



HAL
open science

Spécialisation générique et discursive d'une unité lexicale L'exemple de joggeuse dans la presse quotidienne régionale

Lucie Gianola, Mathieu Valette

► To cite this version:

Lucie Gianola, Mathieu Valette. Spécialisation générique et discursive d'une unité lexicale L'exemple de joggeuse dans la presse quotidienne régionale. JADT 2018, Jun 2018, Rome, Italie. hal-01833610

HAL Id: hal-01833610

<https://hal.science/hal-01833610v1>

Submitted on 9 Jul 2018

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.

Spécialisation générique et discursive d'une unité lexicale L'exemple de *joggeuse* dans la presse quotidienne régionale

Lucie Gianola¹, Mathieu Valette²

¹Université de Cergy-Pontoise – lucie.gianola@u-cergy.fr

²Institut National des Langues et Civilisations Orientales– mvalette@inalco.fr

Abstract

In this paper, we study the distribution of lexical items designating outdoor sport practitioners (*joggeur/joggeuse*, *randonneur/randonneuse*, *runneur/runneuse*, *promeneur/promeneuse*), in order to identify links between gender, semantic themes and genre across press discourse in French. The corpus is sampled from newspaper articles from regional newspapers. In press discourse, we observe a convergence between gender and genre through the actualized semantic classes.

Résumé

Nous étudions dans cet article la distribution d'unités lexicales désignant les pratiquant-e-s de sport de plein air (*joggeur/joggeuse*, *randonneur/randonneuse*, *runneur/runneuse*, *promeneur/promeneuse*) afin d'identifier les corrélations entre genres sexuels, thèmes sémantiques et genres textuels dans le discours journalistique en français. Le corpus est constitué à partir d'un échantillonnage d'articles de la presse quotidienne régionale. Il apparaît que dans le discours journalistique, on observe une convergence entre genres sexuels et genres textuels par le biais des classes sémantiques instanciées.

Keywords: Press discourse, textometrics, semantic class, genre, gender

1. Introduction

Nous proposons une étude de lexicologie textuelle sur la distribution d'unités lexicales choisies dans un corpus de textes de presse. L'étude n'a pas été réalisée dans une perspective *corpus-driven*, comme c'est souvent le cas en textométrie, mais avec une approche *corpus-based* (Biber, 2009) où les observables ont été prédéfinis. Notre objectif est en effet de nous focaliser sur les désignations des pratiquant-e-s de sport de plein air suivant une opposition en genres sexuels : *joggeur vs joggeuse*, *randonneur vs randonneuse*, *runneur vs runneuse*, *promeneur vs promeneuse*. Il s'agit d'identifier les corrélations entre genres sexuels, isotopies et genres textuels dans le discours journalistique de la presse quotidienne régionale française.

2. Problématique

2.1. Somme des isotopies de genres et de discours en signifiés

La lexicologie textuelle consiste en l'analyse du lexique à partir des conditions textuelles de sa production. Elle repose sur l'hypothèse selon laquelle les unités lexicales subissent un ensemble de contraintes intertextuelles et infratextuelles de la même nature que les formes sémantiques diffuses et non lexicalisées et qui en conditionnent les régimes de production et d'interprétation. Dans de précédents travaux, ont été proposées les conditions théoriques d'une analyse textuelle du lexique, principalement focalisées sur l'étude de la néologie sémantique – ou *néosémie*

(Rastier et Valette, 2009) et des formes sémantiques diffuses en voie de lexicalisation synthétique ou *protosémie* (Valette, 2010ab). Il s'agit ici d'étudier l'utilisation systématique d'une unité lexicale donnée dans un genre textuel précis et l'incidence de cette utilisation sur son sémantisme. En effet, tout mot placé dans un texte en reçoit des déterminations sémantiques, qui sont susceptibles de modifier son signifié (afférence de sèmes). Posant l'hypothèse selon laquelle le signifié est une forme sémantique lexicalisée (Valette 2010b), on considérera que les sèmes des isotopies du texte peuvent se propager au signifié d'une unité lexicale par le processus de *sommation* décrit par (Rastier, 2006). L'observation a pu être faite concernant les isotopies de domaine (redomanialisation d'une unité lexicale dans le cas de la néosémie par exemple) mais les isotopies génériques (relatives au genre textuel) ou discursives (relatives au discours) peuvent-elles transformer le signifié d'un mot de la même façon que les isotopies domaniales ? C'est à cette question que nous allons tâcher de répondre ici.

