

# ***Greenspeak* scandinave : explorations co-occurentielles autour de la notion d'« environnement »**

Romuald Dalodiere

Doctorant, Faculté de traduction et d'interprétation (Université de Mons)

[romuald.dalodiere@umons.ac.be](mailto:romuald.dalodiere@umons.ac.be)

## **Abstract**

Among other things, discourse analysis is concerned with lexical recurrence – sedimentation in texts via discursive routines or formula as theorized by e.g. Fiala *et al.* (1987) or Krieg-Planque (2009, 2010, 2013). Such lexical recurrence bears social significance and may as such be subjected to interpretation. Recurring lexical patterns are particularly of interest for critical discourse analysis, which warns against the naturalisation of ideologies (Fairclough, 2010). When it comes to corporate social responsibility (CSR) discourse, several linguistic phenomena pertaining to greenspeak-strategies (Harré *et al.*, 1999) have been highlighted as means to the construction of an environmental ethos. This article represents a preliminary study within a larger body of work that makes use of textometry. It draws on the morphology of Scandinavian languages to get a glimpse of discursive regularity via the analysis of co-occurrences, and more specifically, co-occurrences of compound words of *miljø-/miljö-* (« environment »). The analysis is based on two corpora: one in Danish; the other in Swedish. The texts that make up the corpora are of environmental or societal nature and come from SMEs, whose commitment to such issues can be said to be « voluntary » as these organisations do not fall under the provisions of European directive 2014/95/UE. In addition, the article aims at contributing to a methodology for the extraction of relevant textual data from Internet pages, for textometry-aided discourse analysis.

**Keywords:** Textometry, Discourse analysis, Co-occurrences, Corporate social responsibility

## **Résumé**

Parmi les questions qui intéressent l'analyse du discours se trouvent les formes de récurrences lexicales – figements, routines, formules etc. –, porteuses de signification sociale (Fiala *et al.*, 1987 ; Krieg-Planque, 2009, 2010, 2013). L'inscription durable de tels figements est d'autant plus pertinente pour l'analyse critique du discours, qui met en garde contre la naturalisation d'idéologies (Fairclough, 2010). Dans le cas du discours de responsabilité sociétale des entreprises (RSE), de nombreux phénomènes linguistiques, participant d'une démarche de *greenspeak* (Harré *et al.*, 1999), ont été mis en évidence pour construire le récit de l'implication environnementale de ces dernières. Cet article représente une étude préliminaire au sein d'un

travail plus large mobilisant l'outil textométrique. Il exploite la morphologie des langues scandinaves pour tenter de saisir, par l'analyse co-occurrence, les régularités du discours autour des mots formés de *miljø-/miljö-* ('environnement'). Deux corpus, l'un danois, l'autre suédois, sont convoqués ; les textes concernés relèvent d'une thématique environnementale et/ou sociétale, et proviennent de PME dont l'engagement peut être décrit comme « volontaire » puisqu'elles ne sont pas soumises à la directive 2014/95/UE. Préalablement, l'article cherche à contribuer à une méthodologie pour l'extraction de contenu pertinent à partir de pages Internet, dans un objectif d'analyse du discours outillée par des outils de textométrie.

**Mots-clefs :** analyse du discours, textométrie, co-occurrences, responsabilité sociétale des entreprises

## 1. Introduction

La mise en évidence de séquences récurrentes est l'un des objectifs de la textométrie qui a développé diverses mesures pour apprécier le niveau d'association entre des formes co-occurrences. C'est également un centre d'intérêt de l'analyse du discours qui en fournit plusieurs interprétations. Quantitativement, la fréquence de ces séquences dans un texte n'est pas négligeable, que celles-ci soient catégorisées comme des figements (voir par exemple Krieg-Planque, 2015), des lexies (Fiala *et al.*, 1987), des formules (Krieg-Planque, 2009), des routines (Née *et al.*, 2016) ou des segments répétés (Salem, 1987), en fonction des caractéristiques qu'on leur attribue. Ainsi, Fiala *et al.* (1987) trouvent que ces séquences représentent près de 20% des occurrences de leur objet d'étude, quand elles représentent plus de la moitié du corpus d'Erman et Warren (2000).

Qualitativement, de telles séquences – et leur récurrence – sont un objet d'étude pertinent en analyse du discours. Les figements discursifs qui acquièrent le statut de formule au sens de Krieg-Planque (2009) s'accompagnent d'un sens social et polémique et sont largement réappropriés et dévoyés pour politiser ou dépolitiser une thématique, comme dans le cas de la formule « développement durable » (Krieg-Planque, 2010). De même, les récurrences discursives ont pu être assimilées à de la langue de bois, à une tentative de manipuler l'opinion publique (Fiala *et al.*, 1987, p. 60 ; Krieg-Planque 2013, p. 202). Plus encore, les figements – qui « ne sont pas uniquement des faits de langue, linguistiquement descriptibles, mais aussi des pratiques sociales que des acteurs prennent en charge, dans des logiques organisationnelles ou politiques » (Krieg-Planque, 2015, p. 118) – prennent une dimension rhétorique voire idéologique, en particulier quand ils se produisent sous la forme de références intertextuelles à des textes antérieurs (Krieg-Planque, 2015, pp. 117-118) dont la reproduction

contribue à la « naturalisation » (Fairclough, 2010), par exemple par l'utilisation de cadres sémantiques renvoyant à des discours économiques néoclassiques, comme la notion de « croissance verte » (Stibbe, 2015, p. 50).

## 2. Discours environnemental des entreprises

Ces récurrences peuvent donc se mesurer, mais aussi s'interpréter. Si l'on convoque, dans un esprit multidisciplinaire caractéristique de l'analyse du discours, des domaines de spécialité annexes (en l'occurrence la sociologie), on peut oser un rapprochement entre la notion de naturalisation de Fairclough qui, dans sa version la plus développée, « would be represented by a proposition which was taken as commonsensically given by all members of some community » (Fairclough, 2010, p. 34), et celles de diffusion (« how things flow ») et d'institutionnalisation (« how things become permanent ») décrites par Colyvas et Jonsson (2011, p. 30 pour ces deux citations) : une idée est donc naturalisée (par le discours) dès lors qu'elle est diffusée (elle s'est répandue dans un système social) et institutionnalisée (elle y connaît une implantation durable).

Cette notion offre des perspectives d'analyse pertinentes lorsque l'on s'intéresse au discours environnemental des entreprises. On désigne par *greenspeak* « the whole gamut of linguistic means employed in raising awareness of environmental issues » (Harré *et al.*, 1999, p. 2), dont certaines stratégies lexicales étaient déjà relevées par le *Guardian* en 1992 (Harré *et al.*, 1999, pp. 24-25), notamment l'utilisation de préfixes en *eco-* (voir en français : *écoproduit*, *écolabel*...). Comme le souligne Antelmi (2018, p. 28), « Il discorso “verde” si appoggia ad un vocabolario che, lungi da rappresentare semplicemente la realtà, contribuisce a determinare una visione del mondo. »<sup>1</sup>. Au niveau discursif, ce lexique s'organise sous une multitude d'appellations aux propriétés largement similaires, mais distinguées selon si l'on parle de leur capacité à être « accrocheuses » (*catch-words*), volontairement vagues ou euphémistiques (*fuzzwords*), ou de leur caractère éphémère (*buzzwords*) car liées à un effet de mode (Antelmi, 2018, p. 28). La perspective est résolument socioconstructiviste ; elle implique que de telles stratégies sont révélatrices d'une tendance de fond.

