



**HAL**  
open science

# Lemmatisation et figement lexical : les locutions de type SV

G rard Petit

► **To cite this version:**

G rard Petit. Lemmatisation et figement lexical : les locutions de type SV. Cahiers de Lexicologie, 2003, 82 (1), pp.30-57. hal-00644420

**HAL Id: hal-00644420**

**<https://hal.science/hal-00644420>**

Submitted on 28 Nov 2011

**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 destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.

# Lemmatisation et figement lexical : les locutions de type SV

Gérard PETIT\*

## 0. Introduction

La lemmatisation des locutions de type SV<sup>1</sup> (*tenir le coup, casser les pieds à qqn, manger les pissenlits par la racine...*) ne constitue pas actuellement un objet d'étude pour la lexicologie, ni plus généralement les Sciences du langage<sup>2</sup>. Le fait est imputable en partie à la situation paradoxale d'un tel questionnement : (i) les investigations menées sur les séquences polylexématiques<sup>3</sup> (désormais SP) portent sur l'identification et la formalisation de propriétés sémantiques et syntaxiques (notamment Gross 1996, Mejri 1997, 1998, Blanco et Moreno 1997), et ce dans le cadre général d'une réflexion sur le figement. Ceci a pour conséquence de laisser induire que l'établissement de leur forme lemmatique n'est pas susceptible d'interrogation ou bien constituerait une préoccupation d'intérêt mineur ; (ii) les observations sur la lemmatisation du matériau linguistique sont quant à elles traditionnellement circonscrites au domaine de l'unité lexicale confondue avec l'unité monolexématique. De ce fait les SP se voient reléguées dans un espace intersticiel entre lexicologie et phraséologie (Gross 1996, Rey 1977) où les conditions de leur lemmatisation ne sont pas envisagées.

Si la réflexion sur l'établissement de la structure lemmatique des SP reste un terrain en friche, elle ne constitue pas pour autant un sujet trivial pour la lexicologie. D'abord parce que l'assignation d'un lemme à une séquence consigne l'appartenance de celle-ci à la langue en général, et au lexique en particulier. Les SP ne dérogent pas à la règle des unités monolexématiques. Elles doivent être répertoriées à l'inventaire de la langue et nécessitent pour ce faire une forme identifiante qui permette de les distinguer tout en manifestant leur identité sémiotique. Ensuite parce qu'une pratique de lemmatisation des SP existe déjà, pour l'établissement des nomenclatures lexicographiques ou pour satisfaire au besoin de disposer de formes citationnelles entre autres dans la littérature linguistique et/ou grammaticale. Cette pratique se caractérise par un double impensé théorique : (i) du fait qu'il est intégré à une prédication (texte, article de dictionnaire...), le lemme est soumis aux contraintes rhétoriques qui président à la mise en verbe et lui imposent notamment une linéarité ; (ii) par suite d'une confusion entre adaptation discursive et identité linguistique, la suite lemmatique résultante est

---

\* EA 2290 (SYLED), Université de Paris X

<sup>1</sup> L'appellation de ces séquences est très diversifiée et recouvre des réalités hétérogènes. Sur ce point voir notamment les synthèses de (Mejri 1997, Gross 1996 et Candel 1995)

<sup>2</sup> Exception faite pour les disciplines liées à l'informatique : Intelligence artificielle, Traitement automatique du langage et Lexique grammatical. Toutefois leur objet est moins d'établir l'identifiant linguistique d'une séquence que de mettre au point, *via* des gabarits formels, des procédures d'extraction et d'interprétation automatisées

<sup>3</sup> Nous utiliserons ce terme, emprunté à Mejri 1997, de préférence à *locution, séquence figée, syntagme figé* etc.

considérée non pas comme une forme communicationnelle, mais comme la dénomination réelle de l'entité linguistique qu'elle exprime.

Dans une première partie, nous exposerons les difficultés (et impasses) linguistiques auxquelles conduit la conception linéaire du lemme. Dans une seconde partie, nous lui en opposerons une autre, dont la fonction est d'exprimer l'identité sémiotique<sup>4</sup> de la SP et d'en fournir une description structurale. La SP apparaîtra alors comme une séquence : (i) chaînée, c'est-à-dire scandée de zones de disruptions ; (ii) de densité hétérogène, car constituée de lieux d'opacification et de lieux de compositionnalité sémantique ; (iii) spatiale, les maillons du chaînage pouvant être des structures d'instanciation de variables.

Nous serons conduit à distinguer deux entités : le lemme lexical et le lemme citationnel, chacune ayant sa fonctionnalité propre<sup>5</sup> (et ses limites). Nous constaterons ainsi que le lemme lexical d'une SP ne s'établit pas sur les mêmes principes que celui des unités lexicales monolexématiques (où seule la flexion se voit neutralisée), mais nécessite cruciallement la prise en compte des spécificités lexico-syntaxiques des constituants. De fait, s'il respecte une linéarité syntagmatique ce n'est que partiellement, de manière occurrenceielle et non pas constitutive. On l'aura compris, nous considérons les SP comme des lexèmes à part entière.

## 1. Lemmatisation et lexicologie : l'influence du mot

### 1.1. Les apories de la lexicographie

L'enseignement de la lexicologie connaît depuis peu d'un regain d'intérêt, si l'on en juge par la publication rapprochée de quatre manuels : Mortureux (1997), Niklas-Salminen (1997), Lehmann et Martin-Berthet (1998), Rey-Debove (1998). Toutefois seuls Mortureux et Rey-Debove abordent explicitement la question de la lemmatisation du matériau lexical et celle du lemme comme forme identifiante. La description de Mortureux (1997 : 16) est à cet égard la plus complète<sup>6</sup> :

“Tout utilisateur de dictionnaire a pu constater que les entrées se présentent régulièrement sous la forme du singulier pour les noms et les adjectifs (ceux-ci au masculin), et de l'infinitif présent actif pour les verbes. L'opération qui consiste à ramener à une forme unique les formes fléchies des mots variables (noms, adjectifs et verbes, tels qu'ils figurent dans les discours) s'appelle la lemmatisation. C'est une manifestation exemplaire de l'abstraction qui préside à la conversion des vocables en entrées, qui sont elles-mêmes une image des lexèmes. En effet, elle consiste à repérer, pour en faire l'entrée, l'unité formelle commune à toutes les formes variables sous lesquelles il peut apparaître (singulier ou pluriel, masculin ou féminin pour les noms et les adjectifs ; divers modes, temps ou personne pour les verbes).”

Elle reflète, en le condensant, le stéréotype associé à la notion de lemmatisation dans la conscience et la culture linguistiques courantes (pour preuve, voir notamment, Rey-Debove 1998, 1971 ; Collinot et Mazière 1998 ; Martins-Baltard ed. 1995, 1997 ; Fiala, Lafon et Piguet ed. 1997 ; Rey 1977 ; Dubois 1971 ; Wagner 1967), lequel implique que<sup>7</sup> :

---

<sup>4</sup> Dans une perspective benvénistienne, nous considérerons comme sémiotique toute propriété relative au statut linguistique d'une unité, à son mode de signification, à sa structuration lexico-syntaxique (voir Petit 1998 a et 1999 c)

<sup>5</sup> Raison pour laquelle nous utiliserons, comme forme communicationnelle, le lemme textuel des SP

<sup>6</sup> Rey-Debove reste très allusive. Dans la même perspective que Mortureux, elle n'aborde la question qu'en comparant (1998 : 248) l'entrée et l'exemple dans les dictionnaires et en qualifiant ce dernier de séquence délemmatisée, sans toutefois avoir précisé en quoi consistait le caractère lemmatique de l'entrée

<sup>7</sup> Nous n'argumenterons pas sur la nécessité pour le lemme de constituer *l'unité formelle commune à toutes les formes variables sous lesquelles* [l'unité lexicale] *peut apparaître*. Déjà chez (Dubois et Dubois 1971 : 61-64) la question est en débat.

- a. pour le lexicologue, les questions relatives à la lemmatisation trouvent leur domaine de pertinence privilégié dans l'application lexicographique, plus précisément concernant les procédures de codage qui président à l'enregistrement d'un item par les dictionnaire. Dans ces circonstances la réflexion se limite soit à énoncer des règles techniques (celles qui sont jugées les plus saillantes seulement, et de manière condensée) :

“Ainsi l'entrée d'un verbe est la forme de l'infinitif (*chanter*) ; celle-ci recouvre les différentes combinaisons de la racine et des désinences de temps et de nombre (*chantait, chantions, chanteras*, etc.) l'infinitif n'étant qu'une des formes du verbe. [...] lorsque [les noms] ne comportent que des combinaisons morphème + pluriel ou morphème + féminin, la forme de base est le mot dépourvu de marques de nombre et de genre : *mur* (pour *mur* et *murs*), *enfant* (pour *enfant* masculin et *enfant* féminin). Encore faut-il constater que, lorsque des informations graphiques doivent être données sur le pluriel ou le féminin, le mot d'entrée est suivi des variations de la syllabe finale (*paysan, anne*) : autrement dit, il est suivi de l'indication des allomorphes lorsque ceux-ci sont différents *graphiquement*.” (Dubois et Dubois 1971 : 61-64)<sup>8</sup>

et, parfois seulement, les difficultés d'application auxquels elles donnent lieu :

“Cette dénomination de la classe des occurrences d'un mot par une forme arbitraire a eu parfois pour conséquences de constituer comme entrée des termes qui ne figurent pas dans les performances verbales. Ainsi certains verbes ne se rencontrent que sous la forme pronominale (*s'accroupir, s'adonner*, etc.) ; si le lexicographe considère que les formes pronominales ne sont en réalité que des formes conjuguées d'une verbe de base, intransitif ou transitif, ils sera conduit à définir une entrée du type *accroupir* verbe transitif, *adonner* verbe transitif, et à faire suivre cette entrée de l'indication “inusité comme transitif”. Autrement dit, il construit une entrée arbitraire afin de normaliser la règle d'entrée des verbes. Il en est de même pour les noms qui ne sont employés qu'au pluriel : *ténèbre(s)* aura parfois dans un dictionnaire une entrée “*ténèbre (inusité au singulier)*” ; ou pour des formes verbales indépendantes de l'infinitif et cependant regroupées sous cette dénomination : *ci-gît* est renvoyé à *gésir*.” (Dubois et Dubois 1971 : 64)<sup>9</sup>

Soit elle pointe la dimension discursive et historique de la procédure et son marquage idéologique :

“L'entrée lexicographique a une fonction prédictive : elle prévoit, en les objectivant une fois pour toutes, toutes les occurrences à venir de la forme mot modélisée. PIPE est bien le nom dénommant l'ensemble PIPE des occurrences écrites ou orales de PIPE. C'est bien dans cette perspective que l'on consulte un dictionnaire. Qu'il s'agisse d'orthographe ou de signification, c'est à la forme canonique de l'entrée que le lecteur consultant se reportera pour y confronter un emploi contingent de cette même forme. Dans ce sens le dictionnaire est non seulement un prêt-à-bien-parler le mot mais aussi prêt-à-bien-écrire le mot.” (Collinot et Mazière 1998 : 89-90)

Fait remarquable, ces analyses énoncent et stigmatisent dans l'ensemble les mêmes règles : choix *a priori* du singulier et du masculin pour le nom et l'adjectif, de l'infinitif pour le verbe<sup>10</sup>. Toutefois la lemmatisation ne constitue pas pour elles un objet de recherches principal, susceptible de déboucher sur des propositions visant à modifier les règles précitées ou à en énoncer de nouvelles. Rien d'étonnant donc si une réflexion sur les nécessités et contingences liées à l'étiquetage

---

<sup>8</sup> Nous sélectionnons volontairement des références anciennes afin de marquer le lestage historique de la conceptualisation sur laquelle vit la lexicologie actuelle, mais aussi la nécessité pour cette dernière de repenser certaines de ses bases. Paradoxalement (?) ce sont les conceptualisations anciennes (Dubois et Dubois 1971 ; Rey-Debove 1971 ; Rey 1977) qui restent les plus détaillées, la lexicologie contemporaine semblant s'être totalement détachée d'une réflexion globale sur l'identifiant des unités qu'elle reconnaît

<sup>9</sup> Le lecteur est également invité à se reporter à Rey (1977 : *L'unité lexicale et l'unité de traitement lexicographique*)

<sup>10</sup> En ceci on est très loin de la vivacité que la réflexion sur le lemme connaît dans le domaine du TAL ou de la linguistique de corpus (voir à cet égard Habert, Nazarenko et Salem 97)

métalinguistique des unités passe au second rang des préoccupations, quand elles n'est pas tout simplement évacuée.

