

*Pascaline Dury & Patrick Drouin**

L'obsolescence des termes en langues de spécialité : une étude semi-automatique de la « nécrologie » en corpus informatisés, appliquée au domaine de l'écologie

Abstract

Neology, the birth of words or terms, has been studied both in the area of lexicology and terminology for a long time. However, the other end of the life cycle of lexical units, *lexical death* or *necrology* as we call it here, has long been neglected. Corpus-based studies and the availability of electronic corpora have opened up exciting new possibilities to researchers interested in the behavior of lexical units. In this paper, we will present an investigation into lexical *necrology* based on a corpus made of multiple subcorpora representing various diachronic periods in the field of ecology. The corpus, made up of approximately 600,000 words, is divided into 8 diachronic periods ranging from 1950 to 2005.

Résumé

La formation des termes en langues de spécialité, la néologie, suscite depuis longtemps l'intérêt des chercheurs et est à l'origine de nombreux travaux en terminologie et lexicologie. Par contre, l'obsolescence de ces termes et leur disparition d'une langue de spécialité, la « mort lexicale », ou la *nécrologie* reste un aspect négligé du cycle de vie lexical. La recherche présentée ici porte sur l'étude de la *nécrologie* lexicale en corpus spécialisé dans le domaine de l'écologie. Le corpus contient environ 600,000 mots et couvre un éventail chronologique allant de 1950 à 2005.

1. Introduction

La formation des termes en langues de spécialité, la néologie, suscite depuis longtemps l'intérêt des chercheurs et est à l'origine de nombreux travaux en terminologie et lexicologie (par exemple par Boulanger 1989, 1990 et Humbley 1994 et 2006). Par contre, l'obsolescence de ces termes et leur disparition d'une langue de spécialité, la « mort lexicale » (*lexical death*, Grzega, 2002), ou la **nécrologie terminologique** reste un aspect négligé du cycle de vie lexical. La recherche présentée ici s'appuie sur l'hypothèse de travail que l'étude de la **nécrologie** en corpus peut être tout aussi riche d'enseignement que celle de la néologie, car elle permet d'observer en détail les mouvements d'un lexique spécialisé en permettant par exemple d'évaluer le rythme auquel se renouvelle son vocabulaire et ainsi de mesurer l'évolution des connaissances.

L'article qui suit rend donc compte des premières étapes et des premiers résultats obtenus dans le cadre de ces recherches sur la **nécrologie** en corpus spécialisé. Le projet de travail initial s'appuie sur les objectifs suivants :

1. D'abord identifier une liste d'indices linguistiques fiables qui permettent de repérer l'obsolescence ou la nécrologie des termes dans une langue de spécialité, ici l'écologie terrestre,
2. Pouvoir exploiter ces indices semi-automatiquement, autrement dit à l'aide d'outils informatiques testés pour le projet, dans un corpus informatisé d'écologie tout d'abord, puis

* *Pascaline Dury*
Centre de recherche en terminologie et
traduction (CRTT)
Université Lumière Lyon 2
France
pascaline.dury@univ-lyon2.fr

* *Patrick Drouin*
Observatoire de linguistique Sens-Texte (OLST)
Université de Montréal
Canada
patrick.drouin@umontreal.ca

dans d'autres corpus scientifiques constitués pour l'occasion ensuite,

3. Dresser la liste des différents types de nécrologie et tester différents outils informatiques pour les repérer semi-automatiquement dans un corpus spécialisé,
4. Enfin, mettre au point une méthode linguistique outillée - transposable à tous les domaines de spécialité- qui favorise et accélère la mise à jour des bases de documentations scientifiques, de bases de connaissances, des dictionnaires spécialisés, et des ontologies informatiques, qui sont souvent conçus pour inclure les termes nouveaux, mais qui ne reflètent pas toujours la disparition des termes, et conservent par conséquent des formes linguistiques désuètes, peu ou plus utilisées par les scientifiques d'un domaine.

2. Description de la nécrologie

Dans le cadre des recherches présentées dans cet article, la nécrologie correspond aux phénomènes suivants :

- la disparition d'une forme lexicale,
- la disparition d'un ou de plusieurs affixes à l'intérieur d'une forme lexicale,
- un changement de nombre ou de catégorie grammaticale au fil du temps,
- la disparition du sens d'un terme.

