à 50 pour le français et l'anglais). Cela implique que nous ne retenons que les collocatifs les plus pertinents des verbes néerlandais, à savoir les collocatifs dont la valeur de LLR est supérieure à 100. Verhogen (3186 occurrences) et toenemen (2722 occurrences) sont plus fréquents que vergroten (410 occurrences) et ont une vingtaine de collocatifs pertinents ; vergroten n'en a aucun.

### *3.2.2. Visualisation des verbes de hausse en fonction de leurs collocatifs*

Comme on l'a vu ci-dessus pour les profils de traduction, les analyses MDS permettent non seulement de situer les verbes les uns par rapport aux autres en fonction des collocatifs partagés, c'est-à-dire en fonction de leur profil combinatoire, elles permettent aussi de montrer la répartition des collocatifs et de dresser une comparaison entre les collocatifs dans les trois langues.

La visualisation des verbes français en fonction de leurs collocatifs (figure 7) montre trois pôles ou trois verbes qui occupent une position isolée. Ces verbes se caractérisent par un profil combinatoire particulier. La position périphérique du verbe *agrandir* s'explique par la prépondérance de quelques collocatifs très fréquents (*image*, *dépêches*…) (figure 8 pour la visualisation des collocatifs français). Tout comme pour la première série d'analyses, ce cas montre que la polysémie des mots interfère de temps à autre et qu'il faudrait nettoyer davantage encore les données brutes. *Augmenter* se retrouve seul, principalement en raison du LLR de ces collocatifs, qui est beaucoup plus élevé que celui des collocatifs des autres verbes. Cela s'explique en partie par la fréquence plus élevée du verbe même. En outre, quelques-uns des collocatifs (*taille* et *texte*) sont très prépondérants (dans « augmenter la taille du texte ») et ne s'observent pas pour les autres verbes. Ces observations plaident donc également pour la constitution de corpus plus vastes et plus équilibrés pour l'extraction de collocatifs pertinents. *Progresser* n'occupe pas de position isolée : il rejoint le cluster des autres verbes en fonction de son profil combinatoire. Son profil de traduction, par contre, était plus particulier et le plaçait dans une position plus périphérique (voir ci-dessus).

Figure 7

**Visualisation des verbes français en fonction de leurs collocatifs**

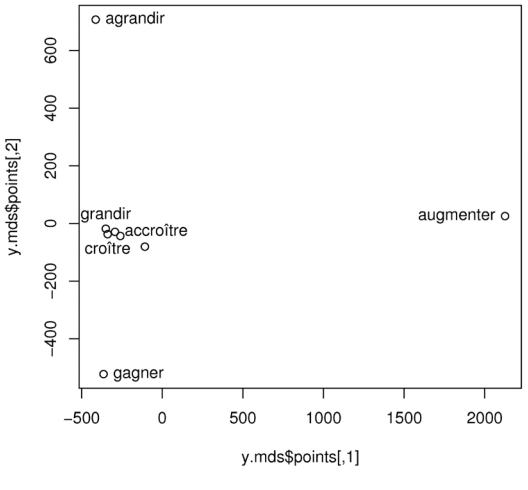
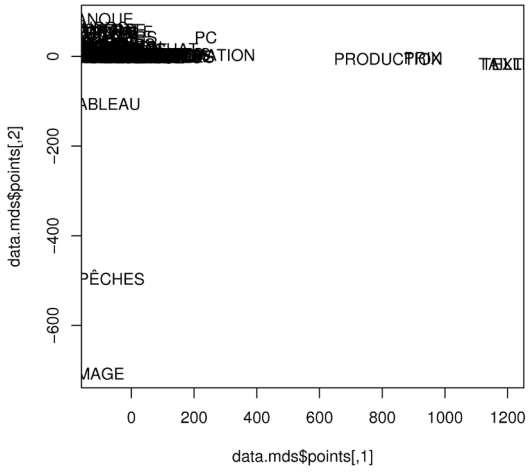
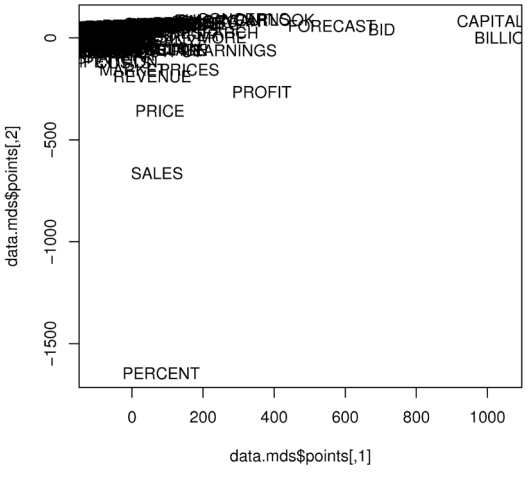


