



HAL
open science

Le défigement : approche hybride linguistique et informatique

Lichao Zhu, Gaël Lejeune

► **To cite this version:**

Lichao Zhu, Gaël Lejeune. Le défigement : approche hybride linguistique et informatique. Razky, Abdelhak; Sfar, Inès; Soutet, Olivier; Mejri, Salah. De la variation dans les langues : universaux partagés et idiomaticité dynamique, Letraria, pp.325-336, 2020, 978-65-86562-08-8. hal-04353668

HAL Id: hal-04353668

<https://hal.science/hal-04353668v1>

Submitted on 19 Dec 2023

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.

Le défigement : approche hybride linguistique et informatique

Lichao Zhu, Sorbonne Université

Gaël Lejeune, Sorbonne Université

Dans cet article, nous proposons un traitement hybride du défigement, en utilisant des outils et des méthodes empruntés à la linguistique et à l'informatique. L'objectif est double : linguistiquement, nous explorons l'aspect mécanique du défigement afin de traiter le défigement purement linguistique, celui de la langue ; informatiquement, nous présentons un prototype de générateur de candidats au défigement.

1. Introduction

Les points de vue des linguistes divergent au sujet du défigement. Ce terme, qui n'est pas exclusif du domaine linguistique, est perçu tantôt comme un procédé tantôt comme un phénomène. En effet, si toute opération de désolidarisation volontaire constitue un défigement, le phénomène du défigement peut se résumer à un ensemble des opérations réalisables ayant l'intention de détourner les formes figées dans la langue et dans le discours. De ce point de vue, nous ne pouvons pas dissocier le procédé (opérations linguistiques) et le phénomène qui se réunissent dans la notion de défigement. Néanmoins, les « opérations défigeantes », c'est-à-dire l'ensemble des manipulations à caractère transformationnel, qui consistent à « resémantiser » une séquence figée, sont limitées linguistiquement. Nous énumérerons dans cet article une dizaine de procédés linguistiques récurrents qu'utilisent les locuteurs, à travers l'analyse d'exemples tirés d'un corpus de défigements¹. Parmi ces procédés, la substitution est le plus utilisé : plus des deux tiers des opérations défigeantes ont recours à la substitution. Notre traitement ici donc cette opération en particulier.

Les travaux sur le défigement proprement parlé sont assez rares en TAL définis avant, alors que le figement suscite de plus en plus d'intérêt en traitement automatique, notamment parce qu'il est lié à la question de détection des unités multi-mots (multi-word units (MWU) ou multi-word expression (MWE)) et plus généralement à la problématique du découpage en unités analysables ou « tokenisation »². De l'autre côté, le défigement est relégué au second plan et ne sert que comme accessoire pour vérifier la fixité et la solidité d'une séquence figée. La seconde raison derrière ce désintérêt est sans doute un manque de méthodes transposables, implantables, informatiquement. Il n'y a pas de méthodes informatiques robustes pour traiter le défigement, en raison de l'imprévisibilité des opérations « défigeantes »³. Les formes multiples de notre objet d'étude qu'est le

¹ Un corpus constitué de 450 exemples tirés du *Canard enchaîné*, classés et décrits linguistiquement (Zhu, 2013).

² En effet, la tokenisation sépare les séquences figées en différents éléments.

³ Cette forme participiale du verbe « défiger » est utilisée pour pallier l'absence de la forme adjectivale dans la description linguistique du défigement.

défigement troublent et dissuadent en quelque sorte la communauté scientifique, par souci de l'exhaustivité et la validité universelle du traitement.

Nous proposons dans cet article un regard croisé, né dans les multiples discussions et les maintes tentatives de modélisation des séquences défigées. Nous présentons ici la première ébauche de notre travail.

