

Les prépositions complexes en français : pour une méthode d'identification multicritère

Dejan Stosic, Benjamin Fagard

▶ **To cite this version:**

Dejan Stosic, Benjamin Fagard. Les prépositions complexes en français : pour une méthode d'identification multicritère. Revue Romane, 2019, Les prépositions complexes dans les langues romanes. halshs-02194104

HAL Id: halshs-02194104
https://shs.hal.science/halshs-02194104

Submitted on 25 Jul 2019

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.

****version auteur avant épreuves****

Les prépositions complexes en français : pour une méthode d'identification multicritère

Dejan Stosic & Benjamin Fagard

Université Toulouse Jean Jaurès & CLLE-ERSS (CNRS & UT2J) / Lattice (CNRS, ENS & Sorbonne Nouvelle, PSL & USPC)

Abstract

In this paper, we develop a new method of identification of complex prepositions, in French. We include well-known semantic and morpho-syntactic tests, and introduce a few others – together, these tests make up a multi-variable grid which we believe can help identify complex prepositions. We run these tests on a list of 75 sequences, most of which are considered to be complex adpositions. However, we also include sequences which share the same patterns of word formation but do not have prepositional uses, in order to check the validity of our grid. Our results show that the semantic and morpho-syntactic clues used to identify complex adpositions, combined with corpus data exploration, are on the whole quite effective, but to varying degrees depending on patterns of formation.

Keywords: Prepositions, Complex adpositions, Lexicalization, French, Corpus analysis

Introduction

Contre toute attente et en dépit de nombreux travaux sur la question, le premier défi à relever pour un linguiste travaillant sur les prépositions complexes est d'en dresser la liste. Cet article s'inscrit dans un projet à long terme visant à produire une telle liste pour le français moderne. La difficulté tient d'une part au fait qu'il s'agit d'une sous-classe de mots particulièrement difficile à définir, et pour laquelle il n'y a pas encore de consensus ; d'autre part, à leur nombre : tandis qu'on compte généralement quelques douzaines de prépositions simples dans les langues qui ont cette catégorie, les estimations sont bien plus élevées pour les prépositions complexes – plusieurs centaines au moins (voir par exemple Borillo 1991, 1997 et Le Pesant 2006 pour le français, Huddelston & Pullum 2002 pour l'anglais). Ceci explique sans doute en partie pourquoi il y a si peu de listes de prépositions complexes en français ; nous avons tout de même pu prendre comme point de départ les quelques listes existantes (Borillo 1991, 1997, Melis 2003, Fagard, Stosic & Mardale 2016, Stosic en préparation), que nous avons confrontées aux données d'un corpus de langue moderne, incluant des textes littéraires et non littéraires, constitué à partir des bases Frantext et frTenTen.

L'article s'organise en quatre parties. Dans la première section, nous revenons sur la définition de la classe des prépositions. Partant d'un aperçu des principales caractéristiques de la préposition simple (§ 1.1), nous proposons une caractérisation morphologique,

distributionnelle et sémantique de la préposition complexe (§ 1.2). La deuxième section est consacrée aux critères d'identification des prépositions complexes. Après un rapide rappel de leurs principaux schémas de construction (§ 2.1), nous discutons les critères d'identification couramment utilisés dans la littérature et élaborons une nouvelle méthode d'identification basée sur 21 tests d'ordre qualitatif et quantitatif, en recourant pour ces derniers à deux grandes bases textuelles (§ 2.2). Dans la troisième partie, la grille en question est mise à l'épreuve de 75 expressions, dont la plupart appartient à la classe des prépositions et un tiers aux formations libres. Dans la quatrième et dernière partie, nous proposons une discussion des résultats obtenus.[1]

1. Définir la préposition : de la préposition simple à la préposition complexe

Pour définir une classe de mots, il est d'usage de faire appel de façon concomitante à trois types de critères : distributionnel (ou syntaxique), morphologique et sémantique. Cela permet de prendre en compte un large éventail de propriétés linguistiques lorsqu'il s'agit non seulement de circonscrire les contours d'une catégorie de mots mais aussi de classer une unité lexicale ou grammaticale dans telle ou telle catégorie. Cette méthode d'identification est censée permettre de circonscrire toute catégorie de mots, y compris celle des prépositions. L'essor des recherches sur les prépositions dans différentes langues les quatre dernières décennies a permis d'en saisir les principaux traits de fonctionnement distributionnels, sémantiques et morphologiques et d'aboutir ainsi à une définition assez consensuelle de la catégorie en question (cf. § 2.1. ci-dessous). Cependant, comme mentionné dans l'introduction, tout un pan de la problématique reste flou et insuffisamment exploré. En effet, si les prépositions dites « simples » (du point de vue morphologique) semblent plutôt bien répertoriées et bien couvertes par la définition, c'est loin d'être le cas des prépositions morphologiquement complexes. Celles-ci, d'une part, défient les contours de la classe et la définition elle-même, d'autre part, s'avèrent extrêmement récalcitrantes à un relevé exhaustif, les deux difficultés étant étroitement liées. En outre, autant les prépositions simples constituent une liste à très faible évolution, véhiculent des sens variés et souvent abstraits, et penchent donc du côté de la grammaire par leur fonctionnement global, autant les prépositions complexes sont nombreuses et mal répertoriées, moins polysémiques et intègrent une composante lexicale très nette. Tout en gardant le rôle et la distribution caractéristique des prépositions au niveau syntaxique, les prépositions complexes s'éloignent clairement par les propriétés citées des membres prototypiques de la classe. La définition globale s'avère ainsi inopérante pour une grande part de prépositions complexes, ce qui empêche de fait leur relevé, et par voie de conséquence, leur description. Nous aborderons les deux problèmes respectivement dans les sous-sections 1.1. et 1.2.

1.1. Prépositions simples : aperçu

La préposition est habituellement définie comme un mot invariable qui établit un lien de dépendance entre le syntagme dont elle est la tête et un autre constituant. Aussi bien l'élément introduit – le complément/régime de la préposition –, que l'élément « recteur » – élément auquel se rattache la préposition – peuvent être de nature variée. En effet, le terme recteur

[1] Nous tenons à remercier Basilio Calderone pour son aide dans le traitement statistique des données.

peut correspondre à un verbe (ex. ***compter** sur ses amis*), à un nom (ex. ***un cadeau** pour toi*), à un adjectif (ex. ***dépendant** de l'ordinateur*), à un adverbe (ex. ***parallèlement** à la visite du président*) ou bien à une proposition (ex. *malgré tous ses efforts, **le problème n'est pas résolu***). En ce qui concerne le complément de la préposition, la diversité est encore plus grande. En français, le SN est le régime prototypique de la préposition (ex. *devant **sa maison**, fier de **Marie**, fâché contre **elle***). Il existe cependant nombre d'autres expressions et structures susceptibles d'être introduites par la préposition : un nom nu (ex. *prendre quelqu'un pour **guide***), une relative sans antécédent (ex. *c'est clair pour **qui sait lire***), un adverbe (ex. *pour **demain***), un infinitif (ex. *tout faire pour **décrocher***), une proposition subordonnée introduite par *que* (ex. *tout faire pour **qu'ils partent***), une proposition subordonnée circonstancielle (ex. *ce sera fait pour **quand tu reviendras***) ou un syntagme prépositionnel (ex. *le sujet est à faire pour **dans trois semaines***). La préposition apparaît ainsi comme un « relateur », à savoir un élément de relation (Pottier 1955) qui n'est rien d'autre qu'une marque de dépendance d'un élément A (recteur) par rapport à un élément B (complément). Le fait que dans son fonctionnement la préposition mette en jeu une structure ternaire (A R° B) est un élément de définition fondamental pour cette classe des mots[2] (cf. De Mulder & Stosic 2009).

La relation établie à l'aide d'une préposition entre un élément recteur et un complément est de nature à la fois syntaxique et sémantique. En effet, on s'attend à ce que les prépositions en tant que mots, et conformément à la conception saussurienne de la nature du signe linguistique, véhiculent un sens, chaque mot étant une association d'un signifiant et d'un signifié (cf. Melis 2003 : 90). C'est ce qui apparaît assez clairement dans l'exemple (1), où chaque préposition spécifie une relation spatiale particulière, ce qui ne serait pas possible si chacune d'entre elles n'était pas porteuse d'un sens qui lui est propre. Intuitivement, on dirait que *dans* exprime l'idée de contenance, que *sur* véhicule la notion de support, que *contre* exprime une relation de contact impliquant une opposition de forces, etc.

(1) *L'échelle est **dans / sur / devant / derrière / contre** le garage.*

Une telle évidence ne va néanmoins pas de soi dans les exemples (2)-(4) où il paraît légitime de se demander quelle est la contribution sémantique des prépositions *de*, *à* et *en* :

(2) *la ville **de** Paris / quelque chose **d'**intéressant / encore un problème **de** résolu*
(3) *apprendre **à** lire* (vs *apprendre **ø** une leçon*)
(4) ***en** courant*

Dans ces exemples, la motivation sémantique de l'emploi de *de*, *à* et *en* est difficile à trouver et les possibilités de commutation avec d'autres prépositions sont inexistantes, ce qui suggère des contraintes d'emploi très strictes. Les prépositions en question paraissent ainsi entièrement grammaticalisées dans la mesure où elles se réduisent à des « chevilles syntaxiques », lequel usage – décidément grammatical – correspondrait à la phase ultime de leur évolution. C'est cette impression de vacuité sémantique qui a amené certains linguistes à

[2] Il est à noter que les prépositions ne sont pas les seuls éléments relateurs dans la mesure où l'on trouve à travers les langues, y compris en français, d'autres candidats à cette étiquette : les conjonctions, les postpositions, les marques casuelles…

distinguer des prépositions « vides », dépourvues de signification (ex. *de, à, en*), et des prépositions « pleines », porteuses d'un contenu sémantique (ex. *sur, dans, devant*). On trouve, dans le même esprit, plusieurs autres tentatives de caractérisation nuançant la dichotomie citée : l'opposition entre les prépositions vides (*de, à*), demi-vides (*avec, en, par, pour, sur*) et pleines (toutes les autres) (cf. Brunot et Bruneau 1949), la notion de « préposition incolore » (cf. Spang-Hanssen 1963), ou encore la distinction entre les prépositions fonctionnelles et lexicales (cf. Tremblay 1996, Littlefield 2006, Mardale 2009) (voir Vandeloise 1993 et Melis 2003 pour une synthèse). Le fait est cependant que la réduction à la pure articulation syntaxique n'est observée que dans le cas de certains emplois de quelques prépositions simples extrêmement grammaticalisées, en première instance *de* et *à* (cf. Goyens, Lamiroy & Melis 2002, Fagard & De Mulder 2007, Fagard, De Mulder & Hoelbeek sous presse). La contribution sémantique de la grande majorité des prépositions simples, et surtout complexes, reste en réalité incontestable car nettement perceptible (cf. *dans, pour, pendant, depuis, après,* etc.).