Lorsque l'on parle du discours environnemental des entreprises, il est fréquent de le rapprocher de la « responsabilité sociétale des entreprises » (dorénavant RSE), qui désigne un ensemble de pratiques dont les définitions sont multiples et conflictuelles (voir par exemple Gendron, 2000, p. 13 ou Rasche *et al.*, 2017, p. 10), mais que l'on peut tenter de simplifier en

<sup>1</sup> Notre traduction : 'Le discours environnemental repose sur un lexique qui, loin de simplement représenter la réalité, contribue à construire une vision du monde.'

la ramenant à l'ensemble des mesures d'ordre environnemental et sociétal adoptées par les entreprises. Plusieurs travaux en analyse du discours ont relevé une tendance à l'institutionnalisation de la question environnementale, notamment Livesey et Kearins (2002, pp. 246-247), qui décrivent la façon dont la formation d'une petite communauté discursive d'acteurs puissants réussit à imposer une vision du *reporting* RSE. Le cas est illustré par l'exemple d'une « alliance discursive » mutuellement profitable nouée entre l'entreprise Shell et John Elkington<sup>2</sup>. De la même façon, il a été prédit (Livesey, 2002) que les discours du développement durable, fréquemment amalgamés avec la notion de RSE, seraient marqués par des discours scientifiques, techniques et normatifs. De fait, Bernard (2015, pp. 216-217) constate ainsi que les rapports RSE des grandes entreprises sud-africaines sont fortement influencés par une grille normative, en particulier celle du GRI (Global Reporting Initiative). Loconto *et al.* (2020) remarquent que le modèle de durabilité agricole favorisé par les entreprises du secteur est celui qui est le plus facilement mesurable, tandis qu'il a été montré que des entreprises pétrolières telles que Shell (Livesey, 2001), BP (Fuoli, 2012) ou Total (Garric *et al.*, 2007), suscitant naturellement la méfiance de l'opinion publique, se créent une identité discursive d'experts techniques, dans le sillage d'une tendance à l'utilisation de la langue spécialisée à des fins rhétoriques (Fairclough, 1992, pp. 173-174). Dans cette perspective, l'utilisation répétée d'anglicismes participe d'une telle stratégie et peut servir des objectifs de *greenwashing* (Caimotto & Molino, 2011). Plusieurs formes de récurrences dans le discours RSE ont également été relevées : pour Garric *et al.* (2007), le recours à l'expression « développement durable » par Total dans son rapport RSE pour l'année 2003 ne vise qu'à produire un effet d'annonce, suggérant ainsi que la multinationale pétrolière se repose sur la seule force argumentative du syntagme, conformément à l'analyse de Krieg-Planque (2010) qui voit dans cette formule un « opérateur de neutralisation de la conflictualité ». Le discours RSE est également fréquemment traversé par des constructions figées caractéristiques fonctionnant comme des « métaphores d'intégration » qui véhiculent l'image d'une adoption naturelle de la RSE par les entreprises et d'une compatibilité totale de celle-ci avec leurs impératifs économiques (Catenaccio, 2012, pp. 111-132). De la même façon, on peut y retrouver des patrons syntaxiques traduisant une modalité déontico-volitive, qui servent principalement à construire l'ethos de l'entreprise tout en ne l'engageant qu'au minimum (Bondi, 2016, pp. 71-75).

Tous ces travaux, toutefois, concernent des grandes entreprises, voire des firmes multinationales. Si la RSE des petites et moyennes entreprises (dorénavant PME) fait l'objet

<sup>2</sup> John Elkington a popularisé la notion de « trois piliers du développement durable » en introduisant son concept de *Triple Bottom Line* (dorénavant TBL).

de recherches en études de gestion, par exemple sur la perception qu'elles en ont (Jenkins, 2006 ; Murillo & Lozano, 2006), il n'existe pas, à notre connaissance, de travaux en analyse du discours sur ce sujet. Or, les PME représentent la grande majorité des entreprises de chaque pays : elles constituent en effet 99,9% des entreprises britanniques en 2004 par exemple, selon Jenkins (2006, p. 242), faisant d'elles des « actrices invisibles » mais nombreuses, dont l'appréhension de la question environnementale ne devrait pas être ignorée. Il est également pertinent de s'interroger sur les pratiques (ici discursives) des organisations d'Europe du Nord, puisque la région est réputée pour ses performances en matière de RSE (Strand *et al.*, 2015).

Dans cet article, nous nous intéressons aux communications des entreprises scandinaves, danoises et suédoises en l'occurrence, sur leur site Internet, et des récurrences autour de la notion d'environnement. Pour cela, nous mobilisons un logiciel de textométrie, le Trameur, afin de repérer les co-occurrences autour des formes lemmatisées des termes *miljø/miljö* 'environnement', en danois et suédois respectivement, ainsi que des mots composés à partir de ces mêmes formes : la morphologie des langues scandinaves permet la création de nombreux mots composés, dont beaucoup représentent des hapax, ou à tout le moins de formes rares, dans les corpus. Muller ne s'y trompait pas quand il mentionnait, au sujet des mots composés, « l'effrayante malléabilité du lexique germanique » (Muller, 1968, p. 144). Ses propos portaient alors sur l'incertitude relative à la nature du mot composé, formé « en langue » ou « en discours ». Il est vrai que de telles formes sont susceptibles de poser des problèmes à l'analyste ; en revanche, elles ouvrent aussi la voie à des calculs co-occurrenceux intéressants, notamment quand on considère que ces formes complexes, en tant que telles, peuvent s'interpréter comme des segments répétés<sup>3</sup>. Rechercher leurs co-occurents permet ainsi d'analyser comment des thématiques s'articulent autour d'unités de sens figées. Grâce à la forme très lâche des co-occurrences, et en amalgamant les formes rares, voire hapaxiques sous une forme lemmatisée unique, nous essayons de mettre en évidence des récurrences du discours.

### 3. Constitution du corpus et méthodologie d'analyse

Nous faisons porter notre analyse sur deux corpus constitués à partir d'URLs de sites Internet de PME danoises et suédoises. Nous n'avons pas connaissance de travaux précédents en textométrie qui reposeraient sur des corpus constitués à partir du contenu textuel de pages Web d'entreprises, et proposons à l'occasion de cette section quelques réflexions d'ordre méthodologique.

<sup>3</sup> Nous remercions Dominique Longrée qui nous a suggéré cette perspective au cours d'une conversation.

Deux dimensions doivent être prises en compte : l'utilisation du Web comme ressource textuelle, et la démarche épistémologique animant la recherche. Pour la première, deux approches du Web se dessinent : celle du *Web for corpus*, où « la toile » est perçue comme une source de données exploitables pour créer des corpus, et celle du *Web as corpus*, où c'est l'entièreté du contenu d'Internet qui représente un corpus potentiel (De Schryver, 2002, p. 276). Barats *et al.* (2016, pp. 107-114), qui reprennent cette distinction et soulèvent, pour le premier cas, un certain nombre de problématiques épistémologiques et méthodologiques relatives à la constitution de corpus à partir du Web, n'abordent pas spécifiquement la question de l'extraction de contenu textuel pertinent, qui se pose inmanquablement dès lors que l'on s'intéresse à des sites d'entreprise. Il est alors nécessaire de s'interroger sur la démarche épistémologique : l'extraction automatique de contenu est une problématique du traitement automatique du langage (dorénavant TAL), grand consommateur de corpus gigantesques créés à partir d'Internet ; toutefois, le fait que son épistémologie soit intrinsèquement différente de celle de l'analyse du discours<sup>4</sup> rend ses différentes méthodes inadaptées à la textométrie. Les problématiques que posent l'extraction automatisée de contenu textuel pertinent sont pourtant les mêmes en TAL que dans notre cas : les deux principaux obstacles que sont le *boilerplate* (le bruit généré par les différentes rubriques, bandeaux, etc. présents sur chaque page) et les phénomènes de doublons lors de l'extraction sont au centre de la thèse de Pomikálek (2011) et s'appliquent ici aussi. Pour ce qui est du nettoyage du *boilerplate*, les différentes méthodes automatisées produisent des résultats de qualité variable, affectés notamment par la langue dans laquelle sont rédigés les textes (Barbaresi & Lejeune, 2020). Les tentatives que nous avons effectuées avec la version de démonstration en ligne de JusText<sup>5</sup>, l'algorithme développé par Pomikálek (2011), ont produit des résultats jugés insatisfaisants au regard des impératifs de la textométrie pour l'analyse du discours : à savoir, disposer de textes certainement moins nombreux mais qui ne soient pas amputés d'un ou plusieurs paragraphes, afin de pouvoir effectuer une analyse qui ne soit pas linguistique, mais discursive.