Les néologismes qui ne s'implantent pas dans un lexique et finalement disparaissent de la langue sont aussi considérés comme des nécrologismes. De la même façon, comme le souligne Guilbert ci-dessous, un terme peut disparaître d'un lexique donné à une époque donnée pour réapparaître ultérieurement dans ce même lexique. Ce terme sera alors considéré comme un nécrologisme lors de la période qui correspond à sa disparition –même provisoire- du lexique : « La notion de terme vieilli est très fluctuante. Tel terme, qu'on peut considérer comme sorti de l'usage à une époque donnée, peut donner lieu à des résurgences dans le discours de quelques individus, ou tel vocabulaire particulier » (Guilbert, 1975: 33). Il est important de souligner que dans cet article, l'obsolescence est observée sous l'angle de ses manifestations linguistiques, soit la disparition d'une forme, d'un sens, etc.

3. Description de la méthodologie employée

3.1. La constitution du corpus

Le corpus contient environ 600, 000 mots et couvre un éventail chronologique allant de 1950 à 2005. Il est constitué de textes spécialisés (articles scientifiques, thèses, chapitres d'ouvrages) écrits en anglais et appartenant au domaine de l'écologie terrestre. Les documents vulgarisés, écrits par des anglophones non-natifs et appartenant à d'autres sous-domaines de l'écologie (écologie humaine, écologie marine par exemple) n'ont pas été retenus dans la constitution du corpus.

3.2. Le choix du découpage temporel du corpus

Le corpus a été divisé en 6 sous-parties ou sous-corpus de 10 années chacun (1950-1960, 1961-1970, 1971-1980, 1981-1990, 1991-2000). La dernière période chronologique, qui contient des documents publiés entre 2001 et 2005, ne couvre que 5 années. Ce choix de consacrer le dernier sous-corpus à une période chronologique beaucoup plus courte a d'abord été fait pour des raisons extra-linguistiques : puisque l'écologie est un domaine scientifique particulièrement médiatisé depuis une dizaine d'années, nous posons l'hypothèse que cette surmédiatisation se traduit par un renouvellement lexical plus rapide dans le dernier sous-corpus étudié, et que la présence de nécrologismes y est donc plus importante que dans les autres périodes du corpus. Sur le plan méthodologique ensuite, nous posons l'hypothèse que différents types de phénomènes diachroniques

se manifestent selon la taille de la fenêtre chronologique étudiée (voir à ce sujet Dury et Picton, 2009), et que les types de nécrologie à l'œuvre ne sont peut-être pas les mêmes sur 5 ans que sur 10 ans. Cette dernière sous-partie chronologique courte nous servira donc de mesure de comparaison « diachronique » avec toutes les autres.

4. Analyse du corpus et premiers résultats

Tous les résultats donnés ci-dessous, et notamment les nécrologismes potentiels extraits du corpus à l'aide des indices linguistiques ont été validés par un expert en écologie¹.

4.1. Mise en place d'une liste d'indices linguistiques

Le premier objectif de cette recherche était de lister différents indices linguistiques fiables, exploitables informatiquement, qui puissent contribuer à repérer la nécrologie dans un corpus spécialisé, afin de semi-automatiser la recherche de l'obsolescence en corpus. Nous avons considéré les indices linguistiques suivants comme étant capables de pointer vers un ou plusieurs types de nécrologie, et les avons successivement testés dans le corpus d'écologie à l'étude :

- **Les marqueurs linguistiques**, c'est-à-dire l'utilisation d'expressions et de formulations caractéristiques par les auteurs pour parler d'un terme désuet ou disparu (par exemple : « *formerly called* », « *previously known as* », etc.),
- **La ponctuation et la typographie**, c'est-à-dire l'utilisation des parenthèses, des guillemets et des italiques par les auteurs pour parler d'un terme désuet ou disparu,
- **La distribution**, c'est-à-dire les changements pouvant se produire dans les cooccurrents d'un terme au fil du temps, indiquant alors peut-être un changement de sens de ce terme,
- **La variation synonymique**, c'est-à-dire le foisonnement de termes concurrents qui peut se produire lorsqu'un terme disparaît d'un lexique, tout comme il se produit lorsqu'un néologisme apparaît dans un lexique,
- **La grammaire**, c'est-à-dire le changement de nombre ou de catégorie grammaticale d'un terme au fil du temps (par exemple un terme utilisé comme substantif dans la partie « ancienne » du corpus, ensuite utilisé comme adjectif, et dont la forme substantivée disparaît au fil du temps).
- **La morphologie**, c'est-à-dire la disparition ou la modification d'un ou de plusieurs affixes dans une forme lexicale,
- **La fréquence**, c'est-à-dire l'observation statistique des changements de fréquences d'apparition d'un terme dans l'ensemble des sous-corpus.

4.2. Exploitation informatique de ces indices linguistiques

Le second objectif de cette recherche était d'exploiter informatiquement ces indices linguistiques pour extraire les premiers résultats, et tenter d'identifier les différents types de nécrologie possibles. Chacun des indices cités ci-dessus ont été outillés et exploités tour à tour.