- b. seules sont concernées les unités syntaxiques minimales, N-V-Adj, lesquelles constituent le lexique prototypique tant pour la conscience courante que pour celle du linguiste. Lorsque les unités infralexicales ou polylexématiques sont prises en compte (essentiellement Rey-Debove 1998 : 251-257 ; 1971 : 112-120 ; Rey 1977 : 34-35, mais également Dubois et Dubois 1971 : 39-40) ce n'est que pour signaler leur (in)capacité à intégrer la nomenclature des dictionnaires (avec une exception pour Rey-Debove 1971 : 302-303) ou spécifier leurs relations - plus ou moins problématiques - avec ce lexique prototypique (notamment Lehmann et Martin-Berthet 1998 : 36, 176 ; Niklas-Salminen 1997 : 78 ; Rey 1977 : 183-200).

Si l'on croise cette donnée avec la précédente, seul se voit visé par une réflexion sur son identité lemmatique le signe, *in fine* dans sa capacité à fournir une entrée de dictionnaire. Or si l'on sait par ailleurs que la logique lexicographique repose sur l'unité-mot monolexématique (*chien, manger, joli*), polylexématique à condition que la forme en soit ramenée à celle du mot graphique (*coffre-fort, essuie-mains, bas-jointé*), mais exceptionnellement synaptique (*pomme de terre, chemin de fer*)<sup>11</sup>, la locution, telle que nous l'entendons ici, ne rentre pas dans le format requis pour faire l'objet sinon d'une théorisation, du moins d'un questionnement. Le fait est que ces ouvrages<sup>12</sup> à caractère métalexicographique (à défaut d'être lexicologiques ou métalexicologiques) convoquent, lorsqu'ils traitent des locutions (et ils n'en traitent que concernant leurs propriétés sémantiques, syntaxiques et culturelles), des lemmes reçus, préconstruits, empruntés à des corpus antérieurs donc puisés dans un fond traditionnel et estimé transparent.

Il est patent par ailleurs que les descriptions portant spécifiquement sur la locution (Mejri 1997, Martins-Baltard dir 1995, 1997, Filala, Lafon et Piguët ed. 1997, Gross 1996 pour ne prendre que des références récentes) tendent à délaisser la question du lemme. La seule exception, à notre connaissance, est constituée par (Blanco et Moreno 1997) qui pointent quelques problèmes posés dans le domaine de la lexicographie bilingue. Concernant l'établissement de la forme lemmatique, ces ouvrages empruntent au même fond commun que les précédents.

Les études lexicales sont les victimes (et le théâtre) d'une inertie :

- qui remet les questions relatives à la forme du signifiant aux mains de la phonologie, et parfois de la morphologie, se déchargeant des autres aspects sur la lexicographie ou la didactique (notamment pour les questions d'orthographe) et plus résiduellement sur la graphématique ;

- qui recourt à un impensé théorique touchant la forme citationnelle des items : le fond commun ne se voit jamais interrogé (par la lexicologie) sur sa constitution (lexicologique). Ce dernier point est peut-être d'une importance secondaire concernant les lexèmes simples (mais pas les lexèmes construits : Kerleroux 1999, 1996 ; Villoing 1999 ; Amiot 1997). Il ne l'est pas dès que l'on aborde le domaine des unités polylexématiques.

---

<sup>11</sup> La pratique actuelle s'est quelque peu assouplie depuis Dubois et Dubois (1971 : 62), qui constatait : “[...] les termes composés dont les composants ne sont pas réunis par le trait d'union et qui sont ainsi séparés par un blanc ne sont pas reconnus comme des entrées : ainsi *pomme de terre* et *chou rave* (contrairement à *chou-fleur*), *compte rendu* (contrairement à *compte-gouttes*).” Voir également les remarques de (Rey-Debove 1971 : 112-118)

<sup>12</sup> Dans une visée purement métalexicographique : Collinot et Mazière (1998), Rey-Debove (1971), Dubois et Dubois (1971), Wagner (1967). De manière composite entre la métalexicographie et la lexicologie : Rey-Debove (1998), Mortureux (1997) et Rey (1977). Les approches purement lexicologiques (Lehmann et Martin-Berthet 1998, Niklas-Salminen 1997) n'abordent pas les questions de lemmatisation.

## 1.2. Les certitudes de la métalexigraphie

Seule<sup>13</sup> Rey-Debove (1971 : 302-306) propose un exemple de faire explicité en matière de lemmatisation de locution, au chapitre 7.2.3 *Traitement métalinguistique de l'exemple*, § 7.2.3.2 *La neutralisation*. Ces intitulés prêtent eux-mêmes à commentaire : la locution n'est perçue par l'auteur (mais également par les dictionnaires Robert dans leurs diverses éditions) que comme une variante de l'exemple lexicographique<sup>14</sup>. Elle n'est pas considérée comme une forme du matériau lexical (contrairement à ce que propose Rey 1977), mais comme une séquence qui ressortit prioritairement à la syntagmatique de l'énonciation. Pour cette raison elle voit sa fonction réduite au rôle d'illustration.

Envisagée comme neutralisation de fragments discursifs, la lemmatisation n'est pas reconnue comme telle (*i.e.* comme mise au format d'une unité linguistique de première articulation), mais comme modalité d'intégration d'un matériau *a priori* indifférencié (lexical, phraséologique, etc.) : non pas construction d'une identité sémiotique, mais déconstruction de variables énonciatives. À la décharge de l'auteur on observera que même concernant la mise aux normes des items lexicaux, les termes *lemmatisation* ou *lemme* ne sont pas non plus employés dans l'ouvrage précité.

La démonstration de Rey-Debove s'appuie sur le corpus suivant :

- 1. Les soldats de Verdun ont opposé une farouche résistance à l'ennemi
- 2. Le criminel opposera sûrement une résistance acharnée aux policiers
- 3. Quelqu'un oppose une vive résistance à son adversaire
- 4. Opposer une vive résistance à l'ennemi
- 5. Il oppose une résistance à son chef
- 6. Opposer une résistance à quelqu'un

où 1, 2 et 5 sont supposés figurer des énoncés attestés sous forme de phrases complètes. La relation entre 3 et 1, 2 est justifiée (p 303) par le fait que "ces deux phrases sont en quelque sorte les variables d'une phrase canonique caractérisée par une structure syntactique et des classes sémantiques permanentes", structure et classes que 3 figurera : "c'est en cherchant la partie commune aux deux phrases 1 et 2 que l'on produit l'énoncé 3" (*ibid.*).

Le passage de 3 à 6 s'opère d'abord par réduction de la flexion verbale : "cependant *oppose* n'est pas tout à fait satisfaisant, la forme du verbe la plus neutre étant l'infinitif, mais qui ne peut fonctionner avec les sujets. La solution est alors la suppression du sujet" (*ibid.*) d'où la phrase 4. L'énoncé 6 se déduit du rapprochement de 4 à 5, ce dernier étant une variable de 1 et 2. Et l'auteur de conclure :

"Cette dernière phrase [l'énoncé 6] est le modèle commun aux cinq phrases précédentes, qui permet de réaliser ces cinq variables, et bien d'autres. On appellera NEUTRALISATION cette recherche d'un modèle de phrase grâce auquel on puisse générer quantité de phrases particulières. Ces phrases particulières seront obtenues par le rétablissement d'un sujet, par celui de la forme conjuguée, par la détermination des constituants (adjectifs, adverbes, compléments) et par le remplacement des mots généraux à sémème pauvre par tous les mots à sémème plus riche qu'ils incluent logiquement." (*ibid.* p 304)

Les présupposés d'une telle démarche, pour partie harrissiens et pour partie générativistes, ne visent pas à établir prioritairement la forme lemmatique d'une SP mais *le modèle commun* à une série de

---

<sup>13</sup> Les ouvrages consacrés explicitement aux SP (Mejri 1997, Gross 1996, Martins-Baltar 1995, 1997, Fiala, Lafon et Pignet ed 1997) ne proposent pas de description, ni de théorisation de la mise au format d'une locution

<sup>14</sup> La sémiotique des dictionnaires Robert est à cet égard ambiguë : ne pas distinguer entre locutions et fragments de discours conduit à traiter les SP en exemples, et à dénaturer la fonction de l'exemple et réciproquement

performances. C'est précisément l'extraction de ce modèle qui importe pour nous dans la mesure où elle (et il) s'applique(nt) également aux SP, la lemmatisation constituant à la fois une préservation et une réduction du signifiant.

La procédure s'appuie sur trois opérations : le maintien, l'effacement et la reformulation. Sont maintenus :

- des items lexicaux : *opposer, résistance*
- la prédétermination du N2 *une*
- le joncteur *à* qui exprime la syntaxe du constituant.

En revanche sont supprimés :

- les marques de flexion verbale
- le SN1 : *les soldats de Verdun, le criminel, quelqu'un, il*
- la détermination du N2 : *farouche, acharnée, vive.*

La reformulation porte quant à elle sur :

- des constituants maintenus : *l'ennemi, [les] policiers, son adversaire, son chef > quelqu'un*
- mais aussi des constituants effacés dans une étape ultérieure : *les soldats de Verdun, le criminel, il > quelqu'un ; farouche, acharnée > vive.*

Le maintien de constituants est la manifestation d'une régularité structurale entre un modèle et ses variables d'instanciation, mais aussi du figement (à quelque degré qu'il se présente) inhérent à une SP. À cet égard il est vrai que *résistance*, dans certains de ses emplois prédicatifs, se construit comme argument d'*opposer* (et seulement de ce verbe) et intègre un SPREP régi par *à*<sup>15</sup>. La séquence *opposer Det résistance* est lexicalisée et se comporte comme un prédicat nominal à verbe support, dans la terminologie grossienne. Pareillement la réalisation de Det par un indéfini est obligatoire, sauf en cas de détermination du N par un superlatif absolu, une relative, un adjectif qui n'exprime pas le haut degré ou un Sprep en *de* :

- 7. opposer (une / \*la) résistance
- 8. opposer des résistances / certaines résistances / ? quelques résistances
- 9. opposer la résistance la plus vive
- 10. opposer (une / la) résistance qui convenait à la gravité de la situation / (une / la) résistance appropriée / (une / \*la) résistance (acharnée / de héros<sup>16</sup> / de plusieurs mois)

L'effacement procède de manière moins homogène. Il porte sur des éléments qui occupent une place non contrainte (*i.e.* non lexicalisée) dans la chaîne :

- 11. une (*farouche / vive / vaine / courageuse*) résistance (*acharnée / vaine / courageuse*)

ou des éléments qui n'appartiennent pas à la séquence, comme les constituants de la phrase (*sûrement*). Il porte également sur tout élément rendu incompatible avec la forme du lemme. La mise à l'infinitif du verbe bloque la présence d'un SN1 :

- 12. \*quelqu'un opposer une résistance à quelqu'un

La reformulation s'opère par inclusion logique, selon l'expression même de Rey-Debove, c'est-à-dire par la recherche sinon de l'hyperonyme du moins du superordonné le plus puissant. C'est ainsi que *vif* reformule *farouche* et *acharné*, mais aussi que *quelqu'un* est substitué à *ennemi, policier, adversaire* et *chef*.

