

# Les marqueurs de proverbes : analyse contrastive (français-italien) basée sur corpus

---

 [dorif.it/reperes/vincenzo-lambertini-les-marqueurs-de-proverbes-analyse-contrastive-francais-italien-basee-sur-corpus/](http://dorif.it/reperes/vincenzo-lambertini-les-marqueurs-de-proverbes-analyse-contrastive-francais-italien-basee-sur-corpus/)

**Vincenzo LAMBERTINI**

---

**Vincenzo Lambertini**

Université de Turin

[vincenzo.lambertini@unito.it](mailto:vincenzo.lambertini@unito.it)

---

## Résumé

Cet article vise à analyser les *marqueurs de proverbes* (MDP) en français et en italien. Après une définition de proverbe et de MDP, nous utilisons deux corpus de référence comparables (l'un de langue française, frWaC, et l'autre d'italien, itWaC) pour détecter les meilleurs MDP dans les deux langues. L'analyse linguistique des MDP peut fournir plus de détails sur la façon dont les locuteurs ordinaires perçoivent les proverbes. Nous nous servons, en outre, de deux corpus de discours prononcés en français et en italien lors des séances plénières du Parlement européen afin de vérifier si les MDP sont utilisés de manière différente dans un cadre politico-institutionnel multilingue et multiculturel.

## Abstract

This paper is aimed at investigating the *markers of proverbs* (MOP) in French and Italian. Firstly, proverbs and MOP are defined. Secondly, two comparable reference corpora (one French, frWaC, and one Italian, itWaC) are examined in order to detect the best MOPs in French and Italian. The linguistic analysis of MOPs can provide more details on how ordinary speakers perceive proverbs. Thirdly, two corpora of speeches delivered in French and Italian during plenary sessions of the European Parliament are studied in order to check whether MOPs are used differently in a multilingual and multicultural political-institutional setting.

---

## Introduction[1]

---

La notion de *marqueur de proverbes* (MDP), sur laquelle porte cet article, ne fait pas l'unanimité chez les parémiologues. Selon certains, les MDP sont des traits stylistiques (et plus précisément, des phénomènes formels et sonores) caractérisant les proverbes (VILLERS 2017 : 80). D'autres, les MDP (désignés le plus souvent par des termes génériques, tels que *formules*) accompagnent l'énonciation des proverbes eux-mêmes (SCHAPIRA 2000 : 89-92), en remplissant la fonction de *marqueurs médiatifs* (voir à ce titre, ANSCOMBRE 2011 ; GÓMEZ FERNÁNDEZ 2018).

Dans la présente recherche, ayant un focus contrastif (français-italien), nous retiendrons la seconde acception. Nous nous concentrerons, ainsi, sur les possibilités que les MDP offrent aux chercheurs lorsqu'ils sont appliqués à des corpus : en effet, ces marqueurs, associés à l'énonciation des proverbes, loin d'en assurer un repérage exhaustif, permettent d'en obtenir des échantillons représentatifs et authentiques. En d'autres termes, les MDP, que nous allons déterminer selon des approches basées sur corpus (cf. § 3 ; 3.1 ; 3.2), permettent un repérage *corpus-driven*[2] de proverbes non détectés *a priori* par les chercheurs. Cela peut marquer un tournant en parémiologie. En outre, l'analyse linguistique des MDP donne la possibilité d'étudier la manière dont les locuteurs profanes perçoivent les proverbes.

Après avoir défini le proverbe, nous étudierons les MDP dans des corpus relevant du discours ordinaire[3] et appelés frWaC (le corpus de langue française) et itWaC (le corpus de langue italienne). Par la suite, afin de comprendre si le milieu multilingue et multiculturel des plénières du PE influe sur l'utilisation des MDP et sur les proverbes qu'ils accompagnent, nous appliquerons des méthodologies d'analyse comparables à des corpus relevant de la communication politico-institutionnelle en français et en italien et composés de discours prononcés au sein des séances plénières du Parlement européen (PE).

Au niveau théorique et méthodologique, les approches sur lesquelles repose cette recherche relèvent non seulement de la linguistique, mais aussi de l'analyse de corpus. Nous précisons également que tous les exemples montrés dans cet article seront tirés des corpus utilisés et copiés fidèlement. Lorsque des erreurs ou des coquilles figurent dans les données linguistiques originales, elles ne seront pas corrigées, mais elles seront signalées par le symbole '\*'. [4] Il s'agit d'un choix pondéré, qui se veut un gage d'authenticité, visant à démontrer, entre autres, que les proverbes sont l'apanage de toute la communauté linguistico-culturelle, sans distinction d'âge ou de niveau d'éducation.

## 1. Qu'est-ce qu'un proverbe ?

---

Contrairement à ce que l'on pourrait penser, il n'est pas simple de répondre à cette question. Faute de place, nous chercherons à le faire, en nous concentrant seulement sur les conditions nécessaires des proverbes, qui représentent, par voie de conséquence, leurs traits définitoires. Selon nos recherches (cf. LAMBERTINI 2022a), le proverbe est une phrase ON-sentencieuse, à savoir une phrase ayant un « énonciateur-premier qui est un ON-énonciateur » et acceptant « la combinaison avec *comme on dit, comme le dit la*

*sagesse des nations*, etc. » (ANSCOMBRE 2000 : 12). Le proverbe, en outre, est une phrase générique portant sur l'homme (KLEIBER 2000 : 45-48) et connue de la communauté linguistique de référence (SCHAPIRA 2000 : 85-88).

Dans tout proverbe, on détecte au moins deux significations : un sens compositionnel (qui est fonction des significations de ses parties et des règles sur la base desquelles elles se combinent ; cf. CASADEI 1996 : 14) et un sens parémiologique (la signification globale des proverbes ; cf. LAMBERTINI 2022a) qui est générique et dont les référents sont humains (indépendamment de la signification compositionnelle du proverbe).