Dès lors, nous avons pris le parti d'utiliser un outil dédié, semi-automatisé, pour ne conserver que le « contenu pertinent » des pages : le Détricoteur, un outil d'extraction du contenu textuel des pages Internet. À partir d'une liste d'URLs générée au préalable, il ouvre dans une fenêtre la première URL de la liste dont il reproduit le contenu textuel, accompagné des balises HTML qui le structurent (normalement masquées pour les utilisateurs lors de la

<sup>4</sup> Laquelle est une finalité des premiers développements de la lexicométrie, « discipline auxiliaire de l'analyse du discours » (Maingueneau, 1996, p.53), et de ses appellations ultérieures.

<sup>5</sup> <https://nlp.fi.muni.cz/projects/justext/>

navigation). Depuis cette interface, l'analyste a la possibilité de constituer une liste d'exclusion en sélectionnant les balises qui doivent être ignorées lors de l'extraction. Ainsi, le contenu textuel régi par les balises HTML spécifiquement exclues ne sera pas extrait. Une fois la liste d'exclusion générée, elle s'appliquera à toutes les URLs de la liste, dont le contenu textuel sera extrait selon ces modalités (une liste des règles d'exclusion automatiquement générée par le programme permet une reproduction de la démarche). Dans la mesure où chaque site Internet connaît une architecture HTML qui lui est propre, la procédure doit être adaptée pour chaque nom de domaine.

Nous nous intéressons à la communication **volontaire** des entreprises en matière d'actions environnementales et sociétales : concrètement, il s'agit de travailler sur des entreprises qui ne sont pas soumises à la directive européenne 2014/95/UE (qui oblige les entreprises d'une certaine taille à communiquer en matière de RSE), afin d'analyser l'institutionnalisation de ces types de discours au sein d'entités bien plus petites, et, éventuellement, moins mondialisées. La liste des sites Internet candidats a été constituée à partir de bases de données spécifiques<sup>6</sup> afin de ne retenir que des PME. Pour chacun des sites, une liste d'URLs constitutives a été générée au moyen du programme Xenu's Link Sleuth<sup>7</sup>. Chacune de ces listes a été soumise au Détricotier. Toutes les pages extraites ont fait l'objet d'un premier nettoyage : seules les pages contenant au moins 100 mots ont été retenues. Ce seuil, fixé de manière arbitraire, permet notamment d'effectuer un « pré-nettoyage » de grande ampleur<sup>8</sup>. Une sélection manuelle des pages au contenu environnemental et/ou sociétal a ensuite été opérée<sup>9</sup>.

Afin d'identifier formellement les doublons apparus lors de l'extraction<sup>10</sup>, nous avons eu recours à l'indice de Jaccard, par lequel le niveau de similitude entre deux ensembles est

<sup>6</sup> En l'occurrence les bases de données de Bureau van Dijk (BvD), société travaillant dans la collecte d'informations financières d'entreprises.

<sup>7</sup> Xenu's Link Sleuth est un programme originellement conçu pour vérifier la présence de liens cassés au sein d'un nom de domaine. Nous l'avons utilisé pour sa capacité à générer une liste d'URLs valides à partir d'un nom de domaine, étape essentielle à la suite de notre processus d'extraction par le Détricotier. Il faut toutefois préciser que, compte tenu du caractère extrêmement volatil du contenu sur Internet, la liste de pages générées et leur contenu ne sont valides qu'au moment de leur extraction, et sujettes à changement par la suite. Les URLs constitutives de notre corpus ont été extraites entre le 16 et le 26 mai 2020.

<sup>8</sup> Beaucoup de pages après passage dans le Détricotier ne contenaient que peu, voire pas de mots, par exemple des pages ne contenant que des images (qui ne sont pas extraites par le Détricotier).

<sup>9</sup> Les pages au contenu uniquement social et / ou économique, deux autres piliers de la RSE, n'ont pas été incluses dans le corpus.

<sup>10</sup> Ce *contenu dupliqué interne* peut connaître plusieurs origines : les sites miroirs d'autres sites, les styles d'affichage multiples d'un même site ou encore la reprise à l'identique d'actualités entre différentes plateformes par exemple (Pomikálek, 2011, p. : 55).

mesuré à l'aide du rapport entre leur intersection et leur union<sup>11</sup>. L'indice de Jaccard est fréquemment utilisé en statistique textuelle (Brunet, 2004). Cet indice se calcule très facilement pour chacun des textes dès lors que l'on dispose du « tableau lexical entier » des textes d'un corpus au format tableur, ce que Lexico 5<sup>12</sup> permet de faire par exemple. Le résultat du calcul de l'indice de Jaccard est compris entre 0 (absence totale de formes graphiques partagées) et 1 (identité parfaite entre deux textes). Le calcul a été opéré chaque fois que deux textes paraissaient être des doublons : lorsque l'indice était supérieur à 0,9 (la valeur a été fixée arbitrairement), les (quasi-)doublons étaient supprimés. Enfin, les chercheurs en textométrie recommandent que les textes des corpus étudiés soient compris dans un rapport de grandeur allant de 1 à 10 au maximum (Leblanc, 2016, p. 54) pour permettre la validité de certaines mesures statistiques. Avec un nombre de mots minimal fixé à 100, seules les URLs de moins de 1000 mots peuvent être retenues. Le spectre ainsi obtenu est cohérent avec la réalité des pages à caractère sociétal ou environnemental, de quelques centaines de mots en général.

Le nombre d'entreprises retenues pour l'analyse est de 31, avec un nombre maximal de trois URLs par entreprise (une sélection aléatoire étant opérée dès lors qu'une entreprise contenait plus de trois URLs). Les détails des deux corpus sont les suivants (décompte Le Trameur) :

Danemark : 31 sites, 59 URLs, 19 614 formes (« mots ») au total

Suède : 31 sites, 59 URLs, 20 359 formes au total

### 3.1. Norme de dépouillement

Une norme de dépouillement préalable a été appliquée au corpus. On entend par « norme de dépouillement » « une exigence de standardisation provisoire des textes contenus dans un corpus. Cette standardisation est destinée avant tout à les rendre comparables, à les stabiliser le temps d'une expérience » (Habert *et al.*, 1997, p. 187). En d'autres termes, il s'agit de déterminer une norme d'harmonisation des pratiques orthographiques ou typographiques au sein du corpus. La procédure est nécessairement arbitraire, et apporte son lot d'insatisfactions. Il y a un demi-siècle déjà, Muller prévenait : « Le caractère le plus assuré d'une norme, c'est

<sup>11</sup> L'indice de Jaccard s'exprime sous la forme  $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ , où  $|A \cap B|$  désigne l'ensemble formé par l'intersection de A et de B, et  $|A \cup B|$  l'union des ensembles A et B. Ainsi,  $|A \cup B|$  s'exprime aussi comme la somme des ensembles A et B dont on soustrait l'intersection, soit  $|A \cup B| = |A| + |B| - |A \cap B|$ , et le calcul de l'indice de Jaccard peut donc se réécrire sous la forme :  $J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$ .

<sup>12</sup> Lexico 5 est un logiciel de textométrie développé par André Salem, téléchargeable à cette adresse : <https://lexico.com/>.



de ne satisfaire personne, à commencer par son auteur » (Muller, 1968, p. 144). Les choix opérés ont un coût statistique, et aucun n'est neutre : comme l'écrivent Lebart et Salem (1994, p. 34), « Pour un même texte, les différentes normes de dépouillement ne conduisent pas au[x] mêmes décomptes. ». Les remarques sur les normes de dépouillement abondent dans la littérature, et, si les progrès techniques ont permis de clore certains débats (notamment la question de la lemmatisation), d'autres continuent d'être vivaces. Il serait fastidieux de faire une liste des questions que notre cas particulier a suscité, mais deux d'entre elles peuvent être relevées à titre d'illustration : la première est la présence, dans certains textes, d'adresses e-mail (exemple@exemple.com) ou adresses Internet (www.exemple.com), qui doivent être modifiées car le point qu'elles contiennent est identifié comme un délimiteur par les programmes. La seconde touche aux listes à puces, particulièrement fréquentes et susceptibles de poser de nombreux problèmes, notamment lorsque les puces sont constituées de deux phrases ou plus, séparées par des caractères délimiteurs classiques (le point, le point d'exclamation ou d'interrogation...) qui redessinent de nouvelles frontières de la phrase avec les puces précédentes.