2.2. Présentation du corpus

Le corpus est donc constitué suivant deux axes, lexical et discursif : nous avons utilisé 8 formes considérées comme des mots-clés pour collecter des textes exclusivement issus du discours journalistique et, plus précisément, de la presse quotidienne régionale, sans considération de genre textuel. Le corpus a été collecté de manière semi-automatique à l'aide d'un script d'aspiration de pages web puis nettoyé et dédoublonné manuellement, afin d'écarter des articles constitués de reprises de dépêches AFP qui se retrouvent d'un titre à un autre. Le script, basé sur la commande Linux *cURL*, est alimenté par une liste d'URL collectées sur les sites des titres de presse à l'aide de requêtes effectuées sur le moteur de recherche Google (*site:nomdusite forme*, modulée par un inhibiteur *-blade* dans le cas de « *runner* » afin d'écarter les articles à propos du film *Blade Runner*). Entre 100 et 130 URL ont été collectées pour chaque forme. La phase de nettoyage a permis de supprimer les en-têtes, sommaires, liens annexes, légendes d'images, etc., pour ne conserver que le titre et le corps de l'article. Le corpus est organisé en huit sous-corpus correspondant aux 8 formes étudiées : *Joggeur*, *Joggeuse*, *Promeneur*, *Promeneuse*, *Randonneur*, *Randonneuse*, *Runner*, *Runneuse*, dont les statistiques sont présentées dans le tableau suivant.

Sous-corpus	Nombre de mots
<i>Joggeur</i>	40 671
<i>Joggeuse</i>	48 285
<i>Randonneur</i>	35 162
<i>Randonneuse</i>	31 931
<i>Promeneur</i>	44 497
<i>Promeneuse</i>	31 009
<i>Runner</i>	22 212
<i>Runneuse</i>	31 367
Total	285 134

Les articles sont issus principalement de titres de la presse quotidienne régionale comme *Nice Matin*, *Ouest-France*, *L'Est Républicain*, *La Dépêche du Midi*, *La Montagne*, *Corse-Matin*, *La Provence*. La collecte n'a pas été orientée sur une rubrique en particulier mais sur l'ensemble des titres, et nous n'avons pas défini de limite temporelle.

