

La proximité linguistique entre le français et l'italien : le cas des collocations

Lorenza Russo

Département de linguistique

Université de Genève

<Lorenza.Russo@etu.unige.ch>

Résumé

Éloy (2004: 7) définit la proximité linguistique comme le fait qu'entre un idiome A et un idiome B existent des ressemblances phonologiques, lexicales, morphologiques, syntaxiques, voire une différence seulement partielle, ou encore une zone d'intersection ou de coïncidence plus ou moins importante. Autrement dit, on peut définir les langues proches comme des langues typologiquement et génétiquement apparentées (Dabène & Degache 1996) qui partagent des caractéristiques lexicales, morphologiques et syntaxiques (Hédiard, 1989; Arcaini, 2000).

La proximité linguistique entre le français et l'italien semble une notion largement acceptée, mais différentes études abordent cette notion sans pour autant clairement la définir (Cartoni, 2008: 77). Dans cet article, nous analysons la proximité linguistique entre le français et l'italien du point de vue des expressions polylexicales et, plus particulièrement, des collocations. Bien que de nombreux aspects des collocations aient fait l'objet de recherches — pensons, par exemple, aux efforts en lexicographie pour les cataloguer, recenser et interpréter tout comme aux efforts en traitement automatique du langage pour les reconnaître, extraire, analyser et traduire —, à notre connaissance, personne n'a jusqu'à maintenant examiné comment la proximité linguistique se manifeste dans les collocations et dans quelle mesure. Nous discutons les détails de deux études que nous avons menées sur le couple de langues proches français-italien et nous mettons les résultats obtenus en comparaison avec ceux obtenus pour le couple de langues éloignées français-anglais.

Mots clés: Proximité linguistique entre le français et l'italien, collocations, traduction automatique, collocations transparentes.

1. Pourquoi les collocations

De nombreux chercheurs ont essayé de définir la nature des expressions polylexicales: Sag et al. (2002: 1) les définissent comme des « idiosyncratic interpretations that cross word boundaries (or spaces) » ; Mel'ëuk (2013: 129) les définit comme des « énoncés multilexémiques non libres »,

ces derniers étant une « configuration de deux ou plus lexèmes syntaxiquement liés »; Wehrli (2013 : 38) les définit comme des « complex units » et, plus en détail, comme des « fixed or semi-fixed lexical or syntactic units made up of several words » (Wehrli, 2001 : 273); Calzolari et al. (2003, 2002 : 1934) définissent une expression polylexicale comme « a sequence of words that acts as a single unit at some level of linguistic analysis »; Monti et al. (2018 : 1) affirment que « multi-words units or expressions are meaningful lexical units made of two or more words in which at least one of them is restricted by linguistic conventions in the sense that it's not freely chosen ». On pourrait résumer toutes ces définitions, en disant que les expressions polylexicales sont des unités linguistiques qui peuvent être décomposées en plusieurs lexèmes et qui présentent une idiomaticité au niveau lexical, syntaxique, sémantique et/ou pragmatique.

On peut distinguer plusieurs types d'expressions polylexicales¹ :

- 1) les mots composés — unités de catégorie lexicale — dont tous les constituants sont contigus et dont la signification n'est pas nécessairement compositionnelle², comme dans le cas de *c'est-à-dire*, *d'ores et déjà*, *chemin de fer* ;
- 2) les expressions idiomatiques — unités de catégorie syntaxique — dont la signification est habituellement non compositionnelle et qui peuvent montrer un niveau variable de flexibilité syntaxique. Pensons, par exemple, à certaines expressions telle que *rendre un hommage à quelqu'un* qui accepte la passivation (*Un hommage lui a été rendu*) alors que d'autres, telles que *casser sa pipe* ou *faire peur* ne l'acceptent pas (**La peur lui a été faite*) ;
- 3) les collocations, qui sont des associations conventionnelles³ de mots

1 La distinction que nous donnons ici vise à donner une première différenciation entre les expressions polylexicales. Les articles de Wehrli (2001), Nerima et al. (2006) et Wehrli (2013), par exemple, suivent les schémas que nous proposons, alors que d'autres articles — comme, celui de Sag et al. (2002) ou de Constant et al. (2017) — donnent une classification plus complexe des expressions polylexicales.

2 Il faut considérer la compositionnalité comme la « possibilité de calculer le sens d'un prédicat d'après l'organisation d'éléments qui le constituent » (Leoni de Léon, 2008 : 14). Le contraire de la compositionnalité est la non-compositionnalité ou l'opacité.