### 3.2. Méthodologie d'analyse des co-occurrences

La co-occurrence peut se définir comme « la co-présence régulière de deux unités linguistiques dans une fenêtre contextuelle choisie » (Poudat & Landragin, 2017, p. 201). La définition de la co-occurrence et son positionnement vis-à-vis de la collocation fait l'objet d'attitudes diverses de la part des chercheurs, et son acception varie d'ailleurs entre les mondes anglo-saxon et francophone (Martinez, 2003, pp. 119-123, Mayaffre, 2008, p. 811). Pour distinguer entre collocation et co-occurrence, Martinez propose de définir la seconde négativement par rapport à la première, c'est-à-dire, de faire l'inventaire de ce qu'elle n'est pas : « la co-occurrence est une combinaison lexicale qui n'est pas orientée, qui associe des éléments dont le nombre pas limité à deux, qui ne sont pas obligatoirement contigus et qui n'entretiennent pas nécessairement une relation d'ordre sémantique ou syntaxique » (Martinez, 2003, p. 123). La recherche de co-occurrences permet de « jouer en même temps sur la séquentialité et sur la répétitivité du discours sans se limiter à une successivité stricte des formes » : elle offre en outre de faire porter l'analyse sur « une unité plus lâche que le segment répété » (Née *et al.*, 2017, p. 112 pour ces deux citations). Ainsi, si l'on considère que les mots composés des langues germaniques (en l'occurrence du danois et du suédois) peuvent se comprendre comme des segments répétés, le fait de travailler sur des co-occurrences de segments répétés eux-mêmes particulièrement flexibles (du fait de la méthode retenue pour l'étiquetage des mots composés,

voir ci-dessous) permet de saisir le discours dans ses contours les plus généraux. Il s'agit de révéler les enjeux phraséologiques, discursifs et thématiques qui habitent le discours environnemental, quand celui-ci se produit sur une plateforme ouverte au lectorat le plus large possible et émane d'organisations qui, sans aucun doute, n'ont pas les mêmes moyens que les grandes entreprises en matière de professionnalisation de leur communication.

Le programme que nous avons utilisé est le Trameur, qui permet un étiquetage automatique du danois et du suédois (via le module TreeTagger), et propose en outre de modifier, exporter et réimporter cet étiquetage. L'étiquetage n'est toutefois pas infallible, comme le souligne Pomikálek (2011, p. 12), et échoue notamment, dans le cas des langues scandinaves, à étiqueter correctement nombre de mots composés.

Pour chacun des corpus, nous avons procédé en cinq étapes, chacune divisée en deux opérations identiques : le calcul de co-occurents suivi d'un retour au texte à l'aide du concordancier du Trameur. Tous les calculs de co-occurrences ont été menés avec un indice de spécificité minimum fixé à 5 (une probabilité de 1 sur 100 000 (Poudat & Landragin, 2017, p. 172)), et une fenêtre contextuelle limitée à la phrase (avec le point, point d'exclamation et point d'interrogation comme caractères délimiteurs). La première étape a consisté en la recherche des co-occurents des formes lemmatisées de *miljø* (DK) / *miljö* (SV) afin de repérer les associations thématiques ainsi que les figements éventuels.

Les quatre étapes suivantes consistent en le réétiquetage des formes lemmatisées des mots formés de *miljø-/miljö-* sous une forme commune (*MILJØX/MILJÖX*) à différents niveaux d'occurrences dans le corpus : cinq occurrences ou moins, dix occurrences ou moins, toutes les occurrences, et plus de dix occurrences. Pour le décompte du nombre d'occurrences d'une forme donnée, nous avons choisi de ramener sous un lemme unique toutes les formes fléchies<sup>13</sup> d'un même mot formé de *miljø-/miljö-*, mais également ses formes dérivées. Ainsi, si *miljøcertificeret* ('certifié environnementalement') est exclu de la seconde étape car il connaît six occurrences dans le corpus danois, c'est également le cas de sa forme dérivée *miljøcertificering* ('certification environnementale', deux occurrences) et des formes fléchies de cette dernière, *miljøcertificering[-er/-erne]* (formes du pluriel indéfini et défini respectivement, une occurrence pour chacune). En revanche, si *miljøledelse* ('gestion environnementale') est exclu de la deuxième étape à cause de ses six occurrences, les formes *miljøledelsessystem[-er]* ('système[s] de gestion environnementale', formes indéfinies du singulier et du pluriel

<sup>13</sup> La lemmatisation doit normalement neutraliser la flexion, mais les mots composés (en danois et suédois au moins) sont typiquement mal identifiés par TreeTagger, ce qui implique que de nombreuses formes fléchies d'un même mot sont comptabilisées comme autant de lemmes.

respectivement, deux et une occurrences dans le corpus) sont incluses dans cette seconde étape car elles constituent un nouveau « mot ».

#### 4. Résultats

Dans le sillage de Krieg-Planque (2015, p. 117), nous comprenons le figement comme « une occasion de resserrement de la créativité verbale, le principe d'imprévisibilité inhérent à l'activité de langage (récursivité, commutabilité, ouverture de la combinatoire...) se trouvant alors restreint ». Si on considère que les mots composés représentent finalement des segments répétés, il apparaît qu'en s'intéressant à leurs co-occurents, on regarde finalement le voisinage de ces mêmes segments, dont peuvent se déduire des interprétations qui dépassent le seul cadre linguistique : « le figement et le défigement ne sont pas uniquement des faits de langue, linguistiquement descriptibles, mais aussi des pratiques sociales que des acteurs prennent en charge, dans des logiques organisationnelles ou politiques » (Krieg-Planque, 2015, p. 117).

Nous présentons tour à tour les résultats pour les deux corpus. Nous ne nous intéressons qu'aux co-occurents que nous qualifions de « saillants », c'est-à-dire, qui connaissent un indice de spécificité supérieur ou égal à cinq, et se produisent chez cinq entreprises au moins (et non cinq URLs), et qui sont pertinents pour l'analyse. Ces seuils ont été fixés arbitrairement.

Pour plus de facilité et de lisibilité, nous reproduisons dans le Tableau 1 ci-dessous les résultats tels qu'ils apparaissent dans le Trameur. Les chiffres indiquent, dans l'ordre, l'indice de spécificité, le nombre de contextes (dans le cas présent, des phrases), le nombre d'URLs et le nombre d'entreprises concernés.

	DANEMARK	SUÈDE
<b>Miljø / miljö</b>	<i>fælles</i> (‘commun’) 6 ; 6 ; 5 ; 5	<i>arbetsmiljö</i> (‘environnement de travail’) 7 ; 12 ; 8 ; 7
	<i>mulig</i> (‘possible’) 5 ; 9 ; 8 ; 8	<i>kvalitet</i> (‘qualité’) 18 ; 35 ; 21 ; 20
<b>Formes composées de miljø- / miljö- Cinq occurrences ou moins</b>	-	<i>leverantör</i> (‘fournisseur’) 5 ; 10 ; 9 ; 7
	-	<i>kontinuerlig</i> (‘continuuel’) 5 ; 6 ; 6 ; 6
<b>Formes composées de miljø- / miljö- Dix occurrences ou moins</b>	<i>iso</i> 10 ; 17 ; 14 ; 11	<i>leverantör</i> (‘fournisseur’) 5 ; 13 ; 11 ; 9
	<i>14001</i> 10 ; 15 ; 12 ; 10	
<b>Formes composées de miljø- / miljö- Toutes les occurrences</b>	<i>iso</i> 9 ; 18 ; 14 ; 11	<i>leverantör</i> (‘fournisseur’) 5 ; 17 ; 14 ; 12
	<i>14001</i> 9 ; 16 ; 12 ; 10	<i>möjlig</i> (‘possible’) 7 ; 17 ; 14 ; 12
		<i>minimera</i> (‘minimiser’) 5 ; 13 ; 12 ; 9
		<i>kvalitets-</i> (ellipse d’une forme composée de ‘qualité’) 16 ; 20 ; 11 ; 10
<b>Formes composées de miljø- / miljö- Plus de dix occurrences</b>	<i>social</i> (‘social ») 5 ; 8 ; 7 ; 7	<i>aktiv</i> (‘actif’) 5 ; 8 ; 7 ; 6
	<i>økonomisk</i> (‘économique’) 6 ; 5 ; 5 ; 5	<i>negativ</i> (‘négatif’) 5 ; 5 ; 5 ; 5

Tableau 1. Co-occurents saillants pertinents de *miljø / miljö*, et de leurs formes composées connaissant cinq occurrences ou moins, dix occurrences ou moins, à toutes les occurrences, et plus de dix occurrences<sup>14</sup>.

