



HAL
open science

L'activité prise au mot : le traitement des néologismes en entreprise

Frédéric Erlos

► **To cite this version:**

Frédéric Erlos. L'activité prise au mot : le traitement des néologismes en entreprise. P. Dury, J. Carlos de Hoyos, J. Makri-Morel, F. Maniez, V. Renner, M. Belen Villar Diaz. La néologie en langue de spécialité - Détection, implantation et circulation des nouveaux termes, CRTT, pp. 13-36, 2014, Travaux du CRTT, 978-2-9533061-0-1. halshs-01306459

HAL Id: halshs-01306459

<https://shs.hal.science/halshs-01306459>

Submitted on 23 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'activité prise au mot : le traitement des néologismes en entreprise

Frédéric ERLOS
Université Paris 3 – SYLED-CLA2T
ferlos@orange.fr

Résumé

Le sociolecte en usage dans les organisations mêle différents degrés de spécialisation de la langue et se présente rarement sous la forme d'un lexique formalisé faisant autorité. Afin de prendre la mesure de ces caractéristiques, nous proposons de limiter à certaines situations de communication la collecte des nouveautés relatives à un tel code fonctionnel. En effet, l'utilisation récurrente des mêmes genres du discours produit des normes de sélection lexicale, dont les évolutions peuvent faire l'objet d'un suivi dans le temps. Les logiciels de textométrie permettent alors d'outiller l'exploitation de corpus organisés sous la forme de séries textuelles chronologiques, en vue du repérage de formes candidates au néologisme. Ces formes sont ensuite évaluées depuis trois points de vue complémentaires : leur place dans le référentiel servant de cadre à la communication, leur fonctionnement comme support d'un objet de discours, et enfin, leur appartenance à l'une des composantes du parler d'entreprise. Une collecte de néologismes, réalisée durant plusieurs mois à partir d'un corpus de circulaires internes, sert à illustrer le propos et met en lumière l'importance du renouvellement de la terminologie d'une entreprise.

Mots clés

Néologisme ; néonyme ; entreprise ; langue de spécialité ; parler d'entreprise ; terminologie d'entreprise ; référentiel ; genre du discours ; objet de discours ; corpus ; linguistique de corpus ; série textuelle chronologique ; textométrie.

Introduction

Une entreprise crée et modifie en permanence les objets qui sont utiles à son fonctionnement : infrastructure, organisation, production et commercialisation de produits et de services. Les connaissances nécessaires à la réalisation de cette activité sont transmises à l'intérieur de l'organisation sur les supports et avec les formats les plus divers : présentations orales, publications, messagerie électronique, sites intranets, espaces de travail collaboratifs. Le néologisme, défini à ce stade comme « *une unité nouvelle, de nature lexicale, dans un code linguistique défini* ¹ », manifeste et véhicule le mouvement impulsé par l'activité, en même

¹ A. REY (1976 : 4).

temps qu'il représente un élément perturbateur pour l'état des connaissances précédant son apparition. Dès lors, le repérage et le référencement de ces unités apparaissent comme étant de nature à renforcer la clarté et l'efficacité de la communication nécessaires à l'activité. En effet, durant la période d'instabilité qui accompagne l'apparition d'un mot nouveau, des tâches telles que l'explicitation d'une notion, l'association d'un sigle à son développé, ou encore, la description d'une entité, sont étroitement associées à la mise en cohérence des contenus, à l'harmonisation terminologique ainsi qu'à la traduction. Cependant, nous constatons que ces traitements terminographiques se heurtent à plusieurs difficultés portant à la fois sur le choix des unités néologiques et sur la spécification de la nouveauté qui les caractérise. Nous décrivons une méthodologie destinée à surmonter ces obstacles, au moins partiellement, et à asseoir sur des bases consistantes un dispositif textométrique opérationnel, que nous présentons et dont nous discutons les résultats, obtenus à partir d'un corpus de documents réservés à l'usage interne.

1. Le choix des unités néologiques dans le contexte de l'entreprise

La mise en place d'un traitement des néologismes en entreprise nécessite qu'un périmètre de collecte adapté soit défini. Pour cela, il faut au préalable tenter de situer le sociolecte propre à une organisation entre les deux limites représentées par le lexique de la langue commune et l'idiolecte de chaque locuteur. D'autre part, le traitement des unités nouvelles ayant pour but de faciliter les échanges, il convient de veiller à ce qu'il soit réalisé au plus près du contexte réel des communications verbales.

1.1. Détermination de la portée de la collecte

1.1.1. Néologismes et spécificités d'un sociolecte

Les discours produits au sein d'une entreprise mobilisent différentes strates de la langue : au fonds commun s'ajoutent, d'une part, les langues spécialisées en usage dans les secteurs où l'activité se déploie et, d'autre part, la terminologie propre à toute organisation. Dardo de Vecchi a proposé de rendre compte de cette réalité à l'aide de la notion de « parler d'entreprise »,

pour désigner l'ensemble des processus linguistiques qui actualisent les répertoires linguistiques des membres d'une communauté, définie en fonction de l'appartenance à une entreprise. Autrement dit, la cristallisation linguistique de tout moyen de communication mis à la disposition d'une entreprise, pour des conceptualisations ayant des origines diverses².

L'adoption du parler d'entreprise comme norme de référence pour la collecte des néologismes conduit à privilégier un code fonctionnel situé³. Certes, un tel sociolecte agrège des composantes générales – langue commune et certaines de ses spécialisations –, mais sa spécificité s'enracine dans un foyer générateur de termes dont les productions deviennent

² D. DE VECCHI (1999 : 316).

³ A. REY (1976 : 15).

centrales, en raison même du particularisme qu'elles manifestent. Les unités caractéristiques du parler d'entreprise doivent par conséquent faire l'objet d'une attention particulière lors de la collecte.

Cependant, le contexte de l'entreprise n'est généralement pas très favorable au développement d'une tradition qui aurait préalablement mis au net sa terminologie spécifique. Certes, le mouvement permanent que l'activité imprime à toutes les tâches qu'elle mobilise n'empêche pas l'apparition de glossaires ou de lexiques. Cependant, il faut constater que ces instruments de travail sont émiétés en fonction de besoins *ad hoc* et qu'ils sont traités avec un soin et un niveau de détail très variables. Non seulement ils ne sont que très rarement à jour, mais surtout, aucun d'entre eux ne peut prétendre représenter un état satisfaisant du parler d'entreprise. Or, l'hétérogénéité des normes vient compliquer le repérage des néologismes, qui nécessite que l'on puisse s'appuyer, entre autres, sur une image du code en usage pour une période donnée. Dans le but de réduire cette complexité, nous proposons, dans un premier temps, de renoncer au cadre constitué par une organisation prise dans sa globalité.

1.1.2. Néologismes et référentiel

Afin de répondre à cette exigence, il nous a semblé fructueux d'aborder la question à partir d'un modèle facilement adaptable et ouvert aux problématiques liées à l'activité. C'est ce qu'offre la notion de référentiel proposée par l'épistémologue Ferdinand Gonseth :

(...) le mot référentiel en vient à signifier l'ensemble (explicite ou implicite) des préalables faute desquels telle ou telle activité systématique ne pourrait pas avoir lieu. Le référentiel peut alors s'offrir comme un cadre où cette activité prendra place. C'est ainsi que tout ce qui fait pour nous, en telle ou telle circonstance, la « réalité » d'un terrain de jeu peut servir de référentiel commun au joueur et au spectateur⁴.

Le cadre offert par le référentiel pouvant s'appliquer à toute forme d'activité, il permet de délimiter un contexte de collecte plus fin que celui offert par l'entreprise ou par les secteurs d'activité. Cette notion possède également la vertu de mettre en avant l'articulation nécessaire entre les référentiels individuels et collectifs, ce qui permet de poser la question du cadre de la collecte en termes de connaissances mobilisées par et pour la réalisation de l'activité.

En effet, quelle que soit la situation, la réalisation de l'activité présuppose l'existence d'un ensemble de connaissances partagées, au moins jusqu'à un certain point. Par conséquent, toutes les unités faisant appel à la compétence cognitive des acteurs, plus qu'à leur compétence linguistique⁵, doivent faire l'objet d'un traitement terminographique dans le cadre d'un référentiel donné. Dès lors que l'enjeu cognitif prédomine, la collecte des unités doit être étendue dans deux directions. La première concerne, comme nous l'avons vu plus haut, les différentes strates linguistiques dont le parler d'entreprise est composé. La seconde porte sur la diversité des natures d'unités susceptibles d'être collectées : lexie, syntagme lexicalisé,

⁴ F. GONSETH (1975 : 22).