1.2. Prépositions complexes : éléments de définition

Pour ce qui est des prépositions complexes, elles apparaissent comme équivalents fonctionnels des prépositions simples du point de vue distributionnel et sémantique, mais s'en distinguent par leurs propriétés morphologiques étant donné qu'elles présentent une structure interne composite (cf. Melis 2003: 112-114). Ainsi, dans l'exemple (5), à l'instar de *pour*, l'expression *à cause de* : a) introduit un complément, b) rattache le constituant ainsi formé au verbe recteur *se fâcher*, et c) spécifie une relation sémantique entre la structure intégrante (*les deux hommes se sont fâchés*) et l'élément intégré (*une petite mesquinerie*).

(1) a. Les deux hommes se sont fâchés **pour** une petite mesquinerie
b. Les deux hommes se sont fâchés **à cause d**'une petite mesquinerie

Mais, d'après Fagard *et al*. (soumis a), à la différence d'une préposition simple, en l'occurrence *pour*, qui cumule les trois fonctions, une préposition complexe prototypique distribue celles-ci à trois éléments différents : une préposition simple finale lie le complément (*de*), une autre préposition simple initiale rattache l'ensemble au constituant recteur (*à*), et un élément central de nature lexicale est porteur de la spécification sémantique de la relation établie (*cause*). Si la grande majorité des prépositions complexes en français relève de ce schéma de construction (ex. *à côté de, à la fin de, à l'égard de, de la part de, par le biais de, sous l'égide de, par rapport à*), il y en a un certain nombre qui, au moins du point de vue synchronique, dérogent à la règle et peuvent de ce fait être considérées comme non-canoniques[3] : *à travers, face à, à même, grâce à*, etc. (cf. Fagard *et al*. soumis b). La présence de l'élément lexical dans la structure des prépositions complexes entraîne une spécification sémantique plus grande (cf. Borillo 2002 : 144), ce qui réduit considérablement leur degré de

[3] Cela ne signifie cependant pas que de tels items ne méritent pas le statut de préposition complexe, tout au contraire. En règle générale, il s'agit d'unités qui ont atteint un degré de lexicalisation et/ou de grammaticalisation plus avancé, mais d'autres raisons peuvent être à l'origine de leur éloignement du modèle canonique. Faute de place, cette question ne pourra pas être abordée ici.

polysémie et leur périmètre d'usage : le sémantisme des prépositions complexes s'avère ainsi moins général et moins abstrait que celui des prépositions simples et leurs contextes d'emplois sont *a priori* moins variés et moins nombreux. Les prépositions complexes remettent ainsi sérieusement en cause la nature grammaticale de la catégorie des prépositions ; il serait certainement plus exact de considérer, comme l'affirment Fagard & De Mulder (2007 : 26), que « toutes les prépositions ont des emplois lexicaux et des emplois grammaticaux, mais que le poids respectif de ces emplois peut être différent d'une préposition à l'autre ». Notons pour terminer sur ce point qu'en dépit de leur sémantique distribuée et d'une structure composite, les prépositions complexes présentent une certaine cohésion interne qui en fait des unités linguistiques à part entière, plus ou moins figées, i.e. ayant des propriétés sémantiques et syntaxiques qui ne sont pas nécessairement prédictibles à partir des éléments qui les composent (cf. Gross 1996, 2006, Melis 2003 : 110-115, Brinton & Traugott 2005, Fagard & De Mulder 2007). Elles sont de ce fait susceptibles d'avoir leurs propres contraintes de sélection, différentes de celles du noyau lexical (par exemple, on peut être au bord de la faillite sans que l'on puisse parler réellement du bord de la faillite ; cf. Melis 2003 : 113).

Nous retiendrons cette caractérisation (minimale) des prépositions complexes pour la suite de l'article, tout en sachant qu'elle est loin d'être entièrement satisfaisante, et ceci pour deux raisons principales : d'une part la délimitation avec les prépositions simples n'est pas sans poser de problème dans certains cas (cf. *depuis* et *envers* qui sont considérés comme prépositions simples, alors que leur complexité morphologique ne fait pas doute : *de+puis* et *en+vers*), d'autre part, il n'est pas sûr qu'elle permette toujours de trancher si une séquence ayant la forme canonique et relativement figée et fréquente en discours mérite d'être classée ou non comme préposition complexe : *dans la région de* (*Paris*) vs *dans l'espace de* (*trois semaines*)[4].

2. Critères d'identification des prépositions complexes : un vieux débat... toujours d'actualité

La difficulté de circonscrire avec précision les frontières d'une classe ou d'une sous-classe de mots est un des problèmes récurrents en linguistique. Elle est cependant particulièrement manifeste dans le cas des prépositions complexes, en raison de l'ambivalence quasi-systématique des résultats de tests d'identification utilisés, qu'ils soient d'ordre syntaxique, sémantique ou morphologique. Si les tentatives de circonscription de la sous-classe en français ne manquent pas, force est de constater :
- qu'elles restent toutes partielles, pour une raison ou une autre,
- que les tests utilisés s'avèrent systématiquement non concluants, ne serait-ce qu'en partie, et
- qu'à ce jour aucune méthode de description ne permet de dresser un profil global satisfaisant de la sous-classe des prépositions complexes faute de paramètres de classabilité fiables.

[4] Dans Frantext (sous-corpus « contemporain » : 1960 à aujourd'hui, 978 textes et 59 918 650 mots), on trouve ainsi 236 occurrences de *dans la région de* (3,9 par million de mots), 11 de *dans le périmètre de* (0,2 par million), 201 de *dans l'espace de* (3,4 par million). Pour comparaison, on trouve dans le corpus frTenTen 12 488 occurrences de *dans l'espace de* (1 par million), 3 906 de *dans le périmètre de* (0,01 par million), 96 927 de *dans la région de* (8 par million).

D'autre part, les linguistes ayant travaillé sur la question sont d'accord :

- que les schémas de construction des prépositions complexes sont nombreux et variés,
- qu'il est illusoire de chercher à dégager des frontières nettes de la sous-classe des prépositions complexes : ce qui ressort très clairement de tout un ensemble de travaux sur la question est qu'il existe plutôt un continuum entre les séquences entièrement figées et sémantiquement opaques entretenant des relations systématiques avec des prépositions simples d'une part, et les séquences syntaxiquement très libres et sémantiquement transparentes très proches de constituants libres. Entre les deux, on trouve « une zone dans laquelle se situent des suites plus ou moins figées et plus ou moins opaques qui entretiennent des relations plus ou moins systématiques avec les prépositions simples » (Melis 2003 : 115).

Partant de ces constats qui semblent faire l'unanimité, nous nous proposons d'élaborer une méthode d'identification multicritère des prépositions complexes qui :

1) tienne compte de l'ensemble des schémas de construction pour assurer une couverture maximale de la sous-classe,
2) vise à positionner chaque structure candidate sur une échelle de prototypicalité allant des meilleurs représentants, placés au centre de la catégorie, jusqu'aux candidats limites situés à la périphérie.

2.1. Les schémas de construction des PrépComp en français

Pour réaliser le premier objectif, nous partirons de la typologie de Melis (2003 : 107), qui regroupe toutes les structures reconnues comme PrépComp en cinq types, selon la classe de « leur élément saillant », i.e. du noyau lexical. En prenant en compte deux paramètres additionnels (la présence d'une préposition en position initiale et en position finale), l'auteur dégage au total onze schémas de construction, que nous reproduisons ici :

A. structures comportant un nom
   - A1 : N P (ex. *face à*)
   - A2 : P_1 N P_2 (ex. *en face de*)
   - A3 : P_1 *le/la/les* N P_2 (ex. *au bord de*)

B. structures comportant une forme verbale
   - B1 : P_1 Inf. P_2 (ex. *à partir de*)
   - B2 : P *ce* relative (ex. *en ce qui concerne*)

C. structures comportant un adjectif
   - C1 : (*le*) ADJ *de* (ex. *le long de*)
   - C2 : *au plus* ADJ *de* (ex. *au plus profond de*)

D. structures comportant un adverbe
   - D1 : ADV *de* (ex. *lors de*)
   - D2 : P ADV (ex. *à même*)

E. structures comportant une préposition
   - E1 : P_1P_2 (ex. *d'après*)
   - E2 : N-P_1-N P_2 (ex. *vis-à-vis de*)

Il est généralement admis que la grande majorité des PrépComp a un noyau nominal, sans que les proportions soient établies. La description de la sous-classe entière est de ce fait souvent réduite aux items relevant de ce schéma de construction, ce qui n'en donne pas une image complètement correcte. C'est pourquoi il nous semble que, pour être satisfaisante, la méthode d'identification doit viser l'ensemble des schémas possibles, qui en réalité sont encore plus nombreux que ne le laisse croire la typologie de Melis (2003). Dans le cadre de notre étude, nous partirons d'une liste de patrons à granularité plus fine :

N. structures à noyau nominal

N1 :	P Dét N P	(ex. *au bord de*)
N2 :	P N P (de)	(ex. *en face de*)
N3 :	P N P (#de)	(ex. *par rapport à*)
N4 :	N P	(ex. *face à*)
N5 :	P Dét Adj N de	(ex. *au fin fond de*)
N6 :	Dét N P	(ex. *à cause de*)
N7 :	N V P	(ex. *compte tenu de*)
N8 :	N P N P	(ex. *vis-à-vis de*)
N9 :	P N	(ex. *à travers*)

V. structures à noyau verbal

V1 :	V P	(ex. *s'agissant de*)
V2 :	P_1 V P_2	(ex. *à partir de*)
V3 :	autre	(ex. *en ce qui concerne*)

Adj. structures à noyau adjectival

Adj1 : P_1 Dét *plus* Adj P_2	(ex. *au plus profond de*)

Adv. structures à noyau adverbial

Adv1 : Adv P :	(ex. *lors de*)
Adv2 : P_1 Adv P_2	(ex. *en dehors de*)
Adv3 : P Adv	(ex. *à même*)

P. structures à noyau prépositionnel

P1 :	P_1 P_2	(ex. *d'après*)

Pour établir une liste d'indices raisonnée, nous ferons une observation préliminaire des prépositions complexes relevées à ce jour dans la littérature en nous appuyant notamment sur : Borillo (1997, 2002), Melis (2003), Le Pesant (2006), Stosic (en préparation).