#### 4.1. Résultats pour le Danemark

Le corpus danois présente peu d’instances de récurrences autour de la notion d’environnement, que celles-ci prennent la forme de figements (ou même de semi-figements) « formels » ou plus « thématiques » – c’est-à-dire qui ne se limiteraient pas à la « successivité stricte des formes »,

<sup>14</sup> Lecture : *arbetsmiljö*, que l’on peut traduire par ‘environnement de travail’ en français, co-occure avec la forme lemmatisée *miljö* dans le corpus suédois avec un indice de spécificité de 7. Cette co-occurrence se retrouve dans 12 contextes (ici 12 phrases), réparties dans 8 URLs différentes rassemblant 7 entreprises

pour reprendre la citation de Née *et al.*, 2017, p. 112, citée supra. Avec *miljø*, deux co-occurents pertinents peuvent être relevés. Le premier, *fælles* (« commun, partagé »), prend dans quatre cas sur cinq la forme d'un figement formel (*vores fælles miljø*, « notre environnement à tous »). Ce figement peut être interprété comme une référence intertextuelle au rapport Brundtland, qui popularisa la notion de « développement durable » en 1987 intitulé *Our common future* et dont le titre danois est *Vores fælles fremtid*. Le second co-occurent pertinent de *miljø* ne donne pas lieu à des figements stricts. Il s'agit de *mulig* (« possible »), dont les réalisations en contextes produisent dans six contextes sur neuf des récurrences syntaxiques très lâches, évoquant l'idée du « moins/mieux possible » ou du « autant que possible » mises en relation avec l'environnement :

- (1) [...] *der påvirker miljø og helbred mindst muligt*<sup>15</sup>  
 [...] *der påvirke-r miljø og helbred mindst muligt-t.*  
 qui affecter-PRS environnement et santé moins.SUP possible-ADV  
 'qui affectent le moins possible l'environnement et la santé.' (k-a-1)
- (2) [...] *så bæredygtigt som muligt for det lokale miljø*  
 [...] *så bæredygtig-t som mulig-t for det lokal-e miljø.*  
 aussi durable-ADV que possible-N pour DET local-DEF environnement  
 'de façon aussi durable que possible pour l'environnement local.' (scanomat-1)

De façon notable, et contrairement au corpus suédois, les formes lemmatisées des mots formés de *miljø*- qui connaissent cinq occurrences ou moins dans le corpus ne se réunissent autour d'aucun co-occurent saillant.

Ce sont les calculs opérés sur les mots composés connaissant dix occurrences ou moins qui donnent à voir une récurrence thématique : il s'agit des formes *iso* et *14001*, qui co-occurent 17 et 15 fois avec une forme composée de *miljø*- respectivement. *14001* est alors toujours précédé de *iso*, pour former « iso 14001 », une norme dédiée à la gestion environnementale. Cette co-occurrence de formes composées de *miljø*- et de *iso 14001* se réalise chez 10 entreprises, soit un tiers des entreprises du corpus. Dans le détail, les formes de *miljø*- qui co-occurent avec *iso (14001)* présentent relativement peu de variété et y renvoient pour faire référence à sa dimension certifiante (*certificering / certificeret*) ou à l'idée de « gestion » (*ledelse / styring*) environnementale qu'elle implique (le total du nombre de formes composées de *miljø*- (18) dépasse les 17 co-occurrences avec *iso* en raison d'un contexte où deux formes composées cohabitent dans la même phrase) :

<sup>15</sup> Tous les exemples ont été glosés (Leipzig Glossing Rules) afin de faciliter la compréhension des lecteurs qui ne seraient pas familiers du danois ou du suédois.

Mot formé de <i>miljø-</i>	Nombre d'occurrences
<i>Miljøcertificering</i> / <i>miljøcertificeret</i> ('certification environnementale' / 'certifié environnementalement')	6 occurrences
<i>Miljøledelse</i> 'gestion environnementale'	3 occurrences
<i>Miljøledelsessystem(-er)</i> 'système(s) de gestion environnementale'	2 occurrences
<i>Miljøstyringssystem</i> 'système de gestion environnementale'	4 occurrences
<i>Miljøstyring</i> 'gestion environnementale'	1 occurrence
<i>Miljøstandard</i> 'norme environnementale'	1 occurrence
<i>Miljøpræstation</i> 'résultat environnemental'	1 occurrence

Tableau 2. Détail des formes composées de *miljø-* connaissant moins de dix occurrences dans le corpus qui co-occurrent avec (*iso*) 14001.

Les retours au texte montrent un lien direct entre la dimension environnementale et la norme en question, par exemple :

(3) ***bj er miljøcertificeret i henhold til iso 14001***

*bj er miljø-certificere-t i.henhold.til iso 14001*  
 bj être.PRS environnement-certifier-PRF selon iso 14001  
 'bj dispose d'une certification environnementale conformément à la norme ISO 14001'  
 (bjgear-2)

(4) ***vores miljøledelsessystem er baseret på iso 14001***

*vores miljø-ledelse-s-system er basere-t på iso 14001*  
 notre environnement-gestion-INTF-système être.PRS reposer-PRF sur iso 14001  
 'notre système de gestion environnementale repose sur la norme ISO 14001' (ceas-1)

Si ces co-occurents ne donnent pas à voir de figements discursifs, ils n'en présentent pas moins un intérêt en matière d'analyse du discours : en adoptant une approche sociologique du discours par la théorie institutionnelle, on envisage dans la prévalence du recours à la norme ISO 14001 une forme d'isomorphisme susceptible de découler de pressions normatives, voire de processus mimétiques (DiMaggio & Powell, 1983) : la convergence des pratiques des entreprises pourrait alors traduire, dans le premier cas, une uniformisation de la formation des responsables de politiques RSE, qui finalement adoptent les mêmes pratiques et deviennent interchangeables. Dans le second, il pourrait s'agir d'une stratégie de réduction des incertitudes par l'imitation. Les processus d'isomorphismes, écrivent Meyer et Rowan, ont plusieurs conséquences importantes pour les organisations, parmi lesquelles celle-ci : ils les poussent à incorporer des éléments « which are legitimated externally, rather than in terms of efficiency » (Meyer & Rowan, 1977, p. 348). Cet article n'a pas pour objectif d'explorer

davantage les raisons d'une adoption aussi large de la norme ISO 14001 par les PME danoises, mais la question présente un intérêt afin de comprendre comment les petites et moyennes entreprises envisagent leur démarche environnementale. En dehors de l'aspect purement institutionnel, la référence à un système de normes externe, indépendant et reconnu représente également un argument d'autorité, au sens de Perelman et Olbrechts-Tyteca (2008, pp. 410-417).

L'analyse des co-occurents de toutes les formes composées de *miljø-* ne donne pas davantage de pistes : « iso » et « 14001 » continuent d'être les seuls co-occurents pertinents, et, s'ils gagnent chacun un contexte, ils perdent tous les deux en spécificité : probablement justement parce qu'ils ne gagnent qu'un seul contexte au regard de la quantité de nouvelles occurrences qui représentent autant de candidats potentiels à de nouvelles co-occurrences.

