

Les proverbes et la pauvreté : méthodologie d'analyse linguistique et *corpus-driven* des proverbes sur la pauvreté en français et en italien

Vincenzo Lambertini, Université de Bologne, Italie

Citation: Lambertini, V. (2019), "Les proverbes et la pauvreté : méthodologie d'analyse linguistique et *corpus-driven* des proverbes sur la pauvreté en français et en italien", *mediAzioni* 24, <http://mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

1. Introduction

Dans le passé, les études sur les proverbes ont principalement suivi deux tendances : d'une part, elles ont préféré adopter des approches folkloriques visant à mettre en lumière les traditions d'une communauté linguistique plutôt que des démarches plus strictement linguistiques ; d'autre part, elles ont souvent opté pour des approches *descendantes* (ou *top-down*), qui analysaient les proverbes issus de listes ou de dictionnaires de proverbes, si bien que leurs conclusions n'étaient valables que pour les échantillons de proverbes préétablis.

Quant aux chercheurs qui ont étudié les proverbes en les repérant dans des contextes authentiques (et en particulier dans des textes écrits), ils se sont limités à certains domaines (comme par exemple la politique) ou à certains auteurs. À notre sens, les raisons de cette démarche étaient essentiellement de nature pratique : étant donné que pour repérer un proverbe il fallait lire le texte en entier, une sélection des matériaux à analyser s'imposait afin de rendre possible ce type de recherche.

Or, des méthodologies, des approches, voire des disciplines novatrices, comme la linguistique de corpus, ont vu le jour grâce aux nouvelles technologies. C'est précisément du fait de ces innovations qu'il est possible aujourd'hui d'adopter des approches *ascendantes* (ou *bottom-up*), et plus précisément, en empruntant une expression utilisée en linguistique des corpus, une approche *corpus-driven*¹, c'est-à-dire guidée par les corpus. Cette démarche présente l'avantage de repérer les données à étudier directement dans les corpus linguistiques (ce qui permet d'avoir des échantillons plus représentatifs) pour formuler des hypothèses sur la base des observations produites.

Cet article entend présenter une méthode d'analyse de type *corpus-driven* des proverbes sur la pauvreté en italien et en français. Pour ce faire, nous définirons d'abord *le proverbe* d'un point de vue purement linguistique, afin de mettre en lumière ses conditions nécessaires. Puis nous nous pencherons sur les différences entre une approche descendante et une démarche ascendante, ce qui mettra en évidence le potentiel de l'approche *corpus-driven* pour l'étude des proverbes. Enfin, nous établirons une liste des proverbes sur la pauvreté les plus fréquemment énoncés de nos jours, ce qui nous permettra de formuler une série de considérations non seulement sur ces proverbes, mais aussi sur leur traduction et sur leur apprentissage.

2. Linguistique des proverbes

Avant de commencer notre analyse, il est nécessaire de préciser ce que nous entendons par proverbe. Cette précision n'est pas superflue, car bien que les études sur les proverbes aient une longue tradition, les chercheurs ne sont pas tous d'accord sur ce qu'est un proverbe. Dans cette étude, nous tenterons de trouver un compromis entre les différentes définitions, en tenant compte de la prémisse méthodologique suivante : étant donné le caractère linguistique et non folklorique de notre recherche, nous prendrons en considération seulement les conditions linguistiquement nécessaires des proverbes. Nous considérons donc

¹ En linguistique des corpus, l'approche dite *corpus-driven*, et adoptée dans le cadre de la présente étude, est opposée à l'approche nommée *corpus-based*. Pour plus de détails sur ces deux démarches, voir Tognini-Bonelli (2001 : 65-99).

que les proverbes sont au moins des phrases², de type ON-sentencieux (Anscombe 2000 : 11-12), à savoir des phrases sentencieuses dont l'énonciateur premier n'est plus reconnaissable. En outre, nous nous basons sur le fait que les proverbes sont des phrases génériques avec restriction à l'homme (Kleiber 2000 : 45-48 et Schapira 1999 : 69-78). Par conséquent, la première condition nécessaire des proverbes est qu'ils doivent être des phrases : il s'agit d'une affirmation d'apparence banale mais qui définit de manière claire et indéniable leur caractéristique principale, tout en expliquant pourquoi les proverbes sont différents d'autres formes linguistiques comme les expressions figées. Vu sous un autre angle, nous pourrions dire que les proverbes ont la capacité d'entrer dans la communication sans qu'aucun changement ou ajout s'imposent. En effet, cette dimension phrastique fait en sorte que le proverbe a une signification et est en même temps capable de transmettre un message, ce qui n'est pas le cas pour les expressions figées, qui ne sont que des constituants et qui, sans autre ajout, sont incapables de véhiculer un message.