## 2. Corpus utilisés dans cette étude

---

Premièrement, dans cette recherche, nous aurons recours à deux corpus de référence appelés frWaC (le corpus de langue française) et itWaC (le corpus de langue italienne).[5] Il est question de deux corpus comparables, à savoir construits sur la base des mêmes critères (sans pour autant que l'un soit la traduction de l'autre), de très grandes dimensions (chacun est composé d'environ deux milliards de mots), formés de textes repérés automatiquement sur la Toile (ils contiennent pour la plupart des articles, des blogs et des forums publiés sur le web au cours des premières années 2000), étiquetés sur le plan morphosyntaxique (ce qu'on appelle en linguistique de corpus *POS tagging*) et lemmatisés. Ces deux corpus peuvent être repérés à ce lien (<https://www.clarin.si/noske/wacs.cgi/>), où le logiciel en libre accès NoSketchEngine[6] en permet la sélection et le traitement. Nous pouvons considérer ces deux corpus comme des corpus de référence, voués à représenter les principales variétés d'une langue (CHIARI 2007 : 51) et donnant accès au discours ordinaire.

Deuxièmement, afin d'analyser les MDP dans la communication politico-institutionnelle en milieu européen, nous nous servirons d'un corpus de discours français et italiens issus des séances plénières du PE. Pour constituer ce corpus, nous avons utilisé la plateforme de consultation des plénières du PE[7] qui nous a permis de sélectionner un total de 492 journées[8] de juillet 2014 à juin 2022. Pour écarter toute influence des processus de traduction et d'interprétation sur nos données, nous avons décidé de limiter notre recherche aux seuls discours originaux prononcés en français et en italien, en évitant ainsi d'inclure leurs traductions officielles et leurs interprétations simultanées.

## 3. Les MDP en linguistique de corpus

---

Tout comme pour les marqueurs de métaphore (cf. DEIGNAN 2009 : 16), les MDP peuvent être exploités en linguistique de corpus pour repérer des proverbes dans de grands corpus. Le raisonnement qui est à la base de ce constat est intuitif : si les MDP sont des expressions employées lorsqu'un locuteur énonce un proverbe, leur utilisation est censée en indiquer la

présence. Toutefois, afin que les MDP puissent produire des résultats cohérents, ils doivent permettre d'obtenir le maximum de résultats pertinents (à savoir le nombre le plus élevé de proverbes) en engendrant le moins de résultats non pertinents.

Partons d'un constat : ANSCOMBRE (2011 : 28) estime que des expressions telles que *comme on dit* ou *comme dit le proverbe* sont compatibles avec l'énonciation des proverbes. C'est d'ailleurs ce que pense SCHAPIRA (2000), qui souligne tout de même que le plus souvent ces « formules introductrices » ne sont pas utilisées, car « le proverbe s'insère directement dans le discours » (*ibid.* : 90). En effet, comme nous l'avons mis en évidence dans l'introduction de cet article, les MDP ne permettent pas de repérer tous les proverbes présents dans un corpus, mais plutôt d'en détecter des échantillons représentatifs et non déterminés *a priori*.

### 3.1. Détection des MDP

---

Tout d'abord, vérifions si les expressions mentionnées *supra* peuvent être considérées comme des MDP. Commençons par *comme on dit*, qui, une fois cherché dans le corpus frWaC, fournit 5 900 résultats. L'analyse des 590 premiers résultats (qui correspondent à 10% des résultats totaux) non ordonnés alphabétiquement est à notre avis éloquente : seuls 31 proverbes (soit environ 5% de l'échantillon examiné) peuvent être repérés sur la base de ce marqueur, alors que le reste des résultats ne contient pas de proverbes. En effet, d'après notre analyse, ce marqueur est plutôt un *marqueur médiatif*, étant donné qu'il signale l'emprunt à la parole d'autrui (cf. ANSCOMBRE 2011 : 13), comme le suggèrent les exemples suivants :

- (1) L'état s'en lave les mains, *comme on dit*.
- (2) La cerise sur le gâteau, *comme on dit*.
- (3) Cela risque de devenir très difficile de joindre les deux bouts *comme on dit* !
- (4) C'est « consumer-friendly » *comme on dit* dans le milieu.
- (5) C'est cool, *comme on dit* de nos jours!
- (6) Quand il faut y aller il faut y aller, *comme on dit*.

Dans certains cas (exemples 1-3), *comme on dit* s'applique à l'énonciation d'expressions figées et locutions, peu importe si d'origine cultivée (comme *s'en laver les mains*, issue de l'Évangile de Matthieu) ou populaire (*la cerise sur le gâteau, joindre les deux bouts*).<sup>[9]</sup> Dans d'autres cas, le même MDP peut signaler la présence de termes ou d'expressions particulièrement utilisés dans certains milieux ou dans des domaines spécifiques mais qui n'appartiennent pas forcément au langage figuré et qui ne présentent pas d'écart entre la signification compositionnelle et la signification globale (cf. exemples 4-5). En revanche, dans l'exemple (6), nous avons une phrase qui est une tautologie plutôt qu'un proverbe (cf. ANSCOMBRE 2008, 2012) mais qui est un bon exemple des va-et-vient entre les différents niveaux culturels d'une communauté (allant d'une culture plus élevée à une culture plus populaire ou « partagée » ; cf. GALISSON 1991 : 117). En effet, cette phrase a été

« recyclée » pour forger le titre d'un film américano-italien, *Quand il faut y aller, faut y aller*, connu en Italie sous le titre original de *Nati con la camicia*, qui repose à son tour sur l'expression figée *Nascere con la camicia* (correspondant à l'expression française *Être né avec une cuillère d'argent dans la bouche*).

De toute manière, nous soulignons que même si ce marqueur n'est pas spécifiquement un MDP (comme le montre l'analyse quantitative des résultats qu'il permet d'obtenir), nous pouvons remarquer, au niveau qualitatif, que le peu de proverbes qu'il génère sont encore en usage de nos jours, comme en témoignent les exemples (7-10). Il ne faut pas oublier, en effet, qu'en parémiographie, le risque de repérer des proverbes désuets est toujours à l'affût.