⁵ J. HUMBLEY (2006 : 122).

terme simple ou complexe, mais aussi unité brachygraphique⁶ et dénomination propre⁷. Les critères de sélection des néologismes, que ceux-ci correspondent à une nouvelle unité – les emprunts faits à une autre langue et les changements de nature grammaticale sont rangés dans cette catégorie –, ou à la modification d'un concept ou d'une description déjà associés à une unité existante, doivent intégrer cette diversité et favoriser sa restitution.

1.2. Spécification de la nouveauté des néologismes

Dès lors que l'actualité des échanges impose son rythme à la collecte, celle-ci se déroule à l'échelle d'une micro-diachronie ayant pour toile de fond l'évolution d'un code fonctionnel situé. Nous avons vu que celui-ci ne fait pas l'objet d'un recensement susceptible de matérialiser la norme linguistique pour une période donnée. En revanche, il est possible d'aborder cette question d'un autre point de vue, en recherchant le moyen de capter l'évolution du code au fur et à mesure qu'elle se produit. Les néologismes apparaissant dans les discours, cela revient à se demander dans quelle mesure la sélection lexicale et phraséologique opérée lors de la production des discours peut également fonctionner comme substitut d'un code fonctionnel de référence. Cependant, le vocabulaire des discours ne peut être utilisé comme norme que si nous pouvons avoir prise, d'une part, sur les facteurs extralinguistiques qui sont susceptibles de le modifier et, d'autre part, sur la manière d'utiliser les propriétés d'un tel vocabulaire, afin qu'il fonctionne à son tour comme une norme.

1.2.1. Le genre du discours comme norme située

Un référentiel sert de cadre à des communications verbales dont les caractéristiques principales varient en fonction de la situation, des diverses compétences des interlocuteurs, des propriétés de l'univers du discours et du genre du discours utilisé. Celui-ci permet d'articuler une spécification fine des critères extralinguistiques caractérisant la situation de communication avec l'observation des usages linguistiques :

L'utilisation de la langue s'effectue sous la forme d'énoncés concrets uniques (oraux et écrits) qui émanent des représentants de tel ou tel domaine de l'activité humaine. L'énoncé reflète les conditions spécifiques et les finalités de chacun de ces domaines, non seulement par son contenu (thématique) et son style de langue, autrement dit par la sélection opérée dans les moyens de la langue – moyens lexicaux, phraséologiques et grammaticaux –, mais aussi et surtout par sa construction compositionnelle. Ces trois éléments (contenu thématique, style et construction compositionnelle) fusionnent indissolublement dans le tout que constitue l'énoncé, et chacun d'eux est marqué par la spécificité de la sphère d'échange. Tout énoncé pris isolément est, bien entendu, individuel, mais chaque sphère d'utilisation de la langue élabore ses types relativement stables d'énoncés, et c'est ce que nous appelons les genres du discours⁸.

⁶ R. KOCOUREK (1991 : 93-94).

⁷ I. TAMBA, B. BOSREDON, (1994 : 123) : « (...) toutes les formes fixes de dénominations d'un référent unique, françaises ou étrangères, sans considération de statut grammatical, (...) ».

⁸ M. BAKHTINE (1984 : 265).

Cependant, si tous les discours produits⁹ dans une telle sphère d'échange sont, en théorie, dignes d'intérêt pour la collecte des néologismes, les contraintes pratiques du recensement et de la description obligent à privilégier les échanges qui, d'une part, laissent des traces facilement exploitables, donc plutôt écrites, et d'autre part, empruntent des formes relativement stabilisées dans la durée. En effet, c'est à travers ses réalisations discursives que le code fonctionnel mobilisable dans le cadre d'un référentiel donné peut être observé. Si les conditions de production des discours remettaient en jeu à chaque échange la totalité d'un parler d'entreprise, c'est la possibilité même de l'observation des néologismes qui serait menacée de dissolution. Afin de prendre en compte ces contraintes, le repérage des néologismes nécessite la mise en place d'un dispositif qui repose, en définitive, sur l'identification de foyers d'énonciation stabilisés observables à partir de leurs productions écrites.

Par ailleurs, l'intérêt présenté par tel ou tel genre du discours varie en fonction des finalités de la collecte¹⁰. De manière générale, et pour les néologismes apparaissant dans un parler d'entreprise en particulier, les genres du discours utilisés pour diffuser des informations devant passer d'une population initiée à une autre qui l'est moins, constituent de bons postes d'observation. Il peut s'agir, par exemple, des informations internes à transmettre d'une communauté de locuteurs vers d'autres métiers, d'une maison mère vers ses filiales ou de l'entreprise vers des destinataires externes (pouvoirs publics, analystes, journalistes, clientèle, etc.). Les indications explicites destinées à signaler les nouveautés sont plus souvent présentes dans les genres du discours d'interface que dans ceux utilisés pour les échanges entre pairs. D'autre part, le genre du discours d'autorité vise à imposer à ses lecteurs l'obligation de prendre en compte l'information nouvelle en même temps que la bonne manière de la dénommer. À défaut de présenter un terrain riche en termes concurrents, ces discours livrent une version officielle de la dénomination, qui est destinée à occuper le statut de vedette dans les référentiels terminologiques. Il apparaît ainsi que l'attention prêtée au genre du discours permet de situer la collecte et de prévoir ses lacunes, mais aussi d'observer le fonctionnement des unités dans leur cotexte et de privilégier, comme nous le verrons plus loin, celles dont le référent fonctionne comme un objet de discours¹¹.

1.2.2. La série textuelle chronologique comme norme endogène

Les garanties offertes par la sélection d'un ou de plusieurs genres du discours adaptés gagnent à être renforcées pour que le vocabulaire d'un corpus puisse fonctionner comme une norme. En effet, les occurrences d'un type d'énoncé peuvent respecter plus ou moins leur modèle ;

⁹ A. BORZEIX, B. FRANCKEL (2005 : 231-261).

¹⁰ J. PEARSON (1998), A. PICTON (2009).

¹¹ J.-B. GRIZE (1996), F. SITRI (1998). Dans la logique naturelle de Jean-Blaise Grize, l'objet de discours est à la fois une notion associée dans les discours à une ou plusieurs expressions nominales et le résultat d'une schématisation qui « (...) est une représentation discursive orientée vers un destinataire de ce que son auteur conçoit ou imagine d'une certaine réalité. » (1996 : 50). Si l'attention de cet auteur se porte principalement sur les opérations logiques utilisées dans l'élaboration des schématisations, d'autres auteurs, dont Frédérique Sitri, ont cherché à mettre en lumière les moyens linguistiques mis en œuvre pour la construction de ces objets, dans et par les discours.

elles sont également susceptibles de provenir de pôles différents au sein d'une même sphère d'échanges ; enfin, leur production peut être plus ou moins espacée dans le temps. Ce sont là encore autant de propriétés externes dont les variations sont susceptibles de perturber l'étude des effets de la variable temps sur l'utilisation d'un vocabulaire. Il convient donc de les maîtriser dans la mesure du possible. C'est pourquoi les corpus que nous utilisons ont été constitués sous la forme de séries textuelles chronologiques :

Nous réserverons donc l'appellation de série textuelle chronologique aux séries périodiques homogènes constituées par des textes produits dans des situations d'énonciation similaires, si possible par un même locuteur (individuel ou collectif) et présentant des caractéristiques lexicométriques comparables¹².

Dans ces corpus, chaque texte est en relation avec la norme que constitue le vocabulaire de la série textuelle, c'est-à-dire l'ensemble formé par tous les mots attestés au moins une fois dans l'un des textes de la série. Lorsqu'un nouveau texte est ajouté à un corpus de suivi (*monitoring corpus*), cette relation peut revêtir des aspects variables répartis entre deux extrêmes théoriques : une absorption complète, le nouveau texte ne contient que des mots déjà utilisés, et une originalité totale, la dernière occurrence de la série n'est constituée que de mots nouveaux. Un tel dispositif promeut le vocabulaire mobilisé dans le corpus au statut de norme de référence pour un contexte précis. Les mots qui viennent enrichir la norme revêtent alors le statut de formes candidates pour le repérage des néologismes, soit en raison de leur nouveauté, soit parce que leur fréquence dans le texte ajouté à la série est particulièrement élevée par rapport à la fréquence attendue pour l'un quelconque des autres textes du corpus.