2.2. Pour une échelle de prototypicalité

Pour ce qui est du deuxième objectif, qui consiste à mieux circonscrire la sous-classe des prépositions complexes en français en évaluant le degré de correspondance des séquences candidates aux membres prototypiques, il faut donc choisir les tests les plus adaptés, pour le schéma principal comme pour d'autres schémas observables, et qui couvrent les aspects distributionnels, morphologiques et sémantiques. L'efficacité des tests existants ayant déjà été examinée dans plusieurs travaux (voir notamment Borillo 1997, 2002, Adler 2001, Melis 2003, Fagard & De Mulder 2007), nous ne les réinterrogerons pas ici : nous procéderons

plutôt à des adaptations nécessaires et, pour certains, à leur décomposition en indices plus fins pour mieux évaluer le degré de prototypicalité des séquences candidates à la sous-classe (cf. § 3.2.1). En outre, de nouveaux indices de classabilité seront introduits. En particulier, nous retiendrons un autre critère de figement, essentiel mais souvent délaissé parce qu'il est difficile à manier : la fréquence d'apparition des suites candidates en discours (cf. § 3.2.2). Notre démarche combinera ainsi le recours à l'introspection, qui est indispensable pour la manipulation des indices formels et sémantiques établis en amont, et la consultation des corpus en vue de l'étude de la fréquence d'usage des structures retenues pour analyse.

2.2.1. *Critères d'identification morphologiques, syntaxiques et sémantiques*

La littérature sur les PrépComp offre tout un ensemble de critères qualitatifs conçus dans l'objectif d'évaluer le degré de figement de différentes séquences candidates à la classe. Il s'en dégage deux types d'indices : les uns sont d'ordre formel, les autres de nature sémantique, comme le résume Melis (2003 : 110) : « Sur le premier plan, il est question de figement au niveau syntagmatique – la structure interne de la locution, ses propriétés transformationnelles – et au niveau paradigmatique – la possibilité de remplacer certaines composantes par des termes proches. Sur le second plan, il est question de non-compositionnalité ou de l'impossibilité de calculer le sens de la locution à partir du sens des composantes ». Au lieu de reprendre d'une manière systématique tous les critères proposés jusqu'à présent, nous partirons de la définition de ce qui pourrait être une PrépComp canonique afin d'évaluer dans quelle mesure différentes expressions candidates au statut de PrépComp se rapprochent (ou s'éloignent) du prototype d'une PrépComp en français. Il s'agit d'un changement de perspective considérable dans l'utilisation des critères en question : ceux-ci ne seront pas considérés comme formant une série de conditions nécessaires et suffisantes pour l'identification des PrépComp mais comme fournissant des paramètres d'évaluation de la ressemblance des séquences testées aux membres centraux de la classe des PrépComp (cf. § 3.2.2. ci-dessous).

Pour définir le prototype d'une PrépComp en français, nous nous appuierons à la fois sur la définition avancée ci-dessus (cf. § 2.2., et Fagard et al. soumis) et sur un ensemble de traits qui semblent caractériser une série d'expressions qui sont communément admises au statut de PrépComp, autant par des lexicographes que par des linguistes. Ainsi, on peut dire qu'une PrépComp prototypique en français est une structure morphologiquement composite formant une unité et ayant la distribution et le fonctionnement sémantique d'une préposition simple (cf. § 2.2. ci-dessus). Le fait que ces éléments présentent des traits communs correspond à un trait récurrent des phénomènes de grammaticalisation, à savoir la paradigmaticalisation (Lehmann 1995).

Plus précisément, une PrépComp canonique comporte une préposition initiale, un noyau lexical et une préposition finale. La présence des prépositions initiale et finale est donc à traduire en critères : du fait de comporter une préposition initiale, les séquences *à côté de, en dessous de* et *à partir de* peuvent être considérées comme de meilleures candidates à la classe que *le long de, lors de* et *s'agissant de* qui n'en comportent pas. De même, la présence d'une préposition finale dans les séquences *au sujet de, en dehors de* et *à compter de* en fait de meilleures candidates que *à travers, par-dessus* ou *en ce qui concerne* où n'apparaît aucune préposition à la fin. Ensuite, il est possible de tenir compte de la préposition même qui apparaît à l'initiale ou à la fin de la séquence : une observation préliminaire des PrépComp consensuelles fait apparaître que la grande majorité des PrépComp comportent à l'initiale la préposition *à* et en position finale *de* (cf. notamment Borillo 1997, Melis 2003, Stosic en

préparation). Cette propriété peut être prise en compte dans le calcul du degré de prototypicalité des séquences candidates : pour ne prendre que quelques exemples, *à l'instar de, au lieu de, à travers* correspondraient mieux, de ce point de vue, au modèle de PrépComp en français que *sous la direction de, par respect pour* ou encore *sans égard à* où ce n'est pas *à* qui figure en position initiale. Il en est de même pour la préposition finale : *par défi pour, en rivalité avec* et *quant à* s'éloignent de la PrépComp canonique, contrairement à *le long de, au bord de* ou *hors de* qui s'en rapprochent. Enfin, l'application de ces critères ne peut concerner que les séquences candidates comportant une préposition au début ou à la fin, les autres séquences en étant exclues.

Ensuite, les séquences candidates peuvent être évaluées par rapport à leur (non-)conformité avec les règles de la syntaxe du français (cf. Melis 2003 : 110). La gamme de ces irrégularités est très large et il peut s'agir d'un critère d'évaluation très général consistant à constater si irrégularité il y a ou non. L'idée sous-jacente à ce critère est qu'une structure violant des régularités syntaxiques de la langue a des chances d'être figée et donc d'appartenir au stock des PrépComp. Ainsi, du fait de l'absence de déterminant dans leur structure, *à côté de* et *à base de* ont atteint un stade de lexicalisation plus avancé que *du côté de* et *sur la base de* dont la structure est en tous points conforme à la syntaxe régulière. Ce critère peut s'appliquer avec plus ou moins de difficultés à toute expression candidate. Dans le même ordre d'idée, l'absence de déterminant, ou plus précisément d'article, peut être considérée comme un indice de figement supplémentaire dans le cas de structures à noyau lexical nominal.

Il est également possible d'utiliser différents paramètres transformationnels largement décrits dans la littérature pour évaluer le degré d'affinité d'une expression candidate avec la classe des PrépComp (cf. Borillo 1997, Melis 2003 : 110-111). La variation du déterminant permet d'opposer *du côté de* et *au bord de*, le premier admettant l'article indéfini (cf. *d'un côté de*), le second non (cf. **à un bord de*). Ainsi *au bord de* apparaît-il plus figé que *du côté de*. L'anaphorisation du régime à l'aide d'un déterminant possessif ou démonstratif (ou son blocage) est également utilisée comme critère d'évaluation du figement (cf. Borillo 1997, Adler 2001, 2007 Fagard & De Mulder 2007). C'est ce qui permet d'opposer par exemple *à côté de* qui autorise l'anaphorisation de son régime (cf. *à son côté*) et *au lieu de* qui ne l'autorise pas (cf. **à son lieu*), et qui apparaît donc comme plus avancé dans son figement. On peut inclure dans ce groupe de critères la possibilité de coordination de la séquence candidate avec une autre préposition, simple ou complexe. L'existence d'une telle possibilité pour une expression donnée signifie en principe qu'elle relève de la même classe distributionnelle et se rapproche donc des membres centraux de la classe (cf. *se mettre du côté et au niveau du genou blessé, sauces variées à base de et de poivre japonais, intervenir à cause de et malgré la complexité de la situation*).

La cohésion d'une séquence candidate peut aussi être examinée à l'aide de tests d'insertion qui consistent à introduire divers types de modifieurs en son sein. Il est ainsi courant d'observer les possibilités d'expansion des PrépComp au noyau nominal à l'aide de modifieurs adjectivaux (cf. Borillo 1997 : 180-181). Une telle possibilité existe par exemple pour *du côté (gauche/visible) de, en contact (direct) de* et *sous contrôle (permanent/total) de*, mais non pour *par rapport à, au cœur de* ou *à cause de* qui présentent, de ce point de vue, une cohésion nettement plus grande, interdisant ce type d'expansion. Bien que de nature entièrement différente, l'insertion de certains types de modifieurs adverbiaux comme *notamment, justement, autant, même, non pas, aussi, plutôt, surtout*, etc. permet de dégager des oppositions intéressantes entre certaines expressions comme *du côté (notamment/tant) de, à partir (non/ justement/ autant) de, à cause (précisément/ peut-être/ aussi) de* qui sont susceptibles d'accueillir des adverbes et d'autres comme *à travers, le long de, quant à, à l'instar de* qui n'admettent aucune insertion d'adverbiaux. Ce critère, qui mériterait une étude

à part à lui-même, peut s'appliquer de manière systématique à toute séquence candidate à la classe des PrépComp. Par ailleurs, nous prendrons en compte dans notre étude le degré de variation des modifieurs adjectivaux et adverbiaux lorsque leur insertion s'avère possible : certaines séquences candidates admettent en effet un grand nombre d'items lexicaux, d'autres n'en accueillent que quelques-uns.

Enfin, le dernier groupe de critères qualitatifs auxquels nous aurons recours dans cette étude sont de nature sémantique. Tout d'abord, il nous semble pertinent de prendre en compte la nature sémantique du nom dans le cas des séquences à noyau lexical nominal (cf. Borillo 1997 : 179) ; il est en effet bien connu, et l'examen du relevé des PrépComp le confirme, que la très grande majorité des PrépComp comporte un nom exprimant la partition ou l'orientation spatiales (cf. Svorou 1994, Borillo 1997, Stosic en préparation). Les séquences comme *au bord de, au milieu de, au sommet de* peuvent ainsi être considérées comme plus proches des membres prototypiques de la classe que les expressions *en rivalité avec, par respect pour* ou *aux bons soins de*. Rappelons qu'il ne s'agit pas d'exclure ces dernières de la classe, mais simplement d'indiquer que les séquences de ce type paraissent moins centrales compte tenu de la nature sémantique de leur noyau nominal. Un autre critère sémantique que nous retenons est celui de l'opacité : en effet, les PrépComp en tant qu'unités figées sont censées avoir un sens opaque, qu'il n'est donc pas possible de déduire du sens des éléments qui les composent, contrairement aux séquences libres qui, elles, ont par définition un sens compositionnel. Encore une fois, les choses sont loin d'être aussi tranchées : on observe plutôt des degrés d'opacité et de transparence variables que les comportements tranchés. La référentialité du noyau lexical est un autre critère possible, étroitement lié au précédent. Il est en effet intéressant d'évaluer dans quelle mesure le noyau lexical d'une PrépComp donnée garde ses capacités référentielles dans ses différents emplois. Par exemple, si dans l'expression *au fond de*, le nom *fond* est en emploi référentiel dans certains cas, plutôt limités (cf. *au fond du lac / de la baignoire*), le noyau nominal de l'expression *sous la direction de* est, lui, pratiquement toujours en emploi référentiel. Ou encore, la référentialité des noms *travers* et *insu* est nulle dans les séquences *à travers* et *à l'insu de*. Le dernier critère sémantique que nous retiendrons ici concerne l'existence, pour une séquence candidate, d'extensions sémantiques (cf. *en face de mes fenêtres* vs *en face de mes péchés*, *à portée de la bouche* vs *à portée de la raison*). Celles-ci témoignent en effet d'un stade de lexicalisation et/ou de grammaticalisation assez avancé (cf. Borillo 1997 : 182-183) : le processus de figement étant enclenché, le noyau lexical connaît une désémantisation ce qui laisse à l'expression la possibilité d'étendre ses emplois bien au-delà des capacités combinatoires du noyau lexical. La diversification des usages est ainsi un signe de figement de la séquence.