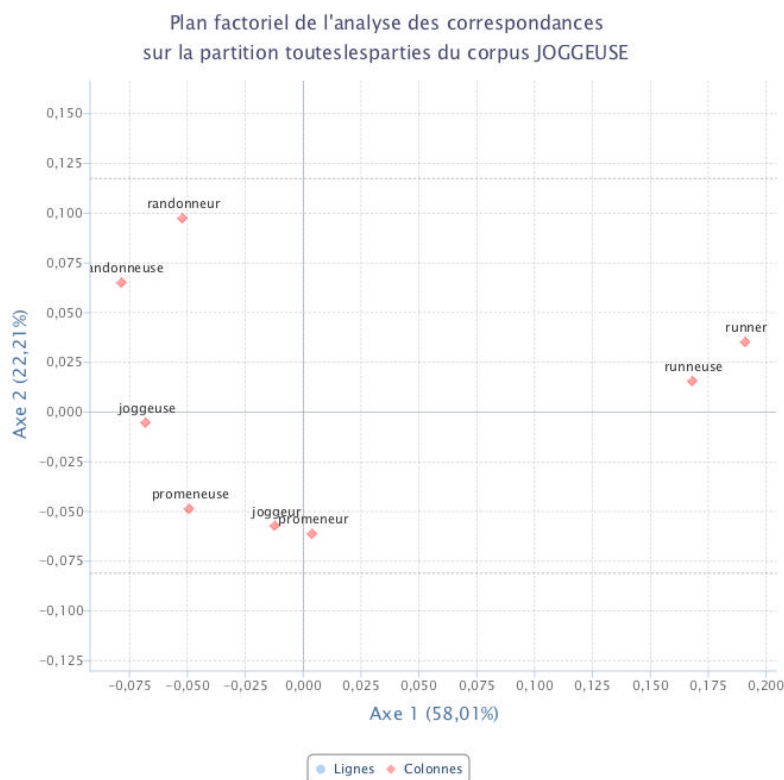


Figure 1 : Analyse factorielle des correspondances sur les parties du discours

3. Analyses¹

3.1. Observations générales

Une analyse factorielle préliminaire (figure 1) portant sur les seules parties du discours montre une opposition marquée sur l'axe 1 entre les sous-corpus *Runner* et *Runneuse* et les autres sous-corpus. Cet écart s'explique par les genres textuels des sous-corpus considérés. En effet, comme l'ont montré les travaux pionniers de (Biber, 1988) et, à leur suite, ceux de (Malrieu et Rastier, 2001), les variables locales que constituent les parties du discours sont des marqueurs de genre particulièrement stables. Ici, il apparaît que *Runner* et *Runneuse* relèvent du genre du compte rendu d'événements sportifs tandis que les 6 autres sous-corpus sont composés en grande majorité de faits divers. Autrement dit, la plupart des unités lexicales choisies pour nos requêtes, qui correspondent à des pratiques sportives de plein air, n'appartiendraient pas – ou alors à la marge – au vocabulaire des genres sportifs du discours journalistique.

L'analyse factorielle des correspondances sur les formes, dont la fréquence est au moins égale à 10 occurrences, offre à voir une distribution très différente. *Runner* et *Runneuse* sont toujours très proches mais il en est désormais de même de *Randonneur* et *Randonneuse* (désormais *Randonneur-se*) (figure 2). Les sous-corpus *Joggeur*, *Promeneur* et *Promeneuse* se situent à la croisée des axes et seront étudiés individuellement, mais *Joggeuse* se singularise.

¹ Le corpus a été analysé au moyen du logiciel de textométrie TXM (<http://textometrie.ens-lyon.fr/>) (Heiden *et al.* 2010).