Pour cette étude, nous avons constitué un corpus ouvert de circulaires internes, diffusées par l'organe central d'un groupe bancaire vers d'autres entités appartenant au même groupe, à raison de deux cents par an environ¹³. Ces circulaires viennent modifier, ajuster ou enrichir les processus en cours durant l'année. Élaborées par les différents services de cet organe central, elles se présentent sous la forme de quelques pages obéissant à un formalisme assez strict (cartouche introductif normalisé, différentes relectures avant publication, signature de l'émetteur).

2. Un filet à papillons textométrique

L'exploitation d'une série textuelle chronologique sous la forme d'un corpus ouvert commence par le formatage des textes devant faire l'objet des traitements informatiques. Nous nous limitons ici à mentionner les opérations textométriques¹⁴ fondamentales

¹² A. SALEM (1988 : 107).

¹³ Dans le cadre d'un autre travail, nous avons utilisé un corpus de rapports d'activité dont la publication est annuelle. Les exigences réglementaires qui imposent ce type de bilan annuel valent aussi bien pour l'extérieur que pour l'intérieur des organisations de taille importante. F. ERLOS (2009).

¹⁴ La textométrie recouvre le champ de la lexicométrie, celle-ci étant définie comme un « ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire d'un corpus de textes » (L. LEBART et A. SALEM, 1994 : 314).

nécessaires et suffisantes pour le but poursuivi. Celles-ci peuvent ensuite être complexifiées en fonction des finalités recherchées et des possibilités des logiciels¹⁵.

2.1. Évolution diachronique du vocabulaire dans un corpus ouvert

Lorsqu'il a été rendu exploitable par le logiciel de textométrie, le corpus est segmenté en formes graphiques, c'est-à-dire en suites de caractères encadrées par deux espaces ou par des caractères délimiteurs, comme les signes de ponctuation. Toutes les formes graphiques recensées à ce stade sont associées à une fréquence, correspondant au nombre de leurs occurrences dans le corpus, et à un numéro d'ordre lexicométrique, obtenu en triant les fréquences par ordre décroissant. En fonction de la partition retenue pour le corpus, dans le cas le plus simple chaque texte représente une partie, le logiciel construit le tableau lexical entier. Dans ce tableau à double entrée, le lexique des formes graphiques représente la bordure verticale gauche et la succession des parties la bordure horizontale haute. Le tableau offre sur chaque ligne les fréquences d'une forme graphique dans les différentes parties du corpus, et dans chaque colonne, les occurrences des formes graphiques dans l'une des parties du corpus.

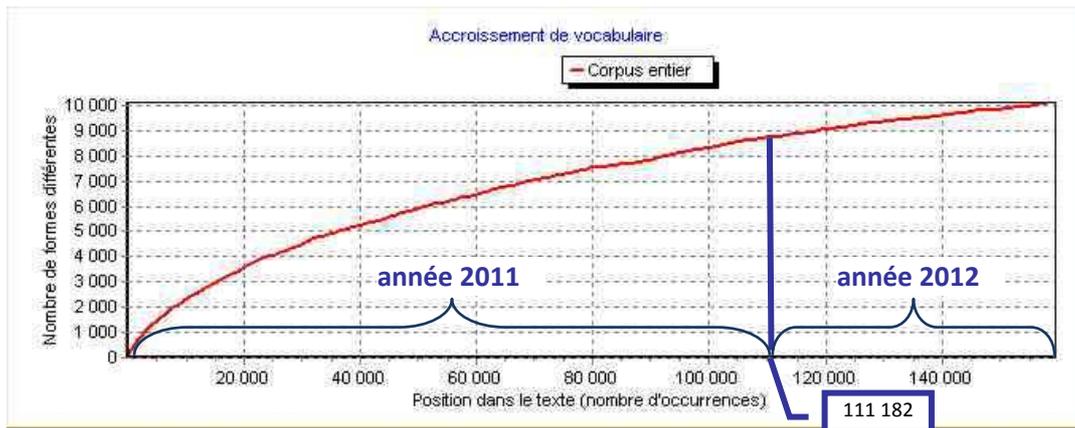
| Ordre lexicométrique | Formes | Partie 1 | P 2 | P 3 | P 4 | P 5 |
|----------------------|------------|----------|-----|-----|-----|-----|
| 1 | de | 20 | 36 | 14 | 21 | 21 |
| 2 | la | 6 | 10 | 3 | 15 | 16 |
| 3 | des | 2 | 22 | 9 | 22 | 7 |
| 4 | et | 4 | 15 | 5 | 16 | 6 |
| 5 | à | 3 | 7 | 4 | 4 | 15 |

Tableau 1 : *Extrait du tableau lexical entier (TLE) du corpus de circulaires 2011-2012 présentant les cinq premières formes et les cinq premières parties*

À partir de ces données fondamentales, l'exploitation textométrique d'un corpus fournit des indications, à la fois générales et particulières, sur la manière dont la temporalité affecte l'évolution d'un vocabulaire. Pris dans sa globalité, le vocabulaire du corpus s'enrichit de formes nouvelles au fur et à mesure que de nouveaux textes sont ajoutés. L'exploitation informatique de ces données permet de préciser dans quelle proportion se fait cet accroissement.

¹⁵ Liste non limitative de logiciels de textométrie :

<http://adcost-elliadd.univ-fcomte.fr/astadiag/> (Astadiag). http://www.dtmvic.com/05_SoftwareF.html (DmtVic). http://ancilla.unice.fr/?redirected_from=ancilla.unice.fr/~brunet/pub/hyperbase.html (Hyperbase). <http://www.tal.univ-paris3.fr/lexico/> (Lexico 3). <http://www.ling.uqam.ca/ato/sato/> (Sato 4.2). <http://textometrie.ens-lyon.fr/> (plateforme TXM 0.6). <http://www.xaira.org/> (Xaira 1.25). On pourra consulter pour d'autres approches, par exemple, P. DROUIN et al. (2006) et A. PICTON (2009).



Graphique 1 : *Accroissement du vocabulaire d'un corpus de circulaires*

Commentaire : le graphique ci-dessus représente la manière dont les 10 000 formes graphiques du lexique d'un corpus ouvert de circulaires (année 2011 et les cinq premiers mois de 2012), se répartissent sur les 160 000 vocables dont se composent les textes du corpus. En mai 2012, le corpus comportait 269 parties, ce qui représente une moyenne de 600 occurrences par texte. Le graphique met en évidence un tassement de l'accroissement du vocabulaire au fur et à mesure que le nombre d'occurrences s'accroît. Alors que 40 000 occurrences ont été suffisantes pour faire apparaître 5000 formes différentes, plus de 100 000 occurrences supplémentaires ont été nécessaires pour doubler cette quantité. Le graphique indique également, pour les cinq premiers mois de l'année 2012, une progression d'un millier de formes nouvelles réparties sur 83 textes, soit en moyenne une douzaine de formes nouvelles par circulaire. Le calcul a été réalisé avec le logiciel Lexico 3.

La mesure de l'accroissement du vocabulaire fournit ainsi un ordre de grandeur pour la quantité de formes nouvelles à laquelle il faut s'attendre dans chaque partie nouvelle ajoutée au corpus, compte tenu de sa position au sein de la série textuelle chronologique. Cet accroissement participe d'un mouvement plus profond, caractéristique des séries textuelles et qu'André Salem a proposé de nommer « temps lexical ». Il consiste en un renouvellement progressif du vocabulaire du corpus :

tout émetteur produisant des textes sur une période de temps assez longue utilise sans cesse de nouvelles formes de vocabulaire qui viennent supplanter, du point de vue fréquentiel, d'autres formes dont l'usage se raréfie. (...) Il s'ensuit que les vocabulaires de parties correspondant à des périodes consécutives dans le temps présentent en général plus de similitudes entre eux que les vocabulaires correspondant à des périodes séparées par un intervalle de temps plus long¹⁶.

Dans le corpus de circulaires que nous utilisons, le « texte » est à considérer ici globalement comme une année de production de circulaires. L'année 2011 sert de référence pour l'année 2012 en cours. Il aurait également été possible de descendre au niveau de chaque service émetteur de circulaire mais, compte tenu du but poursuivi, l'émetteur considéré ici est l'organe central, pris comme locuteur collectif s'adressant aux autres entités du groupe bancaire.

¹⁶ A. SALEM (1988 : 107).