Cette considération revêt une importance majeure y compris pour ce qui est du fonctionnement sémantique des proverbes. Dans ce cadre, et à partir des observations de Schapira (1999 : 78-87), nous avons établi trois niveaux sémantiques distincts : un *sens compositionnel*, qui relève des significations des mots composant les proverbes ; un *sens parémiologique*, qui correspond au sens générique ou « idiomatique » des proverbes ; enfin, le *message* véhiculé par les proverbes et qui relève de leur contexte. La prise en compte de ces trois niveaux sémantiques aboutit à une série de considérations : tout d'abord, la distinction entre proverbes littéraux et proverbes métaphoriques n'a pas grande valeur, étant donné que ce n'est pas le proverbe en soi qui est littéral ou métaphorique, mais c'est son interprétation qui peut être soit littérale soit métaphorique. En effet, un même proverbe peut transmettre, selon le contexte, aussi bien un message qui coïncide avec son sens compositionnel qu'un message qui ne correspond pas à son sens compositionnel. Ce sont alors les éléments contextuels qui indiquent le type d'interprétation qu'il faut attribuer à un énoncé (voir à ce sujet

² L'unité phrastique devient le seuil minimal du proverbe, qui ne pourra pas être un constituant comme les expressions figées mais qui pourra éventuellement réunir plusieurs phrases et créer ainsi un texte.

Prandi 2004 : 13)³. En second lieu, si nous considérons le niveau d'énonciation des proverbes, à savoir celui qui correspond à leur message, nous pouvons comprendre aisément que le contexte d'énonciation des proverbes acquiert un intérêt non négligeable lors de toute étude parémiologique, ce qui rend nécessaire une prise en compte des contextes grâce, par exemple, à la linguistique de corpus.

3. Proverbes, pauvreté et linguistique de corpus : repérage et élaboration des données

Considérons maintenant la typologie des corpus utilisés, que nous avons choisis sur la base des finalités de notre étude.

Premièrement, pour effectuer notre étude bilingue (français-italien), il a été nécessaire de travailler sur deux corpus, l'un français et l'autre italien, qui soient comparables et non pas parallèles. Les premiers corpus sont composés de textes, en deux ou plusieurs langues, ayant des caractéristiques comparables alors que les seconds sont des corpus comprenant des textes en langue source et les traductions correspondantes en langue cible⁴. Même si au départ la traduction des proverbes nous intéressait particulièrement, nous avons décidé d'éliminer dans cette première phase de repérage les données issues de textes traduits, car notre but était de mener une étude statistique et comparée de proverbes italiens et français basée sur des données authentiques et non pas influencées (et, donc, potentiellement faussées) par la traduction.

Deuxièmement, la nature de notre étude nous a imposé de disposer d'un corpus de très grande dimension permettant de compenser la faible fréquence d'utilisation des proverbes et donc d'en repérer le plus grand nombre, bien qu'en littérature parémiologique il soit difficile de trouver des données certaines sur leur

³ C'est pourquoi certains proverbes considérés normalement comme « littéraux » sont susceptibles d'avoir également des lectures métaphoriques, comme par exemple le proverbe *Tel père, tel fils* (voir Kleiber 2000 : 47).

⁴ Pour plus d'informations, voir Tognini-Bonelli (2001 : 6-7) et Chiari (2007 : 53-54).

fréquence d'utilisation moyenne⁵ ou sur leur disparition⁶. Cependant, pour des raisons statistiques élémentaires, nous avons supposé que puisque les proverbes sont des phrases et non pas de simples mots ou des constituants, ils doivent forcément être peu fréquents.

Troisièmement, pour mener une analyse parémiologique tous azimuts qui ne soit pas restreinte à certaines situations de communication, nous avons dû nous servir de corpus généraux, comprenant différentes typologies de textes, différents registres linguistiques, voire des textes ayant un certain degré de spontanéité. Nous cherchions donc des corpus contenant des textes situés à la frontière entre la langue écrite et la langue orale. Toutefois, si, d'une part, nous ne voulions pas baser nos recherches sur des enquêtes spécifiques menées auprès de certains locuteurs, pour éviter de tomber dans le piège du paradoxe de l'observateur (énoncé par Labov 1972 : 209 et évoqué également par Traverso 1999 : 22), d'autre part, nous ne pouvions pas prendre en considération des corpus oraux, en raison de leur taille trop restreinte et donc insuffisante pour mener une analyse parémiologique exhaustive.

C'est pourquoi nous nous sommes servi de deux corpus linguistiques, l'un français et l'autre italien, nommés respectivement frWaC et itWaC⁷. Il s'agit de deux corpus comparables de vastes dimensions (chacun d'entre eux contient approximativement deux milliards de mots) ; en outre, ils sont composés de textes repérés de façon automatique sur le Web, raison pour laquelle ils prennent en compte, pour la plupart, des articles, des blogs et des forums publiés sur

⁵ Norrick (1985 : 6) par exemple affirme que dans un corpus de conversations en anglais (qui est le suivant : *A corpus of English Conversation*) constitué de 43 165 lignes, qui correspondent à 891 pages, on ne peut repérer qu'un seul proverbe.

⁶ Nos études montrent, au contraire, que les proverbes sont utilisés de nos jours et qu'ils sont même fréquents. Si nous excluons la conversation et la communication ordinaires, nous retrouvons les proverbes dans bien des domaines : dans la presse et dans les médias, mais aussi dans la publicité, qui fait un usage intensif des proverbes, très souvent détournés, pour attirer l'attention des consommateurs.

⁷ Voir Baroni et al. (2008). Voir en bibliographie les liens vers les deux corpus qui sont consultables gratuitement en ligne.