- (7) Mais comme on dit, *il ne faut jamais vendre la peau de l'ours avant de l'avoir tué*.
- (8) Et puis comme on dit si bien... *A chaque jour suffit sa peine!!!*
- (9) Mais bon comme on dit *quand \*ont aime on ne compte pas !* [erreur présente dans l'exemple original]
- (10) *La fin justifie les moyens* comme on dit.

Pour produire plus de résultats pertinents, il est possible d'utiliser l'autre MDP évoqué plus haut : *comme dit le proverbe*. D'un point de vue qualitatif, la quasi-totalité des résultats obtenus sont appropriés (nous comptons seulement quelques candidats-proverbes inventés ou repris à partir de citations ou d'aphorismes, comme « Une bonne grippe \*à nouvel an vous \*guérit de tous les maux pour le reste de l'année » ; « Les vacances, c'est tout un travail de repos très difficile »). En outre, nous remarquons que parmi les proverbes obtenus, 12 ne sont pas français mais locaux (nous avons trouvé, par exemple, un proverbe breton) ou étrangers (notamment chinois et africains). Au niveau quantitatif, nous observons que les résultats obtenus ne sont pas nombreux, étant donné que nous avons trouvé seulement 99 occurrences de ce MDP. Il est donc nécessaire de trouver d'autres MDP qui puissent permettre de parvenir à un nombre plus élevé de résultats pertinents. Selon une étude exploratoire que nous avons menée sur les corpus frWaC et itWaC, et consistant à chercher le mot *proverbe* et *proverbio* dans les corpus respectifs, les expressions qui introduisent les proverbes sont très variées. Voici un petit échantillon de MDP issus du corpus frWaC (et dont les équivalents italiens figurent dans le corpus itWaC) : *comme (le) dit/dirait/affirme/souligne/veut le/un proverbe ; comme dit un autre proverbe ; comme le dit si bien un/le fameux/célèbre/vieux proverbes ; à en croire le proverbe/s'il faut en croire le proverbe ; dit/affirme/assure/rappelle le/un proverbe*. Elles peuvent en outre concerner des questions (*tu connais/vous connaissez le proverbe... ? ; y a pas un proverbe du style/du coin qui dit : ... ?*) ou bien des incises (*un proverbe – plutôt vrai ! – dit que ; s'il est vrai, selon le proverbe, que ; le proverbe, vous le connaissez tous, c'est*). Dans ces expressions, nous observons, en outre, de petites coquilles, comme les suivantes : *le proverbe commun qui \*dict ; comme beaucoup j'ai eu à méditer la véracité de \*se proverbe*.

Les exemples précédents montrent que des MDP constitués par des séquences lexicales trop longues risquent de ne pas fournir un nombre satisfaisant de résultats. Il convient, donc, de repérer des MDP plus brefs et qui ne présentent que les éléments stables normalement présents au sein de tous les MDP repérés. Le procédé est assimilable au repérage du plus grand commun diviseur en arithmétique. D'après les exemples mentionnés plus haut, nos 'plus grands communs diviseurs' sont le mot *proverbe* pour le français et *proverbio* pour l'italien.

Toutefois, comme plusieurs auteurs (ANSCOMBRE 2000 ; FRANCESCHI 2008 ; MIEDER 1993) signalent que les étiquettes associées aux proverbes sont multiples, nous en avons pris en compte d'autres aussi, pour en vérifier la validité et la fiabilité. ANSCOMBRE (2000 : 8) rappelle que dans la catégorie du proverbe figurent également les *adages* et les *dictons*. FRANCESCHI (2008 : 364), de son côté, signale que le terme italien souvent utilisé pour désigner les proverbes est *detto*. *Adagio* est également utilisé en italien : d'un point de vue diachronique, il a été longtemps utilisé pour désigner les proverbes en Europe (pensons par exemple aux *Adagiorum Collectanea* d'Erasmus de Rotterdam, qui ne réunissaient pas seulement des adages, mais aussi des expressions figées, des locutions, des épithètes, etc.).

Les marqueurs italiens *adagio* et *detto* posent problème lorsqu'ils sont repérés dans des corpus. En effet, il est nécessaire de désambigüiser le nom commun *adagio* de l'adverbe *adagio* ('lentement', 'doucement'), et le nom commun *detto* du participe passé du verbe *dire* et de l'adjectif *detto* (qui correspond aux adjectifs 'ledit', 'audit', 'dudit', etc. selon la préposition qui le précède). Comme nous l'avons souligné *supra*, le corpus itWaC est doté d'un étiquetage morpho-syntaxique qui permet, par exemple, de distinguer entre nom, adjectif, adverbe, préposition, etc. Toutefois, nous avons remarqué que cet étiquetage n'est pas toujours fiable. Nous avons ainsi cherché les trois supposés MDP d'abord tous seuls, et ensuite en demandant au logiciel NoSketchEngine[10] de ne mettre en évidence que les MDP appartenant à la catégorie 'nom commun'. Enfin, nous avons combiné les trois MDP avec des articles, adjectifs ou des prépositions. Les résultats de cette étude sont synthétisés dans les deux tableaux suivants.

Type de recherche	Proverbe	Adage	Dicton
MDP simple (proverbe, adage, dicton)	3537	2364	1517
MDP avec indication de 'nom commun'	4991	2449	1684
Adjectif + MDP correspondant[11]	327	589	202
Article (défini ou indéfini) + MDP correspondant[12]	1737	1392	867
Préposition simple + MDP correspondant[13]	42	11	3

Préposition contractée + MDP correspondant[14]	259	0	58
Adjectifs démonstratifs + MDP correspondant[15]	296	266	144
Adjectifs possessifs + MDP correspondant[16]	47	24	24

**Tableau 1 : Différentes recherches des trois MDP (*proverbe, adage, dicton*) dans le corpus frWaC.**

Type de recherche	Proverbio	Adagio	Detto
MDP simple (proverbio, adagio, detto)	3706	4337	544195
MDP avec indication de 'nom commun'	4324	3324	0
Adjectif + MDP correspondant[17]	881	819	3088
Article (défini ou indéfini) + MDP correspondant[18]	1802	353	6186
Préposition simple + MDP correspondant[19]	97	20	17364
Préposition contractée + MDP correspondant[20]	211	86	4946
Adjectifs démonstratifs + MDP correspondant[21]	164	27	196
Adjectifs possessifs + MDP correspondant[22]	29	6	94