Par ailleurs, une année d'antériorité s'est révélée suffisante pour accéder aux variations du temps lexical de la série textuelle chronologique prise comme exemple. En effet, les variations lexicales ont des causes multiples que l'on peut essayer de ramener à deux des caractères du genre du discours proposés par Mikhaïl Bakhtine : le contenu thématique et la sélection opérée dans les moyens de la langue. Le premier est susceptible de varier avec chaque circulaire, mais nous observons qu'un grand nombre de sujets ont un caractère cyclique, correspondant dans la majorité des cas à un rythme annuel. Une année d'antériorité permet ainsi de repérer les changements, y compris lorsqu'ils se font à l'intérieur d'un cadre préconstitué depuis plusieurs années. La seconde cause de variation ajoute ses effets à la précédente. Elle dépend cette fois de chaque rédacteur. Celui-ci introduit dans le texte de la circulaire son propre usage de la langue, ainsi que des choix lexicaux dépendant de sa plus ou moins grande maîtrise des codes du genre et de la terminologie du secteur concerné. Il en résulte, entre autres, l'apparition de variantes dont il faut veiller à contenir la portée, compte tenu du renouvellement des rédacteurs. Là encore, le choix d'une année d'antériorité pour le repérage des néologismes a constitué un compromis adapté à ce genre du discours. Enfin, comme nous le montrons plus loin, si la norme de vocabulaire peut être ramenée à une année d'antériorité lors du repérage des néologismes, les dimensions diachronique et générique du corpus gagnent à être étendues avant l'enregistrement du néologisme dans un référentiel terminologique.

2.2. Formes originales, formes spécifiques et néologismes

Au fur et à mesure que de nouvelles circulaires sont diffusées et intégrées au corpus, le dispositif mis en place permet d'interroger l'apport d'un nouveau texte à partir de deux mesures complémentaires : les formes originales et les formes spécifiques. Nous présentons et discutons chaque méthode à partir des néologismes collectés dans le corpus de circulaires.

2.2.1. Les formes originales ne sont pas toutes des néologismes et réciproquement

La première méthode utilisée pour le repérage et l'établissement des néologismes consiste à isoler, dans le vocabulaire du nouveau texte, les formes originales, c'est-à-dire celles qui lui appartiennent exclusivement par rapport au reste de la série. Elles constituent, en quelque sorte, l'apport net du nouveau texte à l'accroissement du vocabulaire du corpus et, à ce titre, elles sont de bonnes candidates pour le repérage des néologismes. Ces formes sont obtenues en testant les lignes du tableau lexical entier. Celles pour lesquelles la somme des fréquences observées dans l'ensemble des parties du corpus, la dernière partie ajoutée comprise, est égale à la fréquence observée pour cette dernière partie de la série, correspondent à des formes graphiques originales. Les données retenues, à titre d'exemple, proviennent d'un échantillon de circulaires émises durant le premier trimestre de l'année 2012, soit cinquante-quatre textes.

Nous avons vu plus haut que chaque partie nouvelle ajoutée au corpus apporte en moyenne une douzaine de formes nouvelles pour une longueur moyenne de 600 occurrences. Nous observons que les formes originales ne sont pas toutes associées à la même fréquence. Sans qu'il soit possible de dire qu'il s'agit d'une règle générale, l'observation montre que les

formes originales possédant la fréquence la plus élevée sont fréquemment associées à des objets de discours, alors que les hapax relèvent plutôt de la mention simple. Si ces formes candidates se révèlent être des néologismes, leur fonctionnement dans les textes comme dénomination d'un objet de discours permet d'établir la nouveauté avec plus de certitude, dès ce premier stade. En effet, le cotexte fournit dans ce cas des indications utiles pour élaborer une définition ou une description associée à l'unité terminologique. Dans l'exemple ci-dessous, la dénomination « Amplia » correspond à un nouveau produit, et les formes « Bonne » et « Compréhension » font partie du terme complexe « Recueil de Bonne Compréhension », désignant un formulaire, rempli par le souscripteur d'un produit financier, et permettant à l'établissement bancaire de recueillir la preuve que son client a bien compris la nature du produit vendu et les risques qui lui sont associés.

| Formes originales, partie 187 | Fréquence |
|-------------------------------|-----------|
| Amplia | 3 |
| AMPLIA | 2 |
| Bonne | 1 |
| Compréhension | 1 |
| Première | 1 |

Tableau 2 : *Formes originales de la partie 187*

Cependant, les formes originales détectées par le test ne sont pas toutes pertinentes pour le repérage des néologismes, loin s'en faut. Ces formes ont diverses provenances, qui peuvent être réparties en fonction de la typologie suivante :

- les coquilles ; elles sont rares, mais elles ne sont pas absentes, malgré les correcteurs d'orthographe et les relectures ;
- les variantes graphiques (capitale attribuée à un mot dans le cours d'une phrase, voire écriture du mot en majuscules) ; elles sont plus nombreuses que les coquilles, en raison de l'absence de normalisation du corpus sur ce point, mais elles sont pertinentes pour le repérage des néologismes, les rédacteurs utilisant fréquemment ce type d'artifice typographique pour signaler une nouveauté. Les graphies « MEMO » et « OBJECTIVITE » sont, par exemple, utilisées pour désigner, dans le premier cas, un format de fiche explicative destinée à la clientèle, et dans le second, un engagement commercial. Dans le tableau 2, la forme « Première » doit sa majuscule à la position occupée par le mot en tête de phrase.
- les coordonnées spatio-temporelles et celles relatives à l'organisation apportent une certaine quantité de dénominations propres, présentes sous la forme d'une simple mention (personnes, lieux, unités organisationnelles).
- enfin, la thématique propre à la circulaire entraîne la sélection de mots qui font à cette occasion une apparition ponctuelle. Par exemple, pour la présentation d'un service proposé aux collectivités publiques, les mots « usagers », « habitants » et « loisirs » font leur entrée dans le vocabulaire du corpus, mais cela ne manifeste rien d'autre qu'un usage banal de la langue commune. Il s'agit alors d'un phénomène de rareté relative, liée à la taille réduite du corpus, ainsi qu'au genre du discours auquel les circulaires appartiennent.

En deuxième lieu, il apparaît que ce test doit être complété par la recherche des formes spécifiques car, d'une part, il ne permet pas de détecter toutes les formes pertinentes pour le repérage des néologismes (un quart des néologismes relevés dans l'échantillon lui ont échappé), et d'autre part, il ne fournit pas d'indications très riches pour guider l'exploration des contextes où les formes apparaissent. Or, cette vérification des contextes est assez contraignante : il faut la recommencer au moins une dizaine de fois par circulaire, alors que 40% des textes se révèlent improductifs en matière de néologisme.

2.2.2. Néologismes et calcul des spécificités

La seconde mesure utilisée pour repérer et établir les néologismes correspond à un jugement porté en probabilité sur la fréquence des formes présentes dans une nouvelle partie ajoutée au corpus. Le calcul des spécificités permet d'évaluer si la fréquence réelle observée pour telle forme dans telle partie du corpus correspond à une surreprésentation, une sous-représentation ou à une valeur proche de la valeur modale, c'est-à-dire de la valeur la plus probable¹⁷. Pour un seuil donné en deçà duquel on estimera sortir de la banalité, par exemple 5% de chances, une forme spécifique positive correspondra à une surreprésentation, une forme spécifique négative équivaudra à une sous-représentation et une forme banale sera une forme dont la fréquence est statistiquement attendue. L'intérêt de la spécificité positive est d'attirer l'attention sur le suremploi d'une forme graphique (originale ou non). Celui-ci fonctionne alors comme un bon indice pour repérer un changement intervenant, soit dans la signification d'un mot ou d'un syntagme lexicalisé existant, soit dans la description d'une entité, ou bien encore, pour détecter l'apparition d'un terme ou d'une dénomination propre complexe¹⁸.

De façon plus générale, le calcul des spécificités fait émerger les principaux éléments thématiques caractéristiques de la partie étudiée¹⁹. Ce sont alors les formes graphiques dotées des coefficients de spécificité les plus élevés qu'il est intéressant de retenir. Dans tous les cas issus de l'échantillon de textes retenus pour cette étude, le calcul des spécificités a fourni les formes candidates pour le repérage des néologismes qui n'avaient pas été détectés à partir des seules formes originales. Ce calcul sélectionne également les formes originales, comme le montre l'exemple donné ci-dessous, mais il a tendance à les noyer parmi les autres formes spécifiques positives. Les limites de ce calcul, lorsqu'il est appliqué au repérage des néologismes, sont discutées dans la troisième section. Nous constatons que lors de la phase de repérage, cet inconvénient est compensé par le fait que les formes graphiques possédant un coefficient de spécificité positif élevé entrent pratiquement toujours dans la caractérisation du néologisme, c'est-à-dire dans la définition du terme ou la description de l'entité nommée.