3 Par « association conventionnelle » nous entendons le fait qu'un locuteur natif

pas forcément contigus dont la signification est généralement transparente. Ces associations constituent des unités syntaxiques telles que des syntagmes verbaux (VP) — comme *caresser l'espoir, apporter un soutien*, etc. — ou des syntagmes nominaux (NP), comme *haute technologie, chemin de croix*, etc.

Plus en détail, les caractéristiques des collocations sur lesquelles la plupart des chercheurs se trouvent d'accord sont les suivantes (Seretan, 2011 : 14-17) :

- les collocations sont des unités préfabriquées du moment qu'elles sont apprises telles quelles par le locuteur qui d'habitude n'est pas conscient des mots qui les constituent ou des règles linguistiques derrière leur formation ;
- elles sont des combinaisons typiques et caractéristiques de mots ;
- elles sont imprévisibles, elles ne peuvent pas être déduites de l'application de règles linguistiques ;
- elles sont récurrentes et formées par deux ou plusieurs mots.

En effet, les collocations peuvent être soi-disant binaires — composées par deux mots, comme *haute technologie* — mais aussi composées par plus que trois éléments. Quelques exemples de collocations complexes sont *arme de destruction massive* (Nerima et al., 2010 : 634), *ligne de partage des eaux*, *système de gestion de base de données* et encore *force de maintien de la paix* ou *organisation de protection de l'environnement* (Wehrli et al., 2010 : 30).

Nous avons choisi de focaliser notre attention sur les collocations parce que, comme l'affirme Mel'čuk (2003), elles sont omniprésentes. En outre, elles se révèlent intéressantes du moment qu'elles peuvent subir différentes transformations syntaxiques (Nerima et al., 2006 : 97). Si l'on considère la collocation *décerner un prix*, elle peut subir les transformations suivantes :

- a) modification adjectivale : *décerner un important prix* ;
- b) passivation : *le Prix Nobel de la paix 2005 a été décerné hier* ;
- c) relativisation : *le prix qui lui a été décerné l'année passée* ;
- d) clivage : *c'est le prix le plus important qui sera décerné demain soir* ;

reconnaît ces combinaisons de mots comme la manière correcte d'exprimer un concept particulier. Si un mot dans l'association est remplacé par un synonyme, la collocation n'est plus perçue comme naturelle ou complètement compréhensible (Wehrli, 2001 ; Ramish, 2015).

- e) interrogation : *quels prix ont été décernés lors de ce festival ?* ;
- f) pronominalisation : *ils ont été décernés* (quand on parle de prix).

D'un point de vue computationnel, l'interprétation correcte des collocations est cruciale pour différentes tâches — de l'étiquetage morphosyntaxique à la désambiguïsation lexicale, de l'extraction et recherche d'information au résumé automatique de textes et à la traduction automatique (Monti et al., 2018). Surtout dans ce dernier cas, si un système de traduction automatique n'arrive pas à correctement identifier et analyser une collocation, il y a de bonnes chances qu'il n'arrivera pas non plus à correctement la traduire. Et cela est souvent la cause pour des traductions automatiques incompréhensibles ou erronées (Seretan, 2015 ; Barreiro et al., 2013 ; Monti et al., 2011).

2. Notre hypothèse

Dans l'optique de la traduction automatique, Wehrli et al. (2010 : 34) affirment qu'une traduction littérale d'une collocation est rarement la traduction la plus appropriée. De leur côté, Monti et al. (2018 : 2) citent les deux tests linguistiques de Ramish (2015) qui visent à détecter si une combinaison de mots est une expression polylexicale ou pas, le deuxième test mentionné consistant à traduire littéralement une expression : si la traduction obtenue semble bizarre, pas naturelle ou même incorrecte, alors l'expression dans la langue source est bien une expression polylexicale. D'un autre côté, Seretan (2015) montre aussi que le nombre de traductions automatiques erronées de collocations est significatif et que l'on pourrait obtenir de bien meilleurs résultats de traduction automatique si les collocations pouvaient être traduites correctement.