2. Caractère scalaire et contraintes du défigement

2.1. Formes multiples des séquences défigées

La caractéristique scalaire (Zhu 2016 ; Mejri 2013) du figement se manifeste à travers les diverses formes de séquence figée exige un traitement lexical global. Partant de la combinaison libre d'unités lexicales à une unité polylexicale ayant la fixité absolue, les séquences de degrés de fixité différents cristallisent chacune à leur tour des traits de fixité différents (morphologiques, syntaxiques, sémantiques et pragmatiques). Par conséquent, le défigement, en lien direct avec le figement, est également protéiforme.

Les slogans publicitaires, les mêmes (Gautier et Siouffi 2016) sur les réseaux sociaux, etc, exploitent souvent la fixité à caractère extralinguistique sous forme de "moule locutionnel" (Zhu 2016 et 2017). Par exemple,

- Touche pas à mon Despote !
- Touche pas à mon porc !
- Pas touche à l'aspartame !
- Touche pas à ma bonne !

sont des défigements résultant du slogan de l'association « SOS racisme » : « Touche pas à mon pote ». D'autres formes sont plus ancrées dans la langue telles que les formes complexes qui sont complètement figées : "*Prendre le taureau par les cornes*", "*Avoir un chat dans la gorge*", "*Tirer les vers du nez*" ou encore "*Les carottes sont cuites*", "*Les dés sont jetés*" et "*Chassez le naturel, il revient au galop*". La plupart des séquences figées, comme le remarquent Lamiroy et Klein (2008), sont des séquences « semi-figées » qui sont sémantiquement déductibles et syntaxiquement ordinaires telles que "*lever le pied*", "*faire d'une pierre deux coups*", etc qui traduisent les degrés de fixité syntactico-sémantiques variables. Or, la banalité syntaxique d'une séquence figée ne signifie pas que son sens est banal et transparent, et vice-versa. *Poser un lapin* est un syntagme verbal ayant une construction banale : V Dét N, mais il est sémantiquement opaque; "*Chat échaudé craint l'eau froide*" est syntaxiquement atypique (le nom *chat* n'est pas précédé d'un déterminant), mais sémantiquement transparent : son sens global peut être déduit du sens de ses composantes.

2.2. Contraintes du défigement

Les liens étroits entre le figement et le défigement se manifestent d'abord par des contraintes successives à différents niveaux linguistiques. La prédominance de la substitution dans les procédés du défigement s'explique par la facilité de la réalisation de cette opération et son faible coût cognitif. La grammaticalité est l'un des indicateurs du « coût » de la substitution. Il y a fort à parier que dans un procédé de substitution du défigement, la majorité des locuteurs substituent une unité par une autre ayant la même catégorie grammaticale, de façon à ce que la reconstruction ne soit pas coûteuse cognitivement et que le lien entre la séquence défigée et sa source soit immédiatement reconnu. Cette contrainte grammaticale permet d'éviter tout conflit d'interprétation, les amalgames lexicaux et les

syntagmes et phrases agrammaticaux⁴, ce qui rend le traitement de la séquence plus économique cognitivement parlant et donne la possibilité de limiter le nombre de “candidats” au défigement.

Indépendamment de leur composition lexicale, les séquences figées sont par définition reconnues par les locuteurs. La question qui se pose est de savoir ce qui est formellement reconnaissable et ce qui est défigeable pour un coût cognitif acceptable, car la forme et le sens sont tous les deux des éléments discriminants dans les opérations défigeantes.