¹⁷ Les probabilités sont calculées à partir du modèle hypergéométrique qui prend en compte la taille du corpus (le nombre total d'occurrences), la taille de la partie étudiée, la fréquence totale de la forme retenue et la fréquence de celle-ci dans la partie étudiée.

¹⁸ Voir, entre autres, F. ERLOS (2009 : 659-764).

¹⁹ Cette méthode est également utilisée pour repérer les différences pouvant affecter les réseaux de co-occurrences associés à une forme. Pour ces explorations complémentaires, les logiciels de textométrie offrent en outre la possibilité d'observer finement les cotextes dans lesquels une forme apparaît. Les classiques concordances sont alors sollicitées, de même que les possibilités offertes par des subdivisions plus fines que les textes, comme les paragraphes, afin de mettre en lumière des marqueurs discursifs destinés à signaler la nouveauté.

L'indication fournie par la fréquence de certaines formes originales, afin d'identifier les unités servant à désigner un objet de discours, trouve ici un complément utile pour l'exploration du cotexte où ces formes apparaissent. Dans l'exemple ci-dessous, nous avons grisé les formes utilisées comme unités pour la définition du néonyme Amplia, qui désigne un fonds à formule éligible au CTO (compte-titre ordinaire), au PEA (plan d'épargne en actions), ainsi qu'à l'assurance-vie.

| Forme | Frq. Tot. | Frq. Par. | Coeff. |
|-----------------------------|-----------|-----------|--------|
| PEA | 45 | 14 | 20 |
| CTO | 57 | 14 | 18 |
| fonds | 101 | 16 | 17 |
| Assurance | 55 | 12 | 15 |
| Vie | 54 | 11 | 14 |
| formule | 28 | 8 | 12 |
| commercialisation | 69 | 9 | 10 |
| Une | 66 | 8 | 8 |
| Amplia | 3 | 3 | 7 |
| 10 formes spécifiques (...) | | | |
| AMPLIA | 2 | 2 | 5 |
| 39 formes spécifiques (...) | | | |
| Bonne | 1 | 1 | 3 |
| Compréhension | 1 | 1 | 3 |
| 22 formes spécifiques (...) | | | |
| Première | 1 | 1 | 3 |

Tableau 3 : *Formes spécifiques positives de la partie 187*

Commentaire : le tableau ci-dessus représente un extrait des formes spécifiques positives de la partie 187 du corpus (colonne Forme). Le nombre de formes spécifiques séparant les formes originales mentionnées dans le tableau 2 a été indiqué sur une ligne coupant les colonnes. Fréquence totale (Frq.Tot.) correspond à la fréquence de la forme dans l'ensemble du corpus, Fréquence dans la partie (Frq. Par.) correspond à la fréquence observée dans cette partie du corpus, et Coeff. correspond à la puissance de 10 utilisée pour la notation scientifique de la probabilité attribuée par le calcul hypergéométrique à la fréquence d'une forme dans la partie. Par exemple, pour une probabilité de 0,000005, notée 5 E-6, l'exposant « 6 » est reporté dans la colonne « coefficient ». Plus la probabilité est faible, plus la puissance de 10 est élevée et plus la fréquence de la forme est jugée caractéristique de la partie étudiée. Dans le tableau, seuls les résultats du calcul hypergéométrique ayant produit une probabilité inférieure à 5% (0,05 ou 5 E-02) pour la fréquence observée sont affichés. Le coefficient étant positif, il s'agit ici de spécificités positives correspondant à la présence de fréquences plus élevées que celle de la valeur modale. Les spécificités négatives ne sont pas affichées. Le calcul a été réalisé avec le logiciel Lexico 3.

La complémentarité des deux approches (formes originales et formes spécifiques) est illustrée à l'aide de l'exemple donné ci-dessous.

| Forme | Frq. Tot. | Fréquence | Coeff. |
|-----------------------------|-----------|-----------|--------|
| collectivités | 37 | 13 | 20 |
| transactions | 12 | 8 | 16 |
| e | 24 | 9 | 15 |
| publiques | 26 | 7 | 11 |
| 69 formes spécifiques (...) | | | |
| loisirs | 1 | 1 | 3 |
| 24 formes spécifiques (...) | | | |
| Collectivités | 14 | 2 | 3 |

Tableau 4. *Formes spécifiques positives de la partie 194*

Le néologisme « e-transactions Collectivités publiques », qui correspond au nom d'un service commercialisé par le groupe bancaire, n'avait pas été détecté lors du test des formes originales. De fait, toutes les formes qui le composent sont déjà assez bien représentées dans le reste du corpus. Cependant, le calcul des spécificités pour la partie 194 a mis en exergue leur forte spécificité positive. Le repérage a alors d'autant plus de chances d'être efficace que le néologisme fonctionne dans le texte sélectionné comme un objet de discours, par opposition à la simple mention, déjà signalée à propos des coordonnées spatio-temporelles ou organisationnelles.

3. Présentation des résultats et discussion

Nous pouvons récapituler les points acquis dans les deux précédentes sections en proposant, pour le contexte étudié, une définition du néologisme librement inspirée de celle donnée par Alain Rey²⁰ : il s'agit d'une unité appartenant à un parler d'entreprise, dont la forme signifiante ou la relation signifiant-signifié n'était pas réalisée au stade immédiatement antérieur du code fonctionnel utilisé dans le cadre d'un référentiel donné. Ce code est assimilé au vocabulaire d'une ou de plusieurs séries textuelles chronologiques, constituées à partir d'occurrences textuelles relevant de genres du discours en usage dans la sphère d'échanges où l'activité de l'entreprise se réalise. De cette conception découle le dispositif textométrique destiné au repérage de formes candidates, mais aussi l'évaluation du caractère néologique des unités correspondant à ces formes. En effet, le dispositif de repérage ne permet pas, à lui seul, de vérifier les principaux critères habituellement adoptés, afin d'identifier le caractère néologique d'une unité²¹. La prise en compte de la diachronie, à travers l'évolution du vocabulaire des textes rassemblés dans une série textuelle chronologique, permet d'obtenir des formes candidates, mais elle ne fournit pas encore d'unités néologiques. Afin d'atteindre ce but, le dispositif mis en place doit permettre la vérification des autres critères que sont

²⁰ A. REY (1976 : 17) : « Ainsi, le néologisme est une unité du lexique, mot, lexie ou syntagme, dont la forme signifiante ou la relation signifiant-signifié, caractérisée par un fonctionnement effectif dans un modèle de communication déterminé, n'était pas réalisée au stade immédiatement antérieur du code de la langue. »

²¹ M.-T. CABRÉ (1998 : 252 sq.).

l'absence de recensement dans les sources lexicologiques et terminologiques de référence, l'instabilité formelle et sémantique de l'unité, ainsi que la perception de sa nouveauté chez les locuteurs²².

3.1. Les néologismes d'un parler d'entreprise

3.1.1. Critères de sélection des néologismes

Le repérage des formes candidates (formes originales et formes spécifiques positives associées à un coefficient élevé pour un texte donné), sert de point de départ à une série de vérifications destinées à sélectionner les unités néologiques. Ces vérifications sont combinées entre elles afin d'étayer le jugement porté sur le caractère néologique de l'unité évaluée. Trois niveaux d'analyse sont pris en compte. Le premier, de caractère pragmatique, traite de manière indirecte le sentiment de nouveauté éprouvé par le locuteur, en le rapportant au référentiel de la situation de communication. Le second critère tient compte du fonctionnement discursif de l'objet désigné par l'unité étudiée. Enfin, un dernier critère s'appuie sur la place occupée par l'unité étudiée parmi les strates du parler d'entreprise. Son appartenance à la langue commune, aux terminologies des secteurs d'activité ou à la terminologie de l'entreprise déclenche, pour chaque cas, des investigations différentes dans les ressources disponibles.