La question que nous nous posons ici est de comprendre jusqu'à quel point tout ce qu'on a affirmé ci-dessus peut s'appliquer à des langues proches. Si l'on considère, par exemple, la collocation *jeter l'éponge*, elle peut poser des problèmes de traduction automatique vers l'anglais, pouvant être traduite soit littéralement en tant que *to throw the sponge*, soit par *to resign* (Constant et al., 2017 : 871). Mais ce problème ne se pose pas quand on la traduit vers l'italien, du moment que *gettare la spugna* se comporte d'une manière tout à fait équivalente à la collocation française. Peut-on donc affirmer que la proximité linguistique entre le français et l'italien se manifeste aussi lors de la traduction des collocations ? Pour répondre à cette question, nous faisons l'hypothèse que si la traduction d'une collocation peut être littérale, cela

signifie que la collocation en français se comporte syntaxiquement de la même manière que la collocation en italien, en ayant, par exemple, la possibilité de subir (ou pas) les mêmes transformations de passivation, clivage, pronominalisation, etc. Dans la suite de ce travail, nous appelons ces cas *collocations transparentes*.

3. Étude préliminaire

Pour mieux comprendre le phénomène des collocations dans les langues proches, nous avons extrait de la base de données lexicale d'Its-2 (Wehrli et al., 2009) des collocations tout en suivant, comme critère d'extraction, le type de collocation indiqué dans la base de données. La base de données d'Its-2 est formée par plusieurs types de lexiques dont, entre autres, le lexique des collocations (Wehrli & Nerima 2009). Ce lexique donne au système de traduction automatique Its-2 des informations sur la configuration syntaxique d'une collocation, sur une éventuelle préposition présente dans l'expression ainsi que sur certains traits de figement, comme par exemple si la collocation présente ou pas un complément figé⁴. Cela est ce que nous appelons *type de collocation*.

Nous avons extrait cinquante collocations en français ainsi que leurs traductions en italien et en anglais pour huit types de collocations, pour un total de 400 collocations. La Table 1 montre en détail les types extraits.

Table 1 : Types de collocations extraites de la base de données lexicale.

Type de collocation	Exemple de collocation
nom-préposition-nom	<i>excursion en bateau</i>
nom-adjectif	<i>spéculation immobilière</i>
verbe-objet	<i>mettre un frein</i>
nom-nom	<i>assurance bagages</i>
verbe-préposition-argument	<i>mettre au courant</i>
verbe-adjectif	<i>tomber malade</i>
verbe-verbe	<i>laisser passer</i>
verbe-adverbe	<i>conserver précieusement</i>

Ce que nous avons remarqué lors d'un travail comparatif sur les 400 collocations que nous avons extrait peut être résumé en deux points :

⁴ Ce dernier est le cas, par exemple, de *avoir une peur bleue*. On dit que la collocation est figée parce qu'il n'est pas possible de dire, par exemple, *avoir une peur très bleue*.

1. dans la majorité des cas analysés, la collocation en italien est une traduction littérale de celle en français ;
2. si l'on considère les collocations en anglais, l'anglais présente une tendance à traduire par un seul mot la collocation en français.

Si l'on revient plus en détail sur notre affirmation au point 1, elle se révèle vraie pour tous les types de collocations pris en considération, y compris les collocations qui comportent des verbes. De plus, ces dernières présentent aussi la même sous-catégorisation en français tout comme en italien. À la Table 2 nous donnons un exemple de collocation traduite littéralement entre le français et l'italien pour tous les types de collocations qui font l'objet de cette étude.

Table 2 : Exemples de collocations françaises traduites littéralement en italien.

Type de collocation	Exemple de collocation en français	Traduction en italien
nom-préposition-nom	<i>date de parution</i>	<i>data di pubblicazione</i>
nom-adjectif	<i>nœud routier</i>	<i>nodo stradale</i>
verbe-objet	<i>faire une tentative</i>	<i>fare un tentativo</i>
nom-nom	<i>mère poule</i>	<i>mamma chioccia</i>
verbe-préposition-argument	<i>être en désaccord</i>	<i>essere in disaccordo</i>
verbe-adjectif	<i>se tenir prêt</i>	<i>tenersi pronto</i>
verbe-verbe	<i>faire taire</i>	<i>far tacere</i>
verbe-adverbe	<i>entrer furtivement</i>	<i>entrare furtivamente</i>

Sur les 400 collocations analysées, seulement trois exceptions ont été trouvées dans nos listes (Table 3). Dans ces trois cas, la sous-catégorisation des verbes en français est différente de celle des verbes en italien.

Table 3 : Différences dans la sous-catégorisation des verbes français et italien dans des collocations de type verbe-verbe.