2.2.2. *Critères d'identification basés sur la fréquence*

L'importance de la fréquence dans l'émergence de nouvelles constructions est évidente. Si le lien entre fréquence et figement est connu depuis longtemps (Zipf 1936), et a été souligné en particulier pour le processus de grammaticalisation (Bybee 2007), on peut se demander dans quelle mesure le lien entre fréquence et figement est spécifique à la grammaticalisation (Feltgen 2017), et surtout quel est précisément ce lien. Concernant les prépositions complexes, en tant que séquences *a priori* figées ayant un rôle grammatical dans la phrase (cf. § 2.1. ci-dessus), elles sont supposées avoir une fréquence d'emploi plus élevée que les structures libres relevant des mêmes patrons de construction. Hoffmann (2005) note cependant qu'il y a des cas de figement de prépositions complexes avec des fréquences très faibles, et qu'il n'y a donc pas de corrélation parfaite entre fréquence et degré de figement.

Ainsi, pour les prépositions complexes de structure [P dét N P], on ne peut se fier à la fréquence de *dans la région de* et *à l'instar de*, puisque c'est la seconde qui semble la plus figée, alors que sa fréquence est bien moindre. On voit donc que la fréquence d'une séquence donnée ne suffit pas pour évaluer son degré de figement, mais sa prise en compte peut donner des indices de figement intéressants.

Par ailleurs, la mise au point d'approches lexicométriques permet d'obtenir un certain nombre d'informations complémentaires, que le chercheur serait bien en peine de mettre en évidence 'à la main', comme le profil combinatoire et en particulier le score d'« information réciproque » (Blumenthal et al. 2005), qui indique ce que l'on pourrait appeler *l'affinité* entre deux termes dans un corpus. Plus techniquement, il s'agit de la fréquence de cooccurrence entre les deux termes, pondérée par la fréquence de chacun des deux termes indépendamment de l'autre. Ce type d'information, que l'on peut calculer automatiquement à l'aide de certains outils comme PrimeStat et Txm[5], peut aider à trancher dans certains cas : *instar* n'étant attesté en français moderne que dans la construction *à l'instar de*, on peut s'attendre à ce que les scores d'information mutuelle montrent l'affinité entre *instar* et les autres termes de la construction (préposition *à* et déterminant défini à gauche, préposition *de* à droite).

Si on fait le test sur le corpus BTLC-Le Monde, le calcul de l'information mutuelle (calculée sur une fenêtre de 5 mots) montre effectivement que sont associés à *instar* (44 occurrences) la préposition *à* (à gauche), le déterminant défini (à gauche) et la préposition *de* (à droite). Si on trouve bien la préposition *dans* et le déterminant défini (à gauche) dans les termes associés à *région* (3856 occurrences), lui sont également associés un grand nombre de modifieurs, notamment des adjectifs comme *séparatiste* ou *autonome* et des adjectifs relationnels comme *parisienne*, ce qui montre bien qu'il n'y a pas figement malgré la fréquence de *dans la région de*.

Cependant, même ainsi, cela ne suffit pas, dans la mesure où il faudrait pouvoir le calculer pour des groupes de mots. Or c'est à notre connaissance encore impossible. Par ailleurs, ce type de calcul fonctionne très bien sur des gros corpus homogènes du point de vue de la graphie, mais moins bien sur des corpus anciens, surtout s'ils ne sont pas lemmatisés. Nous gardons donc le calcul de la fréquence, et de l'information mutuelle, comme des indices supplémentaires de figement, sans en faire des éléments décisifs. Ainsi, pour chaque séquence candidate, nous évaluerons la fréquence d'emploi dans un corpus de 10 milliards de mots (frTenTen) de même que leur fréquence d'usage avec des modifieurs adjectivaux et adverbiaux insérés en leur sein. Le score d'information réciproque sera également calculé pour chaque expression.

2.2.3. *Établissement d'une grille d'identification multicritère*

L'ensemble des critères établis ci-dessus nous permettent de construire une grille d'identification des prépositions complexes comportant au total 21 indicateurs, dont 16 sont d'ordre qualitatif et 5 d'ordre quantitatif. Pour évaluer l'appartenance des structures candidates à la classe des PrépComp, nous associons à chaque critère une échelle numérique d'amplitude réduite permettant de mesurer le degré de correspondance de la structure candidate au profil d'une PrépComp canonique. Plus une séquence candidate a des propriétés prototypiques d'une PrépComp canonique, plus la valeur sera petite. Au final, en cumulant les valeurs des différents critères, chaque séquence pourra être située sur une échelle numérique

[5] Sascha Diwersy, http://romanistik.phil-fak.uni-koeln.de/sdiwersy.html?&L=2 ; http://textometrie.ens-lyon. fr/.

d'amplitude plus importante correspondant au total des valeurs assignées. Nous nous attendons en effet à ce que les meilleures candidates aient des scores minimaux alors que les séquences proches des structures libres aient des scores les plus élevés (cf. § 3.2.4 ci-dessous). Le tableau qui suit présente l'ensemble des critères retenus et leurs échelles numériques respectives :

Ordre	Critères (indices de classabilité)	Echelle numérique	Amplitude de l'échelle	Nature du critère
1	Prés_Prép_Initiale	présence : **0** ; absence : **1**	0 / 1	morphologique
2	Prep_Initiale	à, en : **0** ; de, dans, sur, par, sous : **1** ; autres : **2**	0 / 2	morphologique
3	Prés_Prép_Finale	présence : **0** ; absence : **1**	0 / 1	morphologique
4	Prep_Finale	de : **0** ; à, avec : **1** ; pour : **1.5** ; autres : **2**	0 / 2	morphologique
5	Irrégularité syntaxique	irrégularité : **0** ; régularité : **1**	0 / 1	syntaxique
6	Article	absence : **0** ; présence : **1**	0 / 1	morphologique
7	Reprise du régime par le possessif	impossible : **0** ; possible : **1**	0 / 1	morphologique
8	Variation du déterminant	impossible : **0** ; possible : **1**	0 / 1	morphologique
9	Nature sémantique du nom	partition et orientation spatiales : **0** ; reste espace et actions : **1** ; relations logico-sémantiques : **1.5** ; autres : **2**	0 / 2	sémantique
10	Référentialité	jamais référentiel : **0** ; rarement référentiel : **0.5** ; rarement non référentiel : **1** ; toujours référentiel : **2**	0 / 2	sémantique
11	Insertion du modifieur adjectival	impossible : **0** ; possible : **1**	0 / 1	syntaxique
12	Variation du modifieur adjectival	<2 Adj : **0** ; 2-10 Adj : **1** ; 11-100 Adj : **2** ; > 100 Adj : **3**	0 / 3	syntaxique
13	Fréquence du modifieur adjectival	0% : **0** ; ≤0,1% : **0.5** ; ≤0,3% : **1** ; ≤1% : **1.5** ; ≤5% : **2** ; ≤10% : **2.5** ; >10% : **3**	0 / 3	fréquence
14	Insertion du modifieur adverbial	impossible : **0** ; possible : **1**	0 / 1	syntaxique
15	Variation du modifieur adverbial	0 Adv : **0** ; 1-2 Adv : **1** ; ≥ 3 Adv: **2**	0 / 2	fréquence
16	Fréquence du modifieur adverbial	0% : **0** ; ≤0,1% : **0.5** ; ≤0,3% : **1** ; ≤1% : **1.5** ; ≤5% : **2** ; ≤10% : **2.5** ; >10% : **3**	0 / 3	fréquence
17	Coordination avec un autre SP	possible : **0** ; impossible : **1**	0 / 1	syntaxique
18	Extension sémanitque	observé : **0** ; non-observé : **1**	0 / 1	sémantique
19	Opacité	observé : **0** ; non-observé : **1**	0 / 1	sémantique
20	Fréquence relative / million de mots	>50 : **0** ; 30-50 : **1** ; 10-29 : **2** ; <10 : **3**	0 / 3	fréquence
21	Score d'information réciproque	élevé : **0** ; faible : **1**	0 / 1	fréquence

Tableau 1 : Table des critères

L'avantage principal de cette méthode d'identification des PrépComp est qu'elle permet d'éviter les contradictions auxquelles on aboutit inévitablement lorsque les critères sont utilisés comme conditions nécessaires et suffisantes. Pour ne donner qu'un exemple, les séquences *à l'insu de* et *à l'instar de* sont sémantiquement opaques du fait de comporter des éléments qui n'ont (plus) aucune autonomie dans la langue (cf. Melis 2003 : 112), mais elles admettent toutes deux l'anaphorisation à l'aide du déterminant possessif (cf. *à son insu*, *à son instar*), contrairement à *à base de*, *à cause de* ou *au fond de* qui, bien que sémantiquement transparentes, ne donnent pas lieu à l'anaphorisation. Il est alors difficile de trancher si de telles séquences appartiennent ou non à la classe des PrépComp. Notre méthode d'identification évite cet écueil dans la mesure où les critères permettent, grâce à un score

total, d'évaluer l'adéquation d'une expression donnée avec le modèle de PrépComp et non de l'inclure dans la classe ou de l'en exclure.

2.2.4. *Plusieurs matrices pour une meilleure comparabilité des scores de classabilité*

Comme cela a déjà été dit dans ce qui précède, la spécificité de certains critères ne permet pas leur application à toutes les séquences candidates. Cela signifie que les scores totaux maximaux varieront en fonction du nombre de critères pertinents pour chaque séquence. Afin d'assurer la comparabilité des totaux des scores obtenus, il est important de tenir compte de ces écarts potentiels, ce qui nécessite d'identifier pour chaque schéma de construction les critères qui y sont applicables. Le degré d'adéquation à la sous-classe des PrépComp des séquences testées pourra ainsi être évalué non par rapport à une valeur absolue, identique pour toutes les candidates, mais par rapport au score total des seuls tests applicables, en vue d'une meilleure standardisation des données. Les batteries de tests utilisées seront donc adaptées au profil morpho-syntaxique respectif des séquences candidates.