<sup>15</sup> Nous n'aborderons pas ici la question de savoir si le Sprep se rattache au N *résistance* ou au V *opposer*

<sup>16</sup> Sauf si celui-ci ne comporte pas l'article zéro : *la résistance (du héros, des héros)*

Ce panorama des procédures de neutralisation (en fait, de lemmatisation) pourrait n'appeler aucun commentaire tant il vise à l'ergonomie de l'expression et qu'il semble atteindre son but par les voies les plus rationnelles qui soient (et les plus linguistiques semble-t-il). Il ne semble toutefois exempt de critiques que si on le prend pour ce qu'il est, c'est-à-dire la recherche d'une séquence textuelle - le lemme - admissible comme dénomination d'une entité linguistique indifférenciée puisque considérée *a priori* comme fragment de discours. En effet, les reformulations qu'il propose interviennent sur des places faiblement contraintes sur le plan lexical, au point de pouvoir pour certaines être subsumées par un pronom : *quelqu'un*. Or celui-ci est la version textuelle de la formalisation SN [+hum]. Si son instanciation dans le Sprep à *quelqu'un* est possible, son évacuation de la position sujet n'est rendue nécessaire que par la dimension textuellement correcte du lemme ainsi isolé :

- 12. \*quelqu'un opposer une résistance à quelqu'un
- 13. SN [+hum] opposer une résistance à SN [+hum]

Le marquage du sujet est obligatoire dans la mesure où le trait [-hum] sur SN1 bloque l'acceptabilité de l'énoncé. Autre reproche, la procédure elle-même de reformulation n'est que partiellement exploitée. Ne sont pas réalisés :

- 14. il a opposé une vive résistance
- 15. il lui oppose une résistance acharnée

qui présentent une structure avec effacement du Sprep et une avec sa pronominalisation. Ne sont pas envisagés non plus les cas où la détermination de *résistance* rendrait admissible l'article défini (voir plus haut).

D'un autre côté, l'effacement d'éléments s'accompagne d'une (et provoque une) suture textuelle de la séquence. Cette homogénéisation du matériau masque la présence d'interstices (possibilité d'insertion d'un Adj ou d'un pronom) dans le continu du signifiant. De plus elle ne prédit pas que l'instanciation de la séquence semblera plus naturelle avec une détermination sur *résistance* :

- 5. ? il oppose une résistance à son chef
- 16. il a opposé une (vive, courte) résistance (acharnée / de plusieurs mois) à son chef

ni que cette détermination est de nature à autoriser l'effacement du Sprep (voir plus haut), donc qu'il existe une concurrence entre ces composants.

Autre difficulté, soulevée par Blanco et Moreno (1997 : 174), le format dimensionnel du lemme :

“ [...] l'inclusion ou l'exclusion d'une des composantes de ces unités joue un rôle décisif pour ce qui est de sa catégorisation grammaticale et de sa localisation par rapport à la nomenclature ; ainsi, si le dictionnaire choisit de présenter comme lexie complexe *la tête la première*, l'on aurait affaire à une lexie complexe adverbiale très probablement située sous *tête*. Par contre, si l'on décide de mentionner comme lexies *tomber la tête la première* ou *s'y jeter la tête la première*, il faudrait parler de lexies complexes verbales, qui auraient en plus de fortes chances d'apparaître sous *tomber* ou *jeter*.”

La question n'est pas triviale<sup>17</sup> : l'intégration ou la non prise en compte d'éléments en position de limite gauche ou droite du signifiant modifient crucialement l'identité lexicale de la SP elle-même et sa dénomination métalinguistique. La séquence *la tête la première* ne se rencontre que dans un contexte gauche occupé par un V intransitif, exprimant un mouvement horizontal (y *aller...*) ou

---

<sup>17</sup> Nous ne nous arrêtons pas sur les problèmes relatifs au classement de la SP à la nomenclature du dictionnaire. Ils concernent l'organisation lexicographique elle-même et sont étrangers à notre objet



vertical descendant (*tomber, se jeter...*), préférentiellement borné à une de ses extrémités (*y aller / tomber / arriver / sauter / ?venir / \*aller*). Cette contrainte est supportée par un paradigme assez large de verbes, ce qui rend l'instanciation par un seul item extrêmement problématique, voire impossible. Le choix du dictionnaire a été de ne pas mentionner l'élément, d'une part parce qu'il n'est pas possible d'en faire figurer un (et un seul), mais aussi probablement parce que la paradigmatisme du verbe a induit pour le lexicographe que la position qu'il occupe n'est pas contrainte. Cet exemple démontre le caractère non linguistique (et passablement néfaste) des contraintes textuelles qui pèsent sur l'identification d'une SP.

### 1.3. les flottements de la lexicographie et de la linguistique

Les dictionnaires ne respectent pas systématiquement le schéma de lemmatisation exposé par (Rey-Debove 1971). De fait, les critiques de Blanco et Moreno prennent une résonance particulière. Un corpus collecté dans l'édition 1997 du Petit Larousse<sup>18</sup>, à la lettre C, révèle une hétérogénéité des modes d'enregistrement (mais surtout de constitution) du lemme lexicographique de la SP.

Bien entendu, le signifiant de certaines SP est établi conformément à la description de Rey-Debove :

- 17. faire le tour du cadran
- 18. se faire sauter le caisson
- 19. donner campo à qqn

Pour 17 et 18 la structure formelle est rigide et n'admet pas de reformulation, contrairement à 19 (*à qqn*). Toutefois cette rigidité ne va pas jusqu'à exclure la présence de zones intersticielles susceptibles d'accueillir, en discours, des éléments ici effacés :

- 20. faire (tout / rapidement) le tour du cadran
- 21. se faire (certainement / à tous les coups) sauter (instantanément) le caisson
- 22. donner (perfidement) campo à qqn

D'une manière générale, le lemme se présente comme un continu saturé sur chacune de ses positions (pas de variables sur les positions lexicales exprimées) et produit l'image d'une dénomination formellement homogène de l'entité linguistique concernée. À l'opposé, on assiste fréquemment à deux types de perturbations : un surplus d'une part, et de l'autre une incomplétude.

Le surplus de matériau lemmatique est lié à la présence de variables d'expression : une virgule interrompt le continu signifiant et entraîne une nouvelle formulation du constituant syntaxique gauche. Deux positions peuvent être affectées : le V tête (23 à 29) ou le SN subséquent (30 à 34) :

- 23. doubler, passer le cap
- 24. mettre, jouer cartes sur table
- 25. donner, laisser carte blanche à qqn
- 26. mettre, envoyer à la casse
- 27. revenir, retourner à la case départ
- 28. perdre, gagner au change
- 29. démarrer, prendre un virage sur les chapeaux de roues
- 30. cacher son jeu, ses cartes
- 31. caresser une idée, un rêve, etc.
- 32. faire cas, grand cas de
- 33. casser la tête, les oreilles

---

<sup>18</sup> Qui reste le dictionnaire général en un volume le plus diffusé en France et dans l'espace francophone

- 34. être cause de, la cause de

La reprise de V sous un autre item marque l'existence d'une relation de synonymie (23. *doubler / passer* ; 24. *mettre / jouer* ; 25. *donner / laisser* ; 26. *mettre / envoyer* ; 27. *revenir / retourner*), d'antonymie (28. *perdre / gagner*) ou d'altérité dans l'expression (29. *démarrer / prendre un virage*)<sup>19</sup>. Même remarque pour SN (30. *son jeu / ses cartes* ; 33. *la tête / les oreilles* ; 31. *une idée / un rêve*<sup>20</sup>) avec la différence toutefois que la reprise peut ne consigner qu'une variation de détermination, donc de contour sémantique (32. *cas / grand cas* ; 34. *cause de / la cause de*).

La reprise d'un constituant ouvre une brèche dans la linéarité postulée du lemme d'une SP. Ouvrant la porte à la variation, elle induit également que certains constituants entrant en paradigmes, certaines positions de la chaîne signifiante sont paradigmatiques. Ceci n'est pas sans incidence sur le mode d'interprétation de la SP<sup>21</sup> et la détermination de son degré de figement. Nous aborderons plus bas la structuration sémantique de ces paradigmes et son incidence sur la réalisation du lemme. Pour l'heure nous constaterons que ceux-ci sont saturés à deux unités. Il ne s'agit pas là nécessairement de l'expression d'une propriété linguistique des SP, mais d'une limitation imposée par la dimension textuelle de la représentation lemmatique. De la variation, oui, mais pas trop. L'énoncé 31 est à cet égard révélateur :

- 31. caresser une idée, un rêve, etc.

où l'abréviation *etc.* manifeste l'ouverture du paradigme et sa non-limitation à deux items, mais aussi, dans le même temps, que les contraintes de formulation imposent de ne pas dépasser un format prédéfini. De fait le surplus lemmatique, s'il est l'indice d'un débordement, se révèle, du fait de son calibrage, linguistiquement lacunaire.

Autre interrogation : en cas de représentation textuelle d'une variable, selon quel principe l'ordre d'apparition des composantes paradigmatiques est-il déterminé ? Si l'on en juge par les énoncés 23, 25, 31, 32, 34 l'ordre alphabétique prévaut. Toutefois il est concurrencé dans les autres exemples par la règle non-explicite de la saillance (l'item présentant plus forte probabilité d'apparition en discours est placé en tête du binôme). La représentation oscille donc entre exigences programmatiques et contraintes linguistiques.

Le lemme textuel peut présenter localement une incomplétude, une composante obligatoire étant omise dans la représentation. Il s'agit notamment du Sprep dont le SN peut être exprimé par un pronom (*quelqu'un, quelque chose*). Si certains sont effectivement réalisés :

- 35. se cacher de qqch
- 36. ne pas faire de cadeau à qqn

d'autres, sans justification linguistique possible, ne le sont pas :

- 37. casser les pieds
- 38. renvoyer aux calendes grecques

<sup>19</sup> Nous n'entrerons pas ici dans le détail de l'analyse sémantique des SP. Nous remarquerons cependant que la relation lexicale en jeu, synonymie ou antonymie, peut être fondée en langue (*perdre / gagner; revenir / retourner*). Dans d'autres cas elle peut être la spécification d'une hiérarchie-être (*passer > doubler; mettre > jouer; mettre > envoyer; laisser > donner*). Dans ce dernier cas le subordonné fournit son interprétation au superordonné.

<sup>20</sup> Plus sensiblement qu'avec les verbes, la cooccurrence de deux N à la même position d'une unique SP, sans modification d'interprétation ou de relation référentielle pour elle, induit entre ces items une relation de synonymie fonctionnelle lorsqu'une synonymie n'est pas déjà attestée dans la langue (*tête / oreilles*), et ce quel que soit le degré de proximité des référents impliqués (*idée / rêve*)

<sup>21</sup> Sur ce point, voir Petit (1998a)

- 39. mettre en capilotade
- 26. mettre, envoyer à la casse

On relève également des formulations alternes, mais déviantes elles aussi :

- 40. se cacher de (qqch)
- 41. carburer à
- 42. donner carrière (à)
- 43. jouer la carte de (suivi d'un nom)

Ces hésitations sont l'indice d'une spécificité sémiotique du lemme à cet endroit. Le fait qu'il s'agisse dans tous les cas de la limite textuelle droite de la SP et de sa non-saturation par un item lexical spécifique ou à défaut un paradigme d'items lexicaux doit être pris en considération. Elles révèlent la présence d'une zone de fragilité dans le signifiant. Ceci expliquerait l'instabilité observée dans sa représentation par les dictionnaires. Pour résumer, certains lemmes présentent un déficit local d'homogénéité sémiotique. Ils concentrent sur une partie seulement de la chaîne signifiante, celle réalisée par des items contraints, le potentiel d'identification de la SP. Ce point est à mettre en relation avec la structure sémantique de la SP, les zones d'arrière plan correspondant à une compositionnalité accrue de l'interprétation.