Par ailleurs, on constate que les choix du locuteur sont limités. Prenons l'exemple de “prendre le taureau par les cornes”. Pour prouver la solidité de cette séquence, il serait naturel pour un locuteur de remplacer « taureau » par une autre espèce bovine, ou un autre mammifère — un terme qui partagerait le même archiséme avec « taureau ». Or, ce procédé est assimilable à une « faute » selon Gross (1996, 19-20) : « *Le seul jeu possible consiste à introduire dans un domaine d'arguments, caractéristique d'un prédicat donné, un substantif qui ne fait pas partie de la classe sémantique en question.* » Ainsi, *prendre le watusi par les cornes, prendre le bouc par les cornes, prendre le rhinocéros par les cornes...* sont assimilables à des fautes de langue par les locuteurs, car ils partagent des traits sémantiques avec le « taureau » : la proximité sémantique entre ces unités est telle que la permutation synonymique pourrait être interprétée comme une lacune dans la maîtrise de la langue⁵. C'est pourquoi le défigement est par définition intentionnel et que l'intention se manifeste très souvent par le biais du contexte du défigement, car la forme simple d'une séquence défigée est insuffisamment caractérisée pour démontrer l'intentionnalité et la compétence langagière du locuteur. *A contrario*, le locuteur doit fournir un contexte, nécessairement discursif, pour prouver ses compétences langagières et persuader son interlocuteur qu'il s'agit là d'un défigement volontaire non d'une faute.

D'autres formes du défigement sont également susceptibles de mettre en doute les compétences du locuteur. Les jeux phoniques tels que le métaplasme n'échappent pas non plus à la suspicion. Ils sont acceptables sémantiquement, car ils ne visent pas *a priori* le sens global de la séquence, mais visent la phonie d'une composante individuelle de la séquence. Les séquences défigées (potentielles) de *prendre le taureau par les cornes* qui sont susceptibles d'être créées, à l'exemple de “prendre le tarot par les cornes” sont sujettes d'un jeu phonique plutôt d'un jeu sémantique, ce qui lui permet d'être mieux reconnue en tant que séquence défigée bien que le coût cognitif pour “trouver un sens” à la séquence défigée soit important. En effet, une opération défigeante fondée sur les ressemblances phoniques se détacherait du sens et révélerait l'intentionnalité du jeu qui ne serait pas d'ordre sémantique.

2.3. Raisonnement par l'absurde

Le raisonnement par l'absurde défini comme “à l'intérieur d'une théorie donnée, une certaine thèse 0, démontre que la négation de cette thèse entraîne, au terme d'un certain nombre d'inférences, soit deux conséquences, a et non-a, contradictoires l'une de l'autre, soit deux conséquences, a et b, dont on connaît simplement l'incompatibilité logique.” (Gardies 1991, 9), semble être le biais auquel

4 Nous évoquons les défigements agrammaticaux dans §3.

5 “Le watusi”, “le bouc” et “le rhinocéros” remotivent la séquence auparavant figée et occultent la fixité de la séquence.

les locuteurs ont majoritairement recours pour sortir des normes imposées par la langue.

D'un point de vue sémantique, cette déconnexion de la logique justifie en quelques sortes la validité du principe de la pertinence. Comme l'expliquent Dan Sperber et Deirdre Wilson : " Tout acte de communication ostensive communique la présomption de sa propre pertinence optimale" (Sperber & Wilson 1989, 237), un défigement "réussi" présuppose donc une intentionnalité qui est toujours pertinente de la part du locuteur. En termes d'économies du langage, le défigement est par définition économique cognitivement et doit se faire comprendre par l'interlocuteur à la fois au niveau des opérations effectuées et au niveau de l'intention. Dans l'esprit du concept de "Fonction muette du langage" formulé par Jacques Coursil (Coursil 2000), on peut considérer que c'est le sujet qui écoute qui défige. Par conséquent, il n'existe pas de défigement sans reconnaissance par un locuteur de l'aspect défigé, et donc intentionnel, de la séquence produite. Le défigement n'existe que par ce qu'il est reconnu comme tel. Il faut donc à la fois que le sujet entendant ait connaissance de la séquence figée de référence et puisse reconnaître le procédé de transformation.