Les 17 patrons de construction retenus (cf. § 3.1) peuvent ainsi être regroupés dans 9 matrices différentes, chacune comportant une combinaison de critères d'identification différente. Elles sont présentées dans le tableau qui suit :

n° Matrice	Patron	Exemple	Critères																					Score Maximal
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
1	P Dét N P	*à l'instar de*	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	40
	P N P(de)	*à base de*	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	
	P Dét Adj N de	*à l'autre extrémité de*	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	
2	P N P(#de)	*en conformité avec*	ok	ok	ok	ok	ok	ok		ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	39
3	N P	*grâce à*	ok		ok	ok	ok	ok			ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	36
	N V P	*compte tenu de*	ok		ok	ok	ok	ok			ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	
	N P N P	*vis-à-vis de*	ok		ok	ok	ok	ok			ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	
4	Dét N P	*le long de*	ok		ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	38
5	P N	*à travers*	ok	ok	ok		ok	ok			ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	37
6	Adv P	*auprès de*	ok		ok	ok	ok									ok	ok	ok	ok	ok	ok	ok	ok	20
	V P	s'agissant de	ok		ok	ok	ok									ok	ok	ok	ok	ok	ok	ok	ok	
7	P P	*d'après*	ok	ok	ok	ok	ok									ok	ok	ok	ok	ok	ok	ok	ok	22
	P Adv P	*en dehors de*	ok	ok	ok	ok	ok									ok	ok	ok	ok	ok	ok	ok	ok	
	P V P	*à partir de*	ok	ok	ok	ok	ok									ok	ok	ok	ok	ok	ok	ok	ok	
8	P Adv	*par-dessus*	ok	ok	ok		ok									ok	ok	ok	ok	ok	ok	ok	ok	21
	autres	*en ce qui concerne*	ok	ok	ok		ok									ok	ok	ok	ok	ok	ok	ok	ok	
9	P Dét plus Adj P	*au plus bas de*	ok	ok	ok	ok	ok	ok	ok	ok						ok	ok	ok	ok	ok	ok	ok	ok	25

Tableau 2 : Répartition des patrons de construction dans des matrices

3. Application de la grille à un échantillon de séquences candidates

Afin d'examiner la pertinence de la grille multicritère établie dans la section précédente, nous l'appliquons à deux séries d'expressions. La première est constituée d'un ensemble de 49 séquences qui sont généralement admises au statut de PrépComp. La deuxième comporte 26

séquences pour lesquelles nous faisons l'hypothèse qu'elles relèvent de la combinatoire libre. Chacun des dix-sept schémas de construction établis dans la section 2.1 est couvert par une ou plusieurs séquences de la classe des PrépComp, et dix d'entre eux sont également représentés par une ou plusieurs séquences *a priori* non figées. Le tableau qui suit donne notre échantillon d'étude :

Noyau lexical	**Patron de construction**	**Hypothèse**	
		Prépositions complexes	**Structures libres**
Nom	P Dét N P	*à l'instar de, au lieu de, dans l'attente de, dans la direction de, du côté de, sous la direction de*	*sur le bord de, dans la région de, sur la totalité de, après l'annonce de, pour le bien de, au lendemain de*
	P N P (de)	*à base de, à cause de, en direction de, sous contrôle de*	*à valeur de, en présence de, par excès de*
	P N P (#de)	*en conformité avec, en rivalité avec, par comparaison avec, par rapport à, par respect pour, en avance sur*	*par confusion avec, par lâcheté devant*
	N P	*face à, grâce à, suite à*	
	P Dét Adj N de	*à l'autre extrémité de, à l'autre bout de, au beau milieu de, au fin fond de, aux bons soins de*	*au grand bénéfice de, au premier jour de*
	Dét N P	*le long de*	*le discours de*
	N V P	*compte tenu de*	
	N P N P	*vis-à-vis de*	
	P N	*à travers*	
Adverbe	Adv P	*contrairement à, quant à, relativement à, auprès de, lors de*	*rarement sur, encore dans*
	P Adv P	*en deçà de, en dehors de, en dessous de, en plus de*	*avec plus de, de là dans*
	P Adv	*à même, de derrière, par-dessus*	*avec toujours*
Verbe	V P	*s'agissant de, en passant par*	*en partant de, en communiquant avec*
	P V P	*à partir de*	*à commencer par*
	autre	*en ce qui concerne*	
Préposition	P P	*d'après, de chez, dès avant, hors de*	*jusque dans, sauf pour, depuis chez*
Adjectif	P Dét plus Adj P	*au plus bas de*	*au plus intime de, au plus sombre de*

Tableau 3 : Échantillon des prépositions complexes étudiées

A chacune de ces séquences a été appliquée la batterie de tests préalablement définie comme adéquate en fonction de sa matrice d'appartenance (cf. Tableau 2), ce qui nous a permis d'obtenir un score pour chaque expression candidate. Ces scores n'étant pas directement comparables du fait des différences entre le nombre des tests applicables, un ratio a été calculé entre le score obtenu et le score maximal possible de la matrice dont relève l'expression. Ce sont ces valeurs, exprimées en pourcentage, qui nous permettent de situer les séquences candidates sur une échelle de prototypicalité. Le tableau 4 donne une première

répartition des séquences selon l'hypothèse concernant leur appartenance ou non à la classe des prépositions complexes.

	Hypothèse de départ		**Total**
Quartiles	Prépositions complexes	Séquences libres	
Moins de 80%	3	16	**19**
Moins de 52%	13	6	**19**
Moins de 41%	23	4	**27**
Moins de 21%	10	0	**10**
Total	**49**	**26**	

Tableau 4 : Répartition des séquences candidates en fonction de l'hypothèse de départ et du score obtenu

On voit que 46 des 49 expressions considérées comme PrépComp présentent un score en dessous de 52% et que, inversement, 16 expressions sur 26 considérées comme structures libres selon notre hypothèse présentent un score égal ou supérieur à 52% (voir l'Annexe 1). De même, les deux extrémités de l'échelle sont clairement marquées (voir l'Annexe 2) : d'un côté, les trente séquences aux scores les plus bas (3%-34%) sont toutes des PrépComp (ex. *à travers, à partir de, au lieu de, face à, à cause de, d'après, le long de, en ce qui concerne, par rapport à*), de l'autre côté les séquences aux scores les plus élevés (68%-80%) sont bel et bien des structures libres (ex. *jusque dans, le discours de, sauf pour, avec toujours, en communiquant avec, depuis chez, rarement sur, encore dans*). Entre les deux, on trouve une large zone de transition comportant des séquences qui, tout en présentant un degré de lexicalisation certain, ne semblent pas des membres parfaits du paradigme des prépositions. Tel est le cas de plusieurs prépositions lexicales comme *sous contrôle de, contrairement à, sous la direction de, par respect pour*, ainsi que de quelques séquences qui, bien que difficilement acceptables dans la classe des prépositions, sont assez fréquentes en discours et ont une structure et une distribution proches de celles des PrépComp (ex. *au plus intime de, par lâcheté devant, au lendemain de, par confusion avec*). Cependant, leurs propriétés sémantico-référentielles et leurs latitudes combinatoires les éloignent du fonctionnement prototypique des PrépComp. Notons enfin que le score moyen des séquences supposées être PrépComp est de 33%, alors que le score moyen de celles qui sont supposées être des structures libres est de 55%.

Un autre point mérite d'être souligné à propos de la position de certaines séquences sur l'échelle de prototypicalité. En effet, le graphique de l'Annexe 2 fait clairement apparaître que la position de certaines expressions sur l'échelle n'est pas en conformité avec les attentes initiales. Par exemple, les séquences *par excès de, à valeur de, en présence de*, qui ne sont jamais répertoriées comme PrépComp, apparaissent avec des scores relativement bas (respectivement 36%, 38% et 41%) côtoyant ainsi les expressions *en dessous de, du côté de* ou *de chez*, dont le statut de PrépComp semble plutôt bien établi. Inversement, certaines séquences considérées comme PrépComp se retrouvent très haut dans l'échelle, avec des scores assez élevés (ex. *s'agissant de* – 65%, *en passant par* – 58%)[6]. Ces écarts peuvent être interprétés de deux façons. La première consisterait à se conformer à la grille en retenant les séquences aux scores bas comme PrépComp et en éliminant celles aux scores trop élevés de la

[6] Cf. Stosic (2012) pour la séquence *en passant par*.

classe des PrépComp ; cette solution nous paraît tout à fait envisageable par exemple pour *à valeur de* qui semble effectivement un bon candidat à la classe mais peu propice pour *au premier jour de* (40%) et *par lâcheté devant* (41%) dont les propriétés sémantico-référentielles et les latitudes combinatoires les classent clairement du côté des structures libres. De même, sortir *s'agissant de* de la classe des prépositions n'a rien d'évident. La deuxième façon d'interpréter les écarts observés pourrait correspondre à un réexamen détaillé des écarts afin de s'assurer que la grille est bien adaptée à tous les types d'expressions. On peut par exemple se demander légitimement si notre grille est suffisamment discriminante pour les séquences issues de formes gérondives : non seulement l'écart entre *s'agissant de* (65%) et *en partant de* (63%) n'est pas significatif, mais il est défavorable à la première, contrairement à notre hypothèse.

4. Discussion

Si la méthode de description ici esquissée ne semble pas en mesure de distinguer à coup sûr les PrépComp des séquences libres, les résultats obtenus nous paraissent globalement concluants compte tenu de nos objectifs et du caractère graduel du processus de figement. Tout d'abord, elle échappe à un des principaux biais des travaux antérieurs sur la question qui se cantonnaient généralement aux séquences à noyau nominal, cherchant à trancher sur l'appartenance de telle ou telle expression à la classe des PrépComp. Grâce à une assez large gamme de critères de classabilité pondérés et de nature différente, la grille élaborée a une bonne couverture de la classe en termes de schémas de construction. Deuxièmement, elle permet de positionner chaque structure candidate sur une échelle de prototypicalité allant des meilleurs représentants, placés en bas de l'échelle, jusqu'aux candidats limites situés à l'autre extrémité. La modélisation de la classe des PrépComp résultant de notre méthode de description paraît ainsi assez fidèle à l'image du continuum qui lui est souvent associée ; il ne s'agit donc pas de dire si une séquence candidate est ou n'est pas une PrépComp, mais d'évaluer dans quelle mesure elle correspond aux membres prototypiques de la classe. Ceci faisant, elle permet de rendre convenablement compte du caractère graduel du processus de figement qui affecte les PrépComp.