Les lacunes peuvent, plus curieusement, affecter la borne gauche, à savoir le verbe. Les séquences :

- 44. un cautère sur une jambe de bois
- 45. la cinquième roue du carrosse
- 46. sur le carreau

telles qu'elles sont libellées, sont contre intuitives dans la mesure où leur production effective en discours nécessite en contexte gauche<sup>22</sup> la réalisation d'un item dont le choix n'est pas libre :

- 47. (poser / mettre / c'est / ? appliquer / ? coller / \*disposer / \*trouver) un cautère sur une jambe de bois
- 48. (être / devenir / rester / ? demeurer / \*réprimander / \*renvoyer) la cinquième roue du carrosse
- 49. (être / rester / mettre / se (re)trouver / ?\* envoyer) sur le carreau

La difficulté est ici analogue à celle rencontrée précédemment. La limite gauche de la SP est relativement ouverte dans la mesure où elle n'est pas instanciée par un item particulier, mais par un paradigme. Ici encore, un déficit d'homogénéité sémiotique peut laisser induire que la position V n'est pas marquée dans le lexique, donc qu'elle est construite exclusivement par la syntaxe (*modulo* le respect des règles de sous-catégorisation) et n'a pas à être renseignée dans le lemme.

Une non prise en compte de la limite gauche peut également se rencontrer quand la position est lexicalement indexée sur un item. Nous n'emprunterons pas notre exemple à la lexicographie, mais... à la sémantique lexicale. Ainsi, Franckel et Lebaud (1992) traitant la SP *faire le lit de P* (où P est un terme prédicatif : *fascisme, intégrisme, terreur...*) excluent *faire* du lemme de la séquence et de sa description sémantique. Seule est tenue pour pertinente la suite *le lit de P*, bien que par ailleurs l'exemplification restitue le verbe. L'enjeu pour les auteurs est de démontrer que la polysémie de *lit* s'appuie sur un schéma organisateur dont *le lit de P* serait l'expression la plus saillante<sup>23</sup>. Ils en concluent que *lit*, dans *le lit de P* effectue une opération de spécification sur P :

<sup>22</sup> Le libellé du lemme lexicographique laisser induire par ailleurs que 44 et 45 peuvent occuper une position sujet

<sup>23</sup> De fait, ils ne conçoivent pas *le lit de P* comme une SP, mais comme une suite dont le statut reste plus ou moins indéterminé

“Faire le lit du fascisme, c’est faire en sorte que le fascisme se développe et prospère : *fascisme* devient vraiment *fascisme*. *Lit* se caractérise par le fait que s’y opère ce que l’on pourrait appeler un centrage.” (1992 : 98)

laquelle se décline dans des configurations différentes selon le contexte droit de *lit* et notamment le caractère prédicatif ou non du N2 (*le lit de la rivière, un lit de roses, de feuilles, d’oignons*) :

“L’hypothèse est que le caractère opératoire de *lit* va plus ou moins s’affaiblir ou se manifester sous des formes détournées.” (1992 : 100)

De fait, *lit* dans son emploi (+meuble) se voit relégué dans les valeurs les moins saillantes, les moins représentatives et les moins organisationnelles de la polysémie de l’item.

Franckel et Lebaud postulent que le sens repéré comme prototypique n’est activé que si effectivement un nom prédicatif complète *lit*. Il existe cependant une seconde condition, non abordée dans leur analyse : *lit* doit être lui-même complément de *faire*. Le déficit d’opérabilité se manifeste précisément dans les contextes où le N n’est plus l’argument obligatoire de ce verbe. Ceci implique que les propriétés considérées comme définitoires de *lit* et organisationnelles de sa polysémie sont pour partie héritées de sa cooccurrence avec un factitif<sup>24</sup>. En d’autres termes les auteurs semblent imputer à *lit* des propriétés induites par *faire*<sup>25</sup>. Le caractère opératoire de l’item ne résulte donc pas exclusivement de l’influence du contexte droit, mais de l’interaction des deux contextes. Enfin, au regard de la question qui nous intéresse, on aura constaté que la description sémantique de *lit*, telle qu’elle est menée par Franckel et Lebaud, porte sur un objet... qui n’existe pas. En effet comment peut-on considérer l’expression *le lit de P* comme emblématique des emplois de *lit*, sachant que cette séquence ne dispose d’aucune autonomie, voire d’aucune existence linguistique, contrairement à *faire le lit de P* qui est seule attestée et seule linguistiquement possible ?

D’une manière analogue, Gross (1996 : 28) considère les séquences (les parenthèses sont le lui) :

- 50. (tomber sur quelqu’un) à bras raccourcis
- 51. (conduire) à tombeau ouvert
- 52. (démarrer) au quart de tour

comme des “suites figées comprenant un nom mais qui ont une distribution adverbiale”. Or précisément à *bras raccourcis*, à *tombeau ouvert*, *au quart de tour* ne se rencontrent que dans des contextes verbaux contraints, respectivement *tomber* (éventuellement *taper*, *frapper*, *cogner*), *conduire* et *démarrer* (éventuellement *partir*). Le verbe fait donc partie intégrante de la SP. Il doit être mentionné dans le lemme<sup>26</sup>, ces SP étant alors non pas des adverbes mais des substituts de verbes dont l’interprétation est partiellement compositionnelle.

---

<sup>24</sup> Celle-ci ne ressortit pas à la phraséologie et n’est pas régie ni calculable par des règles de sous-catégorisation. En effet *lit* n’apparaît pas en emploi autonome, mais exclusivement en description définie. Par ailleurs, il ne négocie pas sa relation sémantique à *faire* sur le mode de la compositionnalité : si l’interprétation du verbe dans la séquence induit bien une valeur factitive, en revanche celle-ci ne peut se paraphraser par *faire*, mais plutôt par *préparer*. Le verbe et le nom renégocient leur valeur sémantique au sein de la SP, raison pour laquelle il est tout à fait licite de postuler que cette cooccurrence procède du figement

<sup>25</sup> La séquence (*faire*) *le lit de P* perd sa légitimité organisationnelle dans la mesure où l’occurrence du N est précisément contrainte par le figement. Si un emploi prototypique de la polysémie doit être recherché, il importe qu’il soit libre syntaxiquement et sémantiquement, c’est-à-dire qu’il puisse apparaître aussi bien en description définie qu’indéfinie et dans des contextes verbaux régis exclusivement par les règles de sous-catégorisation. Par conséquent le schéma explicatif proposé par Franckel et Lebaud s’inverse et perd son caractère contre-intuitif : l’emploi prototypiquement saillant et organisationnel est de ceux où *lit* n’est pas pris en étai entre un factitif et un terme prédicatif

<sup>26</sup> Pour leur lemmatisation, voir chapitre 2

En conclusion de cette première partie deux remarques s'imposent :

- a. l'adoption d'un modèle linéaire et textuel pour le lemme des SP présente des inconvénients majeurs. Elle :

- privilégie la cohésion rhétorique de la séquence sur sa mise en perspective structurelle
- crée l'illusion d'une linéarité de la SP et d'une homogénéité sémiotique (donc d'une homogénéité sémantique). De fait elle subit l'influence (plus ou moins implicite) d'un triple modèle théorique : saussurien (principe de linéarité du signifiant), lexématique (réduction du format du lemme aux lexèmes pivots de la SP) et lexicographique (l'interrogation sur le lemme des SP n'est jugée pertinente que dans le cadre de la confection de dictionnaires), inapproprié à la réalité de son objet
- ne gère pas rigoureusement les lignes de fuites du lemme (limite gauche et droite)

- b. il est nécessaire de s'interroger sur les conditions lexicales (et lexicologiques) d'élaboration du lemme des SP et ce afin de :

- réapproprié un objet linguistique dans le champ d'une discipline (la lexicologie) qui s'en est laissé déposséder méthodologiquement (à son insu, mais avec sa complaisance) parce qu'elle vit encore trop souvent sur l'illusion du signe-mot
- déterminer un modèle de forme qui rende compte des spécificités linguistiques de la SP (spatialité et hétérogénéité sémiotique) : en abandonnant le principe absolu de la linéarité ; en marquant la présence d'éventuelles zones de rupture ; en rendant compte de la variation lexicale sur les maillons de la chaîne concernés ; en spécifiant les places qui, tout en appartenant à la chaîne, opèrent une transition avec le syntagmatique et la phraséologie.

Tel sera l'objet de notre seconde partie.

## 2. Propositions pour l'établissement du lemme lexical des SP de type SV

Nous envisagerons les contraintes lexico-sémantiques qui régissent la distribution des composantes du lemme d'une SP ainsi que la disposition de celle-ci à l'intérieur de la chaîne signifiante. Ce n'est pas prioritairement comme indices d'un degré de figement que nous les penserons, mais dans leur capacité à moduler la linéarité du lemme, à la déconstruire et à régir la densité sémiotique de cette chaîne<sup>27</sup>.

### 2.1. Les limites fonctionnelles du signifiant

L'identité linguistique d'une SP est conditionnée par le format de la séquence lemmatique qui la dénomme. Celui-ci dispose d'une limite gauche, généralement identifiée sur le premier item lexical, en deçà de laquelle on considère que tout matériau réalisé dans un énoncé ou une phrase n'appartient pas à la SP, mais ressortit pleinement à la syntagmatique libre donc à la grammaire. Il dispose également d'une limite (droite dans la linéarité textuelle de la séquence), dont l'identification peut coïncider avec un point d'ancrage fort dans la syntagmatique, c'est-à-dire la compositionnalité (*ne pas faire de cadeau à qqn, casser les pieds de qqn...*).

#### - a. la sous-catégorisation du SN1

Presque toutes les SP peuvent s'interpréter, dans un contexte approprié, comme un syntagme libre. La sous-catégorisation du SN sujet est à cet égard déterminante. Sa mention dans le lemme peut être nécessaire ou souhaitable. Elle peut également ne pas être pertinente. Ce sont les trois cas que nous envisagerons ici.

---

<sup>27</sup> Il ne s'agit donc pas pour nous de proposer une version distributionnaliste supplémentaire

Dans :

- 53. Paul change souvent son fusil d'épaule
- 54. La sentinelle change souvent son fusil d'épaule
- 55. Paul traîne une casserole
- 56. Le chien traîne une casserole

l'interprétation conclura à la présence d'une SP en 53 et 55, mais (préférentiellement) pas en 54 et 56. La SP sera identifiée respectivement sous le lemme *changer son fusil d'épaule* et *traîner une casserole* puisque *changer* et *traîner* sont les premiers items décelables des séquences, et *épaule* / *casserole* les derniers. Le jugement lexico-syntaxique qui assigne des statuts différents à *changer son fusil d'épaule* / *traîner une casserole* en 53 et 55, et en 54 et 56 se fonde sur la sous-catégorisation du SN1. La séquence n'est une SP qu'à la condition que son extérieur gauche réponde à des contraintes sémantiques spécifiques. Traditionnellement la question de la sous-catégorisation est abordée en termes classématiques - le plus souvent par le recours aux seuls traits [ $\pm$ animé], [ $\pm$ hum] - ou bien, dans les descriptions plus fines du lexique-grammaires, en indexations positives par classes d'objets (Blanco et Bonell 1998 ; *Langages* n°131).