Type de collocation	Français	Italien
verbe-verbe	<i>aller chercher</i>	<i>andare a cercare</i>
verbe-verbe	<i>venir chercher</i>	<i>venire a cercare</i>
verbe-adverbe	<i>regarder bouche bée</i>	<i>guardare a bocca aperta</i>

Cela indique que lors de l'utilisation de verbes tels qu'*aller* et *venir* dans des structures qui présupposent un verbe suivi par un syntagme verbal à

L'infinitif, l'italien se différencie du français en utilisant une préposition entre les deux verbes. Cependant, même si l'on tient compte de ces trois exceptions, le nombre de collocations qui présentent une traduction littérale entre le français et l'italien reste élevé. De plus, si l'on considère les collocations extraites du point de vue syntaxique, il nous semble que les collocations que nous avons extraites présentent toutes — sauf les trois cas mentionnés ci-dessus — une structure syntaxique identique et une égale possibilité de modifications syntaxiques. Considérons, par exemple, la collocation *négoçier un accord* : dans la Table 4 nous montrons comment cette collocation en français peut subir les mêmes transformations syntaxiques que son équivalent en italien.

Table 4 : Similitudes entre des collocations en français et en italien.

Type de transformation	Français	Italien
modification adjectivale	<i>Il a voulu négocier un très important accord.</i>	<i>Ha voluto negoziare un accordo molto importante.</i>
passivation	<i>Un accord a été négocié hier.</i>	<i>Un accordo è stato negoziato ieri.</i>
relativisation	<i>L'accord qu'il a négocié n'a pas été respecté.</i>	<i>L'accordo che ha negoziato non è stato rispettato.</i>
clivage	<i>C'est un accord important à négocier.</i>	<i>È un accordo importante da negoziare.</i>
interrogation	<i>Quel accord ont-ils négocié?</i>	<i>Quale accordo hanno negoziato?</i>
pronominalisation	<i>Cet accord est ancien. Il a été négocié au siècle dernier.</i>	<i>Questo accordo è vecchio. È stato negoziato il secolo scorso.</i>

Pour revenir à notre affirmation au point 2 — c'est-à-dire que l'anglais présente une tendance à traduire par un seul mot la collocation en français —, cela se révèle vrai non seulement pour les collocations du type nom-nom ou nom-préposition-nom, mais aussi lorsqu'il y a des collocations avec des verbes (surtout dans le cas de collocations verbe-préposition-argument) (Table 5).

Table 5 : Différences dans la traduction des collocations du français vers l'anglais.

Type de collocation	Français	Italien	Anglais
nom-adjectif	<i>propriétaire terrien</i>	<i>proprietario terriero</i>	<i>landowner</i>
nom-adjectif	<i>chasseur bombar-dier</i>	<i>caccia bombardiere</i>	<i>fighter-bomber</i>
nom-préposition-nom	<i>nombre de lecteurs</i>	<i>numero di lettori</i>	<i>readership</i>
nom-préposition-nom	<i>coup de fusil</i>	<i>colpo di pistola</i>	<i>gunshot</i>
verbe-préposition-argument	<i>partir en croisade</i>	<i>partire in crociata</i>	<i>crusade</i>
verbe-préposition-argument	<i>mettre à la retraite</i>	<i>mettere in pensione</i>	<i>to retire</i>
verbe-préposition-argument	<i>atterrir sur le ventre</i>	<i>atterrare sulla pancia</i>	<i>to belly-land</i>

Au vu de ces résultats, il nous semble que la proximité entre le français et l'italien se traduit par une possibilité de traduire littéralement un nombre élevé de collocations. Pour ce qui est du couple de langues éloignées français-anglais, une traduction littérale des collocations entre ces deux langues n'est pas souhaitable et semble plus risquée qu'entre français-italien.

Pour appuyer avec plus de force le fait qu'une traduction littérale reflète une proximité syntaxique entre le français et l'italien (Table 4), nous avons mené une deuxième étude, d'une plus grande ampleur, sur toutes les collocations disponibles dans la base de données lexicales d'Its-2.

4. Deuxième étude

Afin d'extraire automatiquement parmi toutes les collocations disponibles dans la base de données lexicales d'Its-2⁵ celles qui présentent une traduction littérale, nous avons défini le critère de la *traduction transparente*. Ce critère se base sur la logique suivante : une collocation dans une langue source se traduit de manière transparente dans une langue cible si et seulement si la traduction de ses parties — c'est-à-dire des deux lexèmes formant la collocation — donne le même résultat que la traduction de la collocation entière. En guise d'exemple, considérons le cas de la collocation *légitime*

5 Le nombre total de collocations présentes dans les lexiques d'Its-2 pour les langues objets de notre étude est mentionné à la Table 6. Nous remercions Luka Nerima pour son aide lors de l'extraction des collocations de la base de données.

défense en français : les deux lexèmes qui composent la collocation — *légitime* et *défense* — se traduisent respectivement en italien par *legittima* et par *difesa* ; la collocation tout entière (*légitime défense*), se traduit en italien par *legittima difesa*, c'est-à-dire par les mêmes mots qui sont la traduction directe des deux lexèmes composant la collocation. Cette collocation — et tous les cas similaires à celui-ci — fera partie des collocations extraites de la base de données.