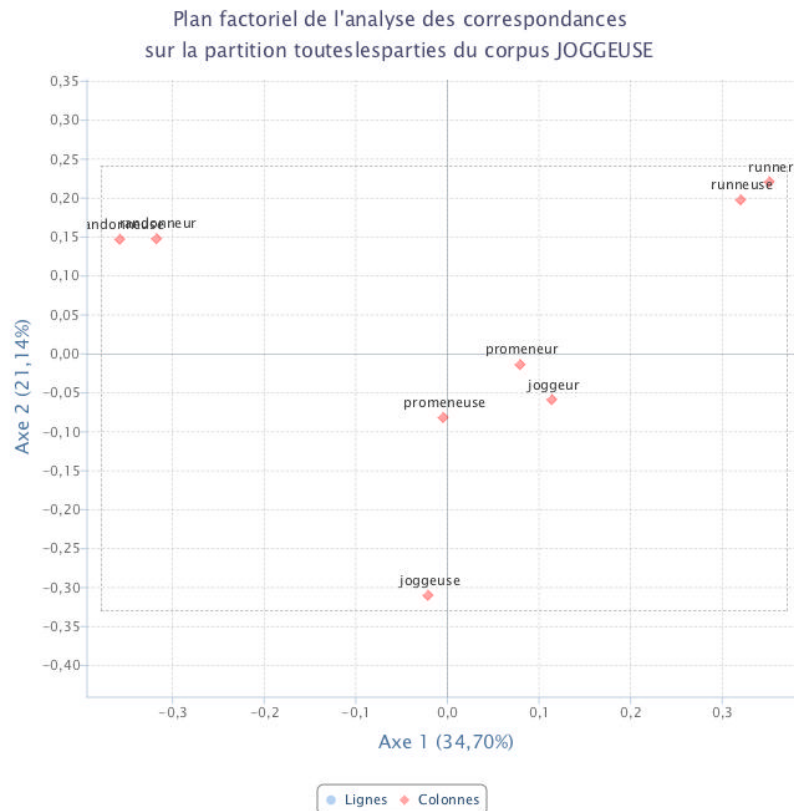


Figure 2 : Analyse factorielle des correspondances sur les mots formes

3.2. Analyses des classes sémantiques constituantes

L'analyse des spécificités (formes) des regroupements ainsi constitués nous indique les contextes d'instanciation des différentes formes.

Le regroupement a priori très homogène *Randonneur-se* offre à voir un vocabulaire associé aux *accidents de montagne*. Le corpus est structuré en 3 classes sémantiques principales,

- celle des accidents : « chute », « mortelle », « mètre », « avalanche », « fracture », « cheville », « hôpital », « blessée », « trauma », « glisser » etc.
- celle des disparitions : « disparu », « alerte », « retrouvé », « emporté », « inquiet », etc.
- celle des secours : « PGHM » (pour Peloton de gendarmerie de haute montagne), « hélicoptère », « Dragon » (un modèle d'hélicoptère) « évacué-e », « pompiers », « CRS », « secouriste », « secteur », « équipe », « sauveteur », « secourir », etc.

Le sous-corpus *Promeneur* et le sous-corpus *Promeneuse* relatent essentiellement 3 types d'événements :

- la promenade : « sentier », « phare », « littoral », « patrimoine », « chemin », etc.
- les accidents : accident de chasse essentiellement : « chasseurs », « chasse », etc.
- les découvertes : « macabres », « corps », « cadavre », « tronc », « jambe », « squelette », « ossement », « obus », « pépite », etc.

Le sous-corpus *Joggeur* ne comporte quant à lui qu'une classe sémantique principale, celle des accidents n'incluant pas de tiers humain : « arrêt, malaise, crise cardiaque », « algues vertes »,

attaques d'animaux (« rapace », « aigle », « buse »), sulfure d'hydrogène, H₂S, intoxication, toxique, gaz. Il est à noter que cette classe ne s'actualise pas dans le sous-corpus *Joggeuse*.

Les deux sous-corpus restants, le regroupement très homogène *Runner* et *Runneuse* (désormais *Runneur-se*) et *Joggeuse* méritent toute notre attention. D'un point de vue ontologique, le jogging comme le running sont des formes similaires de course à pied relevant du domaine du sport. Mais leur usage dans le discours journalistique diffère très sensiblement. Dans le regroupement *Runneur-se*, qui comporte, comme nous l'avons vu, essentiellement des articles relatant des événements sportifs, le vocabulaire est structuré autour des classes sémantiques suivantes :

- définitoire : hyperonyme « sport », synonyme « coureur », etc. Ainsi, le sous-corpus *Runneur-se* est le seul dont le sens correspond à la signification.
- classe de la compétition : « course », « marathon », « semi-marathon », « trail », « triathlon », « championnat », « inscription », « départ », « épreuve », « km », « victoire », « podium », « médaille », « sponsors », etc.
- classe des blessures : « blessure », « foulure », « ampoule », « contracture », etc.

Il comporte également deux classes sémantiques liées aux techniques associées à la pratique :

- classe des équipements : « équipement », « baskets », « chaussures », « brassière », « connectés », « GPS » ou « montre GPS », etc.
- classe des entraînements : « entraînement », « préparation », « fractionné », « cardio », « conseils », « performances », « yoga » (comme activité complémentaire destinée à éviter les blessures), etc.

Il est à noter que le sous-corpus *Runneuse* se singularise par la mention d'événements sportifs caritatifs liés à la lutte contre le cancer du sein : « octobre rose », « prévention ».