Il y a trois formes composées de *miljø-* qui, une fois ramenées à une forme lemmatisée, connaissent plus de 10 occurrences dans le corpus. Avec 17, 16 et 13 occurrences respectivement, ce sont : *miljørigtig* 'bien pour l'environnement', *miljøvenlig* 'respectueux de l'environnement', et *miljømæssig*, dont le suffixe, *-mæssig*, ne peut avoir aucune existence autonome, mais exprime l'idée de relation avec le radical auquel il se raccroche. La recherche de co-occurents sur le rassemblement de ces trois formes sous une forme graphique unique fait ressortir deux co-occurents pertinents : *social* 'social' et *økonomisk* 'économique'. Un retour au texte montre que ces deux adjectifs ne co-occurrent en réalité qu'avec *miljømæssig*, chez huit entreprises au total (neuf contextes). *Miljømæssig* apparaît ainsi avec ces deux co-occurents simultanément (quatre contextes, quatre entreprises) ou un seul (cinq contextes, quatre entreprises) :

(5) ***vi tænker social, økonomisk og miljømæssig bæredygtighed***

*vi tænke-r social, økonomisk og miljømæssig bæredygtig-hed*

1PL penser-PRS social économique et environnemental durable-NMLZ

'nous intégrons les notions de durabilité sociale, économique et environnementale' (ekj-1)

(6) ***i plougmann vingtoft arbejder vi på at implementere miljømæssigt og socialt ansvarlige forretningsmetoder***

*i plougmann vingtoft arbejde-r vi på at implementere*

chez plougmann vingtoft travailler-PRS 1PL à INF mettre en œuvre

*miljømæssig-t og social-t ansvarlig-e forretning-s-metod-er*

environnemental-ADV et social-ADV responsable-PL affaire-INTF-méthode-PL

'chez plougmann vingtoft, nous travaillons à la mise en œuvre de procédures environnementalement et socialement responsables...' (pv-1)

Il y a donc dans le discours un (semi-)figement qui associe les notions d'économie, de société et d'environnement. La co-occurrence de ces trois notions n'est pas le fruit du hasard : il s'agit en fait d'une référence aux trois piliers de la RSE, voire, du développement durable, qui gagna en popularité dans les années 1990 après l'introduction par John Elkington de la TBL consacrant l'analyse combinée des performances économiques, environnementales et sociétales dans le *reporting* des entreprises (Carroll, 2015, pp. 92-93). Il a été avancé que la TBL, qui a bénéficié du soutien d'acteurs institutionnels puissants comme le Global Reporting Initiative, a pu former une communauté discursive avec un noyau dur d'acteurs influents, contribuant ainsi à légitimer et diffuser des perceptions correspondant à des intérêts particuliers et susceptibles d'imposer une forme d'hégémonie discursive (Livesey & Kearins, 2002 ; Livesey, 2002). Dans le sillage de Krieg-Planque (2015), qui suggère que les figements sous la forme de références intertextuelles puissent être corrélés à une forme de reconnaissance de l'autorité, les renvois des PME danoises à une notion popularisée en premier lieu au sein des grandes entreprises pourrait traduire une maîtrise poussée des enjeux de la RSE classique, donc une dilution progressive de ses principes vers les petites entreprises. Il ne s'agit toutefois pas de tirer des conclusions hâtives : il n'est ainsi nulle part fait mention de la TBL dans le corpus danois. Néanmoins, la récurrence de telles formes peut indiquer un fort conformisme normatif dans leur expression de la problématique.

## 4.2. Résultats pour la Suède

Dans le cas de la Suède, le calcul de co-occurrences sur la forme lemmatisée de *miljö* fait ressortir deux co-occurents saillants : *arbetsmiljö* ('environnement de travail') chez sept entreprises avec un indice de spécificité de sept, et surtout *kvalitet* ('qualité') chez 20 entreprises, qui co-occure avec un indice de spécificité élevé de 18. Ces deux co-occurents apparaissent fréquemment ensemble, avec des constructions comme :

- (7) *policy för miljö, arbetsmiljö, trafiksäkerhet och kvalitet*  
*policy för miljö, arbet-s-miljö, trafik-säker-het och kvalitet.*  
 politique pour environnement travail-INTF-environnement trafic-sûr-NMLZ et qualité  
 'politique en matière d'environnement, environnement de travail, sécurité routière et qualité...' (aktab-5)

Le figement *kvalitet och miljö* est majoritaire avec 14 occurrences dans le corpus chez 11 entreprises différentes :



(8) *varje medlem får en utbildning i kvalitet och miljö*

varje medlem få-r en utbildningi kvalitet och miljö  
 chaque membre recevoir-PRS INDEF formation en qualité et environnement  
 ‘chaque membre reçoit une formation qualité et environnement’ (dipart-28)

Ces figements se retrouvent parfois (quatre fois) sous la forme de titres de rubriques, qui tendent à montrer l’intégration de la notion par les entreprises : la notion de *kvalitet och miljö* ou *kvalitet, miljö och arbetsmiljö* est l’équivalent suédois de la QSE / QHSE (Qualité, (Hygiène), Sécurité, Environnement), un ensemble de procédures relatives à la prévention des risques.

Il est significatif que la forme *kvalitets*, dérivée de *kvalitet* (‘qualité’), soit le co-occurent principal des formes lemmatisées des mots formés de *miljö-* pour trois des quatre seuils : lorsque le nombre d’occurrences est inférieur ou égal à cinq (dans ce cas toutefois, le nombre d’entreprises concernées est inférieur à cinq), inférieur ou égal à dix, ou toutes les formes. Même en s’intéressant aux formes qui connaissent un nombre d’occurrences strictement supérieur à dix, il est le deuxième co-occurent le plus fort. Cet état de fait illustre une difficulté qui n’est pas sans conséquences pour l’exactitude des mesures : dans le cas d’une succession de mots composés, la ou les première(s) expressions de la série s’écrit (s’écrivent) en ne conservant que le premier mot de la composition, ce qui produit des formes elliptiques, susceptibles d’isoler des mots qui ne sont pas pris en compte lors des calculs. Par exemple avec des constructions comme *kvalitets- och miljöpolicy*, où *policy* s’applique à *kvalitet* aussi bien qu’à *miljö* mais est éludé pour éviter la répétition. Comme l’ordre de la séquence n’est pas imposé, on retrouve également des constructions dans le sens opposé, sous la forme *miljö- och kvalitetspolicy*, ce qui tend à fausser le nombre d’occurrences prises en compte par le programme.

Pour les mots formés de *miljö-* qui connaissent moins de 5 occurrences, deux co-occurents saillants et pertinents peuvent être relevés : en premier lieu *leverantör* (‘fournisseur’), dix occurrences, sept entreprises), qui par ailleurs apparaissent avec *krav* (‘exigence’) ou une séquence composée de cette dernière forme, à six reprises, chez cinq entreprises dans des tournures variées :

(9) [...] *att välja leverantörer som [...] uppfyller våra miljökrav.*

[...] att välja leverantör-er som [...] uppfyll-er vår-a miljö-krav.  
 PRT.INF choisir fournisseur-PL qui remplir-PRS notre-PL environnement-  
 exigence  
 ‘de choisir des fournisseurs qui correspondent à nos exigences environnementales.’  
 (nordic-house-1)



- (13) **vår verksamhet skall bedrivas med minsta möjliga miljöpåverkan**  
 vår verksamhet skall bedriva-s med minsta möjlig-a  
 notre activité devoir.PRS mener-PASS avec moins.SUP-DEF possible-DEF  
 miljö-påverkan  
 environnement.impact  
 ‘notre activité doit être conduite avec l’impact environnemental le plus faible possible...’  
 (aktab-5)
- (14) **vi beaktar bästa möjliga [...] miljöpåverkan vid inköp**  
 vi beakta-r bäst-a möjlig-a [...] miljö-påverkan vid inköp [...]  
 1PL considérer-PRS bien.SUP-DEF possible-DEF environnement-impact lors achat  
 ‘nous rechercherons le meilleur impact environnemental [...] possible lors de nos achats’ (hnt-8)
- (15) **[...] med så liten miljöpåverkan som möjligt.**  
 [...] med så liten miljö-påverkan som möjlig-t.  
 avec aussi petit environnement-impact que possible-NE  
 ‘avec un impact environnemental aussi faible que possible’ (wiréns-åkeri-1)