Ce critère de traduction transparente est bien utile pour extraire les collocations dans la base bilingue français-italien (Wehrli & Nerima 2009), mais il risque de se révéler trop strict lors de l'extraction de collocations de la base de données bilingue français-anglais. Dans ce deuxième cas, nous avons donc autorisé l'inversion de l'ordre des deux lexèmes formant la collocation, car, par exemple, dans une collocation de type nom-nom on trouvera en français la tête lexicale en première position, alors qu'en anglais elle sera en deuxième position. Considérons, en guise d'exemple, le cas de la collocation française *assurance chômage*. La tête lexicale de cette collocation est bien *assurance*, mais cette collocation se traduit en anglais par *unemployment insurance*, où l'équivalent anglais de la tête lexicale française est bien *insurance*, qui se trouve en deuxième position. Accepter l'inversion en anglais n'influence pas les résultats du moment que le critère de transparence continue d'être appliqué : notre collocation française est constituée par les lexèmes *assurance* et *chômage*, lesquels se traduisent en anglais par les lexèmes *insurance* et *unemployment*.

Pour finir, le dernier critère utilisé pour l'extraction des collocations a été basé sur la préférence de traduction présente dans le lexique bilingue (Wehrli & Nerima 2009) : le lexique peut contenir plusieurs traductions et elles sont ordonnées par ordre de fréquence (de 6 à 1), la traduction la plus courante ayant la fréquence la plus haute. Pour notre extraction, nous n'avons considéré que les correspondances bilingues qui ont la traduction la plus courante aussi bien pour les lexèmes que pour les collocations. Pour mieux expliquer ce dernier critère, considérons la collocation *maison de campagne* en français. Imaginons que dans le lexique bilingue des lexèmes français-italien, *maison* soit traduite de préférence par le mot italien *appartamento*, et *campagne* par le mot *campagna*. Imaginons aussi que dans le lexique bilingue des collocations *maison de campagne* soit traduite par *casa di campagna* en italien. D'après notre critère exprimé auparavant, cette collocation ne va pas rentrer dans notre liste de collocations qui présentent une traduction transparente entre le français et l'italien du moment que la traduction du lexème *maison* tout seul et celle du lexème *maison* dans la collocation diffèrent. Vice versa, si la

traduction de préférence du lexème *maison* était *casa*, alors cette collocation aurait fait partie de notre liste, puisque la traduction aurait été la même dans les deux cas.

Les résultats obtenus de l'extraction des collocations montrent qu'il y a 50% de collocations transparentes entre le français et l'italien alors que les collocations transparentes entre le français et l'anglais ne constituent que 25% du total des collocations disponibles entre ces deux langues (Table 6). Le nombre total de collocations disponibles pour les deux couples de langues étant bien différent – le français et l'italien comptent 2'346 entrées (50% desquelles sont transparentes) alors que le français et l'anglais en comptent 7'760 (25% desquelles sont transparentes) –, nous avons utilisé le calcul statistique en χ^2 pour comprendre si la différence entre le nombre de collocations transparentes en français-italien et celui en français anglais est statistiquement significative⁶. Les résultats de nos calculs montrent que la différence est significative ($\chi^2 = 542,823$, $p < 0.001$). Cela peut être vu comme une confirmation de ce qu'on a déjà remarqué dans l'étude précédente: les collocations entre les langues proches français-italien présentent des traits en communs et, dans un nombre considérable de cas, elles peuvent être traduites littéralement.

Table 6 : Nombre de collocations transparentes entre français et italien ainsi qu'entre français et anglais dans la base de données lexicale.

Païres de langues	Collocations disponibles	Collocations transparentes	% collocations transparentes
FR — IT	2'346	1'186	50.1%
FR — EN	7'760	1'952	25.2%

Si l'on considère les résultats obtenus du point de vue syntaxique, les collocations transparentes trouvées montrent une flexibilité similaire à des transformations syntaxiques. Pour ce qui est, par exemple, des collocations verbes-objets, nous avons constaté que les mêmes transformations syntaxiques sont possibles pour la paire de langues français-italien (Table 7) tout comme pour la paire de langues français-anglais (Table 8).