Internet au cours des premières années 2000⁸.

Avant de nous pencher sur la méthode de repérage semi-automatique des proverbes dans ces corpus, nous tenons à réitérer l'importance d'éviter toute ingérence du chercheur, ainsi que toute influence des ressources parémiologiques sur la phase d'extraction des proverbes pour ne pas conditionner les résultats de la recherche. En effet, si le repérage des proverbes repose entièrement sur les connaissances du chercheur et sur les ressources dont il dispose, le choix des proverbes pourrait se limiter aux seuls proverbes connus par le chercheur et présents dans ses ressources. La recherche serait ainsi orientée vers certaines conclusions, en négligeant dès le début d'autres pistes de recherche.

Pour éviter cet écueil, nous avons donc eu recours à la méthodologie *corpus-driven* : les proverbes ont été repérés non pas à partir de leurs mots, mais en fonction d'éléments (mots ou expressions) qui en indiquaient la présence et que nous avons désignés par l'expression *marqueurs de proverbe* (Lambertini 2016 : 83-88). Avant tout, une précision s'impose pour ce qui est de la valeur attachée au concept de *marqueur* dans cette étude. Cette notion a été empruntée à des études menées dans le domaine de la métaphore et des corpus. En particulier, nous faisons référence aux études de Stefanowitsch (2006)⁹, qui propose de détecter des métaphores à l'intérieur de corpus linguistiques à partir de mots ou expressions qui accompagnent habituellement tout type de métaphore. On citera comme exemples de ces expressions, qu'il appelle *markers of metaphor* (trad. : marqueurs de métaphore) : *kind of, so to speak, etc.* De la même manière, plusieurs chercheurs (comme Schapira 2000, Kleiber 1999, Cram 1983) ont remarqué que souvent les proverbes sont accompagnés de formules telles que : *comme on dit, comme dit le proverbe, etc.* Nous avons ainsi appliqué la notion

⁸ Ces deux corpus ont été construits en suivant les mêmes critères et sont composés de textes ayant été rédigés plus ou moins dans les mêmes années. Plus précisément, nous avons observé que la concordance du mot *proverbio*, dans le corpus itWaC, montre qu'une majorité de textes date de 2005 et que la concordance du mot *proverbe*, dans le corpus frWaC, produit des résultats appartenant à des textes publiés pour la plupart en 2007. A ce propos, voir Lambertini (2016 : 115-119).

⁹ Voir aussi Deignan (2009 : 16-19).

de *marqueur* à ces expressions et nous avons cherché le marqueur de proverbe le plus fréquent et le plus prolifique, à savoir celui qui nous aide à détecter le nombre le plus élevé de résultats pertinents. Ainsi sommes-nous parvenus à la conclusion que le marqueur le plus fréquent (et donc le plus utilisé par des locuteurs dans tous les contextes) est le mot *proverbe(s)* (au singulier et au pluriel) pour le français et le correspondant *proverbio* (singulier) et *proverbi* (pluriel) pour l'italien. Cette stratégie nous a permis d'extraire tous les proverbes définis comme tels par les utilisateurs d'Internet. Certes, comme l'a montré Mieder (1993: 18-37) lors de son enquête auprès de cinquante-cinq personnes auxquelles il a demandé de donner une définition du proverbe, les locuteurs ont des idées très variées sur les proverbes. Nos observations confirment qu'ils considèrent parfois comme un proverbe ce qui est en réalité une expression figée¹⁰ ou un aphorisme. Afin de surmonter cet obstacle, nous avons décidé de retenir seulement les proverbes ayant plus de vingt occurrences. Nous avons fixé ce seuil minimal, en nous inspirant de ce qui est fait en linguistique de corpus : comme le dit Sinclair (2005), pour formuler une description lexicographique d'un mot il faut avoir au moins vingt exemples de ce mot. Certes, les proverbes ne sont ni des mots ni des constituants ; toutefois, faute d'études plus spécifiques, nous avons appliqué cette méthodologie aux proverbes et nous avons effectivement observé que les phrases définies comme « proverbes » par les locuteurs et ayant au moins vingt occurrences réunissaient dans la quasi-totalité des cas les critères linguistiques énoncés ci-dessus.

Pour détecter les proverbes sur la pauvreté dans nos deux corpus, les marqueurs de proverbe ont été associés à des mots appartenant au champ lexical de la pauvreté en italien¹¹ et en français¹², qui ont été repérés par le biais de deux

¹⁰ Néanmoins, comme l'a démontré Conenna (2017), on ne peut pas exclure qu'il y ait une sorte de continuum entre les expressions figées et les proverbes, qui « sont reliés par un lexique commun et par un jeu de renvois et d'allusions » (Conenna 2017 : 154).

¹¹ Voici les mots choisis pour l'italien et présentés par ordre alphabétique : *bisogno/i* (besoin/s), *denaro/i* (argent), *fame* (faim), *mancanza* (manque), *miseria* (misère), *necessità* (nécessité), *pane* (pain), *penuria* (pénurie), *poco/ca/chi/che* (peu/peu nombreux/peu nombreuses), *povero/i* (pauvre/s), *povertà* (pauvreté), *ricchezza/e* (richesse/s), *ricco/a/chi/che* (riche/s), *ristrettezza/e* (étroitesse), *scarsità* (carence).