4.2.1. Les marqueurs linguistiques

Cet indice linguistique repose sur l'hypothèse que les scientifiques s'expriment parfois sur la terminologie de leur domaine, et qu'ils peuvent alors mentionner un terme désuet, disparu, ou en cours de disparition en utilisant des marqueurs linguistiques tels que « *formerly known* », « *previously called* », etc. Un certain nombre de ces indices ont été effectivement identifiés dans le

¹ Il s'agit de monsieur Henri Tachet, Professeur d'Ecologie Emérite, Université Lyon 1 Claude Bernard, Laboratoire d'Ecologie, que nous remercions pour son aide et sa disponibilité.

corpus, comme les exemples donnés ci-dessous le montrent (les marqueurs linguistiques repérés sont indiqués en rouge) :

- (1) *The term superorganism turned out to be unnecessary, not even allowing an always welcome simplification of ecological jargon.* [1991-2000]
- (2) *Associations, formerly known as sociations, however defined, were also likely to afford living places to particular species of animals which had adapted to them and whose presence must have some impact on the vegetation, and various attempts were made to include the names of characteristic animal species in the association descriptions.* [1971-1980]
- (3) *By 1954, we find them talking about units such as biotic associations, which are now called "ecotones".* [1961-1970]
- (4) *The deme-terminology formerly used by ecologists offers no appropriate term for the afore-mentioned concept, the term 'deme', denoting any group of individuals of a specified taxon, being quite as vague as the concept of population. Any prefix only states one characteristic of such a group.* [1961-1970]

La construction de la liste de marqueurs a été effectuée sur la base de l'intuition des chercheurs et de mots clés potentiels comme ceux qui apparaissent dans les contextes précédents. À partir de nos premières observations faites en corpus sur des néologismes potentiels, nous avons aussi pu compléter cette liste initiale de marqueurs et retourner aux documents du corpus pour extraire de nouveaux contextes. Ce processus d'aller-retour entre le texte pour l'identification de marqueurs et la découverte de nouveaux marqueurs est très productif.

4.2.2. La ponctuation et la typographie

Tout comme pour les marqueurs linguistiques, nous posons l'hypothèse que les scientifiques peuvent aussi signaler un terme désuet, disparu ou en cours de disparition en utilisant des parenthèses, des guillemets ou encore des italiques. Cet indice produit beaucoup de bruit car il n'est pas exclusivement « nécrologique » : en effet, de toute évidence, les guillemets, les parenthèses et les italiques peuvent être utilisés par les auteurs pour désigner autre chose que des termes en obsolescence ; ils peuvent être utilisés par exemple pour introduire des néologismes. Néanmoins, un certain nombre d'entre eux sont utilisés dans des cas où les scientifiques pointent vers l'obsolescence, comme le montrent les exemples ci-dessous :

- (1) *For convenience or purposes of study, one may designate 'subcommunities' (now called 'associations') within the continuum based on combinations of dominants or on indicators.* [1950-1960]
- (2) *Indeed, such "ecotones" must be considered places of tension where two organizations meet and exchange their respective components (former boundaries), or as places where stresses of a genetic character, important in evolution, are at work.* [1971-1980]
- (3) *His word "mores", which has an awkward and alien sound, disappeared, but the idea he was seeking to express by it later found an outlet in Elton's "niche" and became a central ecological concept of great intellectual reward.* [1971-1980]

La liste limitée de signes de ponctuation et d'éléments typographiques utilisés à ces fins permet d'obtenir des résultats facilement à partir d'un concordancier qui exploite les expressions régulières. Bien que les résultats soient peu précis et bruités, le langage peut rapidement cibler les contextes intéressants à partir d'une simple inspection visuelle. Une grande partie des contextes identifiés démontre que les auteurs utilisent souvent la typographie pour véhiculer de l'information métalinguistique.

4.2.3. La distribution

Nous posons l'hypothèse qu'un changement au fil du temps dans les cooccurrents associés à un terme peut indiquer un changement du sens de ce terme et de ce fait révéler une nécrologie de type sémantique. C'est ce que montrent les deux schémas ci-dessous, où le terme à l'étude « *province* », change effectivement de sens au cours de la période étudiée dans le corpus. Dans le schéma 1, extrait des premières tranches chronologiques du corpus (1950-1970), « *province* » apparaît avec un sens synonyme de « zone », « territoire écologique », et se rencontre dans le voisinage des collocants suivants : « *biotic province* », « *temperate province* », « *adjacent province* », etc. Par contre, dans le second schéma, extrait des tranches chronologiques plus récentes du corpus (1991-2005), « *province* » apparaît avec un nouveau sens (région géographique), et de nouveaux cooccurrents (« *Quebec province* », « *Canada province* », etc.).