Dans son déroulement, l'analyse se place à chacun des trois niveaux en suivant, sans ordre prédéfini, les indications les plus immédiatement accessibles et en cherchant à les valider ou à les rejeter à l'aide des informations tirées des autres critères. Par exemple, le mot « compréhension », qui apparaît dans les tableaux 2 et 3, n'est pas étudié isolément, mais en tant que partie du terme complexe « recueil de bonne compréhension ». L'identification de la position occupée par cette unité dans les strates du parler d'entreprise va de pair avec l'évaluation de la position de l'objet dans le référentiel ainsi que l'observation de son fonctionnement discursif dans le texte étudié. Cette circulation entre les trois critères, réalisée par l'analyse, est difficile à restituer de façon linéaire. C'est pourquoi, pour plus de clarté, nous présentons les trois critères séparément.

3.1.1.1. Centralité de l'objet désigné et actualisation du référentiel

L'interrogation de la place occupée par l'objet désigné – à l'aide d'une unité néologique –, dans le référentiel associé à la situation de communication, repose sur l'hypothèse que les producteurs des textes du corpus souscrivent au principe de coopération gricéen, ainsi qu'aux maximes conversationnelles qui lui sont associées²³. De ce point de vue, les maximes de manière, qui portent sur la façon de dire ce que l'on a à dire (en règle générale « soyez clair »

²² On sait que L. GUILBERT (1975 : 36-40) a renvoyé dos à dos, avec des accents chomskyens, l'interrogation de la compétence des locuteurs, jugée trop variable, et l'exploration des produits de la performance, jugés incommensurables, au profit de locuteurs-auditeurs idéaux, incarnés par les lexicologues.

²³ H.P. GRICE (1975).

et en particulier, « évitez de tenir des propos obscurs ou ambigus, soyez bref et méthodique »), si tant est qu'elles soient toujours respectées, contraignent l'énonciateur à renseigner son interlocuteur sur un nouvel objet qu'il n'est pas censé connaître. Par conséquent, nous supposons que la prise en charge énonciative de la nouveauté aura d'autant plus de chances d'être explicite, et donc visible, que la place occupée par l'objet en question sera centrale. La caractérisation de cette position, comme centrale ou périphérique, est établie par rapport au référentiel associé à la circulaire. L'activité dont il est question est alors adoptée comme repère. Pour reprendre l'exemple précédent, le « recueil de bonne compréhension » étant étroitement associé à la commercialisation d'un produit financier, et cette forme d'activité étant centrale dans le contexte de production de la circulaire étudiée, la position de cet objet a été jugée comme étant plutôt centrale. À l'inverse, « loisirs », qui apparaît dans le tableau 4, a été jugé périphérique. En effet, le mot renvoie en l'occurrence aux loisirs courants, qui occupent une place secondaire par rapport au service de paiement à distance dont il est question dans la circulaire étudiée.

3.1.1.2. Type de fonctionnement discursif de l'objet de discours et prise en charge de la nouveauté

Le second critère tient compte du fonctionnement discursif de l'objet de discours auquel le néologisme candidat a fourni une expression nominale. L'évaluation porte alors sur la richesse des informations contenues dans la classe-objet ouverte par l'objet de discours. Cependant, ce n'est pas tant l'ouverture d'une classe qui importe ici que le fait qu'une telle ouverture pourrait difficilement passer sous silence la nouveauté de l'objet lorsqu'elle est avérée. Dans la schématisation qui est proposée par le discours d'une circulaire, certains objets peuvent n'être que mentionnés, alors que d'autres ouvrent une classe-objet englobant divers aspects de l'objet évoqué. Ces classes se composent d'ingrédients et d'objets qui, à leur tour, sont susceptibles d'ouvrir de nouvelles classes. Cette description du fonctionnement discursif, telle qu'elle est proposée par la logique naturelle, doit être articulée avec les dispositifs textuels destinés à prendre en charge la nouveauté, ainsi que les unités concernées. Celles-ci sont susceptibles d'être reformulées à l'aide des mécanismes de la cohésion textuelle, telle que l'anaphore. Lorsque l'unité candidate initie une classe-objet dans un texte de circulaire, les informations qui lui sont associées permettent, d'une part, de valider ou d'invalidier le jugement de néologisme (présence d'expressions telles que « nouveau produit », « nouvelle activité »), et d'autre part, de disposer des matériaux utiles pour la construction d'une définition ou d'une description. En revanche, une simple mention ne permet pas de trancher et suspend la qualification du néologisme aux résultats fournis par les autres critères. Par ailleurs, les unités dont les occurrences sont signalées par des signes typographiques, tels que parenthèses ou guillemets, ou bien celles qui sont associées à une périphrase descriptive, connaissent un traitement destiné à faire le départ entre une prise de distance de l'énonciateur et un signalement de la nouveauté, réalisé à l'aide d'une citation ou de tout autre procédé.

3.1.1.3. Strates du parler d'entreprise et choix des normes de référence

Enfin, le troisième critère s'appuie sur la position occupée par l'unité candidate dans les strates du parler d'entreprise. Le rattachement de l'unité à l'une de ces strates entraîne des

vérifications différentes. Les tableaux 2, 3 et 4 fournissent des exemples d'unités appartenant à chacune de ces strates : « Amplia » est une dénomination propre de produit forgée par l'entreprise ; « loisirs » correspond au sens courant qu'il a dans la langue commune ; « recueil de bonne compréhension » est une unité terminologique relevant du secteur de la conformité et appliquée à la commercialisation des produits bancaires, financiers ou d'assurance. Lorsqu'une unité telle qu'« Amplia » répond positivement aux différents critères (forme originale, objet occupant une position centrale dans le référentiel de la circulaire et fonctionnement discursif sous la forme d'une classe-objet), son appartenance à la terminologie de l'entreprise est jugée suffisamment établie pour l'attribution de la qualité de néologisme. En revanche, « loisirs », dont on a vu plus haut qu'il occupe une place périphérique par rapport au référentiel, appartient sans ambiguïté à la langue commune (« *Par métonymie. Activité, distraction à laquelle on se consacre pendant les loisirs.*²⁴ »). Comme par ailleurs cet hapax n'est pas utilisé pour ouvrir une classe-objet, il semble aller de soi que son originalité doit être attribuée à sa rareté dans la série textuelle, compte tenu des sélections appelées par le genre du discours auquel les circulaires appartiennent. La distinction qui doit être opérée entre rareté et nouveauté s'applique systématiquement pour les unités néologiques relevant de la terminologie des secteurs d'activité comme, par exemple, le « recueil de bonne compréhension » dont il a déjà été question. Elle nécessite d'étendre dans deux directions le périmètre des normes de référence.

Une première extension consiste à accroître de quelques années la période prise en compte pour les circulaires (il s'agit de la période allant de 2005 à 2010), et à élargir le champ de la norme générique, afin de prendre en compte des corpus rassemblant des textes appartenant à des genres du discours voisins, produits par le même locuteur collectif. Cela permet de s'assurer du caractère néologique du terme à partir de données textuelles plus étendues que le corpus utilisé pour le repérage. La seconde extension prend en compte les ressources terminologiques disponibles, internes et externes, afin de distinguer les néologismes à proprement parler des simples variantes de termes existants. Ces vérifications peuvent ensuite donner lieu à deux sortes d'arbitrages. Dans le premier cas, lorsqu'un néonyme candidat est bien attesté dans les corpus textuels de vérification, il n'est pas enregistré, car son implantation dans le parler d'entreprise est jugée suffisante, quel que soit son statut dans les ressources terminologiques existantes. En revanche, dans le cas inverse où le candidat est absent des corpus de vérification, mais présent dans les ressources terminologiques, il est considéré comme un néologisme pour le code fonctionnel en usage dans le référentiel pris pour cadre. Dans ce cas, il y a une sorte de « prime » au statut cognitif incertain du terme pour le contexte étudié. Le même raisonnement est appliqué lorsque les néonymes candidats sont absents des deux sources complémentaires de vérification.

Nous illustrons ce processus de qualification, qui fait passer par étapes une unité du statut de forme ou de groupe de formes à celui de néologisme, à l'aide de la dénomination « e-transactions Collectivités publiques », présentée dans le commentaire du tableau 4. La forme

²⁴ *Trésor de la langue française*, article « Loisir ».