FIGURE 8
**Visualisation des collocatifs des verbes français**




En anglais, on retrouve 2 pôles, à savoir **increase** et **raise**. La position isolée du verbe **increase** s'explique par son profil combinatoire particulier, c'est-à-dire par la présence de quelques collocatifs très pertinents, par exemple **percent**, **sales**, **price**, **profit**. Surtout le collocatif **percent** a une valeur de LLR très élevée (1704) (tableau 2 ci-dessus) par rapport aux autres collocatifs fréquents tels que **price** (403). Le verbe **raise** doit sa position périphérique principalement aux collocatifs **billion**, **capital**, **forecast**, **outlook**… (figure 9).

FIGURE 9
**Visualisation des collocatifs des verbes anglais**




Les analyses MDS des verbes néerlandais en termes de profil combinatoire aboutissent à des résultats similaires. Le verbe stijgen est de très loin le verbe le plus fréquent (13 224 occurrences) avec le plus de collocatifs qui ont, par conséquent, un LLR

beaucoup plus élevé que les collocatifs des autres verbes. Ainsi, le collocatif le plus pertinent de stijgen, procent (voir **percent** en anglais), a une valeur de LLR de 11 247, contre une valeur de LLR de 869 et 299 comme collocatif des verbes toenemen et verhogen respectivement. Vergroten est le deuxième pôle en néerlandais, en raison de quelques collocatifs pertinents très spécifiques (koersdoel, advies, winstverwachting).

En conclusion, les verbes de hausse qui occupent une position isolée se distinguent bel et bien des autres par un profil combinatoire particulier. En français, on retrouve 3 pôles, à savoir *augmenter* (le verbe le plus fréquent), *agrandir* (en raison des collocatifs particuliers tels que *image, …*) et *gagner* (collocatifs particuliers, *manque, gagnant, argent*, qui relèvent du contexte financier ou qui font partie d'expressions). La visualisation des verbes anglais permet de retrouver 2 pôles, **increase** et **raise**. Le premier verbe a un collocatif très pertinent (**percent**). Le second se singularise surtout par sa fréquence très élevée, qui entraîne plus de collocatifs (ce qui est le cas pour tous les verbes très fréquents). Toutefois, presque tous les collocatifs pertinents du verbe **raise** sont employés dans le contexte de la hausse. La visualisation des verbes néerlandais permet de formuler la même conclusion : la fréquence d'un verbe, stijgen, dépasse de très loin la fréquence des autres verbes, dont verhogen, qui est le deuxième verbe en termes de fréquence, et son collocatif le plus pertinent (procent) est extrêmement significatif. Les collocatifs les plus pertinents du verbe stijgen sont employés comme sujet, dans un contexte de hausse, tandis que ceux du verbe verhogen sont employés comme objet direct.