Ces critères suffisent au traitement de certaines SP parce qu'ils en permettent la désambiguïisation, comme dans les exemples cités plus haut. Ou bien ils édictent les conditions nécessaires à la grammaticalité des énoncés :

- 57. (SN[+hum] / \*SN[-hum]) faire un brin de causerie avec qqn
- 58. (SN[+hum] / \*SN[-hum]) se faire traîner dans la boue par qqn
- 59. (SN[+hum] / \*SN[-hum]) en avoir sa claquette de qqn / de qqch

Toutefois le recours à SN[+hum] se trouve en lui-même insuffisant. En effet, certains N comme *bébé*, *nourrisson*, a fortiori *foetus*, présentent un comportement particulier au regard de la sous-catégorisation N[+hum] puisqu'ils rejettent des contextes appropriés à l'adulte pour des raisons physiologico-comportementales ou culturelles (*faire un brin de causerie*, *changer son fusil d'épaule*, *manger les pissenlits par la racine*, *avoir sa claquette de*, etc.).

Pareillement la SP *causer chiffons* exige une description spécifiant les traits [+femme] et [+adulte] :

- 60. Marie et Chloé causent chiffons (Marie et Chloé sont des adultes)
- 61. ?? Marie et Chloé / ?? les enfants causent chiffons (Marie et Chloé sont des enfants)<sup>28</sup>
- 62. ? Marie et Paul causent chiffons
- 63. ?? Jean et Paul causent chiffons

Certaines SP qui demandent un SN [-hum], exigent une sous-spécification sinon domaniale, du moins catégorielle. Ainsi *être un cataplasme* réclame un SN1 marqué [+mets] et préférentiellement [+massif] :

- 64. cette purée / ce riz est un cataplasme

bien que des SN [+comptables] soient admissibles :

- 65. cette salade est un cataplasme
- 66. ces spaghettis / haricots sont un cataplasme

---

<sup>28</sup> La question reste entière si les protagonistes sont adolescentes, *a fortiori* si l'une est adulte et l'autre adolescente

car leur référent, envisagé comme mets préparé (et non pas collection d'ingrédients avant préparation), ou bol alimentaire se voit par là-même massifié. Il se présente comme une totalité dont la structuration est relativement homogène, homéomère, ses limites n'étant pas intrinsèques mais fournies par le contenant (Kleiber 1997, 1999, Tyvaert 1997 a et b). De fait, les propriétés (forme, texture) du tout (le mets, le bol alimentaire) occultent dans l'arrière-plan celles des parties individuelles<sup>29</sup>. En revanche, lorsque le SN1 est nettement identifié comme [+comptable] la locution s'applique mal :

- 67. ?? ces pommes de terre / tranches napolitaines / saucisses sont un cataplasme

A la différence des précédentes, une SP comme *avoir le casque*, ou *changer son fusil d'épaule* impliquent une sous-catégorisation négative du SN1. Dans le premier cas celui-ci doit plutôt exclure, faute d'ambiguïté :

- 68. ? le motard / pompier a le casque

tout N[+hum] référant à une profession ou une activité en rapport avec le port du casque (*motard, pilote* (automobile, d'avion), *astronaute, militaire, gendarme, ouvrier de chantier, pompier, fabricant* ou *distributeur* de casques...) et dans le second ceux référant à des personnes susceptibles de manier un fusil (*militaire, chasseur...*) :

- 69. ? le chasseur a changé son fusil d'épaule
- 70. en matière de communication, l'entreprise a changé son fusil d'épaule<sup>30</sup>

Cette exclusion ne doit pas être pensée en termes de grammaticalité. Tout SN marqué [+hum] est par principe admis, moyennant les réserves faites plus haut sur des N comme *bébé, nourrisson*. Toutefois l'interprétation de la séquence comme SP se voit assurée sans ambiguïté ni parasitage si le N tête du SN1 n'appartient pas à un paradigme spécifique. Pour cette raison il est souhaitable que le faisceau de traits sémantiques qui circonscrit l'exclusion soit spécifié dans l'intitulé du lemme.

Enfin, certaines SP restent indifférentes à la sous-catégorisation du SN1. La séquence possède son identité intrinsèque et ne dispose d'aucune instanciation possible comme syntagme libre. Ainsi *tenir le coup* admet indifféremment un SN1 marqué [+hum] qu'un SN1 [-hum] :

- 71. Paul / la courroie tient le coup

un SN concret ou abstrait<sup>31</sup>

- 72. la courroie / son moral tient le coup

un SN comptable ou massif :

- 73. la courroie tient le coup
- 74. la glace tient le coup, on peut marcher dessus

## **- b. Les positions non-agentives comme lignes de fuite phraséologiques**

Dans les SP :

---

<sup>29</sup> Ils satisfont notamment aux tests de la division homogène et de la référence cumulative, comme les massifs. La même remarque vaut pour *frite* et *moule*, dont la discrétisation référentielle est davantage accentuée que celle de *haricot* et *spaghetti*

<sup>30</sup> La question reste toutefois entière avec les Npropres : *Paul a le casque / a changé son fusil d'épaule* reste intrinsèquement ambigu par faute de sous-catégorisation explicite. Celle-ci doit être fournie par le cotexte ou le contexte

<sup>31</sup> Bien que statistiquement les SN [+concret] aient une plus grande latitude d'apparition

- 38. renvoyer qqch aux calendes grecques
- 39. mettre qqch, qqn en capilotade
- 26. mettre, envoyer qqch à la casse

la réalisation du SN2 (*qqch, qqn*) est obligatoire :

- 75. \*il a renvoyé aux calendes grecques
- 76. \*il a mis en capilotade
- 77. \*il a mis / envoyé à la casse

pareillement pour les Sprep :

- 78. opposer une résistance à qqn, qqch / \*opposer une résistance
- 79. se cacher de qqch / \*se cacher
- 80. casser les pieds à qqn / \*casser les pieds

Sur le plan sémiotique ces SPREP appartiennent à la locution. Ils sont irrégulièrement pris en compte dans les lemmatisations lexicographique et linguistique (cf. plus haut) parce qu'elles ressortissent à la syntagmatique libre. Ceci explique le choix fait par certains dictionnaires de ne pas systématiquement les enregistrer. Ils préfèrent prendre en compte le matériau ne renvoyant pas à des classes ouvertes.

Les places saturées par des items lexicaux contraints constituent la colonne vertébrale de la SP, lui assurent sa visibilité et garantissent par voie d'économie sa mémorisation. Toutefois elles ne sauraient à elles seules saturer l'identité de la SP. Les SN ou Sprep précités constituent l'envers de la SP. À la fois dans et hors de la locution, ils constituent le fondu qui rattache, en l'amenuisant, la SP à l'énonciation et à la grammaire. Ils manifestent la part d'elle-même qui échappe au (poly)lexématique, en tant que celui-ci n'est pensé que comme contraint. Ils consignent une des limites du figement tout en marquant que si une SP tire son origine de combinaisons syntagmatiques et compositionnelles, celle qui présente de tels constituants y retourne. Enregistrer ces SN et Sprep est indispensable en tant qu'ils sont pleinement constitutifs de l'identité de la locution.

Certains constituants peuvent ne pas être réalisés sous une forme unique dans toutes les occurrences de la SP. C'est ce que l'on observe avec :

- 81. faire un brin de causette avec qqn
- 82. casser les pieds à qqn

Deux cas de figure sont à distinguer :

- a. le SPREP est effaçable bien qu'obligatoire. La SP n'admet le lemme réduit *faire un brin de causette* que si le SN1 est collectif ou pluriel :

- 83. Nicolas a fait un brin de causette avec Marie
- 84. Nicolas et Marie / les enfants ont fait un brin de causette
- 85. \*Nicolas a fait un brin de causette

La suppression du Sprep n'en est donc pas une, mais un transfert d'informations sur le SN1<sup>32</sup>. De fait le lemme lexical de la SP doit marquer cette propriété :

---

<sup>32</sup> La suppression ne peut jamais être envisagée totalement. On admettrait :

- Nicolas a fait un brin de causette  
à condition de supposer la structure elliptique et de postuler que le complément est impliqué dans l'interlocution



<b>SN1</b>	<i>faire un brin de causette</i>	<b>SPREP</b>
SN1[+hum] [+sing]		<i>avec SN2[+hum]</i>
SN1[+hum] [+plur / cumul]		∅

**SN1** identifie la position sujet, en tant qu'elle comporte plus d'une réalisation possible - [+sing] vs [+plur / cumul] (nous considérerons comme cumulatif un SN qui exprime l'agent et le patient) ;

- **SPREP** indique que le syntagme subséquent ne peut être réalisé que par un SPREP comprenant la préposition *avec*, et à la condition que le SN1 soit marqué [+sing] ; si le SN1 est marqué [+plur / cumul], il y a blocage de la réalisation. En tout état de cause la mention d'un co-agent du procès est nécessaire à l'expression de la locution. Elle appartient à sa structure sémantique et conditionne son identité lexicale ;

- la séparation horizontale (trait fin) sur les chaînons 1 et 3 indique une rection ligne à ligne : l'apparition du Sprep est conditionnée par la sous-catégorisation de SN1 à la première ligne, la non apparition par celle de la seconde ligne.

- b. le constituant n'est pas effaçable. *Casser les pieds à qqn* n'admet pas l'effacement du Sprep, mais sa cliticisation :

- 86. Nicolas casse les pieds à Marie
- 87. Nicolas lui casse les pieds
- 88. ??Nicolas casse les pieds<sup>33</sup>

Le constituant appartient ici encore à la structure sémantique de la SP. Il se caractérise par l'existence d'une variable d'instanciation qui se double d'une alternative dans le positionnement. Le lemme prendra la forme :

<b>SN1</b>	<b>Pron SPREP</b>	<i>casser les pieds</i>	<b>SPREP</b>
SN1 [±hum]	∅		<i>à SN2[+hum]</i>
	Clit.Sprep		∅

où *Pron SPREP* marque le chaînon imparti à la cliticisation de SPREP. Le Sprep et le clitique sont en distribution complémentaire. L'instanciation d'une variable provoque, à distance, la réalisation d'une autre qui lui est corrélée : la cliticisation du Sprep bloque sa réalisation sous la forme à SN. La distribution des ruptures dans la continuité signifiante obéit à la grammaire de la syntagmatique courante.

La pronominalisation des SN et Sprep obligatoires et identifiables par une simple étiquette syntaxique (SN [+hum]...) est de nature à perturber la continuité du lemme, et ce de deux manières. Premièrement en ouvrant un interstice dans le continu signifiant :

- 38. renvoyer qqch aux calendes grecques
- 89. Paul a renvoyé l'examen du dossier aux calendes grecques
- 90. l'examen du dossier, Paul l'a renvoyé aux calendes grecques

peuvent se lemmatiser comme suit :

<b>SN1</b>	<b>Pron SN2</b>	<i>renvoyer</i>	<b>SN2</b>	<i>aux calendes grecques</i>
SN1 [+hum]	∅		SN2 [-hum]	
	Clit.SN2		∅	

<sup>33</sup> Admissible éventuellement avec un aspect duratif ou une valeur itérative, proche de l'état

Contrairement à *casser les pieds à qqn*, si le chaînon Pron n'établit pas de rupture repérable dans le continu signifiant du fait qu'il se positionne avant le premier item lexical contraint, en revanche SN2 impose une rupture entre *renvoyer* et *aux calendes grecques*.