Ainsi, le défigement se débarrasse des normes langagières et entre dans les normes de l'absurde, il devient donc un raisonnement ostensif. "Le raisonnement par l'absurde, une fois retourné en raisonnement ostensif, se diviserait donc en une double démarche dont l'une procéderait à son tour par l'absurde. Cet effet ne surprendra jamais que celui qui aurait mal observé le déroulement du raisonnement apagogique⁶ (...). Sans doute celui-ci se clôt-il par deux résultats contradictoires l'un de l'autre ; mais, si ces deux résultats sont contradictoires, c'est parce que l'auteur de la démonstration établit la vérité de l'un, tandis qu'il ne fait jaillir l'autre que comme une conséquence de la supposition qu'il se propose de réfuter." (Gardies 1991, 46). La réfutation de la fixité conventionnelle entraîne par conséquent la coexistence de deux formes : l'une figée, qui est en filigrane, l'autre défigée, qui est en saillance. Ainsi, le défigement contraint l'interprétant à combiner les sèmes des composantes de la séquence défigée, en concordance avec le contexte ; parallèlement, un rappel systématique à la séquence défigée engendre une seconde somme des sèmes afférents (Nunberg et al. 1994, Zhu 2016). Ces deux procédés de calcul se déroulant simultanément interfèrent l'un avec l'autre dans l'interprétation de la séquence défigée.

Il en ressort que l'intentionnalité et l'absurde seraient des conditions *sine qua non* du défigement. Lorsque l'on évoque une séquence dite « défigée », on présuppose que l'opération à laquelle on recourt se distingue de manière catégorique de la faute qui est, elle, non intentionnelle⁷. Reste à savoir si tous les défigements volontaires sont systématiquement reconnus par le locuteur.

3. Candidats patents au défigement

D'un côté, la rupture des raisonnements normatifs est tout d'abord sémantique avant d'être lexicale. Considérons l'énoncé suivant tiré de *La Rabouilleuse* de Balzac :

⁶ Le raisonnement apagogique part de la proposition elle-même pour ensuite la réfuter, le raisonnement ostensif part de la négation d'une proposition pour la réfuter.

⁷ Les fautes ne sont pas produites inconsciemment, selon certains spécialistes. Frei (2011) démontre linguistiquement (fautes de langue) et Freud (2010) psychologiquement (les lapsus, les erreurs, etc.) certaines corrélations entre les normes et les formes non normatives.

Les bons comtes ont les bons habits, lui dit alors son ami Léon de Lora qui, malgré sa célébrité comme peintre de paysage, n'a pas renoncé à sa vieille habitude de retourner les proverbes, et qui répondit à Joseph à propos de la modestie avec laquelle il avait reçu les faveurs de la destinée : « Bah ! La pépie vient en mangeant ! » (H. Balzac, *La Rabouilleuse*⁸, p. 526)

Dans cet exemple, "*les bons comtes ont les bons habits*" est une référence au proverbe « Les bons comptes font les bons amis. » et *La pépie vient en mangeant*. se réfère à « L'appétit vient en mangeant. ». Les deux illustrent l'usage de l'absurde. La manipulation phonique avec laquelle que Balzac crée la séquence défigée s'intègre parfaitement tant au plan grammatical qu'au plan rythmique dans le texte⁹. Ainsi, "*prendre le canapé par les cornes*" ou "*prendre le taureau par les racines*" sont plus « aptes » à être considérés comme des défigements, car aucun lien ontologique peut *a priori* s'établir entre "taureau" et "canapé", "cornes" et "racines", ce qui dispenserait une justification dans le sens de la logique du système langagier.

De l'autre, la hiérarchisation des concepts (Tamba 1991) conduit à un nombre de choix limités dans les opérations défigeantes. Il nous semble que la prémisse d'une telle action est toutefois basée sur un consensus « primitif » autour du lexique. Les liens qui existent entre les composantes d'une séquence figée de natures lexicale, syntaxique et sémantique reposent sur des liens « intériorisés » dans la langue. L'absurde motive l'intentionnalité des opérations défigeantes et permet au locuteur de surprendre son interlocuteur. En effet, dans un processus communicationnel traditionnel, le locuteur s'attend à ce qu'un énoncé respectant les normes langagières se distingue nettement des pratiques normatives et non normatives (les fautes). Quelles en sont les frontières ? Y a-t-il des candidats patents au défigement qui se révèlent uniquement par le biais du lexique ? Ce sont des questions que nous nous posons et auxquelles nous souhaitons apporter des éléments de réponse de façon expérimentale.