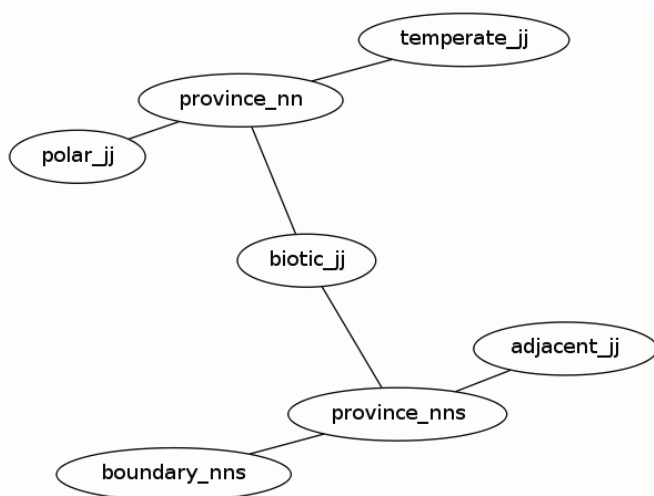


Schéma 1 : cooccurrents pour le terme « *province* » (1950-1970)

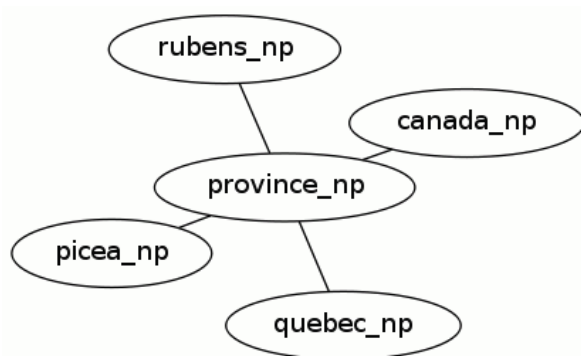


Schéma 2 : cooccurrents pour le terme « *province* » (1991-2005)

L'extraction des cooccurrents a été effectuée à l'aide d'une boîte à outils nommée *Ngram Statistics Package (NSP)* mise librement à la disposition des chercheurs par Banerjee et Pedersen (2003). À l'aide de ces outils, nous avons procédé à l'extraction de l'ensemble des cooccurrents (noms, adjectifs et verbes) des unités nominales du corpus. Les cooccurrents ont été extraits à l'aide d'une fenêtre de trois mots autour du nom ciblé, et les résultats obtenus sont représentés sous forme de graphes pour faciliter la consultation. Nous croyons que la visualisation de graphes comme ceux qui précèdent est plus conviviale pour l'analyste que la lecture de listes de mots à

plat. La conversion du format de fichiers NSP à des graphes a été réalisée à l'aide de scripts rédigés en langage Perl et la création des graphes est effectuée à l'aide de l'outil Graphviz (<http://www.graphviz.org/>).

4.2.4. La variation synonymique

« La démarche terminologique ne serait pas complète en néologie si l'on n'examinait pas toutes les dénominations utilisées pour le même concept. En effet, si l'une finit généralement par s'imposer, c'est seulement après un certain foisonnement néologique » (Humbley, 1994: 709).

En reprenant ici les paroles de Humbley à notre compte, nous posons l'hypothèse que, tout comme en néologie, la démarche terminologique ne serait pas complète en nécrologie si l'on examinait pas toutes les dénominations qui peuvent surgir et entrer en compétition dans un lexique pour remplacer un terme disparu ou en cours d'obsolescence, et que la variation synonymique qui caractérise l'apparition d'un nouveau terme peut aussi accompagner la disparition et le remplacement d'un terme obsolète. Nous avons ainsi trouvé les exemples suivants dans le corpus d'écologie :

- (1) *The term **ecosystem** was first proposed by Tansley in 1935, but the concept is by no means so recent. **Microcosm** (Forbes 1887), **holocoen** (Friederichs, 1930), **biosystem** (Thienemann, 1939), and **bioinert body** (Vemadsky, 1944) are terms which have been used to express similar ideas. [1950-1960]*
- (2) *For convenience or purposes of study, one may designate **subcommunities**, **community units** or "**associations**" within the continuum based on combinations of dominants or on indicators. [1950-1960]*

Les tests initiaux effectués à l'aide d'une analyse distributionnelle visant à faire ressortir la parenté sémantique entre termes n'ont pas conduit aux résultats attendus. Nous avons mis de l'avant une technique similaire à celle décrite dans Tutin (2007) qui consiste à étudier la cooccurrence et à regrouper les formes selon leurs affinités à sélectionner des cooccurrents similaires. L'intuition nous semble bonne puisque les formes qui apparaissent dans des contextes comparables risquent fort de posséder des composantes sémantiques communes. Cependant, la taille limitée de notre corpus nous confronte à des problèmes de rareté de données (*data sparseness*). Le travail d'élargissement du corpus, qui est déjà entrepris, devrait nous permettre d'obtenir des résultats plus probants.