Pareillement :

- 91. ne pas arriver à la cheville de qqn
- 36. ne pas faire de cadeau à qqn

peuvent donner lieu aux réalisations suivantes :

- 92. Paul n'arrive pas à la cheville de Jean
- 93. Paul ne lui arrive pas à la cheville
- 94. Nicolas ne fait pas de cadeau à Marie<sup>34</sup>
- 95. Nicolas ne lui fait pas de cadeau

La cliticisation ouvre une position entre deux composantes spécifiées de la SP (ici l'adverbe de négation *ne pas* et le verbe), c'est à dire à l'intérieur de la structure lexicalement contrainte du lemme :

SN1 SN1 [±hum]	<i>ne...pas</i>	(MODIF) <i>du tout / vraiment..</i>	Pron SPREP Ø	<i>arriver</i>	(MODIF) Ø	<i>à la cheville</i>	SPREP <i>de SN2 [±hum]</i>
		Ø	Clit SPrep		<i>du tout / vraiment...</i>		Ø

SN1 SN1 [+hum]	<i>ne...pas</i>	(MODIF) <i>du tout / vraiment..</i>	Pron Ø	<i>faire</i>	(MODIF) Ø	<i>de cadeau</i>	(MODIF ) Ø	SPREP Ø / à SN2 [+hum]
SN1 [+hum]		Ø	Clit SPrep		Ø		<i>du tout</i>	Ø
SN1 [+cumul]		Ø	<i>se</i> (réciproque)		<i>du tout / vraiment..</i>		Ø	Ø

Le Sprep peut être cliticisé (*Nicolas ne fait pas de cadeau à Marie* > *Nicolas ne lui fait pas de cadeau*), mais aussi le SN1 s'il est cumulatif, c'est-à-dire exprime le sujet et le patient. Dans ce cas la pronominalisation est exprimée par un réciproque (*Nicolas et Marie ne se font pas de cadeau*). En toutes circonstances, le Sprep n'est réalisé que si le SN1 n'est pas cumulatif (*Nicolas et Marie ne font pas de cadeau à Paul*), bien qu'il puisse être omis (*Nicolas ne fait pas de cadeau ; Nicolas et Marie ne font pas de cadeau*).

Second type de perturbation dans la continuité du lemme, la création de positions vides. L'apparition de ruptures n'est pas ici le lieu d'instanciation de variables pleines (*i.e.* de formes lexicales ou grammaticales concurrentes), mais d'une seule variable pleine (une forme linguistique représentée par un pronom ou par un syntagme) et d'une non-réalisation (Ø). La mise en discours opère en surface une saturation de la position vide et rétablit la linéarité du lemme, aucune marque n'étant observable. Il n'en reste pas moins qu'une position Ø est réalisée, donc qu'une rupture du lemme est accomplie.

Une position supplémentaire est construite par un modifieur, MODIF. Elle est facultative, d'où les parenthèses dans notre représentation. Ce chaînon est susceptible d'occuper plusieurs positions

<sup>34</sup> Par commodité, nous ne traiterons pas dans la présente étude l'alternance du singulier et du pluriel sur *cadeau* : *ne pas faire de cadeau / ne pas faire de cadeaux*

alternatives dans une même séquence polylexématique, en distribution complémentaire les unes par rapport aux autres. Une SP peut donc comporter plusieurs systèmes de rection indépendants<sup>35</sup>.

En résumé de ces observations, le format lemmatique de la SP présente une limite gauche invisible (puisque constituée d'un faisceau d'instructions sémantiques descriptif, éventuellement par défaut) et une limite droite qui peut être fuyante, parce qu'elle est susceptible de présenter une mobilité (cliticisation)<sup>36</sup>.

## 2.2. La structuration interne : paradigmatisme et hétérogénéité

Les jeux de variables entrevus aux paragraphes précédents concernent l'entour du lemme de la SP. Traditionnellement, il n'est pas admis, au nom du principe de figement, qu'ils affectent aussi les composants lexicalement contraints, ceux qui n'intègrent pas un paradigme ouvert et dont l'expression ne requiert pas un symbole postiche (SN, Sprep). Une telle conception est réductrice et ne rend pas compte de la réalité linguistique.

### - a. l'instabilité des composants non-lexématiques

La paradigmatisme du matériau lexicalement contraint peut affecter soit des lexèmes, soit des items grammaticaux. Nous commencerons par ces derniers en observant d'abord la variation de la négation, puis celle des prédéterminants.

La SP *ne pas faire de cadeau à qqn* ne s'exprime qu'à la forme négative mais est susceptible d'admettre plusieurs variables :

- 96. *ne (pas / plus / jamais) faire de cadeau à qqn*
- 97. *ne (pas / plus) du tout faire de cadeau à qqn*
- 98. *ne (pas / plus) faire du tout de cadeau à qqn*
- 99. *ne (pas / plus) faire de cadeaux du tout à qqn*
- 100. *ne faire aucun cadeau à qqn*
- 101. *ne plus faire aucun cadeau à qqn*<sup>37</sup>

le lemme s'établissant comme suit :

SN1	NE	PAS	(MODIF) <sup>38</sup>	Pron	<i>faire</i>	PAS'	DE	<i>cadeau</i>	SPREP
SN1 [+hum]	<i>ne,</i> Ø	<i>pas / plus</i>	<i>du tout / vraiment...</i>	Ø		Ø	<i>de</i>		Ø / à SN2 [+hum]
SN1 [+hum]		<i>guère / jamais / plus</i>	Ø	Clit SPrep		Ø			Ø
SN1 [+cumul]		Ø	Ø	<i>se</i> (réciproque)		<i>aucun</i>	Ø		Ø

<sup>35</sup> Dans le cas de *ne pas faire de cadeau à qqn*, si l'on croise les paramètres de la cliticisation et ceux de l'introduction d'un modifieur : *ne pas faire de cadeau à qqn*, *ne pas du tout faire de cadeau à qqn*, *ne pas faire du tout de cadeau à qqn*, *ne pas faire de cadeau du tout à qqn*, *ne pas lui faire de cadeau*, *ne pas du tout lui faire de cadeau*, *ne pas lui faire du tout de cadeau*, *ne pas lui faire de cadeau du tout*, etc. Pour cette raison nous avons adopté une signalétique différente, réservant à un système la séparation horizontale avec trait plein, et adoptant pour l'autre une séparation par pointillés

<sup>36</sup> Également parce qu'elle marque un ancrage fort dans la syntagmatique compositionnelle

<sup>37</sup> Respectivement : *Paul ne fait pas de cadeau à Jean* ; *Paul ne fait (jamais / plus) de cadeau à Jean* ; *Paul ne veut plus du tout faire de cadeau à Jean* ; *Paul ne fait plus du tout de cadeau à Jean* ; *Paul ne veut pas faire du tout de cadeau à Jean* ; *Paul ne fait plus de cadeau du tout à Jean*

<sup>38</sup> Nous n'avons mentionné qu'une fois le modifieur MODIF. Toutefois ce chaînon peut apparaître, moyennant aménagement, derrière *faire* et *cadeau* : *ne pas faire du tout de cadeau*, *ne pas faire de cadeau du tout*, *ne faire aucun cadeau*

où PAS et son corollaire déplacé PAS' réalisent le paradigme des variables du forclusif. Ce paradigme est clos (on peut considérer que *point* est exclu) mais discontinu, c'est-à-dire réparti sur plus d'une place. Celles-ci sont en distribution complémentaire puisque le positionnement de *pas / guère / plus / jamais* est exclu pour *aucun*, et réciproquement<sup>39</sup>.

Sur le plan sémantique ceci implique que la démarche de lemmatisation ne doit pas, contrairement à une habitude ancrée dans la pratique et héritée du présupposé de neutralisation qui lui est imparti, se fixer comme objectif unique de réaliser entre plusieurs variables celle qui sémantiquement est la plus neutre, c'est-à-dire la moins marquée appréciativement et / ou qui représente la donnée commune à l'ensemble des possibles. En revanche, si cet item neutre existe, il peut fournir l'étiquette générique du paradigme dans la mesure où il en représentera le plus petit dénominateur.

Autre particularité de cette SP, le croisement de trois jeux de variables régies :

- celui de la pronominalisation ;
- celui construit par l'expression du forclusif (PAS / PAS')
- celui marqué par la dépendance du modifieur - MODIF - relativement au forclusif.

Ces trois systèmes de variables sont hétérogènes et partiellement indépendants (le troisième n'est pas autonome, contrairement aux deux autres). Les correspondances de niveau (ligne à ligne) s'opèrent par pontage entre chaînons à l'intérieur du même système.

D'une manière similaire *traîner une casserole* et *être dans le caca* admettent des variations sur le prédéterminant :

- 102. traîner une / des / plusieurs / quelques / une certaine / certaines casserole(s)
- 103. traîner une belle / fichue casserole
- 104. traîner la casserole de sa vie
- 105. être dans le caca
- 106. être dans un caca pas possible

SN1 SN[+hum]	<i>traîner</i>	DET	(MODIF) <sup>40</sup>	<i>casserole</i>	(MODIF)
		Det indef	<i>fichue / belle...</i>		Ø
			Ø		Adj / Sprep <sup>41</sup>
		Det def	Ø		<i>de</i> SN2

SN1 SN[±animé]	<i>être</i>	(MODIF) <i>vraiment...</i>	<i>dans</i>	DET	(MODIF)	<i>caca</i>	(MODIF)
				<i>le</i>	Ø		Ø
				<i>un</i>	<i>beau / fichu...</i>		Ø
					Ø		Adj / Sprep

Avec la première SP, et contrairement à ce qui se passe pour la seconde, les possibilités de réalisation sont circonscrites non pas sur des items spécifiques, mais sur l'aspect signifié par le déterminant (défini vs indéfini). C'est à ce niveau qu'est acquis le degré de pertinence suffisant. Si en revanche, comme c'est le cas pour *être dans le caca*, le paradigme se clôt sur des unités spécifiques (*le, un*), qui

<sup>39</sup> Pour la clarté de l'exposition, les variables liées à la pronominalisation de SN1 et SPREP ne sont pas représentées

<sup>40</sup> L'instanciation d'un item dans MODIF est régie par le caractère défini ou non de DET

<sup>41</sup> Ici, comme dans les SP qui suivent, Adj et Sprep dans MODIF renvoient à des axiologiques : *Traîner une casserole épouvantable, une casserole pas possible...*

pour autant ne constituent pas une sous-classe générique de l'étiquette, le lemme les renseignera en extension<sup>42</sup>.

La variation des items grammaticaux est la manifestation d'un phénomène de portée étendue. Entre DET et SPREP, DET et MODIF ou bien entre PAS et PAS', ou PAS' et DE (dans *ne pas faire de cadeau à qqn*), existe une interdépendance homologue de celle repérable entre un clitique et le syntagme qu'il reprend. Homologue seulement car le type de relation en jeu diffère ainsi que son orientation. Sur dernier point la dépendance n'est pas bilatérale mais vectorialisée : SPREP, MODIF et DE sont régis respectivement par DET et PAS. L'expression du lemme doit prendre en compte ces phénomènes notamment s'ils concernent sa limite droite. En effet contrairement à ce que l'on observe avec la cliticisation, la variation sur un chaînon interne - DET - est susceptible de provoquer l'émergence d'un autre chaînon, en position de limite - p.ex. MODIF . En d'autres termes la paradigmatisation de composantes grammaticales peut être aussi une procédure d'amplification du format même de la chaîne. Telle propriété ne se rencontre pas avec les composants lexématiques.

#### - b. et des composants lexématiques

Les composants lexématiques peuvent être également sujets à paradigmatisation. Ainsi les SP :

- 105. être dans le caca
- 107. pédaler dans la choucroute
- 108. avoir la dalle
- 109. recevoir un coup sur le cigare
- 110. être gros comme un cochon

manifestent des variations sur les N, V et Adj :

- 111. (être / rester / se trouver / retomber / se retrouver) dans (le caca / la merde / la mouise)
- 112. pédaler dans (la choucroute / la semoule / le yaourt)
- 113. (avoir / crever) la dalle
- 114. (recevoir / prendre / donner / fiche) un coup sur (le cigare / le citron / la citrouille...)
- 115. être (gros / gras / sale / saoul) comme un cochon

Avec (*être / rester / se trouver*) dans (*le caca / la merde / la mouise*) la configuration laisse apparaître une homogénéité sémantique relative sur le V et le N :

- *être*, *rester* et *se trouver* expriment un état analogue moyennant une perspective sur le référent et un présupposé différents. Ils réalisent la même notion<sup>43</sup>, que l'on peut étiqueter par ÊTRE du fait que ce verbe exprime la donnée sémantique commune aux trois items, mais aussi parce que sa réalisation est statistiquement saillante. *Retomber* et *se retrouver* marquent quant à eux une nouvelle occurrence d'un état qui a déjà été réalisé. On peut les ranger sous l'étiquette ÊTRE DE NOUVEAU. Ils s'articulent aux précédents par leur spécificité aspectuelle (réitération). La notion qu'ils instancient n'est qu'un dérivé d'ÊTRE avec qui elle entretient une relation d'hyponymique (*retomber* et *se retrouver* isolent des sous-types d'ÊTRE).