¹² Voici la liste des mots français rangés par ordre alphabétique : argent, besoin, carence, disette,

dictionnaires de synonymes en ligne : le dictionnaire des synonymes *Treccani*¹³ pour l'italien, et le *Dictionnaire Électronique des Synonymes (DES)* du CRISCO¹⁴.

Ces mots ont été cherchés dans les contextes à droite et à gauche du marqueur de proverbe, à une distance de un à vingt mots, ce qui a permis de les repérer non seulement à l'intérieur des proverbes mais aussi dans leurs contextes. Ce choix s'est imposé en raison de l'importance que revêt le contexte pour la définition du sens parémiologique des proverbes qui, comme nous l'avons vu ci-dessus, ne correspond pas forcément à leur sens compositionnel. Il y a une autre raison à la base de cette décision : si nous considérons les proverbes répertoriés par nos dictionnaires dans les catégories de pauvreté et de richesse, nous remarquons que tous ces proverbes ne présentent pas les mots *pauvreté*, *richesse*, *povertà*, *ricchezza*. Il suffit de mentionner les exemples suivants : *L'argent ne fait pas le bonheur en ménage mais il aide à s'en passer* ; *Les malheureux n'ont point de parents* (Montreynaud *et al.* : 1989) ; *L'abbondanza foriera è d'arroganza* (trad. litt. : L'abondance est porteuse d'arrogance) ; *Meglio poco con onore che molto con vergogna* (trad. litt. : Mieux vaut peu avec honneur que beaucoup avec honte).

4. Comparaison entre approche descendante et approche ascendante : dictionnaires VS corpus

Reste à se demander si une approche *corpus-driven* comme celle que nous venons de formuler peut effectivement apporter des avantages par rapport à des approches descendantes. Pour mieux comprendre la nature des résultats auxquels mènent ces approches opposées, nous avons cherché les proverbes concernant la pauvreté dans six dictionnaires de proverbes italiens et français parmi les plus complets. Certes, comme nous le verrons par la suite, il y a une

famine, insuffisance, insuffisant/e, maigre/s, malheur/eux/euse/euses, manque, misère, nécessité, pain, pauvre/s, pauvreté, riche/s.

¹³ En ligne : <http://www.treccani.it/sinonimi/> (consulté le 14 septembre 2018).

¹⁴ En ligne : <http://crisco.unicaen.fr/des/> (consulté le 14 septembre 2018).

forte disparité quantitative entre les trois dictionnaires italiens et les trois dictionnaires français. Toutefois, étant donné que nous nous rangeons du côté des utilisateurs de ces ressources, et non du côté des concepteurs, nous avons choisi ces dictionnaires sur la base de leur diffusion (sur les sites de commerce en ligne, mais aussi dans les bibliothèques et dans les librairies italiennes et françaises) et de leur fiabilité. Ces dictionnaires sont : le *Dictionnaire de proverbes, sentences et maximes* (Maloux 1980) qui, s'il répertorie un grand nombre de proverbes du monde entier, en recense toutefois peu provenant de France ; le *Dictionnaire des proverbes et dictons de France* (Dournon 1986) et le *Dictionnaire de proverbes et dictons* (Montreynaud et al. 1989) ; *Il grande dizionario dei proverbi italiani* (Guazzotti et Oddera 2006) ; le *Dizionario dei proverbi italiani* (Lapucci 2006) et le *Dizionario dei proverbi* (Boggione et Massobrio 2007). Les méthodes de recherche des proverbes disponibles varient en fonction de la typologie de classification des proverbes proposée par les différents dictionnaires. Ceux qui se basent sur des catégories thématiques nous donnent la possibilité de repérer les proverbes souhaités selon leur appartenance à la catégorie choisie. Dans notre cas, il s'agit notamment de la catégorie « *Pauvreté et richesse* », qui n'est toutefois pas présente dans tous les dictionnaires analysés. Par exemple, dans le dictionnaire de Guazzotti et Oddera (2006) et dans celui de Boggione et Massobrio (2007), nous retrouvons respectivement la catégorie « *Ricchezza e povertà* » et « *Ricchezza/povertà* » (trad. : Richesse et pauvreté ; Richesse/pauvreté), mais le dictionnaire de Montreynaud et al. (1989) propose deux catégories qui concernent la pauvreté, à savoir « *Riche et pauvres* » et « *Les riches, la richesse* ». Nous ne comprenons pas vraiment la raison pour laquelle les auteurs ont établi deux catégories différentes pour des concepts aussi semblables. On peut supposer que ce dictionnaire a une nette orientation de type parémiologique, qu'il attache une grande importance à la catégorisation et qu'il prend peu en compte l'utilisateur, pour qui la systématisation philologique a généralement moins d'intérêt que la praticité d'utilisation (en ce qui concerne la vitesse de repérage et de détection de la catégorie ciblée).