Il nous semble plausible que certains défigements sont interprétables d'office, d'un point de vue lexical, sans un contexte discursif propice, présupposant que l'interlocuteur maîtrise la langue et crée intentionnellement un nouveau schéma d'interprétations, différent de celui de la séquence figée. Ainsi,

1. L'aboi de la jungle (la loi de la jungle) (Canard enchaîné, n°4710)
2. Vermine de rien (mine de rien) (Canard enchaîné, n°4693)
3. Absents pour cent (cent pour cent) (Canard enchaîné, n°4736)

sont des candidats patents, car outre les ressemblances phoniques ("l'aboi" et "la loi", "vermine" et "mine" et "absents" et "cent"), la substitution crée un schéma d'interprétations qui est fondamentalement différent des schémas d'interprétations (global et compositionnel) de la séquence figée. Si la grammaticalité est préservée dans l'exemple 1, les exemples 2 et 3 sont agrammaticaux et les jeux de mots développés sont plus sémantiques que phoniques, ce qui rend évidente l'intention de vouloir défiger. Sans parler du fait que ces exemples sont tirés du journal satirique du *Canard enchaîné*, ce qui incite le locuteur à les interpréter par le raisonnement par l'absurde.

4. Expérience

⁸ *La rabouilleuse*, Garnier Édition, 2008 (1842).

⁹ Comme nous l'avons démontré dans ce paragraphe, le « coût » est important en termes de discours, tandis que le recours à l'absurde (qui présuppose et par conséquent met en valeur la maîtrise de la langue) se défait de la suspicion d'une lacune langagière.

Notre expérience est tout d'abord fondée sur un calcul de proximités phoniques et formelles, en considérant les séquences figées comme des tokens au sein desquels on peut effectuer des substitutions. En procédant à des substitutions, notre ambition est également de savoir à partir de quel degré de substitution le résultat sort des normes langagières et entre dans l'absurde.

L'expérience se déroule en deux temps : d'abord, nous explorons l'aspect mécanique du figement. Notre programme substitue des mots pleins au sein de séquences figées. Par exemple, pour la séquence *il est beau comme le jour*, nous avons généré des séquences défigées comme suit :

- il est beau comme le gour
- il est beau comme le jour
- il est beau comme le joug

La génération de nouvelles séquences est ainsi fondée sur la modification d'un élément (un ou plusieurs mots) de la chaîne du traitement par son équivalent appartenant à la même catégorie grammaticale et possédant une similarité phonique. En nous fondant sur les données de la littérature ainsi que sur les ressources manipulables, nous avons choisi de n'effectuer des substitutions que pour certaines catégories grammaticales, en excluant le déterminant et le pronom. Des parties de discours tels que le nom, le verbe et l'adjectif sont susceptibles d'être remplacés par l'une de ces catégories pour ainsi créer des défigements candidats. Nous effectuons par ailleurs la substitution sur plusieurs paradigmes simultanément.

5. Outils et chaîne de traitements

Les outils que nous présentons ici nous permettent d'identifier les candidats à la substitution les plus pertinents, ci-après *substituts*, en tenant compte de leur proximité grammaticale, graphique et phonétique avec un mot de la séquence à défiger.

Nous utilisons d'abord l'étiqueteur morpho-syntaxique Treetagger (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>) qui a été développé par Helmut Schmid à l'Institut de linguistique informatique de l'Université de Stuttgart. Cet outil nous permet d'identifier pour chaque mot d'un lemme (la forme dictionnaire), sa classe grammaticale et ses informations morpho-syntaxiques telles que le temps (pour les verbes), le nombre et le genre (pour les noms), etc. Ceci nous permet de conserver la grammaticalité des séquences défigées, en privilégiant les substituts de la même catégorie grammaticale.