« transactions » relève *a priori* de la langue commune où elle bénéficie déjà d'une certaine polysémie : acte par lequel on transige, l'accord qui en résulte, accord contractuel, opération commerciale ou boursière, etc., sont mentionnés dans le *Trésor de la Langue Française*. Mais la présence d'un « e » parmi les formes spécifiques et un début d'observation co-textuelle montrent qu'il s'agit de l'unité composée « e-transactions ». Celle-ci est déjà présente dans le corpus et, par ailleurs, elle fait l'objet d'un recensement, depuis 2006, dans le *Grand dictionnaire terminologique* de l'Office Québécois de la Langue Française. Pour celui-ci, « e-transaction » est à proscrire (il s'agit d'un terme déconseillé), et la vedette « transaction électronique », dotée de cinq concurrents, y est définie comme une « transaction sécurisée qui est effectuée, lors d'un achat ou d'un paiement en ligne, par l'intermédiaire du réseau Internet. » Dans le corpus de repérage, la nouveauté correspond à la dénomination d'un service, « e-transactions Collectivités publiques », défini comme une « offre de paiement à distance sécurisé, effectué par carte bancaire et à destination du marché des collectivités publiques ». Absente des corpus de vérification, cette nouvelle dénomination a été enregistrée comme néologisme.

3.1.2. Résultats obtenus

Les tests réalisés à l'aide du dispositif de type textométrique qui a été présenté plus haut ont permis de sélectionner des formes candidates en raison de leur originalité ou de leur fréquence plus élevée dans chaque nouveau texte ajouté à une série textuelle chronologique. Dans l'échantillon de circulaires retenues pour cette étude, 34 néologismes ont été collectés dans 20 textes différents, soit 1,7 néologisme par texte productif, ou 0,6 en moyenne si l'on prend l'ensemble de l'échantillon de 54 textes. Parmi les néologismes collectés, 25 ont pu l'être à partir du seul test des formes originales, soit 75% environ²⁵, le reliquat l'ayant été à partir du test des formes spécifiques.

Parmi les néologismes recensés, les dénominations propres sont les mieux représentées (18 sur 34). Elles se répartissent en noms de produits, noms d'entités ou d'artefacts. Les 16 termes simples et complexes (ces derniers étant classiquement majoritaires) se répartissent à parts égales entre termes bancaires et financiers appartenant à ces secteurs compris au sens large (recueil de bonne compréhension, prêt pour la modernisation de la restauration, escompte des billets à ordre électroniques, seuil des incidents significatifs, prélèvement SEPA / SEPA Direct Debit / SDD, paiement express sans contact / sans contact, décomptabilisation, opération de refinancement à long terme / LTRO / long term refinancing operation)²⁶, et termes propres à l'entreprise. La répartition des unités collectées par rapport aux principales strates du parler d'entreprise est la suivante : les deux tiers relèvent directement de la

²⁵ Il faut préciser que pour les unités complexes combinant au moins deux formes graphiques, nous avons compté qu'une forme originale a permis l'identification d'une unité complexe inclusive en examinant le cotexte à l'aide de concordances. Le caractère néologique a été évalué pour l'unité complexe prise comme un tout.

²⁶ Dans certains cas, la nouveauté réside dans l'utilisation d'un terme anglo-saxon ou français, là où seule l'une de ces deux langues était déjà attestée. Dans ce cas précis, la durée inhabituelle du refinancement (trois ans au lieu de trois mois), a introduit un flottement entre le développé habituellement retenu pour LTRO, « longer term », et celui qui a semblé plus approprié à la nouveauté « long term ».

terminologie de l'entreprise (22 néologismes sur 34), un petit tiers de la terminologie des secteurs d'activité, et une unité correspond à un nom de série télévisée à la mode, qui a été associé à la vente d'un produit.

Ces résultats sont à rapporter au périmètre restreint qui leur a servi de cadre et qui résulte de la combinaison d'une série textuelle chronologique avec un référentiel situé. Les instances discursives rassemblées dans la série textuelle sont représentatives des échanges validés – par opposition aux mises au point effectuées dans les groupes de travail – entre un organe central et une partie au moins des entités qui lui sont rattachées. Les circulaires officialisent des informations, des recommandations ou des instructions et, à ce titre, elles constituent une source pertinente pour le repérage des termes vedettes, validés par les instances de direction de l'entreprise. Dans la perspective d'une collecte du parler d'entreprise qui distinguerait des sous-groupes d'émetteurs au sein de l'organisation, les apports d'un tel corpus gagneraient à être complétés, d'une part, à l'aide d'autres sources moins formelles, afin d'enregistrer les variantes ou les expressions parallèles propres aux métiers concernés, et d'autre part, à partir des travaux préparatoires, qui comportent fréquemment les expressions alternatives utilisées par les premiers initiés pour désigner un même objet (nom de code, surnom, noms d'essais ou provisoires, etc.). Si un tel approfondissement de la collecte était envisagé, il serait probablement nécessaire de limiter son périmètre à tel ou tel secteur précis de l'entreprise.

3.2. L'approche textométrique : un filet adapté mais devant être utilisé avec précaution

3.2.1. Les avantages de la méthode utilisée

L'approche textométrique a été retenue dans la mesure où les textes rassemblés en séries textuelles chronologiques constituent la matière première utilisée pour le repérage et l'établissement des néologismes. En effet, nous avons vu qu'ils jouent à la fois le rôle de norme de référence pour les formes candidates et celui de contexte d'apparition des unités néologiques. Lorsqu'une unité nouvelle est utilisée pour désigner un objet de discours, le contexte fournit, pour les définitions, les caractéristiques du concept, et pour les descriptions, un premier ensemble d'informations de nature encyclopédique.

Par ailleurs, la segmentation du texte en formes graphiques brutes, sans étiquetage morphosyntaxique, présente des avantages pour la collecte des unités nouvelles. De cette manière, non seulement certains usages différenciateurs pour les néologismes sont conservés (orthographe particulière des majuscules initiales, mots en capitales, préférence pour les pluriels, dénominations propres déguisées en syntagmes lexicalisés), mais aussi, l'attribution d'étiquettes erronées pour le marquage des parties du discours ne vient pas perturber l'interprétation. La robustesse de cette technique permet en outre de passer d'une langue à l'autre, sans contrainte particulière. Elle assure également une très bonne portabilité d'un corpus à l'autre, les traitements reposant sur l'exploitation de données similaires.

Enfin, les logiciels de textométrie fournissent un outillage déterminant pour les deux principales étapes de l'identification des néologismes : l'identification de formes candidates

dans un cadre de micro-diachronie, ainsi que l'établissement du caractère néologique des unités étudiées. Dans le premier cas, les outils statistiques permettent de disposer d'une norme endogène pour des données textuelles représentatives d'un genre du discours. Depuis cette approche macroscopique, il est ensuite possible de descendre les paliers textuels successifs allant du groupement de textes d'une même série à un texte particulier, des sections d'un texte, tels que les paragraphes, au cotexte d'une occurrence de forme graphique. Cette navigation facilite grandement le travail du néologue, qui a besoin, pour étayer son jugement, aussi bien d'indications globales que d'attestations précises.

3.2.2. Limites de l'exploitation statistique des données brutes

La manipulation des formes graphiques brutes présente des avantages qui ont été rappelés plus haut, mais elle doit être utilisée avec précaution, aussi bien pour le repérage des formes originales que pour le calcul des formes spécifiques. Dans le premier cas, étant donné que le comptage des formes est réalisé en distinguant les variantes graphiques, un même mot est susceptible d'être compté plusieurs fois, comme dans l'exemple donné plus haut avec la dénomination « *Amplia* », dont les cinq occurrences sont réparties sur deux graphies (tableau 2). Il est raisonnable de penser que pour les formes originales cet inconvénient est limité, dans la mesure où il y a de bonnes chances pour qu'au moins l'une des graphies soit signalée par le test. Le repérage des variantes graphiques est plus délicat avec le calcul des spécificités. Dans le tableau 4, les deux formes du mot « *collectivités* » ont certes été sélectionnées toutes les deux par le calcul, mais elles sont séparées par quatre-vingt quatorze formes spécifiques positives. Cet inconvénient peut être surmonté par une lecture attentive de la liste et en procédant à des regroupements. En revanche, le calcul hypergéométrique utilisé pour la sélection des formes spécifiques peut, dans certains cas, ne pas signaler des formes qu'il aurait été utile d'obtenir pour le repérage des néologismes.