Plusieurs aspects de cette première élaboration mériteraient cependant d'être approfondis et améliorés. Une première piste dans ce sens consisterait à examiner de plus près la performance des critères pour chacune des matrices retenues. On remarque par exemple assez facilement que les tests de la matrice n° 7, s'appliquant aux patrons [P P], [P Adv P], [P V P] ont un très bon pouvoir de polarisation (*d'après, hors de, en plus de* vs *sauf pour, jusque dans, avec plus de*), avec une ligne de démarcation nette se situant à 55%. La répartition est loin d'être aussi tranchée dans le cas de la matrice n° 1 regroupant des patrons à noyau nominal [P Dét N P], [P N P (de)], [P Dét Adj N de], bien qu'elle soit suffisamment fournie autant en nombre de séquences étudiées qu'en nombre de tests applicables. Ce manque de clarté peut s'expliquer en partie par la nature nominale du noyau lexical, nombre de séquences en question oscillant dans leurs usages entre une distribution potentiellement ou clairement prépositionnelle et un degré de lexicalisation insuffisant. C'est d'ailleurs ce type de séquences qui occupe majoritairement la partie centrale du graphique de l'annexe 2, où les PrépComp et les séquences libres (d'après notre hypothèse initiale) sont entremêlées.

Une deuxième façon d'approfondir la méthode de description présentée consisterait à chercher à dégager, à l'aide d'outils statistiques, les variables les plus explicatives pour l'ensemble de l'échantillon, ou bien pour chacune des matrices séparément. Nous avons procédé à une première expérimentation de ce type en utilisant des arbres de décision. La figure 1 présente un arbre de décision construit sur la base des seuls tests s'appliquant à tout l'échantillon, à savoir les tests 1, 3, 5 et 14-21. Dans cette configuration, on voit que c'est la coordination avec un autre SP (test 17) qui apparaît comme variable la plus explicative. Si la coordination avec un autre SP est possible, à l'aide du test 19 (Opacité sémantique), on arrive à identifier avec succès 19% de séquences qui sont des PrépComp. Dans la partie gauche de l'arbre, on voit que, si la coordination avec un autre SP n'est pas possible, la variation du modifieur adverbial (test 14) semble décisive et permet de repérer 13.1% d'expressions qui sont des séquences libres.

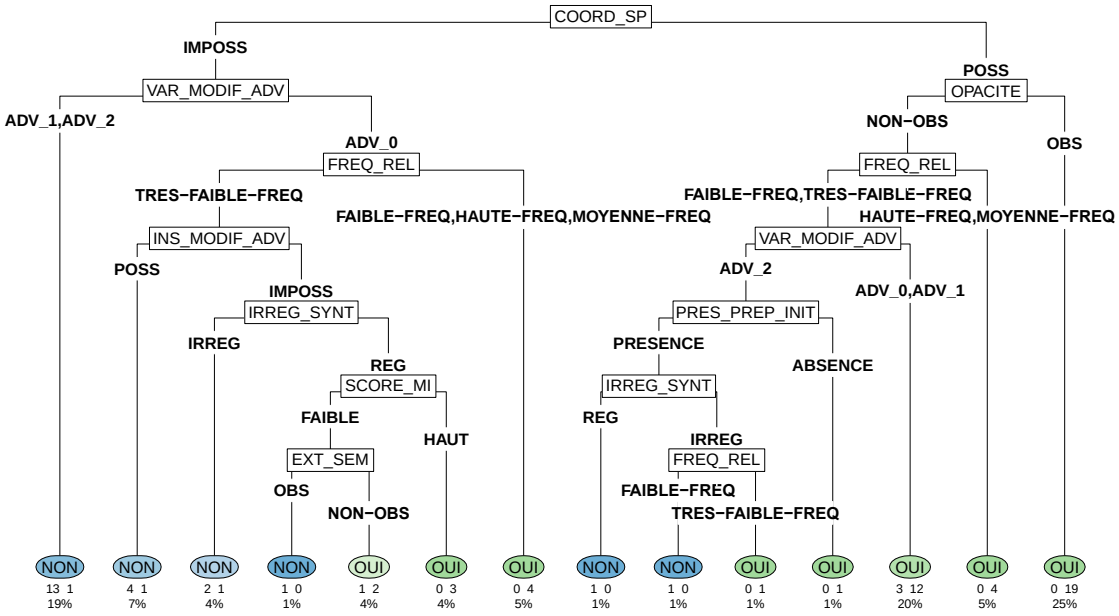


Figure 1 : Arbre de décision construit à partir des seuls tests s'appliquant à tout l'échantillon

Si la même démarche est appliquée à la seule matrice n° 1, à savoir aux patrons [P Dét N P], [P N P (de)], [P Dét Adj N de], c'est un autre critère, celui de la référentialité (test 10), qui semble déterminant (Figure 2). On peut noter que dans le cas de la matrice n° 1, à l'aide de 5 critères (sur 21), on arrive à rendre compte de la totalité des données, ce qui signifie que la nature sémantique du nom (test 9), la possibilité d'insérer ou non un modifieur adjectival (test 11), l'irrégularité syntaxique (test 5) et la préposition initiale (test 2) sont cruciales pour l'identification des PrépComp à noyau nominal relevant de la première matrice.

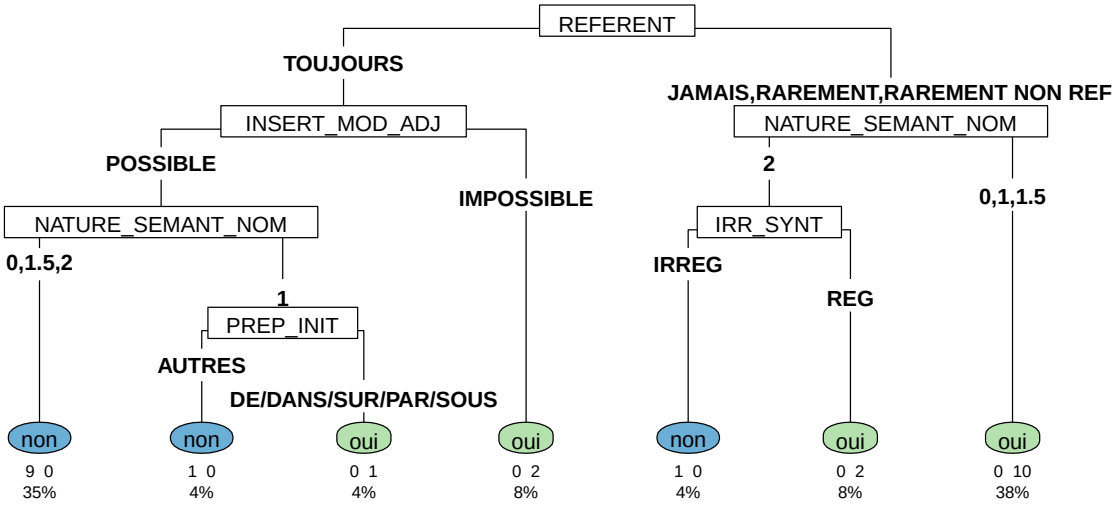


Figure 2 : Arbre de décision pour la matrice n° 1

Il s'agit cependant ici d'une première exploitation statistique des données issues de notre grille multicritère. Cette approche, pour être valide, devra être appliquée à un plus grand nombre de séquences candidates et à un échantillon équilibré entre les PrépComp et les structures libres pour chacune des matrices, et c'est là une troisième possibilité d'élaboration future de la méthode d'identification des PrépComp à peine esquissée dans ce travail.

Conclusion

Partant du constat qu'à ce jour on ne dispose pas de liste de prépositions complexes pour le français moderne, nous revenons tout d'abord sur les raisons de ce manque et sur la définition même de la préposition, et de la préposition complexe en particulier. Après avoir dressé le profil d'une PrépComp prototypique, nous avons développé une méthode d'identification des PrépComp dans l'objectif d'évaluer, pour une séquence candidate donnée, son degré d'adéquation à la PrépComp modèle. Les critères de classabilité, au nombre de 21, sont pour la plupart issus de la littérature existante sur le sujet ; certains d'entre eux ont été décomposés en paramètres plus fins, et plusieurs nouveaux critères ont été introduits pour à la fois assurer une bonne couverture des schémas de construction des PrépComp et essayer de fournir une grille d'identification plus performante. Ainsi, en combinant des critères qualitatifs d'ordre sémantique, morphologique et syntaxique d'une part, et des critères quantitatifs (fréquences et score d'information réciproque dans des corpus) d'autre part, nous situons sur une échelle de prototypicalité chaque expression d'un échantillon de 75 séquences candidates, dont certaines sont bien établies dans la classe des PrépComp, et d'autres des formations libres. La principale originalité de notre démarche vient du fait que les critères retenus ne sont pas considérés comme des conditions nécessaires et suffisantes pour inclure dans (ou exclure de) la classe des PrépComp les séquences candidates, mais comme des paramètres d'évaluation de leur ressemblance aux membres prototypiques de la classe. Cette méthode d'identification, qui reste à parfaire, jette une lumière nouvelle sur la sous-classe des PrépComp en français et devrait permettre à moyen terme d'en cerner (plus) efficacement les contours.

Bibliographie

Adler, S. (2001) : Les locutions prépositives : questions de méthodologie et de définition. *Travaux de linguistique*, 42-43,1, pp. 157-170.

Adler, S. (2007) : Locutions prépositives temporelles et modes d'anaphorisation, in : Trotter, D. (éd.) : *Actes du XXIVe Congrès International de linguistique et Philologies Romanes*. Niemeyer, Tübingen, pp. 495-508.

Blumenthal, P., S. Diwersy & J. Mielebacher (2005) : Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen. *Zeitschrift für romanische Philologie*, 121, pp. 49-83.

Borillo, A. (1991) : Le lexique de l'espace : prépositions et locutions prépositionnelles de lieu en français, in : Tasmowski, L. & A. Zribi-Hertz (éds.) : *Hommage à N. Ruwet*. Communication & Cognition, Gand, pp. 176-190.

Borillo, A. (1997) : Aide à l'identification des prépositions complexes de temps et de lieu. *Faits de langue*, 9, pp. 173-184.

Borillo, A. (2002) : Il y a prépositions et prépositions. *Travaux de Linguistique*, 42-43, pp. 141-155.

Brinton, L. & E.C. Traugott (2005) : *Lexicalization and Language Change*. Cambridge University Press, Cambridge.

Brunot, F. & C. Bruneau (1949) : *Précis de grammaire historique de la langue française*. Masson, Paris.

Bybee, J.L. (2007) : *Frequency of Use and the Organization of Language*. Oxford University Press, Oxford.

Cifuentes Honrubia, J.L. (2003) : *Locuciones prepositivas – sobre la gramaticalización preposicional en español*. Universidad de Alicante, Alicante.