**Tableau 2 : Différentes recherches des trois MDP (*proverbio, adagio, detto*) dans le corpus itWaC.**

Une comparaison entre les données de deux tableaux précédents montre que les MDP les plus « prolifiques » en français et en italien, et ceux qui font parvenir à des résultats comparables du point de vue quantitatif, sont *proverbe* et *proverbio*. Les autres MDP qui pourraient être utilisés pour repérer des proverbes sont *adage* et *adagio*, quoiqu'en italien on remarque des valeurs nettement supérieures, pour toutes les catégories détectées, y compris pour la catégorie 'adagio avec indication de nom commun' et hormis deux catégories : 'article défini/indéfini + adage', qui est en italien presque 4 fois inférieure (peut-être en raison du caractère non obligatoire de l'article en italien), et 'adjectif démonstratif + adage'. Toutefois, si nous nous focalisons sur l'écart entre la recherche simple du mot *adagio*, le même mot limité à la seule catégorie 'nom commun' et les différentes recherches où ce mot est combiné avec d'autres particules appartenant aux catégories morpho-syntaxiques indiquées, nous nous rendons compte que l'étiquetage morpho-syntaxique offert par le système de traitement des deux corpus sélectionnés n'est pas fiable. En effet, si *adagio* 'nom commun' produit plus d'occurrences que 'adage nom commun', la somme des

occurrences obtenues en combinant les deux mots avec d'autres catégories morpho-syntaxiques est plus élevée en français qu'en italien : *adage* précédé d'adjectifs (qualificatifs, démonstratifs, possessifs), articles et prépositions (simples ou contractées) fait état de plus d'occurrences que les mêmes combinaisons avec le mot italien *adagio* (2 282 contre 1 311). Il est clair, donc, que certaines occurrences de *adagio* combiné avec l'indication 'nom commun' sont fausses. En outre, les occurrences de *detto* témoignent de l'imperfection de la machine dans l'attribution de la bonne catégorie morpho-syntaxique des mots. Nous remarquons, en effet, que la recherche de *detto* combiné avec l'indication 'nom commun' ne donne aucun résultat, ce qui est tout à fait impossible. La machine attribuée à ce mot systématiquement la catégorie adjectivale et verbale.

Ces études statistiques montrent que si l'on souhaite mener des recherches comparables entre le français et l'italien en utilisant le même MDP pour les deux langues, il convient d'utiliser les mots *proverbe* et *proverbio*.

Ces deux MDP n'assurent toutefois pas le repérage exclusif de proverbes français et italiens : 24 % des proverbes repérés dans frWaC et 17 % des proverbes repérés dans itWaC sont des proverbes non-français ou non-italiens. Il peut s'agir, par exemple, de proverbes locaux, régionaux ou bien étrangers : ce sont les adjectifs de provenance postposés aux MDP qui permettent d'analyser les différentes origines de ces proverbes.[23] Nous soulignons, en outre, que lorsqu'il n'y a aucun adjectif d'origine placé après les MDP, les proverbes qui suivent ne sont ni étrangers ni locaux.

Nous signalons, enfin, qu'un pourcentage comparable des MDP (14 à 15 %) repérés dans les deux corpus ne sont pas suivis de proverbes.

### 3.2. MDP et perception des locuteurs

---

Les syntagmes formés d'adjectifs suivis des MDP *proverbe* et *proverbio* peuvent signaler les caractéristiques des proverbes qui sont évoquées dans le discours ordinaire. En français, le proverbe est considéré comme[24] : « ancien » (15) ; « beau » (7) ; « bon » (7) ; « célèbre » (16) ; « fameux » (21) ; « joli » (8) ; « nouveau » (8) ; « sage » (4) ; « traditionnel » (2) ; « vieux » (133). En italien : « antico » (213) ; « bel », « bellissimo » (32) ; « buon » (3) ; « celebre » (7) ; « famoso » (98) ; « noto » (65) ; « nuovo » (4) ; « saggio » (21) ; « simpatico » (6) ; « vecchio » (285). L'adjectif qui est le plus fréquemment évoqué avec le nom *proverbe* et *proverbio* est le même : *vieux* en français et *vecchio* en italien, ce qui reflète bien la dimension diachronique du proverbe. Ensuite, ce sont les dimensions doxale et la notoriété du proverbe (l'une de ses caractéristiques définitoires) qui sont mises en évidence grâce aux adjectifs *célèbre*, *fameux* et *celebre*, *famoso*, *noto*.

Il est licite de se demander si l'unité phrastique est retenue par les locuteurs lambda. Cette question est cruciale dans la mesure où le seuil phrastique est l'une des principales caractéristiques définitoires des proverbes, étant donné qu'elle est à même de les distinguer

d'autres formes figées (telles que les expressions figées). Pour répondre à cette interrogation, une étude quantitative concernant les MDP *proverbe* et *proverbio* a été réalisée sur les corpus frWaC et itWaC.

Précisons d'abord que, par une recherche empirique réalisée sur ces corpus (LAMBERTINI 2022a), nous avons démontré que si la même phrase est répétée au moins 3 fois avec le MDP elle a plus de 50% de probabilité d'être un véritable proverbe connu de toute la communauté linguistique.[25] En outre, si les candidats-proverbes sont cherchés tout seuls (sans MDP) et qu'ils ont au moins 20 occurrences, les chances qu'ils soient de véritables proverbes sont quasiment de 100%.