En effet, Pierre Lafon²⁷ estime que le calcul hypergéométrique est particulièrement bien adapté pour toutes les fréquences, à quelques exceptions près. Il attire l'attention sur le fait que ce calcul peut ne plus être utile pour des fréquences inférieures ou égales à trois. Or, nous avons vu qu'il est courant de rencontrer des formes candidates dotées de faibles fréquences. Dans le cas des hapax, le calcul devient alors directement dépendant de la taille de la partie du corpus où ils apparaissent.²⁸ Il est par conséquent nécessaire de tenir compte des caractéristiques des corpus utilisés, afin de déterminer un seuil de probabilité qui ne soit ni trop restrictif, ni trop permissif. En effet, dans le premier cas, il est possible de passer à côté de formes candidates pertinentes, dans l'autre, un résultat trop bruyant risque d'être difficile à exploiter. Les corpus que nous utilisons pour le repérage des néologismes, du fait de leur organisation sous la forme de séries textuelles chronologiques, sont composés de très nombreuses parties possédant des caractéristiques textométriques homogènes. À cette condition, les hapax (formes originales de fréquence 1), sont signalés comme spécifiques, y compris lorsque le seuil de probabilité est fixé à 5%.

²⁷ P. LAFON (1980 : 127-165).

²⁸ L. LEBART, A. SALEM (1994 : 177).

Néanmoins, le calcul hypergéométrique peut également être pris en défaut dans ce type de corpus textuel, lorsque les occurrences d'une forme graphique sont organisées sous la forme d'une progression ou d'un reflux étalés sur plusieurs parties. Dans ce cas, le jugement statistique a tendance à banaliser les fréquences intermédiaires présentes dans les parties situées entre les deux extrémités de la série. Mais cette restriction tombe lorsque les parties étudiées sont la première ou la dernière de la série. En effet, dans ces deux cas, la comparaison des vocabulaires oppose l'une ou l'autre de ces parties à l'ensemble du corpus. Cela permet de tester une fréquence particulièrement faible ou élevée par rapport à toutes les autres valeurs prises par la fréquence d'une forme donnée dans le reste du corpus²⁹. Dans l'approche que nous avons retenue, c'est toujours la dernière partie ajoutée au corpus qui est prise en compte, puisque c'est elle dont il faut tester le vocabulaire pour détecter les formes candidates. Il n'en reste pas moins que le non respect des précautions qui viennent d'être exposées pourrait laisser échapper quelques papillons, que nous aurions cru attraper avec le filet textométrique.

Conclusion

Dans cette étude nous avons adopté une approche résolument pragmatique de la nouveauté, pour laquelle il a été nécessaire de spécifier le caractère relatif de la notion de néologisme lorsqu'elle est appliquée au sociolecte caractérisant une entreprise. Dans ce but, nous avons proposé d'évaluer la nouveauté d'une unité par rapport au vocabulaire habituellement utilisé pour la production de discours relevant d'un même genre et organisés en corpus, sous la forme de séries textuelles chronologiques regroupant des textes aux propriétés homogènes. Un dispositif textométrique s'appuyant sur deux méthodes éprouvées (calcul des formes originales et des formes spécifiques), permet de faire émerger, par contraste avec la norme que représente le vocabulaire de la série textuelle, des formes graphiques nouvelles ou particulièrement caractéristiques pour chaque texte nouveau ajouté au corpus. Ces formes candidates sont alors utilisées pour repérer les unités néologiques, qui peuvent être de différentes natures : expressions nominales simples ou complexes, terminologiques ou non, unités brachygraphiques et dénominations propres. Pour chacune de ces unités candidates, nous procédons à des vérifications au cours desquelles les niveaux pragmatique, discursif et linguistique sont interrogés. En effet, nous cherchons à identifier la prise en charge discursive de leur nouveauté au regard, d'une part, de l'importance de l'objet qu'elles désignent dans le référentiel servant de cadre à la communication et, d'autre part, du traitement de cet objet dans les discours. Par ailleurs, nous évaluons le caractère avéré de leur nouveauté par rapport aux normes de référence disponibles pour la strate du parler d'entreprise dont elles relèvent (lexiques de la langue commune, terminologies des langues spécialisées, corpus faisant usage de la terminologie d'entreprise). Une telle approche permet de documenter de façon

²⁹ Pour une explication détaillée et des propositions de solution pour traiter cette question, voir A. SALEM (1988 : 122-129).

opérationnelle l'apparition des néologismes au fil des discours. Cependant, elle repose sur des choix qui sont de nature à limiter sa portée, au moins dans un premier temps. D'une part, elle suppose que l'on puisse constituer dans tous les cas les séries textuelles chronologiques adaptées au besoin et, d'autre part, elle pose la question d'une mise en commun des fruits des collectes réalisées à partir de corpus distincts. C'est pourquoi, pour ce qui concerne ce dernier point, nous avons proposé ailleurs de consolider les résultats obtenus à partir de telles collectes au sein de référentiels terminologiques adaptables au contexte.

Bibliographie

- BAKHTINE M. (1984), *Esthétique de la création verbale*, traduit du russe par A. Aucouturier, préface de T. Todorov, Paris, Gallimard.
- BORZEIX A., FRAENKEL B. (coord.) (2005), *Langage et travail – Communication, cognition, action*, Paris, CNRS, (1^{ère} éd. : 2001).
- CABRÉ M. T. (1998), *La Terminologie – Théorie, méthode et applications*, traduit du catalan et adapté par M. C. Cormier et J. Humbley, Paris / Ottawa, Armand Colin / Presses de l'Université d'Ottawa.
- DROUIN P., PAQUIN A., MÉNARD N. (2006), « Extraction semi-automatique des néologismes dans la terminologie du terrorisme », In *Actes des 8^{es} Journées internationales d'analyse des données textuelles*, Besançon, Presses Universitaires de Franche-Comté, p. 279-288.
- ERLOS F. (2009), *Discours d'entreprise et organisation de l'information – Apports de la textométrie dans la construction de référentiels terminologiques adaptables au contexte*, Thèse préparée sous la dir. d'André Salem, Université Paris 3, 2 vol.
- GONSETH F. (1975), *Le Référentiel univers obligé de médiatisation*, Lausanne, L'Âge d'Homme, coll. « Dialectica ».
- GRICE H. P. (1975), Logic and Conversation, In Cole P. & Morgan J.L. (eds.), *Syntax and Semantics 3 : Speech Acts*, New York, Academic Press, p. 41-58.
- GRIZE J.-B. (1996), *Logique naturelle et communications*, Paris, PUF, coll. « Psychologie sociale ».
- GUILBERT L. (1975), *La Créativité lexicale*, Paris, Larousse, coll. « Larousse Université ».
- HABERT B., NAZARENKO A., SALEM A. (1997), *Les linguistiques de corpus*, Paris, Armand Colin / Masson, coll. « U ».
- HUMBLEY J. (2006), Terminologie et nom propre, In *Des arbres et des mots – Hommage à Daniel Blampain*, Bruxelles, Éditions du Hasard, p. 107-124.
- KOCOUREK R. (1991), *La Langue française de la technique et de la science – Vers une linguistique de la langue savante*, préface d'Alain Rey, Wiesbaden, Brandstetter Verlag.
- LAFON P. (1980), Sur la variabilité des formes dans un corpus, In *Mots* n° 1, Paris, Presses de la FNSP, p. 127-165.
- LEBART L. et SALEM A. (1994), *Statistique textuelle*, préface de Christian Baudelot, Paris, Dunod.

- PEARSON J. (1998), *Terms in context*, Amsterdam ; Philadelphia, J. Benjamins.
- PICTON A. (2009), *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*, Thèse préparée sous la dir. d'Anne Condamines, Université Toulouse 2.
- REY A. (1976), Néologisme : un pseudo-concept ?, In *Cahiers de lexicologie*, vol. 28 n°1, p. 3-17
- REY A. (1992), *La Terminologie. Noms et notions*, Paris, PUF, coll. « Que sais-je ? ».
- SALEM A. (1988), Approches du temps lexical – Statistique textuelle et séries chronologiques, In *Mots*, n° 17, Paris, Presses de la FNSP, p. 105-143.
- SITRI F. (1998), *Un modèle d'objet de discours dialogique, entre thématization et reprise*, Thèse préparée sous la dir. de Sophie Moirand, Université Paris 3, 2 vol.
- TAMBA I., BOSREDON B. (1995), Titres de tableaux et noms propres, In M. Noailly (éd.), *Nom propre et nomination, Actes du colloque de Brest, 21-24 avril 1994*, Paris, Klincksieck, p. 123-135.
- VECCHI D. M. de (1999), *La Terminologie en entreprise – Formes d'une singularité lexicale*, Thèse préparée sous la dir. de John Humbley, Université Paris 13, 5 vol.
- VECCHI D. M. de (2007), La nomination et son « suivi » : entreprises et pragmatéterminologie, In *Neologica*, n° 1, p. 51-66.