En guise de comparaison, nous avons établi le profil combinatoire de quelques verbes français dans le logiciel *Sketch Engine*[13], à partir d'un corpus de textes français fourni dans le logiciel. Les figures 10 et 11 montrent les collocatifs du profil combinatoire des verbes *augmenter* et *gagner*. Pour augmenter, on retrouve des collocatifs importants du domaine économique, tels que *productivité, prix, revenu*, etc. Les collocatifs très variés et hétérogènes sémantiquement du verbe *gagner* reflètent la polysémie, observée également dans notre étude exploratoire. Quel que soit le corpus analysé, on constate donc une certaine similarité dans les données. On pourra trouver une confirmation de cet état de fait en réalisant une analyse plus poussée des données fournies par notre corpus de taille relativement restreinte et par le *Sketch Engine* en les comparant par exemple à des données provenant d'autres ressources, comme celles que l'on trouve sur le site des *Voisins De Le Monde*[14].

FIGURE 10

**Extrait du profil combinatoire du verbe *augmenter* (*Sketch Engine*)**

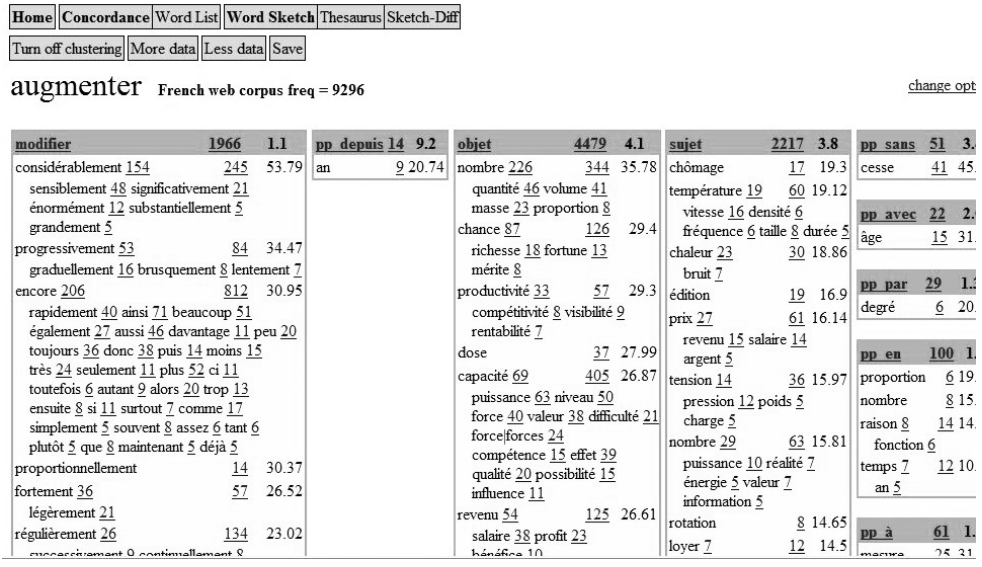
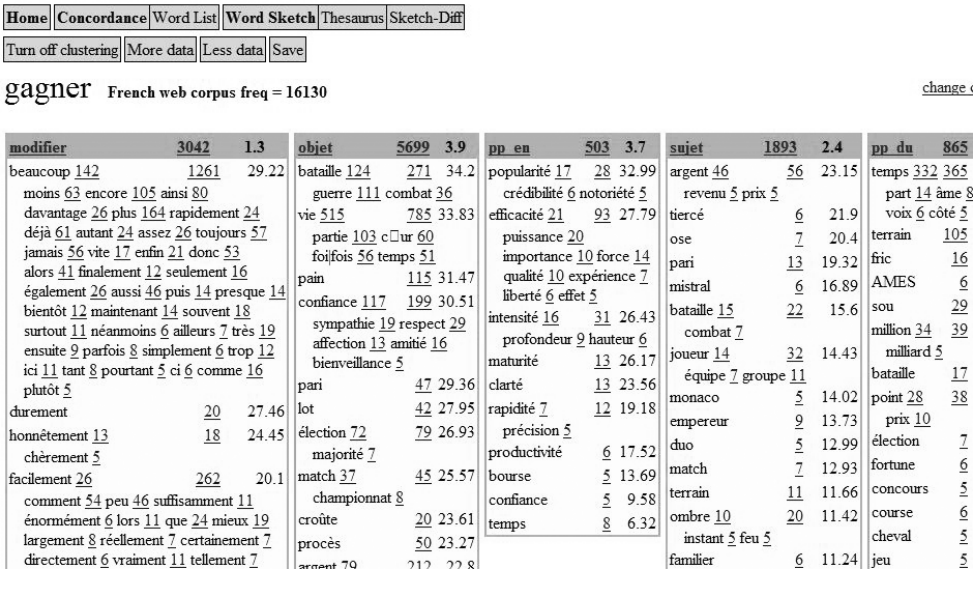