*Minimera* ‘minimiser, réduire’ est le dernier co-occurent pertinent répondant aux critères fixés. Il connaît un indice de spécificité de 5, se retrouve dans 13 contextes chez 9 entreprises, qui pour 6 d’entre elles expriment la nécessité de « réduire l’impact environnemental » dans une collocation avec *miljö-påverkan* ou *miljöbelastning* :

- (16) **[...] kan vi minimera vår miljöpåverkan**  
 [...] kan vi minimera vår miljö-påverkan  
 pouvoir.PRS 1PL minimiser notre environnement-impact  
 ‘nous pouvons minimiser notre impact environnemental’ (hellbergs-1)
- (17) **för att minimera miljöbelastningen**  
 för-att minimera miljö-belastning-en  
 afin-de minimiser environnement-charge-DEF  
 ‘afin de minimiser la charge sur l’environnement’ (swedhandling-6 ; swedhandling-9)

La notion de « fournisseurs », déjà relevée précédemment, continue d’être un co-occurent spécifique : *leverantör* gagne quatre contextes chez trois nouvelles entreprises, mais ils n’enrichissent pas les figements repérés précédemment.

Si enfin on s’intéresse, comme cela avait été fait pour le corpus danois, uniquement aux formes composées qui connaissent plus de 10 occurrences, deux co-occurents saillants ressortent, qui n’apparaissaient pas jusqu’à présent : *aktiv* ‘actif’ et *negativ* ‘négatif’. Le premier,

avec un indice de spécificité de 5, se retrouve dans 8 contextes chez 6 entreprises, et en collocation avec *miljöarbete* ‘travail environnemental’ chez quatre d’entre elles :

(18) ***förlagssystem bedriver ett aktivt miljöarbete***

*förlagssystem bedriv-er ett aktiv-t miljö-arbete*

förlagssystem mener-PRS INDEF actif-NE environnement-travail

‘förlagssystem mène un travail environnemental actif...’ (förlagssystem-1)

(19) [...] ***i kombination med ett aktivt [...] miljöarbete.***

[...] *i kombination med ett aktiv-t [...] miljö-arbete.*

en combinaison avec INDEF actif-NE environnement-travail

‘combiné à un travail environnemental [...] actif.’ (tykoflex-1)

Dans le sillage de la notion de continuité, de telles constructions donnent l’idée d’un processus continu, qui fait de la question environnementale une préoccupation permanente.

*Negativ* quant à lui, co-occure 5 fois chez 5 entreprises avec un indice de spécificité de 5, et toujours avec *miljöpåverkan*, ce qui souligne la versatilité de cette forme et suggère qu’elle ouvre la voie à de nombreux semi-figements.

(20) ***anpassa tillverkningsprocessen så att negativ miljöpåverkan minimeras.***

*anpassa tillverkning-s-process-en så-att negativ miljö-påverkan*

adapter fabrication-INTF-PROCESSUS-DEF afin-de négatif environnement.impact

*minimera-s.*

minimiser-PASS

‘adapter notre processus de fabrication afin de minimiser tout impact environnemental négatif.’ (swedhandling-9)

Qu’en est-il enfin de *kvalitets-* ? Il s’agit d’un co-occurent majeur des formes composées de *miljö-*, de la même façon que *kvalitet* est un co-occurent de *miljö*. Si la récurrence thématique est indéniable, aucun figement sémantique ferme n’est à trouver de ce côté-ci, puisque les compositions de *miljö-* qui co-occurent avec *kvalitets-* sont variables (10 formes différentes selon les règles fixées pour la lemmatisation, sur 20 co-occurrences de *kvalitets-* avec une forme composée de *miljö-* chez 10 entreprises). Les collocats directs de *kvalitets-* parmi les mots formés de *miljö-* se répartissent ainsi :

Mot formé de <i>miljö-</i>	Nombre d'occurrences
<i>miljösystem(-et)</i> '(le) système environnemental'	4 occurrences
<i>miljöfrågor</i> 'questions environnementales'	3 occurrences
<i>miljöplan(-en)</i> '(le) plan environnemental'	3 occurrences
<i>miljöpolicy</i> 'politique environnementale'	3 occurrences
<i>miljöcertifierade</i> 'certifiés environnementalement'	2 occurrences
<i>miljöarbete</i> 'travail environnemental'	1 occurrence
<i>miljöledningssystem</i> 'système de gestion environnementale'	1 occurrence
<i>miljömål</i> 'objectifs environnementaux'	1 occurrence
<i>miljösamordnare</i> 'co-ordinateur environnemental'	1 occurrence
<i>miljösäkringssystem</i> 'système d'assurance environnementale'	1 occurrence

Tableau 3 : Détail des formes composées de *miljö-* collocats directs de *kvalitets-*.

## 5. Conclusion

La comparaison de deux corpus différents, rédigés dans des langues distinctes, impose une grande prudence, à plus forte raison quand leur taille est aussi petite. Des tendances paraissent néanmoins se dégager, dont la généralisation ne demande qu'à être vérifiée. En particulier, le corpus danois est remarquable par son absence de figements, selon les paramètres retenus, autour des formes composées de *miljø-*, à l'exception de deux références intertextuelles et d'une récurrence syntaxique très lâche autour de la notion de possibilité. Ces références intertextuelles et la prévalence de la question des normes ISO 14001 pourraient traduire un certain conformisme de la perception de la RSE et une maîtrise de ses concepts, articulés autour d'un discours, finalement assez peu générique, dans le voisinage de la notion d'environnement. Cette interprétation, toutefois, reste provisoire. À l'inverse, pour le corpus suédois, le discours présente davantage de régularité et de formes semi-figées, ce qui pourrait traduire un manque de créativité lexicale et une forme de « langue de bois », ou de manque de maîtrise des notions relatives à la RSE. Ces différences ouvrent également des pistes sur l'utilisation des mots composés dans les langues scandinaves – certes morphologiquement courants, mais susceptibles de porter une charge rhétorique : les nombreux co-occurents que

connaît la forme *miljöpåverkan* ‘impact environnemental’, qui introduisent des récurrences avec les notions de possibilité, de minimisation, ou une collocation avec ‘négatif’ laisse supposer que les mots composés (ou segments répétés, si on les interprète ainsi), peuvent avoir une valeur rhétorique intrinsèque, du moins en suédois, ce qui les inscrit pleinement dans une démarche de *greenspeak*. Cette perspective ouvre la voie à des recherches complémentaires, quantitatives au moyen d’autres outils de la textométrie, mais aussi qualitatives : les références intertextuelles sont notamment une composante essentielle du framing de Goffman, comme l’explique Gordon (2015, p. 324), et participent pleinement d’une stratégie d’exercice du pouvoir pouvant chercher à « naturaliser » des normes idéologiques (Fairclough, 2010). De telles analyses appellent à continuer de solliciter la théorie institutionnelle des sociologues, dans l’esprit multidisciplinaire caractéristique de l’analyse du discours.