6 Le χ^2 est un test statistique qui permet de déterminer si la différence entre deux fréquences est due au hasard ou est statistiquement significative. Si la différence est réduite, on attribue cela au hasard ; si la différence est grande, elle est considérée comme significative (Greenwood et Nikulin, 1996).

Table 7 : Exemple de similitudes syntaxiques de collocations transparentes en français et en italien.

Type de transformation	Français	Italien
modification adjectivale	<i>Jean a intenté un nouveau procès contre elle.</i>	<i>Jean ha intentato un nuovo processo contro di lei.</i>
passivation	<i>Un procès a été intenté par Jean.</i>	<i>Un processo è stato intentato da Jean.</i>
relativisation	<i>Le procès qui a été intenté par Jean est fini.</i>	<i>Il processo che è stato intentato da Jean è finito.</i>
clivage	<i>C'est le procès intenté par Jean.</i>	<i>È il processo intentato da Jean.</i>
interrogation	<i>Quel procès ont-ils intenté contre Jean?</i>	<i>Che processo hanno intentato contro Jean?</i>
pronominalisation	<i>Ces procès sont nombreux. Ils ont été intentés contre Jean.</i>	<i>Questi processi sono numerosi. Sono intentati contro Jean.</i>

Table 8 : Exemple de similitudes syntaxiques de collocations transparentes en français et en anglais.

Type de transformation	Français	Anglais
modification adjectivale	<i>Ils ont appliqué une nouvelle loi.</i>	<i>They enforced a new law.</i>
passivation	<i>La loi a été appliquée dans tous les cas.</i>	<i>The law has been enforced for each case.</i>
relativisation	<i>La loi qui a été appliquée est récente.</i>	<i>The law that has been enforced is a recent one.</i>
clivage	<i>C'est une nouvelle loi qui a été appliquée.</i>	<i>It's a new law that has been enforced.</i>
interrogation	<i>Quelle loi a été appliquée ?</i>	<i>Which law has been enforced?</i>
pronominalisation	<i>Les lois pénales sont strictes. Elles ont été appliquées par ce juge.</i>	<i>Penal laws are hard. They have been enforced by this judge.</i>

Le même phénomène se reproduit aussi pour les autres types de collocations. Dans la Table 9 nous avons répertorié différents types de collocations dans les trois langues objet de cette étude et toutes présentent la même capacité à des transformations syntaxiques telles que la modification adjectivale ou adverbiale. La Table 9 doit être lue comme suit : la partie de la collocation

qui est en gras est la collocation dans sa forme standard, alors que la partie en italique est une des transformations syntaxiques que la collocation pourrait accepter.

Table 9 : Exemples de transformations syntaxiques similaires des collocations transparentes en français, italien et anglais.

Type de collocation	Français	Italien	Anglais
nom-adj	<i>une longue</i> année sabbatique	<i>un lungo anno</i> sabbatico	<i>a long</i> sabbatical year
verbe-prep-arg	aller <i>volontiers</i> à l'école	andare <i>volentieri</i> a scuola	to go <i>willingly</i> to school
nom-prep-nom	<i>énorme</i> abus de pouvoir	<i>enorme</i> abuso di potere	<i>huge</i> abuse of power
nom-nom	industrie automobile <i>allemande</i>	industria automobile <i>tedesca</i>	<i>german</i> automotive industry
adj-prep-nom	voué à <i>un énorme</i> échec	destinato a <i>un enorme</i> fallimento	doomed to <i>a huge</i> failure
verbe-adj	tomber <i>à nouveau</i> malade	cadere <i>di nuovo</i> malato	to fall sick <i>again</i>
prep-nom	sur <i>gentile</i> demande	su <i>gentile</i> domanda	on <i>kind</i> request
verbe-adv	croître <i>trop</i> rapidement	crescere <i>troppo</i> rapidamente	to grow <i>too</i> fast
pp-modifieur	de longueur <i>très</i> variable	di lunghezza <i>molto</i> variabile	of <i>very</i> varying length
adv-adj	politiquement <i>pas</i> très correct	politicamente <i>non molto</i> corretto	politically <i>not</i> very correct

Ces résultats mettent en évidence trois points :

- 1) si une traduction littérale d'une collocation est possible, cela veut dire aussi que la collocation dans la langue source et celle dans la langue cible partagent la même flexibilité envers les transformations syntaxiques que la collocation peut subir ;
- 2) si l'on regarde les résultats du point de vue des transformations

syntaxiques que les collocations transparentes peuvent accepter, nous avons constaté une similitude entre les deux paires de langues analysées, les collocations transparentes acceptant les mêmes modifications syntaxiques indépendamment de la proximité ou de la distance entre les paires de langues analysées ;

- 3) d'un point de vue quantitatif, il y a plus de collocations transparentes entre le français et l'italien qu'entre le français et l'anglais, cela renforçant l'idée que le français et l'italien sont deux langues proches.