- *caca*, *merde* et *mouise* réalisent une double variation : de registre et d'intensité. *Merde* est très familier, *caca* est familier et *mouise* constitue une hypercorrection en registre standard-soutenu. Les autres items et structures syntaxiques de la SP appartiennent au registre standard ; c'est donc la réalisation d'un des trois N qui confèrera à celle-ci son positionnement locutoire. Le changement de registre s'accompagne également d'une évolution dans la caractérisation du référent. *Être dans la merde* opère une conceptualisation plus crue qu'*être dans le caca / la mouise*, et tend à fonctionner comme la caractérisation d'un état du réel situé à un plus haut degré d'intensité :

---

<sup>42</sup> La même remarque vaut pour les variables d'instanciation de PAS

<sup>43</sup> Par *notion* nous entendons un ensemble de propriétés sémantiques susceptible de caractériser une catégorie référentielle, un type. Cette catégorie peut être un objet, au sens large du terme, un procès ou une propriété. Voir à ce sujet Petit (1999c)

- 116. Paul est dans la mouise / le caca. Il est même franchement dans la merde
- 117. \*Paul est dans la merde. Il est même franchement dans la mouise / le caca

Pour ce qui concerne l'étiquetage, même si *merde* fournit en fait un terme de base, les autres, n'étant que des traductions intralinguales à caractère euphémique, et qu'il est également statistiquement saillant, *caca* et *mouise* confèrent à l'expression un caractère affecté et dérivé, la notation s'effectuera en référence à la notion évoquée : **DIFF** pour 'difficulté très grande', 'situation inextricable'. La notion est structurée sur le mode cohyponymique, du fait de la coréférentialité relative entre les trois items (Petit 1998b et 1999c).

D'après ces remarques, son lemme s'établira comme suit<sup>44</sup> :

<b>SN1</b> SN [±animé]	<b>ÊTRE</b> <i>être / rester / se trouver...</i>	<i>dans la / le</i>	<b>DIFF</b> <i>merde / mouise / caca</i>
	<b>ÊTRE DE NOUVEAU</b> <i>retomber / se retrouver...</i>		

Si l'on y ajoute les variations enregistrées précédemment, on obtient :

<b>SN1</b> SN[±animé]	<b>ÊTRE</b> <i>être / rester / se trouver...</i>	<b>(MODIF)</b> <i>vraiment...</i>	<i>dans</i>	<b>DET</b> <i>la / le</i>	<b>(MODIF)</b> ) Ø	<b>DIFF</b> <i>merde / mouise / caca</i>	<b>(MODIF)</b> Ø
	<b>ÊTRE DE NOUVEAU</b> <i>retomber / se retrouver...</i>			<i>un / une</i>	<i>beau / fichu...</i>		Ø
					Ø		Adj / Sprep...

L'alternance *avoir / crever la dalle* manifeste une variation d'intensité marquée ici non pas par la conceptualisation locutoire (standard vs familier), mais inscrite dans la signification lexicale des unités. Le verbe *crever*, lorsqu'il est construit avec un nom de sentiment ou de sensation, peut être paraphrasé par 'éprouver à un haut degré' (où *éprouver* est synonyme d'*avoir*) :

- 119. crever de faim, de peur, d'envie, de douleur

Un problème important subsiste néanmoins : *crever de dalle* n'est pas attesté, contrairement à *crever la dalle*. Toutefois *crever la faim*, qui entretient avec le précédent une relation d'intensité moindre<sup>45</sup>,

<sup>44</sup> Nous proposons également pour *tomber à bras raccourcis*, *conduire à tombeau ouvert* et *démarrer au quart de tour* (cf plus haut) :

SN [+hum]	<i>tomber</i> <b>FRAPPER</b> <i>frapper, taper, battre, cogner</i>	<i>à bras raccourcis</i>	<i>sur SN</i> SN
SN [+hum]	<b>CONDUIRE</b> <i>conduire, rouler</i>	<i>à tombeau ouvert</i>	
SN [±hum]	<i>rouler</i>		
SN [+hum]	<b>PARTIR</b> <i>démarrer, partir</i>	<b>(MODIF)</b> <i>vraiment...</i>	<i>au quart de tour</i>

est admis. *Crever la dalle* est donc syntaxiquement construit sur le même modèle que *crever la faim*, mais également sur celui d'*avoir*, dans un de ses emplois les moins représentatifs (*avoir la haine / la gagne*, bien qu'attestés ne respectent pas le régime courant du verbe dans ses emplois de support : *avoir faim / peur / raison / tort...*) *avoir la dalle* étant accepté et entretenant lui aussi avec *crever la dalle* une relation d'intensité<sup>46</sup>. Le lemme de la SP aura donc la forme de base :

<b>SN1</b>	<b>AVOIR</b>	<i>la dalle</i>
SN [+animé]	<i>avoir / crever</i>	

et si l'on prend en compte la variation du déterminant sur le modèle d'*être dans le caca* :

<b>SN1</b>	<b>AVOIR</b>	<b>DET</b>	<b>(MODIF)</b>	<i>dalle</i>	<b>(MODIF)</b>
SN [+animé]	<i>avoir / crever</i>	<i>la</i>	Ø		Ø
	<i>avoir</i>	<i>une</i>	Adj		Ø
			Ø		Adj / Sprep

Contrairement à ce que laisse penser sa lemmatisation lexicographique, *recevoir un coup sur le cigare* (n')admet (que) des V réciproques, paraphrasables soit par DONNER, soit par RECEVOIR : (*recevoir / prendre / donner / fiche / mettre*) *un coup sur le cigare*. Par ailleurs l'expression du Sprep peut être saturée par un paradigme, dont *cigare* n'est pas le représentant prototypique<sup>47</sup> : *cafetière, carafe, citron, citrouille, coloquinte, ciboulot, carafon, caillou* sont également admissibles dans ce contexte. Ces N sont les reformulants familiers de *tête* (dans son acception 'crâne'). Le paradigme qu'ils constituent est sémantiquement homogène (contrairement à celui de *semoule / yaourt / choucroute*, cf. plus bas). Ils ne présentent toutefois pas la latitude d'emploi de leur unité source dans sa valeur considérée puisque :

- ils ne disposent pas d'un emploi libre dans la valeur considérée ;
- les réciproques DONNER / RECEVOIR fournissent leur contexte d'emploi privilégié ;
- seul un petit nombre d'entre eux entrent par ailleurs dans une autre SP<sup>48</sup>.

Reste la question de l'étiquetage. Il est vrai que *cigare, citron* et *cafetière* sont les unités dont la probabilité d'apparition est la plus élevée. Toutefois ils ne constituent pas en eux-mêmes une notion, dans la mesure où le référent qu'ils indexent est déjà pourvu d'une dénomination objectale<sup>49</sup> : *tête*. Pour cette raison ce N fournira l'étiquette du paradigme (il en fournit déjà le dénominateur commun sémantique).

<sup>45</sup> À cela s'ajoute une différence aspectuelle, *crever la faim* étant susceptible de recouvrir une valeur durative, contrairement à *crever la dalle*

<sup>46</sup> Si l'on admet que *crever de / la N* signifie 'éprouver à un haut degré / avoir très x' et que *dalle* signifie 'forte faim', *avoir la dalle* s'interpréterait par 'avoir une forte faim' et *crever la dalle* 'éprouver à un haut degré une forte faim', saturant ainsi vers le haut degré l'ensemble des places lexicalement disponibles. *Crever la faim* s'interprétera quant à elle par 'éprouver à un haut degré la faim'

<sup>47</sup> Sur ce point, voir Petit (1998a)

<sup>48</sup> Notamment *ciboulot, citron* et *caillou* forment également une SP avec *avoir* : *ne rien avoir dans le N. Ciboulot* se rencontre également dans la SP *travailler du ciboulot*.

<sup>49</sup> Sur ce point, voir Petit (1999b)

Le lemme prendra donc la configuration suivante :

SN1 SN [+animé]	<b>DONNER</b> <i>donner / mettre / fiche / foutre / envoyer / balancer</i>	(MODIF) <i>vraiment...</i>	<i>un</i>	(MODIF) Ø	<i>coup sur la / le</i>	<b>TÊTE</b> <i>cafetière / carafe / citron / citrouille / coloquinte / ciboulot / cigare / carafon / caillou</i>	<b>SPREP</b> <sup>50</sup> à SN
	<b>RECEVOIR</b> <i>recevoir / prendre / se manger</i>	Ø		<i>bon, vrai, sale...</i>			Ø

Les SP du type *être (gros / gras / sale / saoul) comme un cochon* appartiennent à un formulaire très productif en français : *être X comme un Y*, dont les propriétés sémantiques ont été décrites en détail par (Tamba 1979) et (Mejri 1994). Nous rappellerons seulement que l’instanciation du SN en position Y est déterminée par le choix de l’adjectif en X. Or précisément, dans le cas qui nous intéresse, celui-ci peut ou non appartenir à un groupement notionnel. En effet, *gros* et *gras* peuvent, sur le modèle que nous avons adopté précédemment, être regroupés sous l’étiquette GROS, *gras* ne présentant qu’une variante intensive assortie, il est vrai, de spécificités sémantiques. Il n’existe aucune autre variante, mais surtout aucune intersection avec *sale*, ni avec *saoul*. Ce dernier admet certaines variantes intensives familières (*saoul, bourré, pété, \*rond, \*ivre, \*plein, \*éméché*). En revanche *sale* reste seul à pointer sa propriété référentielle, quel que soit le registre dans lequel on se situe :

- 120. (*sale / \*pourri / \*dégoûtant / \*dégueulasse / \*crasseux / \*malpropre / \*crotté / \*graisseux...*)  
comme un cochon

De fait, la position X est constituée de manière asymétrique, saturée soit par une notion, soit par un item isolé<sup>51</sup> :

SN1 SN[±animé]	<b>ÊTRE</b> <i>être / rester / devenir...</i>	<i>sale</i>	<i>comme un cochon</i>
	<b>ÊTRE DE NOUVEAU</b> <i>redevenir / se retrouver...</i>	<b>GROS</b> <i>gros / gras</i>	
		<b>SAOUL</b> <i>saoul / bourré / pété...</i>	

La configuration de *pédaler dans la choucroute / le yaourt / la semoule* s’organise de manière particulière. Dans les cas précédents, la diversité lexicale recouvrait une identité référentielle au point qu’un regroupement était possible sous des notions et des étiquettes uniques. Si *choucroute, yaourt* et *semoule* désignent bien des mets et découpent leur référent sur le mode du massif, la convergence s’arrête ici et reste un phénomène de surface. Tous les termes marqués [+mets] [+massif] et associés à

<sup>50</sup> Pour ne pas alourdir la présentation, nous n’exposons pas la pronominalisation de ce complément

<sup>51</sup> Une des difficultés de ces SP, que les limites de cet article ne nous permettent pas de traiter est leur organisation en réseau qui implique une spatialisation de la structure lemmatique. Ainsi, *être saoul comme un cochon* commute avec *être saoul comme un polonais, être (rond, pété) comme une queue de pelle, être pété comme un coing, être plein comme une outre*. D’où la nécessité de repérer à chaque fois la SP par rapport à une position donnée. Ici nous avons privilégié X, alors que dans le texte ci-dessus, Y fournissait le point de référence



un référent un tant soit peu consistant ou solide ne sont pas susceptibles d'apparaître derrière *pédaler* :

- 118. \*pédaler dans la purée / le riz / le hacHis / les haricots / le boulgour / la crème / la Chantilly...