## Bibliographie

- Antelmi, D. (2018). *Verdi parole. Un’analisi linguistica del discorso “green”*. Mimesis.
- Barats, C., Leblanc, J.-M., & Fiala, P. (2016). Approches textométriques du web : corpus et outils. In C. Barats (Éd.), *Manuel d’analyse du web en Sciences Humaines et Sociales* (chap. 5, pp. 107-132). Armand Colin.
- Barbaresi, A., & Lejeune, G. (2020). Out-of-the-Box and Into the Ditch? Multilingual Evaluation of Generic Text Extraction Tools. *Proceedings of the 12th Web as Corpus Workshop, Language Resources and Evaluation Conference (LREC 2020, 11-16 mai)*, Marseille, 5-13.
- Bernard, T. (2015). *A critical analysis of corporate reports that articulate corporate social responsibility*. Thèse de doctorat, Stellenbosch University, Stellenbosch.
- Bondi, M. (2016). The future in reports. *Pragmatics and Society*, 7(1), 57-81. <https://doi.org/10.1075/ps.7.1.03bon>
- Brunet, E. (2004). Où l’on mesure la distance entre les distances. *Texto ! Textes et Cultures*, 9(2). [http://www.revue-texto.net/Inedits/Brunet/Brunet\\_Distance.html](http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html)
- Caimotto, M.C., & Molino, A. (2011). Anglicisms in Italian as Alerts to Greenwashing: A Case Study. *Critical Approaches to Discourse Analysis Across Disciplines*, 5(1), 1-16.
- Carroll, A. B. (2015). Corporate social responsibility: The centerpiece of competing and complementary frameworks. *Organizational Dynamics*, 44(2), 87-96. <https://doi.org/10.1016/j.orgdyn.2015.02.002>
- Catenaccio, P. (2012). *Understanding CSR Discourse: Insights from Linguistics and Discourse Analysis*. Arcipelago.
- Colyvas, J. A., & Jonsson, S. (2011). Ubiquity and Legitimacy: Disentangling Diffusion and Institutionalization. *Sociological Theory*, 29(1), 27–53. <https://doi.org/10.1111/j.1467-9558.2010.01386.x>

- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147-160. <https://doi.org/10.2307/2095101>
- Directive 2014/95/UE. Modifiant la directive 2013/34/UE en ce qui concerne la publication d'informations non financières et d'informations relatives à la diversité par certaines grandes entreprises et certains groupes. Parlement européen et Conseil de l'Union européenne. <http://data.europa.eu/eli/dir/2014/95/oj>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, 20(1), 29-62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Fairclough, N. (1992). *Discourse and Social Change*. Polity Press.
- Fairclough, N. (2010 [1995]). *Critical Discourse Analysis: The Critical Study of Language* (2<sup>nd</sup> ed.). Routledge.
- Fiala, P., Habert, B., Lafon, P., & Pineira-Tresmontant, C. (1987). Des mots aux syntagmes. Figements et variations dans la Résolution générale du congrès de la CGT de 1978. *Mots. Les langages du politique*, 14, 47-87.
- Fuoli, M. (2012). Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports. *Discourse & Communication*, 6(1), 55-81. <https://doi.org/10.1177/1750481311427788>
- Garric, N., Léglise, I., & Point, S. (2007). Le rapport RSE, outil de légitimation ? Le cas Total à la lumière d'une analyse de discours. *Revue de l'organisation responsable*, 2(1), 5-19. <https://doi.org/10.3917/ror.021.0005>
- Gendron, C. (2000). *Le questionnement éthique et social de l'entreprise dans la littérature managériale*. Centre de recherche sur les innovations sociales.
- Gordon, H. (2015 [2001]). Framing and Positioning. In D. Tannen, H. E. Hamilton, D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (2<sup>nd</sup> ed., p. 324-345). Wiley-Blackwell.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Armand Colin.
- Harré, R., Brockmeier, J., & Mühlhäusler, P. (1999). *Greenspeak : A Study of Environmental Discourse*. Sage Publications.
- Jenkins, H. (2006). Small Business Champions for Corporate Social Responsibility. *Journal of Business Ethics*, 67(3), 241-256.
- Krieg-Planque, A. (2009). *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Presses Universitaires de Franche-Comté.
- Krieg-Planque, A. (2010). La formule « développement durable » : un opérateur de neutralisation de la conflictualité. *Langage et société*, 134, 5-29. <https://doi.org/10.3917/ls.134.0005>
- Krieg-Planque, A. (2013). Le traitement du « figement » par des locuteurs ordinaires : le sentiment linguistique d'« expression toute faite » dans des contextes de critique du discours politique. *Pratiques*, 159-160, 189-203. <https://doi.org/10.4000/pratiques.2904>

- Krieg-Planque, A. (2015). Construire et déconstruire l'autorité en discours. Le figement discursif et sa subversion. *Mots. Les langages du politique*, 107, 115-131. <https://doi.org/10.4000/mots.21926>
- Lebart, L., & Salem, A. (1994). *Statistique textuelle*. Dunod.
- Leblanc, J.-M. (2016). *Analyses lexicométriques des vœux présidentiels*. Istes.
- Livesey, S. M. (2001). Eco-Identity as Discursive Struggle: Royal Dutch/Shell, Brent Spar, and Nigeria. *The Journal of Business Communication*, 38(1), 58–91. <https://doi.org/10.1177/002194360103800105>
- Livesey, S. M. (2002). The Discourse of the Middle Ground: Citizen Shell Commits to Sustainable Development. *Management Communication Quarterly*, 15(3), 313–349. <https://doi.org/10.1177/0893318902153001>
- Livesey, S. M., & Kearins, K. (2002). Transparent and Caring Corporations? A Study of Sustainability Reports by The Body Shop and Royal Dutch/Shell. *Organization & Environment*, 15(3), 233–258. <https://doi.org/10.1177/1086026602153001>
- Loconto, A., Desquilbet, M., Moreau, T., Couvet, D., & Dorin, B. (2020). The Land sparing – Land sharing controversy: tracing the politics of knowledge. *Land Use Policy*, 96. <https://doi.org/10.1016/j.landusepol.2018.09.014>
- Maingueneau, D. (1996). *Les termes clés de l'analyse du discours*. Seuil.
- Martinez, W. (2003). *Contribution à une méthodologie de l'analyse des co-occurrences lexicales multiples dans les corpus textuels*. Thèse de doctorat, Université Paris-III, Paris.
- Mayaffre, D. (2008, mars). Quand « travail », « famille », « patrie » co-occurrent dans le discours de Nicolas Sarkozy. Etude de cas et réflexion théorique sur la co-occurrence. *JADT 2008 - 9<sup>e</sup> Journées d'Analyse statistique des Données Textuelles*, Lyon, 811-822.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340–363
- Muller, C. (1968). *Initiation à la statistique linguistique*. Larousse.
- Murillo, D., & Lozano, J. M. (2006). SMEs and CSR: An Approach to CSR in their Own Words. *Journal of Business Ethics* 67(3), 227-240. <https://doi.org/10.1007/s10551-006-9181-7>
- Née, É., Leblanc, J.-M., Fleury, S. (2017). Compter dans les textes, quelles unités ? In É. Née (Éd.), *Méthodes et outils informatiques pour l'analyse des discours* (chap. 4, pp. 103-121). Presses universitaires de Rennes.
- Née, É., Sitri, F., & Veniard, M. (2016). Les routines, une catégorie pour l'analyse de discours : le cas des rapports éducatifs. *Lidil*, 53, 71-93. <https://doi.org/10.4000/lidil.3939>
- Perelman, C., & Olbrechts-Tyteca, L. (2008 [1958]). *Traité de l'argumentation. La nouvelle rhétorique* (6<sup>e</sup> édition). Éditions de l'Université libre de Bruxelles.
- Pomikálek, J. (2011). *Removing Boilerplate and Duplicate Content from Web Corpora*. Thèse de doctorat, Masaryk University, Brno.



- Poudat, C., & Landragin, F. (2017). *Explorer un corpus textuel : Méthodes – pratiques – outils*. De Boeck.
- Rasche, A., Morsing, M., & Moon, J. (2017). The Changing Role of Business in Global Society: CSR and Beyond. In A. Rasche, M. Morsing, & J. Moon (Eds.), *Corporate Social Responsibility: Strategy, Communication, Governance* (p. 1-28). Cambridge University Press.
- Salem, A. (1987). *Pratique des segments répétés*. Klincksieck.
- Schryver, G. D. (2002). Web for/as corpus: a perspective for the African languages. *Nordic Journal of African Studies*, 11(2), 266-282.
- Stibbe, A. (2015). *Ecolinguistics: Language, Ecology and the Stories We Live By*. Routledge.
- Strand., R., Freeman, R. E., Hockerts, K. (2015). Corporate Social Responsibility and Sustainability in Scandinavia: An Overview. *Journal of Business Ethics*, 127(1), 1-15. <https://doi.org/10.1007/s10551-014-2224-6>