Les dictionnaires de Lapucci (2006) et de Dournon (1986) ne proposent aucune catégorie thématique concernant la pauvreté ; c'est pourquoi la seule manière de

repérer des proverbes sur ce thème est d'effectuer une recherche des mots présents dans les proverbes. Prenons le dictionnaire de Lapucci (2006), où nous pouvons repérer différents mots liés à la pauvreté et où, pour chaque mot-clé, le dictionnaire nous conseille d'autres mots-clés liés à celui-ci. Commençons par le mot *povero* (trad. : pauvre) qui renvoie au mot *povertà* (trad. : pauvreté) et qui est relié à son tour à plusieurs mots, parmi lesquels *ricchezza* (trad. : richesse). Ce dernier mot, quant à lui, est en relation avec d'autres mots comme *ricco* (trad. : riche), qui est très intéressant puisqu'il s'agit d'un autre cas d'excès de précision quant à la catégorisation des proverbes pour les raisons suivantes : le dictionnaire contient une section de dix-sept proverbes qui sont placés sous le mot *ricco*, considéré comme adjectif, et une autre section composée de trente-et-un proverbes identifiée par le mot *ricco* considéré comme substantif.

Par le biais des deux méthodes choisies, il est possible de repérer des proverbes concernant la pauvreté. Au vu des résultats, nous avons été étonné par le grand nombre de proverbes relatifs à ce sujet. Nous en avons dénombré 197 dans le dictionnaire de Guazzotti et Oddera (2006), 222 dans le dictionnaire de Lapucci (2006) et pas moins de 741 dans le dictionnaire de Boggione et Massobrio (2007). Dans les dictionnaires français sélectionnés nous obtenons un ensemble bien plus restreint de proverbes, ce qui est sans aucun doute dû à la typologie des dictionnaires sélectionnés (qui recensent en général moins de proverbes par rapport aux dictionnaires italiens) mais aussi au fait que « les proverbes italiens sont proportionnellement plus nombreux que les français », comme le souligne Conenna (2017 : 134). Dans le dictionnaire de Dournon (1986) nous recensons vingt-trois proverbes et dans le dictionnaire de Montreynaud et al. (1989) nous en relevons trente¹⁵.

Si nous comparons ces données à celles que nous avons obtenues dans le cadre d'une recherche sur les proverbes concernant la peur, nous pouvons effectivement affirmer que les proverbes sur la pauvreté sont assez nombreux :

¹⁵ Étant donné que le dictionnaire de Maloux (1980) mélange les proverbes des quatre coins de la planète, il est plus difficile de limiter la recherche aux seuls proverbes français de France, c'est pourquoi nous n'avons pas pris en considération ce dictionnaire.

ils sont cinq fois plus nombreux que ceux sur la peur¹⁶. Toutefois, lorsque nous cherchons ces proverbes dans nos corpus, la situation change totalement.

En effet, dans le corpus italien, il n'existe aucun proverbe contenant le mot *povertà* (trad. : pauvreté) et les proverbes où figure le mot *povero* (trad. : pauvre) sont très peu nombreux. Pour rendre plus évident l'écart existant entre les proverbes repérés dans les dictionnaires utilisés et les proverbes recensés dans nos corpus linguistiques, nous avons dressé deux listes de fréquence des mots (*wordlist*), l'une créée à partir des mots les plus fréquents contenus dans les proverbes sur la pauvreté issus des dictionnaires et l'autre rédigée sur la base des mots les plus récurrents présents dans les proverbes concernant la pauvreté repérés dans le corpus itWaC (tous proverbes confondus, sans tenir compte du critère des vingt occurrences).

Substantif	Occurrences dans le dictionnaire de Guazzotti & Oddera	Occurrences dans le corpus itWaC
povero/i (pauvre/s)	37	2
ricco/i (riche/s)	22	1
povertà (pauvreté)	10	0
ricchezza (richesse)	10	4
uomo (homme)	14	1
casa (maison)	9	0
pane (pain)	7	5
amico (ami)	4	2
mercante (marchand)	4	0
nobiltà (noblesse)	4	0
fame (faim)	3	7
bisogno (besoin)	3	4
salute (santé)	0	3
gallina (poule)	0	2

¹⁶ Dans une étude à paraître, nous avons comparé les données issues du dictionnaire de Boggione et Massobrio (2007), seul dictionnaire à consacrer une catégorie thématique à la peur et une autre à la pauvreté. Ce dictionnaire recense 741 proverbes sur la pauvreté et 137 sur la peur. Pour expliquer cet écart, nous pouvons formuler l'hypothèse suivante : comme la peur est une émotion plus abstraite que la pauvreté (considérée comme une situation de manque de ressources et de moyens), on pourrait supposer qu'il y a un nombre plus élevé de proverbes concernant des concepts plus tangibles que des concepts abstraits.

Tableau 1 : Comparaison entre les mots les plus récurrents des proverbes sur la pauvreté repérés dans le dictionnaire Guazzotti et Oddera (2006) et les mots les plus fréquents des proverbes concernant la pauvreté issus du corpus itWaC.