Nous utilisons également la bibliothèque Phonex¹⁰ qui nous permet d'attribuer un score phonétique aux mots de la séquence. Nous pouvons ainsi comparer deux chaînes en comparant leur score phonétique associé. La substitution se faisant selon la proximité phonique entre le mot substitué et les mots candidats substituants, nous pouvons donc déterminer le token à substituer.

Nous exploitons enfin la distance de Levenshtein (Levenshtein 1966) qui nous permet de comparer la proximité graphique entre deux mots. Cette distance mesure le nombre minimal d'opérations (insertion, suppression ou substitution) qu'il faut effectuer en partant du graphique du mot A pour arriver au mot B. A titre d'exemple, pour passer du mot "Chine" au mot "Chinois", il faut substituer un "o" au

¹⁰ <https://github.com/spronkey/phphonex>

“e” puis insérer un “i” et un “s”, la distance de Levenshtein entre ces deux mots est donc de 3. Afin de faciliter la comparaison entre les mots de différentes tailles, nous normalisons cette distance en la divisant par la taille de la chaîne de caractères du token à substituer, et obtenons une mesure relative. Ainsi, est dans l'exemple ci-dessus (entre « Chine » et « Chinois ») la distance relative de 0,6. Plus la distance tend vers 1, plus les mots sont considérés différents ; plus elle tend vers 0, plus les mots sont considérés comme proches.

Pour chaque mot de la chaîne donnée en entrée, nous vérifions qu'il appartient bien à l'une des catégories grammaticales qui nous intéressent. Puis nous déployons les substitutions possibles qui valident au moins une des deux conditions suivantes : une différence de score Phonex inférieure ou égale à 0,002¹¹ ou une distance de Levenshtein d'édition inférieure ou égale à 0,2 (20%).

Nous construisons ensuite toutes les séquences défigées possibles avec les substituts validés à l'étape précédente. Chacune de ces combinaisons reçoit un score qui est la somme des scores des substituts. Le score de chaque élément substitué est obtenu comme suit :

$$\text{— scorePhonex} * (\text{Levenshtein} + 1)$$

Nous conservons les 20 propositions ayant les meilleurs scores, qui sont considérés comme les meilleurs candidats patents au défigement. Le principal problème rencontré dans cette méthode est que le calcul de la distance de Levenshtein est coûteux. Toutefois, cette étape a un effet très positif sur la qualité des candidats générés. Nous proposons ci-dessous un aperçu de candidats pour la séquence “il prend le taureau par les cornes” :

0.000000 : il prend le taureau par les cornes
0.000000 : il prend le taureaux par les carnes
0.000000 : il prend le taraud par les cornes
0.000000 : il prend le taraud par les carmes
0.000000 : il prend le taraud par les carnes
0.000000 : il prend le porreau par les cornes
0.000000 : il prend le porreau par les carmes
0.000000 : il prend le porreau par les carnes
0.000000 : il prend le taureaux par les carmes
0.100379 : il prenne le tares par les cortex
0.100379 : il prenne le taures par les cortex
0.100512 : il prenne le taures par les cardes
0.100512 : il prenne le taures par les carpes

Nous remarquons que la génération automatique de notre outil est d'autant plus efficace si le mot remplacé se trouve à la fin de l'expression. En effet, l'expression figée d'origine est plus facilement identifiée dans cette configuration et le jeu de défigement plus facilement repérable par le locuteur. Un autre aspect important est que si le mot substitué est un mot tabou, l'effet que provoque l'absurde est beaucoup plus frappant.

Conclusion

¹¹ Le choix de ce seuil est quelque peu arbitraire mais s'est avéré pertinent dans les différents essais que nous avons menés.

Nous avons cherché dans notre expérience à nous démarquer des observations « routinières », à partir des mécanismes purement syntaxiques ou phoniques du défigement pour mieux réfléchir aux mécanismes des opérations défigeantes et à leurs limites. En explicitant les contraintes innées du défigement, nous réduisons le coût du traitement et explorons les possibilités des défigements cognitivement productifs et utiles. Les candidats patents au défigement existent bel et bien à condition que les locuteurs respectent le principe de la pertinence et assurent l'intercompréhension.