Néanmoins, pour mieux comprendre si l'unité phrastique est une caractéristique retenue par les locuteurs ordinaires, il convient d'analyser les candidats-proverbes ayant une seule occurrence avec le MDP. Les résultats obtenus de l'analyse des deux corpus sont tout à fait comparables : environ 20 candidats-proverbes sur les 518 de frWaC et 20 sur les 449 d'itWaC ne sont pas des phrases. Tous les autres proverboïdes respectent la condition d'être au moins une phrase, ce qui souligne que cette caractéristique est effectivement considérée comme définitoire par les non-spécialistes. Voyons quelques exemples de non-phrases introduites par les MDP *proverbe* ou *proverbio* :

- (11) [...] comme le dit le proverbe [il] *a le cœur sur la main* il est prêt a tout pour aider des \*personne [...].
- (12) Comme dirait le proverbe : « Elle *tenait à vous le dire!* ».
- (13) C'è un proverbio che dice: “*Fare l'indiano...*” e nel Minnesota ci sono gli indiani...
- (14) Tu sai mio caro, il nostro vecchio proverbio; ‘*restare in piedi anche nel fango;*’ applichamocelo, sfidiamo i rimproveri dei fratelli per servire la causa comune.

Le respect du seuil phrastique des proverbes (qui ne peuvent pas être des constituants) est évident si l'on considère qu'il y a des expressions qui sont 'proverbialisées', à savoir qu'elles sont présentées comme des proverbes :

(15) Comme le dit le proverbe, *chacun voit midi à sa porte*.

Dans (15), l'expression figée *voir midi à sa porte* est utilisée en tant que proverbe : *Chacun voit midi à sa porte* respecte toutes les conditions définitoires du proverbe que nous avons énumérées au § 1. C'est un phénomène récurrent : prenons par exemple l'expression figée française (*Avoir*) *le beurre et l'argent du beurre* (16, 17) qui est utilisée très souvent comme un proverbe (18) :

- (16) Il pensait offrir *le beurre et l'argent du beurre* aux \*français [...].
- (17) C'est le contraire, on souhaite *le beurre et l'argent du beurre*.
- (18) De toute façon *on ne peut pas avoir le beurre et l'argent du beurre*.

D'après nos analyses, les locuteurs ordinaires confondent les proverbes non pas avec les expressions figées mais avec les aphorismes, les citations ou les slogans, comme dans l'exemple suivant :

(19) Car comme le dit le proverbe, « *en France on n'a pas de pétrole mais on a des idées* ».

Dans le corpus frWaC, ce slogan des années 1970 est considéré à trois reprises comme un « proverbe », un « adage » ou une « phrase célèbre ». En réalité, cette incertitude entre proverbes (ou phrases ON-sentencieuses) et aphorismes, citations ou slogans (ou phrases L-sentencieuses ; cf. ANSCOMBRE 2000) n'est pas anodine, car elle renvoie aux étapes que franchit une phrase pour devenir un proverbe : très souvent, les proverbes sont des citations dont on a oublié les origines (cf. VILLERS 2014 : 335-338 ; LAMBERTINI 2022a : 104).

#### **4. Les MDP dans les discours des plénières du PE**

---

Jusqu'ici, nous avons démontré que l'analyse des MDP, menée sur des corpus de référence, peut être précieuse pour étudier les proverbes et la manière dont ils sont perçus par les locuteurs ordinaires.

Parallèlement, il est intéressant d'analyser les proverbes à l'aide de corpus plus spécifiques. Prenons à titre d'exemple le domaine politico-institutionnel du PE, où les orateurs énoncent des proverbes (cf. LAMBERTINI 2022b), tout en sachant qu'ils ne s'adressent pas seulement aux électeurs de leurs pays mais aussi aux électeurs appartenant à d'autres communautés linguistico-culturelles de l'UE. Nous chercherons ainsi à comprendre si le domaine multilingue et multiculturel du PE peut influencer l'usage des MDP (au niveau quantitatif et qualitatif).

Pour mener cette étude, nous avons pris en examen les discours prononcés en italien et en français lors des séances plénières du PE de juillet 2014 à juin 2022 (cf. § 2). Comme les corpus sont plus restreints que les deux corpus de référence utilisés, et qu'ils sont plus spécifiques, étant donné qu'ils relèvent du discours politico-institutionnel en milieu européen, les MDP *proverbe* et *proverbio* fournissent peu de résultats : seuls 8 proverbes en français et 3 en italien. Nous avons ainsi décidé de contempler également les autres marqueurs mentionnés au § 3.1, ce qui nous a permis de repérer 15 proverbes dans les discours français et 7 proverbes dans les discours italiens. Pour élargir l'échantillon de proverbes à analyser, sans pour autant prétendre à l'exhaustivité, nous avons cherché dans les mêmes corpus les proverbes obtenus grâce à frWaC et à itWaC. Sur les 60 proverbes détectés dans frWaC, nous en avons repéré 31 ; sur les 83 proverbes détectés dans itWaC, nous en avons repéré 22. À ce titre, il est intéressant d'analyser les formules qui introduisent tous les proverbes, y compris ceux qui ne sont pas accompagnés de MDP. Ces formules peuvent être, entre autres : *on dit souvent chez nous* ; *je pense que* ; *nous avons constamment*

*rappelé au gouvernement britannique que ; qu'on le veuille ou pas et dicono che ; abbiamo già visto ; come si dice da noi ; e dire che ; è vero ; ricordate colleghi.* En général, c'est l'appartenance des proverbes à la *doxa* qui est mise en relief (et qui est réitérée par l'utilisation des verbes *rappeler* et *ricordare* dans les formules qui introduisent les proverbes), même si les orateurs soulignent parfois qu'il s'agit de proverbes de leur pays, voire de leur région (comme en témoignent ces expressions : *on dit souvent chez nous ; come si dice da noi*). Nous avons également trouvé des exemples (2 dans les discours français et 2 dans les discours italiens) où les orateurs ajoutent une indication de provenance même quand les proverbes ne sont ni étrangers ni locaux (*on a un dicton populaire en France ; un proverbe français disant que ; in Italia c'è un detto che dice ; c'è un detto in Italia che dice*), contrairement à ce que nous avons observé dans les corpus frWaC et itWaC (cf. § 3.1). En outre, les orateurs, qui se servent des proverbes pour des fins de persuasion, tendent à en souligner la vérité irréfutable (*qu'on le veuille ou pas ; è vero*) ou à les présenter comme des pensées originales, en utilisant des verbes d'opinion à la première personne du singulier[26] (*je pense que ; ritengo che*). Enfin, nous mettons en évidence un cas d'attribution du proverbe à un auteur précis (ce qui en réalité contrevient à la condition définitoire de phrase ON-sentencieuse) :

(20) *Molière écrivait: « Qui veut noyer son chien l'accuse de la rage ».* C'est le cas à Hong Kong avec la dramatique disparition du quotidien prodémocratie Apple Daily, accusé de « conspiration et collusion avec des pays étrangers ».[27]

Nous avons constaté la présence de cas pareils dans le discours politique, hors du PE, même en italien (cf. LAMBERTINI 2022a : 115). Notre hypothèse, qui devra bénéficier d'études plus spécifiques, est que la présentation des proverbes en tant que citations d'auteurs illustres a pour but d'augmenter leur force de persuasion, le risque implicite étant que les proverbes soient perçus comme des vérités trop populaires ou banales.