FIGURE 11

**Extrait du profil combinatoire du verbe *gagner* (*Sketch Engine*)**




### *3.3. Profil de traduction versus profil combinatoire*

Il ressort de ces visualisations et analyses exploratoires des corpus parallèles et des corpus monolingues ciblés que les profils de traduction des verbes de hausse correspondent assez bien à leurs profils combinatoires.

Les verbes français *augmenter* et *gagner* se distinguent clairement des autres verbes, d'une part en raison de leurs traductions particulières et d'autre part en raison de leurs collocatifs particuliers. Le verbe *augmenter* se caractérise par des traductions et des collocatifs beaucoup plus fréquents et plus pertinents que ceux des autres verbes, parce que c'est de loin le verbe le plus fréquent, tant dans les corpus parallèles que dans les corpus monolingues ciblés. C'est donc le verbe prototypique pour exprimer la notion de hausse en français. La position particulière du verbe *gagner* s'explique par sa polysémie, qui se reflète aussi bien dans quelques traductions fréquentes très différentes que dans les collocatifs sémantiquement hétérogènes.

Les verbes anglais se caractérisent par la prédominance absolue du verbe **increase** dans les corpus parallèles, qui explique en grande partie son profil de traduction particulier. C'est donc le verbe de hausse par excellence anglais. Par ailleurs, le verbe *augmenter* se traduit presque toujours par **increase**, ce qui renforce leur caractère prototypique. Son profil de traduction particulier est confirmé par son profil combinatoire, c'est-à-dire par quelques collocatifs particuliers très pertinents et par sa fréquence tout de même importante.

## 4. Conclusion

Il est clair que les analyses des traductions et des collocatifs permettent d'enrichir considérablement les descriptions lexicographiques traditionnelles. On pourrait envisager de mentionner non seulement les collocatifs pertinents des verbes, mais également une indication de leur co-fréquence ou de la pertinence statistique de leur co-fréquence. On pourrait même fournir un aperçu de leur profil combinatoire, éventuellement en comparaison avec celui d'un parasynonyme. De telles applications sont implémentées dans la description lexicographique fournie par la *Base lexicale du français*[15] ainsi que dans celles fournies avec les logiciels d'aide à la rédaction *Antidote*[16] et *Cordial*[17], mais manquent encore dans les grands dictionnaires traditionnels.

Bien que cette étude exploratoire soit basée sur un nombre limité de phrases par verbe étudié, elle fait déjà ressortir un certain nombre de collocatifs intéressants et de traductions utiles pour un apprenant (allophone) qui se sert d'un dictionnaire (de traduction). Le dictionnaire papier est limité par des contraintes éditoriales, alors que dans un dictionnaire électronique, rien ne justifie qu'on économise sur les traductions et les collocatifs ou contextes d'emploi fournis. Les trois ressources mentionnées ci-dessus y ajoutent même des liens vers des exemples authentiques tirés de corpus. Il est à noter toutefois que les collocatifs repérés sont largement tributaires du corpus d'où ils sont extraits. Plus le corpus est vaste et équilibré, plus les collocatifs extraits sont fiables et représentatifs de la réalité langagière.

Avec cette étude des verbes dénotant la hausse en anglais, en français et en néerlandais, nous avons voulu illustrer l'apport d'une analyse statistique exploratoire (MDS) à la lexicographie. La combinaison de ce type d'analyse avec des techniques de constitution de corpus toujours plus performantes permet d'effectuer des analyses qui dépassent le mot isolé pour découvrir certaines régularités qui se cachent derrière l'emploi des mots.