De Mulder, W. & D. Stosic (2009) : Présentation. *Langages*, 173, pp. 3-14.

Fagard, B. & W. De Mulder (2007) : La formation des prépositions complexes : grammaticalisation ou lexicalisation ? *Langue française*, 156, pp. 9-29.

Fagard, B., D. Stosic & A. Mardale (2016) : Approaches to Romance complex adpositions. The case of French and Romanian. Atelier *Emergence and Evolution of Complex Adpositions in European Languages*, dans le cadre du *49ème Congrès annuel de la Societas Linguistica Europaea* (SLE 2016), Université de Naples Federico II, Naples (31 août-3 septembre 2016).

Fagard, B., W. De Mulder & T. Hoelbeek (sous presse) : Sémantique des prépositions, in : Marchello-Nizia, C., B. Combettes, S. Prévost & T. Scheer (éds.) : *Grande grammaire historique du français*. De Gruyter Mouton, Berlin & New York.

Fagard, B., J. Pinto de Lima, E. Smirnova & D. Stosic (éds.) (soumis) : *Complex adpositions in Romance: Emergence and variation*.

Feltgen, Q. (2017) : *Statistical Physics of Language Evolution: The Grammaticalization Phenomenon.* Thèse de doctorat, PSL, Paris.

Goyens, M., B. Lamiroy & L. Melis (2002) : Déplacement et repositionnement de la préposition *à* en français. *Linguisticae Investigationes*, XXV,2, pp. 275-310.

Gross, G. (1996) : *Les expressions figées en français. Noms composés et autres locutions*. Ophrys, Paris.

Gross, G. (2006) : Sur le statut des locutions prépositives. *Modèles linguistiques*, 53, pp. 33-50.

Hoffmann, S. (2005) : *Grammaticalization and English Complex Prepositions. A corpus-based analysis*. Routledge, London & New York.

Hopper, P.J. & E.C. Traugott (2003 [1993]) : *Grammaticalization*. Cambridge University Press, Cambridge.

Huddleston, R. & G.K. Pullum (2002) : *The Cambridge Grammar of the English Language.* Cambridge University Press, Cambridge.

Kurzon, D. & S. Adler (éds.) (2008) : *Adpositions. Pragmatic, semantic and syntactic perspectives*. John Benjamins, Amsterdam & Philadelphia.

Lehmann, C. (1985) : Grammaticalization: Synchronic Variation and Diachronic Change. *Lingua E Stile*, 20, pp. 303-318.

Lehmann, C. (1995 [1982]) : *Thoughts on grammaticalization*. Lincom Europa, München & Newcastle.

Le Pesant, D. (2006) : Classification à partir des propriétés syntaxiques. *Modèles linguistiques*, 53, pp. 51-74.

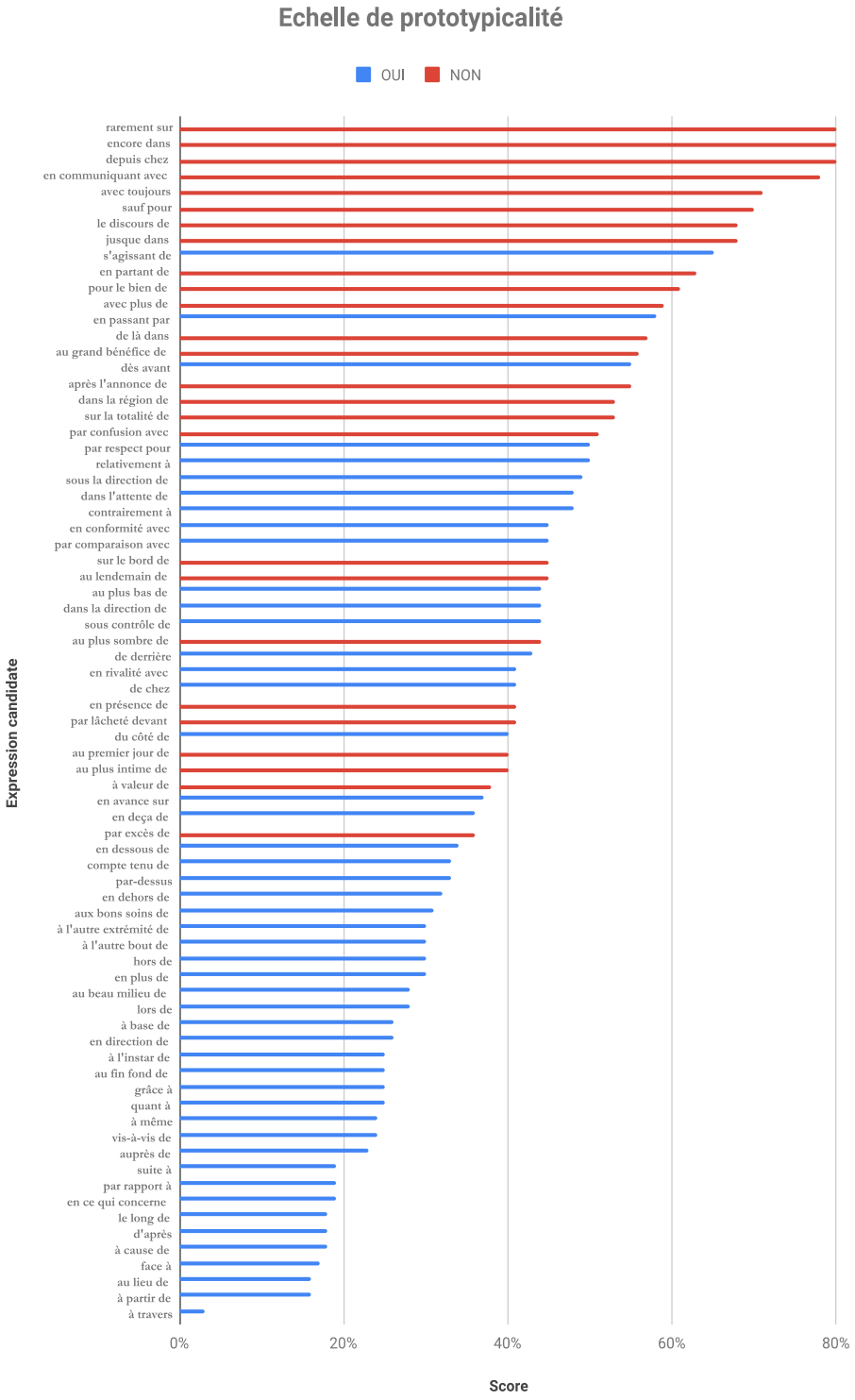
Littlefield, H. (2006) : *Syntax and Acquisition in the Prepositional Domain : Evidence from English for Fine-grained Syntactic Categories*. Thèse de doctorat, Université de Boston.

Mardale, A. (2009) : Sur la distinction entre prépositions lexicales et prépositions fonctionnelles, in : Zafiu, R., B. Croitor & A.-M. Mihail (éds.) : *Studii de gramatică. Omagiu Doamnei Profesoarei Valeria Guţu-Romalo*. Editura Universităţii din Bucureşti, Bucureşti, pp.115-125.
Melis, L. (2003) : *La préposition en français*. Ophrys, Paris.
Pottier, B. (1962) : *Systématique des éléments de relation*. Klincksieck, Paris.
Spang-Hanssen, E. (1963) : *Les prépositions incolores du français moderne*. Gads Forlag, Copenhague.
Stosic, D. (2012) : *En passant par* : une expression en voie de grammaticalisation ? *CORELA*, Numéro thématique *Langue, espace, cognition*.
Stosic, D. (en préparation) : Les prépositions complexes en français.
Svorou, S. (1994) : *The Grammar of Space*. John Benjamins, Amsterdam & Philaldelphia.
Tremblay, M. (1996) : Lexical and non-lexical prepositions in French, in : Di Sciullo, A. (éd.) : *Configurations*. Cascadilla Press, Somerville, MA, pp. 79-98.
Vandeloise, C. (1993) : Présentation. *Langages*, 110, pp. 5-11.
Zipf, G. (1936) : *The Psychobiology of Language*. Routledge, London.