Les résultats de cette comparaison sont évidents : les substantifs présents dans les proverbes issus des dictionnaires et concernant la pauvreté ne sont pas présents dans les proverbes liés au concept de la pauvreté et repérés dans le corpus italien. Cette observation nous mène à la conclusion suivante : les proverbes contenus dans le corpus italien et associés à des contextes qui font référence à la pauvreté ne correspondent pas à ceux qui sont recensés par les dictionnaires dans la catégorie thématique relative à la pauvreté. En effet, aucun proverbe repéré dans le corpus itWaC et lié à la pauvreté ne contient le mot *povertà* (pauvreté), tandis que le mot *povero* (pauvre) a une très faible fréquence : ce résultat témoigne une fois de plus de l'écart entre signification compositionnelle et signification parémiologique, une différence qui doit être prise en compte notamment quand on considère les proverbes utilisés de nos jours. Le tableau 1 montre que le substantif le plus récurrent dans les proverbes issus du corpus itWaC et concernant la pauvreté est le mot *faim* (qui revient sept fois), tandis que ce mot figure seulement trois fois dans les proverbes sur la pauvreté issus du dictionnaire Guazzotti et Oddera (2006).

Il est possible d'effectuer la même comparaison entre les mots les plus récurrents présents dans les proverbes concernant la pauvreté issus des dictionnaires de proverbes français et ceux qui se réfèrent à cette condition et qui ont été repérés dans les proverbes issus du corpus frWaC.

Substantif	Occurrences dans les dictionnaires de Montreynaud et al. et Dournon	Occurrences dans le corpus frWaC
riche/s	18	10
pauvre/s	14	11
Homme	4	0
malheureux	4	0
Pauvreté	4	1
Gueux	3	0

Argent	1	5
malade	0	3
monde	1	2
personne	1	2
poisson	0	2
Vache	1	2

Tableau 2 : Comparaison entre les mots les plus récurrents des proverbes sur la pauvreté repérés dans les dictionnaires Montreynaud et al. (1989) et Dournon (1986) et les mots les plus fréquents des proverbes issus du corpus frWaC et qui font référence à cette condition.

Dans le cas des proverbes français, la situation change : les deux premiers substantifs les plus fréquents sont les mêmes pour les proverbes issus des dictionnaires et pour ceux qui sont contenus dans le corpus de langue française. On pourrait trouver plusieurs explications à cela ; par exemple, on peut supposer qu'en français les proverbes ayant un sens compositionnel portant sur les pauvres et les riches sont plus fréquents qu'en italien. Cet état de fait pourrait avoir des origines historiques, sociales ou culturelles, domaines qui, comme nous l'avons dit au début de cet article, ne font pas l'objet de notre recherche.

Mais force est de constater qu'il y a une correspondance accrue entre les mots les plus fréquents des proverbes sur la pauvreté, issus des dictionnaires français, et les mots les plus récurrents des proverbes repérés dans le corpus français. Notre hypothèse est que cette correspondance est due à la différence quantitative des proverbes présents dans les dictionnaires français et italiens, que nous avons évoquée plus haut. Les dictionnaires italiens recensent un nombre bien plus élevé de proverbes, mais tous ces proverbes ne sont pas utilisés dans la langue courante. Il nous semble donc qu'en parémiologie la quantité ne va pas de pair avec la qualité et qu'il vaudrait peut-être mieux indiquer clairement dans les dictionnaires les proverbes les plus récurrents de nos jours pour faciliter la tâche des linguistes, des traducteurs et des apprenants en langues étrangères.

Après avoir repéré les proverbes concernant la pauvreté à l'aide du marqueur de proverbe en combinaison avec les mots du champ sémantique de la pauvreté, nous avons restreint nos résultats aux seuls proverbes ayant une fréquence supérieure ou égale à vingt occurrences, afin d'éliminer les résultats non

pertinents. Ensuite, nous avons cherché les proverbes repérés dans la première phase de notre recherche sans marqueur de proverbe et en utilisant des clés de recherche spécifiques, dans le langage du logiciel NoSketchEngine¹⁷. Cette démarche s'est imposée pour repérer toutes les variantes des proverbes, y compris leurs formes modifiées ou réduites : en effet, des phénomènes tels que le détournement des proverbes¹⁸ ne peuvent pas être négligés lorsqu'on a affaire aux proverbes énoncés dans des contextes authentiques et spontanés.

Dans les deux tableaux suivants, nous présenterons les proverbes ayant plus de vingt occurrences et repérés sur la base des clés de recherche indiquées dans la colonne du milieu.

Proverbes	Clés de recherche (syntaxe NoSketchEngine)	Occurrences ¹⁹
bisogna fare di necessità virtù	"di" "necess.*" "virt.*"	444
l'appetito vien mangiando	"appetito" "vien"	160
il troppo storpia	"il" "troppo" "st.*"	131
chi si contenta/accontenta gode	".*cont.*" "gode"	120
pecunia non olet	"pecuni.*" "non" "olet"	90
non si vive di solo pane	"viv.*" "di" "solo" "pan.*"	84
chi trova un amico trova un tesoro	"trov.*" "un.*" "amic." "trov.*"	57
chi troppo vuole / nulla stringe ; ma chi poco vuole / certamente finge	"tropp." "vuole"	45
aiutati che il ciel t'aiuta	"aiutati" "che" "il" ".*iel"	24
chi ha i denti non ha il pane, chi ha il pane non ha i denti	"i" "denti" [{"0,3} "non" "ha.*"	22

Tableau 3 : Proverbes issus de itWaC ayant une fréquence supérieure ou égale à vingt.