Il convient, maintenant, de se demander s'il y a des différences typologiques entre les proverbes introduits par les MDP et les proverbes énoncés sans aucune formule.

En analysant les proverbes accompagnés d'un MDP, on remarque qu'il s'agit en général de proverbes locaux, ou appartenant à des communautés linguistico-culturelles différentes par rapport à celle à laquelle appartient l'orateur. En revanche, lorsque les proverbes sont énoncés sans MDP, on trouve en général des proverbes très connus. Prenons le proverbe latin *Errare humanum est, perseverare autem diabolicum* et comparons la manière dont il est énoncé dans les discours en français et dans les discours en italien. Dans les premiers, il est associé au MDP *proverbe* (21), alors que dans les seconds il figure sans marqueur (22, 23), même lorsque la version latine est accompagnée de sa traduction en italien (24).

- (21) Comme dit le proverbe *errare humanum est, perseverare diabolicum, l'erreur est humaine, l'entêtement est diabolique.*
- (22) *E dire che sbagliare è umano ma perseverare è diabolico.*

- (23) Ora, *se errare è umano, perseverare è un'altra cosa*.
- (24) *Errare humanum est, perseverare autem diabolicum*, che tradotto letteralmente significa “commettere errori è umano, ma perseverare è diabolico”.

Notre hypothèse est qu’il peut y avoir une corrélation entre la présence ou l’absence du MDP et la fréquence des proverbes. En effet, dans le corpus frWaC, le proverbe, dans sa version complète,[28] ne présente que 9 occurrences, contre 135 du même proverbe dans le corpus itWaC. Les marqueurs associés à *L’erreur est humaine, persévérer est diabolique* dans le corpus frWaC semblent confirmer la non appartenance de ce proverbe au patrimoine proverbial en usage : nous observons des marqueurs médiatifs génériques[29] (*comme on dit ; on dit souvent que*) ou des MDP accompagnés d’adjectifs d’origine liés à l’Antiquité romaine (*dit l’adage romain ; selon le vieil adage latin*). En revanche, les marqueurs qui accompagnent l’énonciation de *Sbagliare/errare è umano, (ma) perseverare è diabolico*[30] sont de véritables MDP : *come dice il proverbio ; quel ben noto proverbio popolare ; il vecchio proverbio ; con un vecchio adagio popolare ; il detto popolare*.

Les données issues des discours en français et en italien prononcés au PE semblent étayer cette hypothèse : sur les 15 proverbes français introduits par un MDP, seuls 6 présentent plus de 20 occurrences dans le corpus frWaC et seul 1 sur les 7 introduits par un MDP dans les discours italiens présente plus de 20 occurrences. Ce proverbe, accompagné du MDP *detto* (inséré dans l’expression *c’è un detto in Italia che dice*), est : *Non si può avere la botte piena e la moglie ubriaca*, qui correspond à *On ne peut pas avoir le beurre et l’argent du beurre* mais qui est bien plus sexiste en italien (la traduction de son sens compositionnel est plus ou moins ‘On ne peut pas avoir un tonneau plein et une femme saoule’). L’utilisation du MDP pourrait avoir eu pour but la diminution de la charge sexiste de ce proverbe. Nos données semblent ainsi souligner que même dans un contexte multiculturel comme celui du PE, les orateurs préfèrent insérer les proverbes dans leurs discours sans se servir de MDP, sauf lorsqu’ils ont besoin d’en signaler le statut doxal ou citationnel.

## Conclusion

---

La principale conclusion que nous pouvons tirer de cette étude réside dans l’importance des MDP non seulement pour repérer des proverbes selon des approches dites *corpus-driven* dans de grands corpus, mais aussi pour avoir accès à la perception commune qu’ont les locuteurs sur les proverbes. En effet, après avoir illustré la démarche que nous avons suivie pour détecter les MDP les plus prolifiques en français (le mot *proverbe*) et en italien (le mot *proverbio*), voire les plus adéquats pour des études contrastives, dans deux corpus comparables (frWaC et itWaC), nous avons analysé les combinaisons lexicales concernant les MDP, pour mettre au clair la perception des locuteurs ordinaires sur les proverbes. Nous avons ainsi pu constater que l’unité phrastique est généralement respectée lorsqu’ils énoncent une séquence qu’ils appellent *proverbe* ou *proverbio*, même s’ils peuvent confondre les proverbes avec des aphorismes, des citations ou des slogans, et que la nature *doxale* et de notoriété des proverbes est généralement retenue par les locuteurs.

Ensuite, nous avons étudié les MDP sur deux corpus issus des séances plénières du PE qui ont eu lieu de 2014 à 2022. Nous avons pu observer, ainsi, que le milieu multiculturel et multilingue du PE n'influence que le choix des adjectifs de provenance suivant les MDP. En effet, contrairement à ce que nous avons constaté dans les corpus frWaC et itWaC, les orateurs tendent à signaler la provenance nationale des proverbes. Enfin, nous avons posé une hypothèse concernant l'utilisation des MDP dans les discours politico-institutionnels du PE : leur fréquence d'usage semble diminuer en fonction de la notoriété des proverbes énoncés et augmenter lorsque les orateurs souhaitent mettre en évidence le statut *doxal* des proverbes. Des recherches supplémentaires s'imposent, non seulement pour vérifier cette hypothèse, mais encore pour approfondir l'étude des MDP dans le discours politique ainsi que dans d'autres domaines.