5. Conclusion

Dans cet article, nous avons analysé la proximité linguistique entre le français et l'italien du point de vue des collocations. Dans la première étude effectuée, nous avons pu constater que sur un total de 400 collocations analysées, le français et l'italien présentent un nombre très haut de collocations qui peuvent être traduites de manière littérale entre langue source et cible, alors que l'anglais montre une tendance à traduire par un seul mot les collocations françaises. Dans la deuxième étude menée, nous avons trouvé qu'en général il y a plus de collocations transparentes entre le français et l'italien qu'entre le français et l'anglais. Cela renforce l'idée que le français et l'italien sont deux langues proches. De plus, la transparence de traduction entre les collocations semble aussi mener à une flexibilité syntaxique de la collocation par rapport aux transformations possibles que la collocation peut subir. Cela semble vrai non seulement pour le couple de langues proches français-italien, mais aussi pour le couple de langues éloignées français-anglais (même si pour un pourcentage de cas plus bas).

Dans la suite de nos recherches, nous comptons élargir notre analyse à des dictionnaires plus étendus ainsi que tester des systèmes de traduction automatique – tel qu'Its-2 – sur un corpus contenant des collocations transparentes en français et en italien – tout comme en français et en anglais – afin de tester jusqu'à quel point la transparence en traduction peut amener à de meilleurs résultats de traduction automatique. Un bon point de départ pour une telle étude peut être les corpus du groupe PARSEME (PARSIng and Multi-word Expressions) (Savary & Przerpiórkowski 2013 ; Savary et al., 2015) qui comptent aussi un corpus pour l'italien (Monti et al., 2017).

Bibliographie

- Arcaïni, Enrico. 2000. *Italiano e francese, un'analisi comparativa*. Torino : Paravia.
- Barreiro, Anabela, Johanna Monti, Brigitte Orliac & Fernando Batista. 2013. When multiwords go bad in Machine Translation. In *Proceedings of the Workshop on Multiword Units in Machine Translation and Translation Technology*, 26-33.
- Calzolari, Nicoletta, Charles J. Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine Macleod & Antonio Zampolli. 2002. Towards best practice for multiword expressions in collocational lexicons. In *Proceedings of the third LREC (LREC 2002)*, 1934-1940.
- Calzolari, Nicoletta, Alessandro Lenci & Valeria Quonchi. 2003. Towards Multiword and Multilingual Lexicons: Between theory and practice. In *Proceedings of Linguistics and Phonetics 2002 Conference*.
- Cartoni, Bruno. 2008. *De l'incomplétude lexicale en traduction automatique. Vers une approche morphosémantique multilingue*. Thèse de doctorat : Université de Genève.
- Constant, Mathieu, Gülşen Eryigit, Johanna Monti, Lonneke van der Plas, Carols Ramisch, Michael Rosner & Amalia Todirascu. 2017. *Multivord expression and processing: A survey*. *Computational Linguistics* 43(4) : 837-892. DOI : https://doi.org/10.1162/COLL_a_00302
- Dabène, Louise & Christian Degache. 1996. Comprendre les langues voisines. *Études de linguistique appliquée* 104.
- Éloy, Jean-Michel. 2004. Langues proches : que signifie de les enseigner ? *Études de linguistique appliquée* 136 : 393-402.
- Greenwood, Pricilla E. & Mikhail S. Nikulin. 1996. *A Guide to Chi-Squared Testing*. New York : Wiley.
- Hédiard, Marie. 1989. Langues voisines, langues faciles ? In Enrico Arcaïni, Michèle Fourment & Danielle Lévy-Mongelli. *Analisi comparativa: Francese/italiano*, 27-35. Padova : Liviana.
- Leoni de Léon, Jorge Antonio. 2008. *Modèle d'analyse lexico-syntaxique des locutions espagnoles*. Thèse de doctorat : Université de Genève.
- Mel'čuk, Igor. 2003. Les collocations : analyse et traitement. In Francis Grossmann & Agnès Tutin (eds.), *Collocations. Définition, rôle et utilité*, 23-32. Amsterdam : De Werelt.
- Mel'čuk, Igor. 2013. Tout ce que nous voulions savoir sur les phrasèmes, mais.... *Cahiers de lexicologie* 102 : 129-149.
- Monti, Johanna, Anabella Barreiro, Annibale Elia, Federica Marano & Antonella Napoli. 2011. Taking on new challenges in multi-word unit processing for