¹⁷ En ligne : <http://nl.ijs.si/noske/index.html> (consulté le 14 septembre 2018). Ce logiciel de gestion des corpus itWaC et frWaC est disponible gratuitement en ligne. Il permet notamment des recherches sur la base de mots mais aussi en fonction d'étiquettes morphosyntaxiques. Pour plus d'informations, voir Rychlý (2007).

¹⁸ Voir entre autres Schapira (2000).

¹⁹ Pour calculer les occurrences réelles, nous avons éliminé les répétitions des mêmes résultats.

Proverbes	Clés de recherche (syntaxe NoSketchEngine)	Occurrences ²⁰
on est jamais aussi bien servi que par soi-même	"serv.*" "que" "par"	300
ne crachez pas dans la soupe qu'on vous a donnée	"crach.*" "pas" [{"0,5} "soupe"	284
l'argent ne fait pas le bonheur	"argent" [{"0,5} "fait" [{"0,6} ".*heur"	235
un malheur n'arrive jamais seul	"malheur" [{"0,8} "seul"	83
à quelque chose malheur est bon	"à" "quelque" "chose" "malheur"	62
mieux vaut être riche et bien portant	"rich." "et" "bien" "port.*"	48

Tableau 4 : Proverbes issus de frWaC ayant une fréquence supérieure ou égale à vingt.

Ces deux tableaux ont une importance capitale : ils recensent des proverbes qui sont fréquemment énoncés dans des contextes liés à la pauvreté. Ces deux groupes de proverbes, qui sont évidemment très hétérogènes, ont en commun quelques aspects non négligeables, en particulier en ce qui concerne la traduction et l'apprentissage des langues : comme ils ont été repérés dans des contextes relatifs à la pauvreté, nous savons qu'ils transmettent statistiquement des messages portant sur la pauvreté, ce qui nous permet de supposer qu'ils peuvent avoir des significations parémiologiques similaires, malgré des significations compositionnelles distinctes. Mais surtout, nous avons la certitude que ces proverbes sont utilisés de nos jours, voire qu'ils sont répandus, étant donné qu'ils présentent une fréquence d'utilisation élevée. C'est pourquoi nous pouvons estimer qu'ils peuvent être considérés comme des équivalents fonctionnels, et qu'ils peuvent donc être employés pour traduire, dans certains contextes, des proverbes transmettant des messages similaires. Il serait donc intéressant qu'ils soient connus par les apprenants en langues étrangères, qui d'une part pourraient les rencontrer souvent et devraient donc en connaître la signification parémiologique (compétence parémiologique passive) ; d'autre part, une fois identifiés et mémorisés, ils pourront être utilisés par les apprenants dans les contextes les plus appropriés (compétence parémiologique active).

²⁰ Voir note 19.

5. Conclusion

Dans cet article, nous avons mis en lumière les avantages d'une recherche de nature linguistique sur les proverbes et d'une approche ascendante et *corpus-driven* pour leur repérage et leur analyse dans des corpus de grande taille. Nous avons donc pris en considération le sujet de la pauvreté et les proverbes qui sont liés à cette condition, en les repérant d'abord dans les dictionnaires de proverbes (trois italiens et trois français) parmi les plus complets, puis dans des corpus (l'un italien et l'autre français) comparables mais non pas parallèles, ce qui nous a permis de nous concentrer sur des données authentiques et fiables.

Les résultats ont été assez surprenants : les proverbes repérés dans les dictionnaires ne reflètent pas nécessairement la réalité linguistique, où l'on trouve une sélection de proverbes qui ne peuvent être détectés qu'en analysant des données authentiques. En effet, nous avons constaté qu'il y a une *faible présence* (pour les proverbes issus du corpus de langue française), voire une véritable *absence* (quant aux proverbes issus du corpus italien) respectivement des substantifs *pauvreté* et *povertà* dans les proverbes analysés, bien que les dictionnaires recensent de nombreux proverbes contenant ces mots (et donc se référant directement à la pauvreté). La présente étude a ainsi montré que ces proverbes ont quasiment disparu des textes contemporains (plus ou moins spontanés) en italien et en français. Cette réalité pourrait alors souligner l'instabilité du concept de pauvreté, qui subit des changements culturels non négligeables au fil du temps (sa disparition dans le patrimoine parémiologique contemporain par rapport aux proverbes d'antan en est la preuve) et qui est, pour cette raison, difficile à étudier.

Par ailleurs, c'est justement cet écart entre proverbes issus des dictionnaires et proverbes détectés dans les corpus linguistiques qui nous a contraint à repérer les proverbes concernant la pauvreté à l'aide de différents indices lexicaux présents non seulement à l'intérieur des proverbes mais aussi dans leur contexte d'utilisation. Nous avons ainsi été en mesure de détecter les proverbes français et italiens les plus utilisés dans des contextes liés à la pauvreté. Cette approche a également des retombées remarquables pour l'apprentissage de l'italien et du français L2, voire pour la traduction.