## Bibliographie

---

ANSCOMBRE, Jean-Claude, « Parole proverbiale et structures métriques », *Langages*, n. 139, 2000, p. 6-26 ([http://www.persee.fr/web/revues/home/prescript/article/lgge\\_0458-726X\\_2000\\_num\\_34\\_139\\_2377](http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_2000_num_34_139_2377)).

ANSCOMBRE, Jean-Claude, « Les formes sentencieuses : peut-on traduire la sagesse populaire ? », *Meta*, n. 53(2), 2008, p. 253-268 (<https://www.erudit.org/fr/revues/meta/2008-v53-n2-meta2300/018518ar/>).

ANSCOMBRE, Jean-Claude, « L'introduction du pronom neutre dans les marqueurs médiatifs à verbe de dire de type *Comme dit le proverbe / Como dice el refrán* : étude sémantique contrative d'une contrainte polyphonique », *Langages*, n. 184, 2011, p. 13-34 (<https://www.cairn.info/revue-langages-2011-4-page-13.htm>).

ANSCOMBRE, Jean-Claude, « Le problème de l'antonymie dans le champ parémique », in ANSCOMBRE, Jean-Claude, RODRIGUEZ SOMOLINOS, Amalia, GÓMEZ-JORDANA FERARY, Sonia (dirs.), *Voix et marqueurs du discours : des connecteurs à l'argument d'autorité*, Lyon : ENS Éditions, 2012, p. 121-140.

BARONI, Marco, BERNARDINI, Silvia, FERRARESI, Adriano, ZANCHETTA, Eros, « The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora », *Language Resources and Evaluation*, n. 43, 2008, pp. 209-226 ([https://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky\\_2008.pdf](https://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf)).

BEACCO, Jean-Claude (éd.), Représentations métalinguistiques ordinaires et discours, *Langages*, n. 154, 2004.

CASADEI, Federica, *Metafore ed espressioni idiomatiche: uno studio semantico sull'italiano*, Bulzoni, Roma, 1996.

- CHIARI, Isabella, *Introduzione alla linguistica computazionale*, Rome, Bari, Laterza, 2007.
- DEIGNAN, Alice, « Searching for Metaphorical Patterns in Corpora », in BAKER, Paul (éd.). *Contemporary Corpus Linguistics*, Londres, New York, Continuum, 2009, p. 9-31
- FRANCESCHI, Temistocle, *Pagine sparse*, MANCINI Anna Maria, MINIATI Maria Valeria (éd.), Alessandria, Edizioni dell'Orso, 2008.
- GADET, Françoise, *Le français ordinaire*, Paris, Armand Colin, 1989.
- GALISSON, Robert, *De la langue à la culture par les mots*, Paris, CLE, 1991.
- GÓMEZ FERNÁNDEZ, Araceli, « La phraséologie dans le discours du dopage : un langage coloré », *Anales de Filología Francesa*, n. 26, 2018, p. 133-153.
- KLEIBER, Georges, « Sur le sens des proverbes », *Langages*, n. 139, 2000, p. 39-58 ([http://www.persee.fr/web/revues/home/prescript/article/lgge\\_0458-726X\\_2000\\_num\\_34\\_139\\_237](http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_2000_num_34_139_237) 9).
- LAMBERTINI, Vincenzo, *Che cos'è un proverbio*, Rome, Carocci, 2022a.
- LAMBERTINI, Vincenzo, « Les proverbes français et italiens dans la communication orale. Le cas des discours interprétés au Parlement européen », *Lingvisticae Investigationes*, n. 45(2), 2022b, p. 218-239.
- MIEDER, Wolfgang, *Proverbs Are Never Out of Season: Popular Wisdom in the Modern Age*, New York, Oxford, Oxford University Press, 1993.
- PAVEAU, Anne-Marie, « Les non-linguistes font-ils de la linguistique ? Une approche anti-éliminativiste des théories folk », *Pratiques*, n. 139-140, 2008, p. 93-109.
- SCHAPIRA, Charlotte, « Proverbe, proverbialisation et déproverbialisation », *Langages*, n. 139, 2000, p. 81-97 ([http://www.persee.fr/web/revues/home/prescript/article/lgge\\_0458-726X\\_2000\\_num\\_34\\_139\\_2382](http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_2000_num_34_139_2382)).
- TOGNINI-BONELLI, Elena, *Corpus Linguistics at Work*, Amsterdam, Philadelphia, John Benjamins, 2001.
- VILLERS, Damien, *Le Proverbe et les Genres Connexes. Domaines anglais et français*, Sarrebruck, Presses Académiques Francophones, 2014.
- VILLERS, Damien, « Marqueurs stylistiques : leur poids dans la définition et la genèse des proverbes », *Scolia. Revue de linguistique*, n. 31, 2017, p. 79-100.
-

[1] Faute de place, nous signalons que toutes les références web fournies dans cet article ont été vérifiées avant le 20 juin 2023. En outre, pour des raisons de concision et lisibilité, nous n’aurons pas recours à l’écriture inclusive. Lorsque nous parlerons de *locuteurs*, *orateurs*, *chercheurs* nous ferons références aux individus de tous genres.

[2] Les approches *corpus-driven*, ou guidées par les corpus, prévoient que le chercheur interroge un corpus, repère les données à analyser dans le corpus lui-même et met en évidence des *patterns* récurrents, qui lui serviront pour formuler ses théories (cf. TOGNINI-BONELLI 2001).

[3] Nous empruntons l’étiquette *discours ordinaire* aux études développées en analyse du discours française (ADF) sur les représentations métalinguistiques ordinaires (BEACCO 2004) et sur la *folk linguistique* (PAVEAU 2008), sans oublier les recherches sur le *français ordinaire*, qui peut se définir comme « le français familier, dont chacun est porteur dans son fonctionnement quotidien, dans le minimum de surveillance sociale : la langue de tous les jours » (GADET 1989 : 3). C’est notamment cette dernière acception de ‘langue de tous les jours’ que nous retenons ici lorsque nous parlons de *discours ordinaire* et de *locuteur ordinaire*.