4.2.5. La grammaire

Nous posons l'hypothèse qu'un terme peut changer de nombre ou de catégorie grammaticale au fil du temps, et qu'il existe donc un type de nécrologie dite « grammaticale ». C'est le cas du terme « *dominant* » extrait de notre corpus, utilisé comme substantif uniquement dans les sous-corpus couvrant les périodes jusqu'en 1990, puis ensuite employé comme adjectif également, pour ne plus être employé que comme un adjectif dans la dernière sous-partie du corpus :

- (1) *In land communities, plants usually are major **dominants** because not only are they producers but they provide shelter for the great bulk of the organisms in the community. [1950-1960]*
- (2) *Changes in environmental conditions may thus produce appreciable changes in composition and activity of subordinates with little change in **dominants**. [1981-1990]*
- (3) *A shift from annual weed species to **dominant** perennial species with increasing distance from water. [1991-2000]*
- (4) *In early succession, chewing insects, mainly Coleoptera and Diptera were **dominant**. [1991-2000]*

Le corpus ayant été étiqueté morphosyntaxiquement grâce à l'outil TreeTagger (Schmid 1994), il nous est possible d'obtenir la liste des unités lexicales (forme + catégorie morphosyntaxique) pour chacune des tranches du corpus. Une fois ces listes dressées, nous pouvons les comparer afin d'identifier les intersections sur la base de la forme et les disjonctions sur la base de la catégorie morphosyntaxique.

4.2.6. La morphologie

De la même façon que les termes peuvent changer de nombre et de catégorie grammaticale au fil du temps, ils peuvent également voir leur morphologie se modifier et perdre un ou plusieurs affixes, comme dans le cas de « *subdominant / dominant* », « *ecodeme / deme* », et « *genodeme / deme* », exemplifiés ci-dessous :

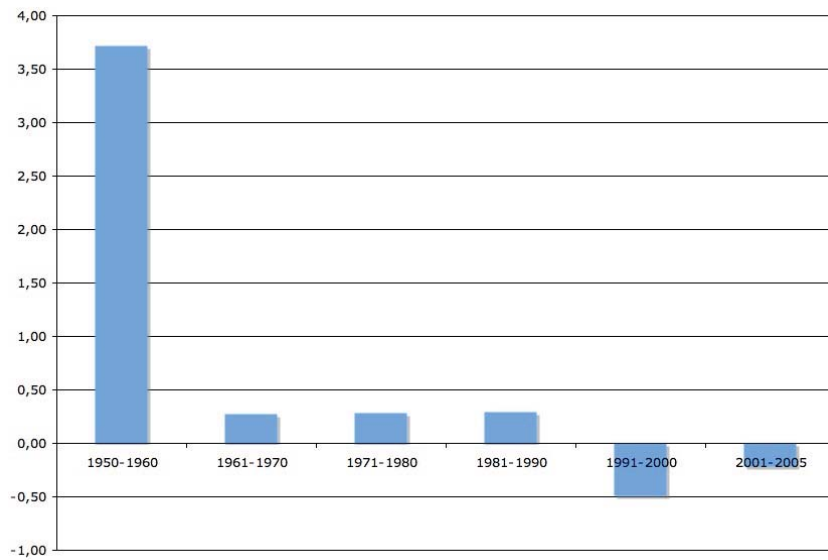
- (1) *The subordinate species exist because they are able to occupy the niche or portions of it that the dominants cannot effectively occupy. The **subdominants** tend to be more specialized in their environmental requirements and more narrow in their physiological tolerances.* [1961-1970]
- (2) *Any prefix only states one characteristic of such a group of individuals: an '**ecodeme**' denoting a deme occurring in a specified kind of habitat; a '**genodeme**', a deme differing from others genotypically; and, accordingly, a **genoecodeme** an **ecodeme** differing from others genotypically.* [1961-1970]

On peut noter au passage que bien que les termes « *deme* » et « *dominant* » sont encore utilisés dans le lexique de l'écologie terrestre, il semble que les termes « *subdominant* », « *ecodeme* », « *genodeme* » et « *genoecodeme* » aient disparu.