L'approche hybride que nous adoptons permet la génération automatique des séquences par substitution phono-morphologique. Toutefois, les candidats générés ne sont pas tous intéressants et génèrent des « bruits » si l'on emploie le terme utilisé en traitement automatique des langues, pour désigner les résultats non-pertinents (ou faux positifs) générés par un programme informatique. Il nous semble donc nécessaire de mettre en place les filtres grammaticaux et sémantiques plus poussés, afin d'être en mesure de sélectionner une plus grande proportion des candidats patents au défigement.

Bibliographie

Cabasino, F. 1999. « Défigements et contraintes syntaxiques », Cahiers de lexicologie, 74, p.99-147.

Coursil, J. 2000. « La fonction muette du langage : essai de linguistique générale contemporaine », Ibis Rouge.

De Foucault, B. 1988. *Les structures linguistiques de la genèse des jeux de mots*, P. Lang.

Eline, J. et Zhu, L. 2014. « Défigement et inférence - cas d'études du Canard enchaîné », avec Joël Eline, *Actes du 4^{ème} Congrès Mondial de Linguistique Française*, SHS Web of Conferences, p. 681- 695.

Frei, H. 2011. *La grammaire des fautes*, PU Rennes, première version en 1929.

Freud, S. 2010. *La psychopathologie de la vie quotidienne sur l'oubli, le lapsus, le geste manqué, la superstition et l'erreur*, Gallimard.

Gardies, J.-L. 1991. *Le raisonnement par l'absurde*, PUF.

Gautier, A. et Siouffi, G. 2016. « Les mêmes langagiers. Propagation, figement, déformation ». *Travaux de Linguistique*, 2 (73), De Boeck Supérieur.

Gross, G. 1996. *Les expressions figées en français : noms composés et autres locutions*, Ophrys.

Grunig, B. 1990. *Les mots de la publicité*, Edition du CNRS.

Habert, B. et Fiala, P. 1989. « La langue de bois en éclat : les défigements dans les titres de presse quotidienne française », *Mots*, vol. 21, no 1, p. 83-99.

Lamiroy B. et Klein J. R. 2005. « Le problème central du figement est le semi-figement », *Linx*, 53, p. 135-154.

- Lejeune G. et Dumonceaux F. 2015, Une approche stylométrique pour la fouille d'opinion, DEfi Fouille de Textes, TALN, p. 12-15.
- Levenshtein, V. I. 1966. "Binary codes capable of correcting deletions, insertions and reversals". Soviet Physics Doklady, 10 (8), p. 707-710.
- Martin, R. 1976. *Inférence, antonymie et paraphrase : éléments pour une théorie sémantique*, Librairie C. Klincksieck.
- Mejri, S. 2013. « Figement et défigement : problématique théorique », *Pratiques*, 159-160, p. 79-97.
- Nunberg G., Sag I. A. et Wasow T. 1994. « Idioms », *Language*, vol. 70, n° 3, p. 491-538.^[1]_[SEP]
- Rastier, F. 1997. « Défigements sémantiques en contexte », *La locution, entre langues et usages*, p. 305-329.
- Sperber, D. et Wilson, D. 1996. *Relevance : Communication and Cognition*, 2^e éd., Wiley-Blackwell.
- Tamba I. 1991. *Organisation hiérarchique et relations de dépendance dans le lexique*. In: *L'Information Grammaticale*, n° 50, p. 43-47.
- Weinreich, U. 1969. « Problems in the analysis of idioms », *Substance and structure of language*, p. 23-81.
- Zhu, L. 2016, « Pour une notion de moule dans le figement », G. Dotoli (dir.), *Les Cahiers du dictionnaire, Classiques Garnier*, n°8, p. 97-109.
- Zhu, L. 2013. « Typologie du défigement dans les médias écrits français », thèse, Université Paris 13.