[4] Dans les exemples montrés, nous ne changerons même pas la ponctuation. Les seules modifications que nous y apporterons concerneront la mise en relief de certaines parties, qui seront en italique.

[5] Pour plus d’informations concernant leur construction, nous renvoyons à BARONI *et al.* (2008).

[6] <https://www.clarin.si/noske/index-en.html>.

[7] Disponible à cette adresse : <https://www.europarl.europa.eu/plenary/fr/debates-video.html#sidesForm>.

[8] En raison de la conception très simple de cette plateforme, qui a été conçue pour une utilisation rapide de la part des citoyens européens plutôt que pour des fins d’analyse linguistique ou statistique, il nous est actuellement impossible de savoir combien de discours (ou combien de mots) en français et en italien ont été prononcés au cours de ces plénières.

[9] Il est difficile (sinon impossible, parfois) de retracer les origines des proverbes et des expressions figées. En outre, à la base d’expressions dites cultivées figurent très souvent des expressions déjà utilisées en milieu populaire et vice-versa.

[10] Voir § 2.

[11] Selon le langage de NoSketchEngine : [tag="A"] “proverbe” ; [tag="A"] “adage” ; [tag="A"] “dicton”.

[12] Selon le langage de NoSketchEngine : [tag="T"] "proverbe" ; [tag="T"] "adage" ; : [tag="T"] "dicton".

[13] Selon le langage de NoSketchEngine : [tag="Ss"] "proverbe" ; [tag="Ss"] "adage" ; [tag="Ss"] "dicton".

[14] Selon le langage de NoSketchEngine : [tag="Sc"] "proverbe" ; [tag="Sc"] "adage" ; [tag="Sc"] "dicton".

[15] Selon le langage de NoSketchEngine : [tag="Pd"] "proverbe" ; [tag="Pd"] "adage" ; [tag="Pd"] "dicton".

[16] Selon le langage de NoSketchEngine : [tag="Ts"] "proverbe" ; : [tag="Ts"] "adage" ; : [tag="Ts"] "dicton".

[17] Selon le langage de NoSketchEngine : [tag="A"] "proverbio" ; [tag="A"] "adagio" ; [tag="A"] "detto".

[18] Selon le langage de NoSketchEngine : [tag="T"] "proverbio" ; [tag="T"] "adagio" ; : [tag="T"] "detto".

[19] Selon le langage de NoSketchEngine : [tag="Sp"] "proverbio" ; [tag="Sp"] "adagio" ; [tag="Sp"] "detto".

[20] Selon le langage de NoSketchEngine : [tag="Spt"] "proverbio" ; [tag="Spt"] "adagio" ; [tag="Spt"] "detto".

[21] Selon le langage de NoSketchEngine : [tag="Dd"] "proverbio" ; [tag="Dd"] "adagio" ; [tag="Dd"] "detto".

[22] Selon le langage de NoSketchEngine : [tag="Ds"] "proverbio" ; : [tag="Ds"] "adagio" ; : [tag="Ds"] "detto".

[23] L'analyse de ces adjectifs a été présentée dans Lambertini (2022 : 106-108). Ici, nous rappelons brièvement que la provenance la plus récurrente des proverbes non-français et non-italiens énoncés dans les deux corpus est représentée par l'Asie (et en particulier par la Chine). Toutefois, il y a une différence remarquable : à la deuxième place des proverbes non-français nous trouvons des proverbes africains ou des anciennes colonies de France, alors que pour les proverbes non-italiens nous observons des proverbes européens et ensuite des proverbes régionaux ou locaux. Cet écart est sans aucun doute lié à l'histoire nationale différente de la France et de l'Italie.

[24] Les adjectifs sont rangés par ordre alphabétique et le nombre d'occurrences est indiqué entre parenthèses.

[25] Nous avons en effet observé que les locuteurs forgent souvent des proverbes (*ex novo* ou en détournant des proverbes existants). Ces « proverboïdes » (sur ce terme, voir SCHAPIRA 2000 : 83) ne peuvent pas être considérés comme de véritables proverbes puisqu'ils ne satisfont pas la condition définitoire de notoriété.

[26] Nous renvoyons à SCHAPIRA (2000) pour une analyse plus approfondie sur les verbes d'opinion introduisant des proverbes et sur les considérations liées au phénomène de la déproverbialisation lorsque les proverbes sont introduits par ces verbes. Ces derniers entraîneraient ainsi une « restitution du proverbe au discours libre, ou en d'autres termes, une occurrence ponctuelle de défigement discursif, dans laquelle le proverbe redevient une simple proposition comme s'il avait été créé librement en discours » (SCHAPIRA 2000 : 93).

[27] [https://www.europarl.europa.eu/doceo/document/CRE-9-2021-07-08-ITM-009-02\\_FR.html](https://www.europarl.europa.eu/doceo/document/CRE-9-2021-07-08-ITM-009-02_FR.html).

[28] Nous précisons que 'l'erreur est humaine' se répète 309 fois dans le corpus frWaC. Il s'agit donc d'une séquence fréquente, relevant du bon sens, issue probablement du proverbe, mais qui a perdu son statut proverbial.

[29] Cf. ANSCOMBRE 2011.

[30] On peut retrouver les différentes variantes de ce proverbe en utilisant cette séquence : ".\*are" [{"1} "umano" [{"0,5} "diabol.\*".

---

Per citare questo articolo:

Vincenzo LAMBERTINI, « Les marqueurs de proverbes : analyse contrastive (français-italien) basée sur corpus », *Repères DoRiF*, numéro hors-série *Varia*, DoRiF Università, Roma, febbraio 2024, <https://www.dorif.it/reperes/vincenzo-lambertini-les-marqueurs-de-proverbes-analyse-contrastive-francais-italien-basee-sur-corpus/>

ISSN 2281-3020

Quest'opera è distribuita con Licenza Creative Commons Attribuzione – Non commerciale – Non opere derivate 3.0 Italia.