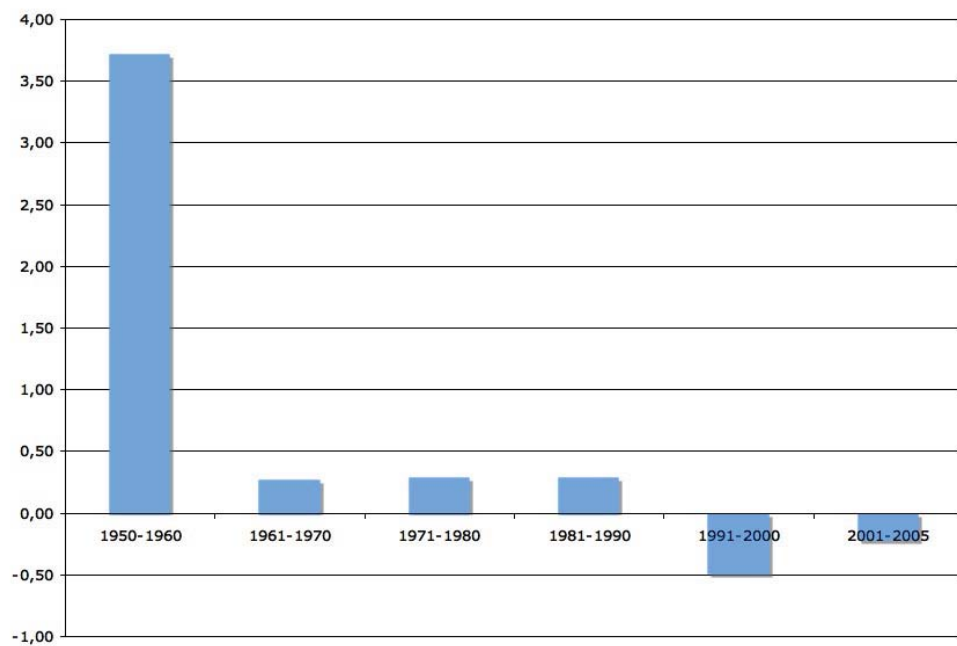
La prise en charge de cette partie du travail à l'aide de traitements informatiques est partielle. Afin d'isoler les éléments reliés morphologiquement, nous avons utilisé une approche relativement simple qui consiste à regrouper les formes qui possèdent des chaînes partagées. Encore une fois ici, nous postulons, tout en étant entièrement conscients des limites et des faiblesses de cette approche, que les formes apparentées morphologiquement sont reliées sémantiquement. L'utilisation d'un analyseur morphologique pourrait conduire à des résultats de meilleure qualité, mais cette technique facile à mettre en place nous a permis de tester rapidement l'idée d'exploiter la morphologie pour le dépistage des nécrologismes.

4.2.7. La fréquence

Pour détecter des nécrologismes à l'aide cet indice linguistique, nous avons décidé de comparer la fréquence d'apparition de tous les termes dans chacun des 6 sous-corpus afin de dégager la **spécificité lexicale** de chacun d'entre eux, c'est-à-dire repérer les formes lexicales qui sont, en termes de fréquences d'apparition, anormalement absentes de l'un des sous-corpus étudiés. Les deux graphes présentés ci-dessous correspondent aux fréquences d'apparition, sur l'ensemble du corpus, des termes « *disclimax* » et « *superorganism* » :



Graphe 1 (« *disclimax* »)



Graphe 2 (« *superorganism* »)

L'ensemble du corpus contient 12 486 formes lexicales substantives. Sur ce total, 127 ont été identifiées comme ayant une fréquence d'apparition significativement plus basse dans au moins un des 6 sous-corpus et une fréquence décroissante dans la dernière tranche chronologique étudiée (2001-2005). Sur ces 127 formes lexicales, 3 ont été validées manuellement comme nécrologismes : « *biogeoogenesis* », « *decreaser* » et « *subassociations* ».

Sur le total de formes lexicales contenues dans le corpus, 99 ont été identifiées comme ayant une fréquence d'apparition significativement plus basse dans au moins un des sous-corpus et une fréquence décroissante sur les deux dernières tranches chronologiques étudiées (1991-2000 /

2001-2005). Sur ces 99 formes lexicales, 6 ont été validées par l'expert comme nécrologismes : « *disclimax* », « *ectocrine* », « *genoecodeme* », « *niche pre-emption* », « *protocooperation* » et « *superorganism* ».

Sur les 12 486 formes lexicales extraites au total, seulement 1 terme, « *protzoon* » a été identifié comme nécrologisme potentiel (ayant une fréquence d'apparition plus basse dans au moins un des sous-corpus et une fréquence décroissante sur les trois dernières tranches chronologiques, 1981-2005), mais après validation de l'expert, n'a pas été conservé comme terme obsolète.