A l'inverse, la joggeuse dans le sous-corpus éponyme n'est nullement une sportive, mais sa caractérisation textuelle est remarquablement précise : elle est une femme agressée pendant son jogging et les classes sémantiques actualisées dans ce sous-corpus relèvent du crime, du droit et de l'enquête judiciaire :

- classe des agressions : « meurtre », « tentative », « agressée », « agression sexuelle », « viol », « enlèvement », « tuée »,
- classe des agresseurs : « homme », « suspect », « meurtrier », « présumé », « portrait-robot », « violeur », « exhibitionniste »
- classe des procédures judiciaires : « enquêteurs », « avocats », « cour », « procureur », « réquisition », « réclusion », « prison », « accusé », « interpellé », « agresseur », « condamné », « procédure », « instruction », « ADN », etc.

3.3. Synthèse

A l'issue de cette analyse, on choisit de se concentrer sur la définition en miroir de la *joggeuse* et de la *runneuse*, laissant de côté les autres unités lexicales détaillées ci-dessus. Les isotopies génériques et discursives qui constituent la trame sémantique des articles dans lesquels occurrent ces deux formes donnent lieu à la construction de deux signifiés antagonistes, par sommation :

La joggeuse apparaît :

1. /isolée/ (elle court seule),
2. /vulnérable/ (elle est sans défense face à un agresseur) et, quoi qu'il arrive, puisque le genre du fait divers l'exige,

3. /victime/ (elle est agressée, violée, tuée).

A l'inverse la runneuse est :

1. /entourée/ (elle court dans le cadre d'événement sportifs collectifs),
2. /sécurisée/ (par la technologie, notamment les montres GPS qui permettent de gérer l'effort et d'optimiser ses performances, par l'entraînement suivi. Les blessures subies apparaissent par ailleurs bénignes par rapport aux risques encourus par la joggeuse),
3. /compétitrice/ (elle participe à des compétitions).

4. Conclusion

Dans cet article, nous avons tenté de montrer comment les fonds sémantiques issus des genres et des discours pouvaient modifier, par sommation, les signifiés des unités lexicales qui sont utilisées. Pour deux unités lexicales partageant a priori un référent identique, celui d'une femme pratiquant la course à pied, l'actualisation en corpus journalistique fait émerger des contenus sémantiques très différents. Il ne s'agit pas de considérer que les joggeuses sont nécessairement des femmes en danger mais la régularité avec laquelle le mot joggeuse est actualisé dans la presse comme une /victime/, /vulnérable/ et /isolée/ pourrait avoir, à terme, une incidence sur la perception d'une pratique dont la réalité médiatique est exclusivement macabre. En d'autres termes, dans le discours de presse, pour les femmes, le jogging est une pratique dangereuse, la joggeuse une victime d'agression, alors que la runneuse une sportive impliquée dans des événements sociaux et le running une pratique sûre et valorisante.

Références

- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge, Cambridge University Press.
- Biber, D. (2009). Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In B. Heine and H. Narrog (editors) *The Oxford Handbook of Linguistic Analysis*, 159–191. Oxford.
- Heiden S., Magué J.-P., et Pincemin B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement, S. Bolasco. editors., *Journées internationales d'Analyses statistiques des Données Textuelles*, vol(2), 1021-1032.
- Malrieu, D. et Rastier, F. (2001). Genres et variations morphosyntaxiques, In *Traitements automatiques du langage*, 42, 2, 547-577.
- Rastier, F. (2006). Passages. In *Corpus*, 6, 125-152.
- Rastier, F., Valette, M. (2009). De la polysémie à la néosémie. In *Le français moderne*, vol. (77), 97-116.
- Valette, M. (2010a). Propositions pour une lexicologie textuelle. In *Zeitschrift für Französische Sprache und Literatur*, vol. (37): 171-188.
- Valette, M. (2010b). Méthodes pour la veille lexicale, In L. Messaoudi, et al. editors *Sur les dictionnaires*, Publication du laboratoire Langage et société, Université Ibn Tofail, Kénitra: 251-272.