Par ailleurs les trois N retenus dans la SP présentent une diversité référentielle importante, appartiennent à des sous-genres de mets trop distincts pour être regroupés sous un étiquetage unique. La notion que leur réunion constituerait est extrêmement diffuse et serait susceptible de donner lieu à une identification trop puissante. De fait l'inscription de ces N dans le chaînage ne recevra pas de dénomination générique, la réalisation lexicale ne correspondant qu'à une réalisation prénotionnelle<sup>52</sup> :

<b>SN1</b> SN[+hum]	<i>pédaler / patiner</i>	<b>(MODIF)</b> <i>vraiment...</i>	<i>dans la / le</i>	<i>semoule / choucroute / yaourt</i>
------------------------	--------------------------	--------------------------------------	---------------------	--------------------------------------

### 3. À titre de conclusion

Conclure l'analyse serait une gageure tant de questions sont restées en suspens. On peut néanmoins dégager des acquis et dresser des perspectives d'étude (dont une partie est actuellement en cours). Concernant les premiers, on doit distinguer entre deux réalisations du lemme. L'une, approximative et linéaire, qui sert de monnaie d'échange dans l'interlocution, que réalisent les dictionnaires et que nous avons utilisée dans le corps de notre texte. Une autre, qui ne correspond plus à un identifiant autonymique, mais à une description structurale et lexicale. Son objet est d'établir la forme de la SP, à savoir son format mais aussi de rendre compte de ses caractéristiques majeures : la variabilité de sa densité sémiotique et sa constitution (spatiale) en réseaux<sup>53</sup>.

Si l'on excepte les séquences totalement figées comme *passer l'arme à gauche, manger les pissenlits par la racine*, etc., la structuration lemmatique des SP présente quatre caractéristiques majeures :

- elle repose sur des items préencodés, non paradigmatiques (*comme un cochon ; un coup sur ; la dalle ; la / le semoule, choucroute, yaourt...*). Ceux-ci sont plus ou moins organisés en groupes syntaxiques. De fait il serait plus exact de parler de segments ;
- elle admet d'être fragmentée par la présence de paradigmes. Ceux-ci conditionnent, quand ils sont l'expression de propriétés syntaxiques, le format dimensionnel de la SP (en termes de quantité de chaînons). La SP peut accueillir plusieurs systèmes paradigmatiques hétérogènes, croisant les données lexicales et syntaxiques. Le figement d'une SP peut être défini en termes de défektivité des paradigmes (plus les paradigmes sont défectifs, plus le figement est élevé) ;
- elle fonde pour partie son identité sur un découpage notionnel du matériau. L'instanciation, sur un chaînon donné, d'un paradigme notionnel ou prénotionnel conditionne étroitement la configuration des segments préencodés. Sur le plan sémantique, cette instanciation confère à la compositionnalité de l'interprétation (Mejri 1997, Petit 1998 a) une densité variable, avec présence de lieux spécifiquement contraints et d'autres dont la pertinence réside dans un faisceau plus ou moins flou d'instructions. De fait la structuration sémantique des SP qui ne sont pas totalement figées s'apparente davantage à chaînage flou qu'à une compositionnalité univoque. Ceci a pour conséquence, on l'aura compris, l'impossibilité d'identifier ces séquences par un lemme textuel unique ;

<sup>52</sup> *Prénotionnelle* et non pas *notionnelle* car c'est bien à l'intérieur d'un champ circonscrit (les mets) que s'opère la sélection des items. Par ailleurs, il est possible de dégager une *gestalt* commune, aussi vague soit-elle. En revanche, la disparité des propriétés spécifiques des items convoqués interdit toute identification sortale

<sup>53</sup> L'observation de la structure des lemmes lexicaux dégagés dans la présente étude démontre qu'ils ne sont pas susceptibles d'une linéarisation, même réduite aux dénominations des chaînons.

- la structure lemmatique des SP est complexe, c'est-à-dire composite. La réalisation effective de la SP en discours nécessite une sélection parmi les variables d'instanciation de la notion. Ceci implique que le lemme lexical, tel que nous cherchons à le dégager doit rendre compte des cheminements disponibles, ouverts ou fermés, qui relient les différentes positions entre elles et au matériau précontraint.

Concernant les pistes à explorer, relevons :

- l'intégration dans la représentation lemmatique d'une éventuelle mobilité des composantes, notamment lorsque la SP satisfait au test de passivation ou de détachement ;
- une étude des structures notionnelles en termes de quantité de notions admissibles pour un même chaînon, de structuration sémantique des notions (des schémas récurrents sont-ils observables ; existe-t-il des isotopies notionnelles, des contraintes générales quantitatives et qualitatives sur la réalisation des notions ?). À prévoir également une étude portant sur la saturation lexicale des notions : certains paradigmes sont clos (p.ex. celui de TÊTE, ÊTRE DE NOUVEAU), d'autres non (p.ex. celui de DONNER / RECEVOIR) ;
- une étude des structures prénotionnelles (description des déficits qui les empêchent de se constituer en notions ; les configurations prénotionnelles peuvent-elles être typologisées ?) ;
- l'extension des paradigmes lexicaux subsumés par une notion : certains sont fermés (p.ex. les reformulants familiers de *tête*) et d'autres tendanciellement ouverts (p.ex. les V qui instancient les notions DONNER, RECEVOIR, ÊTRE...). Il conviendrait de distinguer entre les contraintes liées à l'usage et celles d'origine linguistique et de spécifier les diverses configurations rencontrées ;
- le phénomène de pontage qui assure l'homogénéité de la SP par-delà les ruptures dont sa forme signifiante est susceptible (est-il vectorialisé par un item source ?) ;
- une typologie (à faire) des ruptures dont la linéarité du signifiant est susceptible, et des contraintes qui y sont liées. Elle nécessiterait la prise en compte de celles qui sont liées à l'apparition de modificateurs divers, notamment adjectivaux et adverbiaux (non traitées dans la présente étude). Avec pour toile de fond la recherche de régularités combinatoires entre zones de ruptures et zones de déplacements.

## Références bibliographiques

Amiot D. (1997) *L'antériorité temporelle dans la préfixation en français*. Presses du Septentrion

Blanco X. et Bonell C. (1998) "Vers une structuration syntactico-sémantique de la terminologie médicale. Applications à la traduction espagnol-français" *Cahiers de grammaire*, n°23, Université de Toulouse Le Mirail

Blanco X. et Moreno D. (1997) "Lemmatisation, agencement et catégorisation des lexies complexes dans la lexicographie bilingue français-espagnol" in Fiala, Lafon et Piguet ed.

Candel D. (1995) "Locutions et langues de spécialité" in Martins-Baltar ed.

Collinot A. et Mazière F. (1997) *Un prêt à parler : le dictionnaire*, PUF

Dubois J. et Dubois C. (1971) *Introduction à la lexicographie, le dictionnaire*, Larousse

Fiala P., Lafon P. et Piguet M.-Fr. ed. (1997) *La locution : entre lexique, syntaxe et pragmatique*, Collection Saint-Cloud, INALF, Klincksieck

Franckel, J.-J. et Lebaud D. (1992) "Lexique et opérations. Le lit de l'arbitraire", *La théorie d'Antoine Culioli*, Ophrys

- Gaatonne d. (1981) “Les locutions verbales : pour quoi faire ?”, *Revue Romane*, Vol 16, Copenhague
- Gross G. (1988) “Degré de figement des noms composés”, *Langages*, n°90, Larousse
- Gross G. (1993) “Trois applications de la notion de verbe support”, *L'information grammaticale*, n°59, Paris
- Gross G. (1996) *Les expressions figées en français*, Ophrys
- Gross M. (1993) “Les phrases figées en français”, *L'information grammaticale*, n°59, Paris
- Guiraud P. (1980) *Les locutions françaises*, Que sais-je ? n°903, PUF
- Habert B., Nazarenko A. et Salem A. (1997) *Les linguistiques de corpus*, Colin
- Kerleroux F. (1996) *La coupure invisible*, Presses du Septentrion
- Kerleroux F. (1999) “Sur quelles bases opère l’apocope ?”, *Silexicales* n°2, Presses du Septentrion
- Kleiber G. (1997) “Massif / comptable et partie / tout”, *Verbum*, Vol XIX n°3, *La relation partie-tout*, Presses universitaires de Nancy
- Kleiber G. (1999) *Problèmes de sémantique*, Presses du Septentrion
- Langages*, n°131 *Les classes d’objets*, (1998), Larousse
- Lehmann A. et Martin-Berthet F. (1998) *Introduction à la lexicologie. Sémantique et morphologie*, Dunod
- Martins-Baltar M. ed. (1995) *La locution en discours*, Les cahiers du français contemporain, n°2, Credif, Didier
- Martins-Baltar M. ed. (1997) *La locution entre langue et usages*, ENS éditions, Ophrys
- Mejri S. (1994) “Séquences figées et expression de l’intensité. Essai de description sémantique” *Cahiers de lexicologie*, n°65, INALF, Didier
- Mejri S. (1997) *Le figement lexical. Descriptions linguistiques et structuration sémantique*. Publications de la Faculté des Lettres de la Manouba, Tunis
- Mejri S. (1998) “Le figement et la linéarité du signe linguistique”, *L'information grammaticale*, Numéro Spécial Tunisie, CERES, Tunis
- Mejri S., Gross M, Clas A. et Baccouche T. (1998) dir *Le figement lexical*, Actes des Premières Rencontres Méditerranéennes, les 17-18 et 19 septembre 1998, CERES, Tunis
- Mortureux M.-Fr. (1997) *La lexicologie, entre langue et discours*, Sedes
- Niklas-Salminen A. (1997) *La lexicologie*, Colin
- Petit G. (1998a) “Remarques sur la structuration sémiotique des locutions familières’ in Mejri, Gross, Clas et Baccouche dir.

- Petit G. (1998 b) “Un phénomène d’hybridation sémiotique et sémantique : les noms familiers”, *Le français moderne*, T LXVI n°1, CILF
- Petit G. (1999 a) dir. *Linx*, n°40, *Le statut d’unité lexicale*, Université de Paris X-Nanterre
- Petit G. (1999 b) “La double hybridation de l’unité lexicale”, in Petit G. (1999 a) dir.
- Petit G. (1999 c) (à paraître) “Dénomination et lexique” *Journal of French Language Studies*, Cambridge University Press
- Rey A. (1977) *Le lexique, images et modèles. Du dictionnaire à la lexicologie*, Colin
- Rey-Debove J. (1971) *Étude linguistique et sémiotique des dictionnaires français contemporains*, Mouton
- Rey-Debove J. (1998) *La linguistique du signe*, Colin
- Tamba I 1979 “Signification des figures de comparaison”, *L’information grammaticale*, n°1, Paris
- Tyvaert J.-E. (1997 a) “Continuité référentielle et discrétisation substantivale”, *Verbum*, Vol XIX n°1-2, *Référence et anaphore*, Presses universitaires de Nancy
- Tyvaert J.-E. (1997 b) “Le problème de l’identification des parties”, *Verbum*, Vol XIX n°3, *La relation partie-tout*, Presses universitaires de Nancy
- Villoing F. (1999) “La querelle, au XIXe siècle, autour des mots composés du type [VN]<sup>N</sup> : quels enjeux pour la grammaire ?”, in G.Petit dir.
- Wagner R.-L. (1967) *Les vocabulaires français*, Vol 1, Didier