Aucune forme lexicale n'a été identifiée comme nécrologisme potentiel sur les quatre dernières tranches chronologiques (c'est-à-dire ayant une fréquence d'apparition plus basse dans au moins un des sous-corpus et une fréquence décroissante sur les quatre dernières tranches chronologiques (1971-2005).

Enfin, sur le total de formes lexicales contenues dans le corpus, 26 ont été identifiées comme nécrologismes potentiels (ayant une fréquence d'apparition plus basse dans au moins un des sous-corpus et une fréquence décroissante sur les cinq dernières tranches chronologiques (1960-2005). Sur ces 26 nécrologismes potentiels, 8 ont été confirmés comme effectivement obsolètes ou en cours de disparition : « *associes* », « *biociation* », « *biociations* », « *bioces* », « *life zone* », « *life zones* », « *mesobiota* », « *terrains* ».

Le tableau ci-dessous résume en chiffres les résultats trouvés pour chaque période :

Périodes	1960-2005	1971-2005	1981-2005	1991-2005	2001-2005
Nécrologismes potentiels	26	0	1	99	127
Nécrologismes validés	8	0	0	6	3

Tableau 1 : tableau récapitulatif

La mesure de la spécificité lexicale a été obtenue à l'aide du *calcul des spécificités* de Lafon (1980). Ce dernier permet la comparaison des fréquences des unités lexicales dans un sous-corpus à celles des mêmes unités dans un autre sous-corpus au sein d'un corpus plus large. Pour le présent article, les fréquences d'une des tranches (1950-1960) ont été comparées aux fréquences observées dans le reste des tranches (1961-2005). La technique permet d'isoler trois groupes de formes, celles dont la fréquence est significativement supérieure à celle attendue sur la base des observations effectuées sur les autres tranches (spécificités positives), celles dont la fréquence est inférieure (spécificités négatives) et celles dont la fréquence ne diverge pas des observations (formes banales).

5. Classement des termes obsolètes par types de nécrologie

Les 17 termes validés comme nécrologismes peuvent être classés dans une typologie reprenant les différents types de nécrologie proposés en début d'article (voir partie 2). Chaque type est illustré par quelques exemples :

- Nécrologie lexicale** (disparition d'un terme) : « *niche pre-emption* », « *superorganism* », « *biociation* », « *bioces* », « *ectocrine* ».
- Nécrologie sémantique** (disparition du sens d'un terme) : « *province* », « *macrobiota* », « *mesobiota* ».
- Nécrologie grammaticale** (changement de catégorie grammaticale ou de nombre) : « *dominant* » (utilisé comme adjectif, la forme substantivée à disparu), « *terrains* » (la forme pluriel a disparu).
- Nécrologie morphologique** (disparition d'un ou de plusieurs affixes) : « *biogeocenosis* » (« *biocoenosis* » existe toujours), « *disclimax* » (« *climax* » existe toujours).

Alors que certains des nécrologismes mentionnés dans l'article sont des néologismes qui ne sont finalement pas implantés dans le lexique (« *genoecodeme* », « *ecodeme* », « *subdominant* »), nous n'avons pas rencontré de cas correspondant à la disparition d'une forme lexicale à une période donnée du corpus, avec réapparition à une date ultérieure dans le même corpus.

6. Quelques conclusions

1/L'indice linguistique de la fréquence d'apparition montre que le plus grand nombre de nécrologismes avérés (8) se trouve dans la partie la plus ancienne du corpus (1950-1960). Cependant, notre hypothèse que la langue de l'écologie se renouvelle rapidement depuis les 10 dernières années à cause de la surmédiasation dont elle fait l'objet reste valable, car 9 nécrologismes ont été identifiés dans les deux derniers sous-corpus (dont 3 sur les cinq dernières années seulement). Cela montre également que lorsque l'on constitue un corpus diachronique contenant des documents issus d'un domaine dont les connaissances évoluent rapidement (comme c'est le cas ici de l'écologie, mais aussi de l'informatique, de la cancérologie, etc.), il y a un intérêt à étudier des périodes chronologiquement courtes (moins de 10 ans).

2/Parmi tous les indices linguistiques testés, celui qui concerne les fréquences d'apparition est le plus productif, mais il faut souligner qu'aucun de ces indices n'est suffisamment riche pour être utilisé tout seul et que c'est bien la combinaison de plusieurs de ces indices qui peut donner les résultats les plus significatifs. Ainsi, la plupart des nécrologismes sémantiques détectés dans le corpus (« *province* », « *mesobiota* », « *macrobiota* », etc.) ne peuvent être extraits en utilisant l'indice de la fréquence d'apparition seul ; de la même façon, les nécrologismes grammaticaux (« *dominant* », « *terrains* ») n'auraient pu être trouvés en utilisant l'indice des marqueurs linguistiques uniquement. On peut également noter que l'indice linguistique de la ponctuation et de la typographie génère beaucoup de bruit et que pour qu'il soit interprété correctement, il doit souvent être analysé à la lueur des marqueurs linguistiques (voir le point 4.2.2).