- Machine Translation. In *Proceedings of the Second International Workshop on Free/Open-Source Rule-Based Machine Translation*, 11-19.
- Monti, Johanna, Maria Pia Di Buono & Federico Sangati. 2017. PARSEME-IT Corpus. An annotated corpus of verbal Multiword Expression in Italian. In *Proceedings of the CLIC-it 2017 Conference*, 228-233.
DOI : [<https://doi.org/10.4000/books.aaccademia.2433>]
- Monti, Johanna, Violeta Seretan, Gloria Corpas Pastor & Ruslan Mitkov. 2018. Multiword units in machine translation and translation technology. In Ruslan Mitkov, Johanna Monti, Gloria Corpas Pastor & Violeta Seretan (eds.), *Multiword Units in Machine Translation and Translation Technology*, 1-37. Amsterdam : John Benjamins. DOI : [<https://doi.org/10.1075/cilt.341.01mon>]
- Nerima, Luka, Éric Wehrli & Violeta Seretan. 2010. A Recursive Treatment of Collocations. In *Proceedings of the 8th International Language Resources and Evaluation (LREC 2010)*, 634-638.
- Nerima, Luka, Violeta Seretan & Éric Wehrli. 2006. Le problème des collocations en TAL. *Nouveaux cahiers de linguistique française* 27 : 95-115.
- Ramish, Carlos. 2015. *Multiword Expressions Acquisition: A Generic and Open Framework*. Cham : Springer. DOI : [<https://doi.org/10.1007/978-3-319-09207-2>]
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flinckinger. 2002. Multiword Expressions: A pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CiCLING 2001)*, 1-15.
DOI : [https://doi.org/10.1007/3-540-45715-1_1]
- Savary, Agata & Adam Przepiórkowski. 2013. PARSEME: Parsing and Multi-Word Expressions. Towards linguistic precision and computational efficiency. In *Natural Language Processing. COST Action*, no. IC1207.
- Agata Savary, Manfred Sailer, Yannick Parmentier, Michael Rosner, Victoria Rosén, Adam Przepiórkowski, Cvetana Krstev, Veronika Vincze, Beata Wójtowicz, Gyri Smørdal Losnegaard, Carla Parra Escartín, Jakub Waszczuk, Matthieu Constant, Petya Osenova, Federico Sangati. 2015. PARSEME PARSing and Multiword Expressions within a European multilingual network. In *Proceedings of the 7th Language and Technology Conference (LTC 015)*, 27-29.
- Seretan, Violeta. 2011. *Syntax-based Collocation Extraction*. Cham : Springer.
DOI : [<https://doi.org/10.1007/978-94-007-0134-2>]
- Seretan, Violeta. 2015. Multi-Word Expressions in user-generated content : How many and how well translated? Evidence from a post-editing experiment. In *Proceedings of the Workshop on Multi-Word Units in Machine Translation and Translation Technology*, 50-55.

- Wehrli, Éric. 2001. Parsing and collocations. In Dimitri N. Christodoulakis (ed.), *Natural Language Processing NLP 2000*, 272-282. Cham : Springer. DOI : [https://doi.org/10.1007/3-540-45154-4_26]
- Wehrli, Éric. 2013. Words and lexical units. In Ronaldo Martins (ed.), *Lexical Issues of UNL: Universal Networking Language 2012 Panel*, 35-44. Cambridge : Cambridge Scholars Publishing.
- Wehrli, Éric & Luka Nerima. 2009. L'analyseur syntaxique Fips. In *Actes of the 11th International Conference on Parsing Technologies (IWPT '09)*.
- Wehrli, Éric, Luka Nerima & Yves Scherrer. 2009. Deep linguistic multilingual translation and bilingual dictionaries. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, 90-94. DOI : [<https://doi.org/10.3115/1626431.1626450>]
- Wehrli, Éric, Violeta Seretan & Luka Nerima. 2010. Sentence analysis and collocation identification. In *Proceedings of the Workshop on Multiword Expressions: From Theory to Applications (MWE 2010)*, 28-36.