**REMERCIEMENTS**

Nous tenons à remercier Jörg Tiedemann (Department of Linguistics and Philology, Uppsala University, Suède) et Cédrick Fairon (CENTAL, UCL Louvain-la-Neuve, Belgique) pour leur contribution à la constitution des corpus parallèles et ciblés.

**NOTES**

1. <http://www.natcorp.ox.ac.uk>, consultée le 27 mars 2010.
2. <http://atilf.atilf.fr/frantext.htm>, consultée le 27 mars 2010.
3. <http://www.webcorp.org.uk/>, consultée le 27 mars 2010.
4. <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/>, consultée le 27 mars 2010.
5. <http://www.unice.fr/bcl/spip.php?rubrique38>, consultée le 27 mars 2010.
6. <http://www.lexically.net/wordsmith/>, consultée le 27 mars 2010.
7. <http://www.r-project.org>, consultée le 27 mars 2010.
8. <http://personal.cityu.edu.hk/~davidlee/devotedtocorpora/CBLLinks.htm>, consultée le 27 mars 2010.
9. <http://urd.let.rug.nl/tiedeman/OPUS>, consultée le 27 mars 2010.
10. <http://code.google.com/p/giza-pp>, consultée le 27 mars 2010.
11. <www.xlstat.com>, consultée le 27 mars 2010.
12. On compare des fréquences très basses (absence de traduction = fréquence 1) à des fréquences élevées et même très élevées. La table de contingence entière pour les 5 colonnes comprend 101 rangées ou traductions françaises.
13. <http://www.sketchengine.co.uk>, consultée le 27 mars 2010.
14. <http://www.irit.fr:8080/voisinsdelemonde/>, consultée le 27 mars 2010.
15. <http://ilt.kuleuven.be/blf>, consultée le 27 mars 2010.
16. <http://www.druide.com/antidote.html>, consultée le 27 mars 2010.
17. <http://www.synapse-fr.com/>, consultée le 27 mars 2010.

**RÉFÉRENCES**

Bertels, Ann, Fairon, Cédrick, Tiedemann, Jörg, *et al.* (2009) : Corpus parallèles et corpus ciblés au secours du dictionnaire de traduction. *Cahiers de Lexicologie*. 94(1):199-219.

Blumenthal, Peter (2002) : Profil combinatoire des noms. Synonymie distinctive et analyse contrastive. *Zeitschrift für französische Sprache und Literatur.* 112(2):115-138.

Blumenthal, Peter (2006) : *Wortprofil im Französischen*. Tübingen : Niemeyer.

Fairon, Cédrick (2006) : Corporator : A tool for creating RSS-based specialized corpora. *In* : *Proceedings of the Workshop Web as Corpus* (EACL, Trento).

Hausmann, Franz Josef (1989) : Le dictionnaire de collocations. *In* : Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, *et al.*, dir. *Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin : De Gruyter, 1010-1019.

Hundt, Marianne, Nesselhauf, Nadja et Biewer, Carolin (2007) : *Corpus Linguistics and the Web.* Amsterdam/New York : Rodopi.

Kilgarriff, Adam et Grefenstette, Gregory (2003) : Web as corpus : introduction to the special issue. *Computational Linguistics.* 29(3):333-348.

Paumier, Sébastien (2003) : *De la reconnaissance de formes linguistiques à l'analyse syntaxique*. Thèse de doctorat non publiée. Marne-la-Vallée : Université de Marne-la-Vallée.

Sinclair, John, dir. (1987) : *Looking Up : An Account of the COBUILD Project in Lexical Computing.* London : Collins.

Tiedemann, Jörg et Nygaard, Lars (2004) : The OPUS corpus – parallel & free. *In* : *Proceedings of the Fourth International Conference on Language Resources and Evaluation.* (LREC04, Lisbonne, 2004). Consultée le 27 mars 2010, <http://stp.ling.uu.se/~joerg/paper/opus_lrec04.pdf>.

Verlinde, Serge (1995) : La combinatoire du vocabulaire des fluctuations dans le discours économique. *Cahiers de lexicologie*. 66(1):137-176.