3/Un travail de recherche sur la nécrologie tel que celui qui est présenté ici montre à quel point le rôle de l'expert est important et la validation des candidats-termes extraits essentielle. Il semble également que si les experts peuvent avoir du mal à prendre conscience de l'évolution des connaissances en cours dans leur propre domaine, et peuvent rencontrer des difficultés à valider des termes nouveaux (à ce sujet, voire Rebeyrolle et Soubeilles, 2004), ils peuvent s'exprimer sur la validité ou non d'un nécrologisme avec moins d'hésitation.

7. Ce qu'il reste à faire

1/La taille du corpus anglais doit être augmentée pour atteindre 1 million de mots, et l'éventail des périodes étudiées doit aussi être élargi, en incluant des documents publiés avant 1950. Des corpus semblables en français et aussi en danois seront constitués pour valider la fiabilité des indices linguistiques sur plusieurs langues, mais aussi sur plusieurs domaines. À cette fin, d'autres corpus dans d'autres domaines de spécialité (économie, psychiatrie, criminologie) sont en cours de constitution.

2/D'autres indices linguistiques peuvent être testés informatiquement pour faire apparaître de nouveaux types de nécrologie : en observant par exemples les termes complexes qui disparaissent au fil du temps au profit d'acronymes, ou encore en étudiant les termes, qui devenus trop généraux, se voient remplacés par une série de termes hyponymes, etc.

3/Un travail sur la motivation linguistique des formes obsolescentes pourrait être mené pour observer si les termes non motivés ou mal motivés ont plus de risques – ou non – de disparaître d'un lexique spécialisé que les autres formes lexicales.

4/Enfin, l'objectif final du travail étant de mettre au point une méthode qui permette de repérer semi-automatiquement les nécrologismes, il faut encore développer l'utilisation des techniques statistiques et de celles qui sont issues du traitement automatique de la langue (TAL). En effet, les outils proposés par le TAL et exploités dans le cadre de la présente recherche n'ont pas été in-

tégrés et le processus de travail est essentiellement manuel d'une étape à une autre. L'insertion de l'ensemble des outils au sein d'une chaîne de travail pourrait permettre au langagier d'obtenir semi-automatiquement une liste de nécrologismes candidats. Le recours à des techniques statistiques plus avancées et des observations plus fines des variations de fréquence pourraient nous laisser entrevoir la disparition éventuelle de certaines formes, les nécrologismes en devenir.

8. Bibliographie

- Banerjee, S. et T. Pedersen 2003: « The Design, Implementation, and Use of the Ngram Statistics Package », *Proceedings of CICLing 2003 : computational linguistics and intelligent text processing*, Mexico-Mexico, 370-381.
- Boulanger, J.C. 1989 : « L'évolution du concept de 'néologie' de la linguistique aux industries de la langue », *Proceedings of the Colloque sur l'histoire de la terminologie, Terminologie diachronique* (C.de Schaetzen, ed.), Paris-Bruxelles, 193-211.
- Boulanger, J.C. 1990 : « *La création lexicale et la modernité* », *Le langage et l'homme*, vol. 25, n. 4, 233-240.
- Condamines A., Rebeyrolle J. and A. Soubeille 2004 : « Variation autour de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus », *Proceedings of the Euralex International Congress*, Université de Lorient, Lorient, France, 6-10 juillet 2004, 547-557.
- Dury, P ; Picton A. 2009 : « Terminologie et diachronie : vers une réconciliation théorique et méthodologique ? », *Revue Française de Linguistique Appliquée (RFLA)*, Vol. XIV, 2009-2, La terminologie : orientations actuelles (coord. : J. Humbley).
- Grzega, J. 2002: « Some aspects of modern diachronic onomasiology », *Linguistics*, vol. 40, n. 5, 1021-1045.
- Guilbert, L. 1975: *La créativité lexicale*. Larousse Universités.
- Humbley, J. 1994: « Quelques aspects de la datation de termes techniques : le cas de l'enregistrement et de la reproduction sonores », *Meta*, vol. 39, n.4, 701-705.
- Lafon, P. 1980: « Sur la variabilité de la fréquence des formes dans un corpus », *MOTS*, 1, 128-165.
- Schmid, H. 1994: « Probabilistic part-of-speech tagging using decision trees », *Proceedings of the International Conference on New Methods in Language Processing*, Manchester-UK, 44-49.
- Tutin A. (2007): « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques », *Actes de TALN 2007*, Toulouse, 10 pages..