Annexe 1

Matrice n°	Patron	Hypothèse	Expression candidate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Total	Score MAX	%
1	P Dét N P	oui	*sous la direction de*	0	1	0	0	1	1	1	1	1	2	1	2	2	1	0	1	0	1	1	3	0	**19.5**	40	**49%**
1	P Dét N P	oui	*dans l'attente de*	0	1	0	0	1	1	1	1	1	1	1	2	2	1	0	1	0	1	1	3	1	**19**	40	**48%**
1	P Dét N P	oui	*dans la direction de*	0	1	0	0	1	1	1	0	0	1	1	2	2	1	0	1	0	1	1	3	1	**17.5**	40	**44%**
1	P Dét N P	oui	*du côté de*	0	1	0	0	1	1	1	1	0	1	1	3	2	1	2	0	0	0	1	0	0	**16**	40	**40%**
1	P Dét N P	oui	*à l'instar de*	0	0	0	0	1	1	1	0	2	0	0	0	0	0	2	0	0	1	0	2	0	**10**	40	**25%**
1	P Dét N P	oui	*au lieu de*	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	1	1	1	0	0	0	**6.5**	40	**16%**
1	P Dét N P	non	*pour le bien de*	0	2	0	2	1	1	1	0	2	2	1	2	2	1	1	0	1	1	1	3	1	**24.5**	40	**61%**
1	P Dét N P	non	*dans la région de*	0	1	0	0	1	1	1	1	0	2	1	3	3	1	0	1	0	1	1	3	1	**21**	40	**53%**
1	P Dét N P	non	*après l'annonce de*	0	2	0	2	1	1	1	1	1	2	1	1	2	0	1	0	1	1	1	3	0	**22**	40	**55%**
1	P Dét N P	non	*sur la totalité de*	0	1	0	2	1	1	1	1	2	2	1	0	0	0	1	2	1	1	1	3	0	**21**	40	**53%**
1	P Dét N P	non	*sur le bord de*	0	1	0	0	1	1	1	1	0	2	1	2	2	1	0	0	0	1	1	3	0	**18**	40	**45%**
1	P Dét N P	non	*au lendemain de*	0	0	0	0	1	1	1	1	2	2	1	1	1	1	1	2	0	0	1	3	0	**18**	40	**45%**
1	P N P(de)	oui	*sous contrôle de*	0	1	0	0	0	0	1	0	1	1	1	2	2	1	0	1	1	1	1	3	1	**17.5**	40	**44%**
1	P N P(de)	oui	*à base de*	0	0	0	0	0	0	0	0	0	1	1	1	1	1	2	1	0	1	0	2	0	**10.5**	40	**26%**
1	P N P(de)	oui	*en direction de*	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	2	0	0	**10.5**	40	**26%**
1	P N P(de)	oui	*à cause de*	0	0	0	0	0	0	0	0	2	0	0	0	0	1	2	2	0	1	0	0	0	**7**	40	**18%**
1	P N P(de)	non	*en présence de*	0	0	0	0	0	0	1	1	2	2	1	1	0	1	2	2	0	1	1	2	0	**16.5**	40	**41%**
1	P N P(de)	non	*par excès de*	0	1	0	0	0	0	0	1	2	2	1	0	2	0	0	0	1	1	1	3	0	**14.5**	40	**36%**
1	P N P(de)	non	*à valeur de*	0	0	0	0	0	0	0	0	2	1	0	2	3	1	0	0	1	0	1	3	1	**15**	40	**38%**
1	P Dét Adj N de	oui	*aux bons soins de*	0	0	0	0	1	1	1	0	2	1	1	0	1	0	0	0	1	0	1	3	0	**12.5**	40	**31%**
1	P Dét Adj N de	oui	*à l'autre extrémité de*	0	0	0	0	1	1	1	1	0	2	0	0	0	0	0	0	1	1	1	3	0	**12**	40	**30%**
1	P Dét Adj N de	oui	*à l'autre bout de*	0	0	0	0	1	1	1	1	0	2	0	0	0	0	0	0	1	1	1	3	0	**12**	40	**30%**
1	P Dét Adj N de	oui	*au beau milieu de*	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	1	1	1	3	1	**11**	40	**28%**
1	P Dét Adj N de	oui	*au fin fond de*	0	0	0	0	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	3	0	**10**	40	**25%**
1	P Dét Adj N de	non	*au grand bénéfice de*	0	0	0	0	1	1	1	1	2	2	1	0	2	1	1	3	1	1	1	3	1	**22.5**	40	**56%**
1	P Dét Adj N de	non	*au premier jour de*	0	0	0	0	1	1	0	1	2	2	1	0	0	0	1	2	1	1	1	3	0	**16**	40	**40%**
2	P N P(#de)	oui	*par respect pour*	0	1	0	2	0	0		1	2	1	1	1	2	1	1	2	0	1	1	3	1	**19.5**	39	**50%**
2	P N P(#de)	oui	*en conformité avec*	0	0	0	1	0	0		0	2	2	1	1	2	1	2	1	0	1	1	3	0	**17.5**	39	**45%**
2	P N P(#de)	oui	*par comparaison avec*	0	1	0	1	0	0		1	2	2	1	1	2	1	0	1	0	1	1	3	0	**17.5**	39	**45%**
2	P N P(#de)	oui	*en rivalité avec*	0	0	0	1	0	0		0	2	1	1	1	3	0	0	0	1	1	1	3	1	**16**	39	**41%**
2	P N P(#de)	oui	*en avance sur*	0	0	0	2	0	0		0	1	1	1	0	1	1	1	1	0	1	1	3	1	**14.5**	39	**37%**
2	P N P(#de)	oui	*par rapport à*	0	1	0	1	0	0		0	2	1	0	0	0	1	2	1	0	0	0	0	0	**7.5**	39	**19%**
2	P N P(#de)	non	*par confusion avec*	0	1	0	1	0	0		1	2	2	1	0	3	1	0	2	1	1	1	3	1	**20**	39	**51%**
2	P N P(#de)	non	*par lâcheté devant*	0	1	0	2	0	0		1	2	2	1	0	0	0	0	0	1	1	1	3	1	**16**	39	**41%**
3	N P	oui	*grâce à*	1		0	1	0	0			2	0	0	0	0	1	2	2	0	0	0	0	0	**9**	36	**25%**
3	N P	oui	*suite à*	1		0	1	0	0			1	0	0	0	0	1	2	1	0	0	0	0	0	**7**	36	**19%**
3	N P	oui	*face à*	1		0	1	0	0			0	0	0	0	0	1	2	1	0	0	0	0	0	**6**	36	**17%**
3	N V P	oui	*compte tenu de*	1		0	0	0	0			2	1	1	0	0	0	2	2	0	1	0	3	0	**12**	36	**33%**
3	N P N P	oui	*vis-à-vis de*	1		0	0	0	0			2	0	0	0	0	1	2	1	0	0	0	2	0	**8.5**	36	**24%**
4	Dét N P	oui	*le long de*	1		0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1	**7**	38	**18%**
4	Dét N P	non	*le discours de*	1		0	0	1	1	1	1	2	2	1	3	3	1	0	3	1	1	1	3	0	**26**	38	**68%**
5	P N	oui	*à travers*	0	0	1		0	0			0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	37	**3%**
6	Adv P	oui	*relativement à*	1		0	1	1									1	0	0	0	1	1	3	1	**10**	20	**50%**
6	Adv P	oui	*contrairement à*	1		0	1	1									1	2	1	0	1	1	1	0	**9.5**	20	**48%**
6	Adv P	oui	*lors de*	1		0	0	1									1	1	1	0	1	0	0	0	**5.5**	20	**28%**
6	Adv P	oui	*quant à*	1		0	1	1									0	0	0	1	1	0	0	0	**5**	20	**25%**
6	Adv P	oui	*auprès de*	1		0	0	1									1	1	1	0	0	0	0	0	**4.5**	20	**23%**
6	Adv P	non	*rarement sur*	1		0	2	1									1	2	2	1	1	1	3	1	**16**	20	**80%**
6	Adv P	non	*encore dans*	1		0	2	1									1	2	3	1	1	1	3	0	**16**	20	**80%**
6	V P	oui	*s'agissant de*	1		0	0	1									1	2	2	1	1	1	3	0	**13**	20	**65%**
6	V P	oui	*en passant par*	1		0	2	1									1	2	2	0	0	1	2	0	**11.5**	20	**58%**
6	V P	non	*en communiquant avec*	1		0	1	1									1	2	3	1	1	1	3	1	**15.5**	20	**78%**
6	V P	non	*en partant de*	1		0	0	1									1	2	2	1	1	1	3	0	**12.5**	20	**63%**
7	P P	oui	*dès avant*	0	2	0	2	1									1	1	1	0	0	0	3	1	**12**	22	**55%**
7	P P	oui	*de chez*	0	1	0	2	1									0	0	0	1	0	1	2	1	**9**	22	**41%**
7	P P	oui	*hors de*	0	2	0	0	1									1	1	1	0	0	0	1	0	**6.5**	22	**30%**
7	P P	oui	*d'après*	0	1	0	2	1									0	0	0	0	0	0	0	0	**4**	22	**18%**
7	P P	non	*depuis chez*	0	2	0	2	1									1	2	3	1	1	1	3	1	**17.5**	22	**80%**
7	P P	non	*sauf pour*	0	2	0	2	1									1	2	2	1	1	1	3	0	**15.5**	22	**70%**
7	P P	non	*jusque dans*	0	2	0	2	1									1	2	2	0	1	1	3	0	**15**	22	**68%**
7	P Adv P	oui	*en deça de*	0	0	0	0	1									1	1	1	0	0	1	3	0	**8**	22	**36%**
7	P Adv P	oui	*en dessous de*	0	0	0	0	1									1	1	1	0	0	1	3	0	**7.5**	22	**34%**
7	P Adv P	oui	*en dehors de*	0	0	0	0	1									1	2	1	0	0	1	1	0	**7**	22	**32%**
7	P Adv P	oui	*en plus de*	0	0	0	0	0									1	2	1	0	1	1	0	1	**6.5**	22	**30%**
7	P Adv P	non	*avec plus de*	0	2	0	0	0									1	2	3	1	1	1	2	0	**13**	22	**59%**
7	P Adv P	non	*de là dans*	0	1	0	2	0									1	0	2	1	1	1	3	1	**12.5**	22	**57%**
7	P V P	oui	*à partir de*	0	0	0	0	0									1	2	1	0	0	0	0	0	**3.5**	22	**16%**
8	P Adv	oui	*de derrière*	0	1	1		1									0	0	0	0	1	1	3	1	**9**	21	**43%**
8	P Adv	oui	*par-dessus*	0	1	1		1									0	0	0	0	0	1	3	0	**7**	21	**33%**
8	P Adv	oui	*à même*	0	0	1		1									0	0	0	0	1	0	2	0	**5**	21	**24%**
8	P Adv	non	*avec toujours*	0	2	1		1									1	1	2	1	1	1	3	1	**15**	21	**71%**
8	autres	oui	*en ce qui concerne*	0	0	1		0									0	0	0	1	0	1	1	0	**4**	21	**19%**
9	P Dét plus Adj P	oui	*au plus bas de*	0	0	0	0	1	1	1	1						0	0	0	1	1	1	3	1	**11**	25	**44%**
9	P Dét plus Adj P	non	*au plus sombre de*	0	0	0	0	1	1	1	1						0	0	0	1	1	1	3	1	**11**	25	**44%**
9	P Dét plus Adj P	non	*au plus intime de*	0	0	0	0	1	1	1	1						0	0	0	1	0	1	3	1	**10**	25	**40%**

Annexe 2


Echelle de prototypicalité
OUI
NON
Expression candidate
rarement sur
encore dans
depuis chez
en communiquant avec
avec toujours
sauf pour
le discours de
jusque dans
s'agissant de
en partant de
pour le bien de
avec plus de
en passant par
de là dans
au grand bénéfice de
dès avant
après l'annonce de
dans la région de
sur la totalité de
par confusion avec
par respect pour
relativement à
sous la direction de
dans l'attente de
contrairement à
en conformité avec
par comparaison avec
sur le bord de
au lendemain de
au plus bas de
dans la direction de
sous contrôle de
au plus sombre de
de derrière
en rivalité avec
de chez
en présence de
par lâcheté devant
du côté de
au premier jour de
au plus intime de
à valeur de
en avance sur
en deça de
par excès de
en dessous de
compte tenu de
par-dessus
en dehors de
aux bons soins de
à l'autre extrémité de
à l'autre bout de
hors de
en plus de
au beau milieu de
lors de
à base de
en direction de
à l'instar de
au fin fond de
grâce à
quant à
à même
vis-à-vis de
auprès de
suite à
par rapport à
en ce qui concerne
le long de
d'après
à cause de
face à
au lieu de
à partir de
à travers
0%
20%
40%
60%
80%
Score

Adresses des auteurs :

Dejan Stosic (MCF en Sciences du langage)
CLLE-ERSS
Maison de la Recherche
Université Toulouse Jean Jaurès
5, allées Antonio Machado
F-31058 Toulouse Cedex 9

Benjamin Fagard
Lattice – Ecole Normale Supérieure
1 rue Maurice Arnoux
92120 Montrouge