Cette étude doit alors nous faire réfléchir sur l'importance des études parémiologiques basées sur les proverbes couramment employés dans la communication, mais aussi sur la typologie des ressources concernant les proverbes, qui sont nécessaires pour le repérage, la traduction, voire l'apprentissage des proverbes. Les ressources qui existent actuellement ne reflètent pas forcément les innovations qui ont eu lieu dans le domaine de la recherche linguistique. Les approches utilisées en parémiologie sont en effet encore trop ancrées dans la tradition des études folkloriques sur les proverbes. Le chemin à parcourir pour intégrer les nouvelles technologies dans la parémiologie est encore long, mais, étant donné que ces technologies sont de plus en plus à la portée des chercheurs, nous prévoyons que bien des parémiologues seront prêts à relever ce défi.

Bibliographie

Anscombre, J.-C. (2000) « Parole proverbiale et structures métriques », *Langages*, 139 : 6-26. En ligne :

http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_2000_num_34_139_2377 (Consulté le 14 septembre 2018)

Baroni, M et al. (2008) « The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora », *Language Resources and Evaluation*, 43 (3) : 209-226. En ligne :

<http://wacky.sslmit.unibo.it/doku.php?id=publications> (consulté le 14 septembre 2018)

Boggione, V., L. Massobrio (2007) *Dizionario dei proverbi*, Turin : UTET.

Chiari, I. (2007) *Introduzione alla linguistica computazionale*, Roma : GLF Editori, Laterza.

Conenna, M. (2017) « *Chapeau / cappello*, mots-clés de proverbes français et italiens », *Scolia* 31, Strasbourg : Presses Universitaires de Strasbourg, 133-157.

Cram, D. (1983) « The linguistic status of the proverb », *Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie*, 43 (2), Paris : Classique Garnier, 53-71.

Deignan, A. (2009) « Searching for Metaphorical Patterns in Corpora », in Baker, P. (dir.), *Contemporary Corpus Linguistics*, London : Continuum, 9-31.

Dournon, J.-Y. (1986) *Dictionnaire des proverbes et dictons de France*, Paris : Hachette.

frWaC (French Web), [consulté le 12 février 2018]. En ligne : http://nl.ijs.si/noske/wacs.cgi/first_form?corpname=frwac;align=

Guazzotti, P., M. F. Oddera (2006) *Il grande dizionario dei proverbi italiani* Bologna : Zanichelli.

itWaC (Italian Web), en ligne :

http://nl.ijs.si/noske/wacs.cgi/first_form?corpname=itwac;align= (consulté le 12 février 2018)

Kleiber, G. (1999) « Les proverbes : des dénominations d'un type 'très très spécial' », *Langue française*, 123 : 52-69. En ligne : http://www.persee.fr/web/revues/home/prescript/article/lfr_0023368_1999_num_123_1_6296 (consulté le 14 septembre 2018)

Kleiber, G. (2000) « Sur le sens des proverbes », *Langages*, 139 : 6-26. En ligne : http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_2000_num_34_139_2377 (consulté le 14 septembre 2018)

Labov, W. (1972) *Sociolinguistic Patterns*, Philadelphia : University of Pennsylvania Press.

Lambertini, V. (2016) *Approccio linguistico e corpus-driven al proverbio italiano e francese: alla ricerca della forma perduta*, Thèse de doctorat, Université de Bologne. En ligne : <http://amsdottorato.unibo.it/7727/> (consulté le 14 septembre 2018)

- Lapucci, C. (2006) *Dizionario dei proverbi italiani*, Firenze : Le Monnier.
- Maloux, M. (1980) *Dictionnaire des proverbes, sentences et maximes*, Paris : Larousse.
- Mieder, W. (1993) *Proverbs are never out of season*, New York : Oxford University Press.
- Montreynaud, F. et al. (1989) *Dictionnaire de proverbes et dictons*, Paris : Dictionnaires Le Robert.
- Norrick, N. R. (1985) *How Proverbs Mean : Semantic Studies in English Proverbs*, Berlin : Mouton.
- Prandi, M. (2004) *The Building Blocks of Meaning. Ideas for a philosophical grammar*, Amsterdam : John Benjamins Publishing Company.
- Rychlý, P. (2007) « Manatee/Bonito - A Modular Corpus Manager », in *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, Brno : Masaryk University.
- Schapira, C. (1999) *Les stéréotypes en français. Proverbes et autres formules*, Paris : Ophrys.
- Schapira, C. (2000). « Proverbe, proverbialisation et déproverbialisation », *Langages*, 139 : 81-97. En ligne : http://www.persee.fr/doc/lgge_0458-726x_2000_num_34_139_2382 (consulté le 14 septembre 2018)
- Sinclair, J. McH. (2005) « Corpus and text – Basic principles », in Wynne, M. (sous la dir.), *Developing linguistic corpora: A guide to good practice*, Oxford : Oxbow Books, 1-16.
- Stefanowitsch, A. (2006) « Corpus-based approaches to metaphor and metonymy », in Stefanowitsch, A., Gries, S. Th. (dir.), *Corpus-based Approaches to Metaphore and Metonymy*, Berlin : Mouton de Gruyter, 1-16.
- Tognini-Bonelli, E (2001) *Corpus Linguistics at Work*, Amsterdam : John Benjamins.

Traverso, V. (1999) *L'analyse des conversations*, coll. « 128 Linguistique », Paris : Nathan.