

Le corpus *Per-Fide* et ses applications : une étude de cas sur les expressions idiomatiques

1. Introduction

L'étude dont il est question ici vise précisément à souligner l'impact significatif que la linguistique de corpus peut avoir sur la recherche et l'enseignement dans les domaines touchant à l'étude du langage et à la traduction (Williams, 2005). Nous entendons ici proposer une réflexion sur l'utilité d'un corpus multilingue (tel que le corpus *Per-Fide*) dans le cadre du travail de traduction, sur la base d'un exemple concret, celui des expressions idiomatiques, dans une perspective contrastive français-portugais.

2. Le corpus multilingue *Per-Fide* : nouvel outil de traduction contextuelle en ligne

Les corpus multilingues parallèles sont des outils précieux pour l'analyse contrastive, mais ils sont bien souvent limités au niveau des langues mises en regard et/ou des types de textes considérés. Le projet *Per-Fide* (Araújo *et al.*, 2010) dont nous voulons rendre compte dans le cadre de cette étude trouve précisément son origine dans la nécessité de construire, avec des textes appartenant à des domaines plus diversifiés (littéraire, technico-scientifique, journalistique, juridique et religieux), des corpus alignés à la phrase, qui mettent en parallèle des textes originaux en portugais dans ses différentes variantes (portugais européen, portugais brésilien, portugais africain) et leurs traductions dans six autres langues (espagnol, russe, français, italien, allemand et anglais), ainsi que des textes originaux dans ces six langues étrangères et leurs traductions vers le portugais.

2.1. Mode de recherche monolingue

Le corpus *Per-Fide* peut être consulté librement à partir d'un concordancier (<http://www.per-fide.ilch.uminho.pt/query>), qui permet de procéder à différents types de recherche. Dans cette interface, l'utilisateur peut entrer des critères de recherche non seulement pour la langue source mais aussi pour la langue cible ou bien pour les deux à la fois afin de visualiser rapidement l'élément recherché dans différents corpus simultanément :

The screenshot shows the Per-Fide search interface. On the left, there are three main sections: 'Select Type' with a dropdown set to 'bilingual', 'Select language' with a dropdown set to 'PT-EN', and 'Select corpora' with a list of corpora including Comboni, DGT-TM, ECB, EMEA, EurLex-v1, EuroParlV5, JRC-Acquis-v3, PressEU, Shakespeare, SoftwarePO, and Vatican. A red box highlights the 'Select corpora' list. Below this, there are input fields for 'PT query' (containing 'festa') and 'EN query', each with a 'pdt' button. A 'Search' button is also present. At the bottom, there is a 'Entries per page' dropdown set to 20. On the right, there are three text boxes with annotations: 'Recherche monolingue ou bilingue (choisir la paire de langues)', 'Typologie de textes diversifiée (religieux, littéraire, technique, ...)', and 'Accéder: -au PTD ou -à la concordance bilingue du terme recherché (search)'.

Figure 1 : Interface de recherche du corpus *Per-Fide*

Pour obtenir la concordance monolingue d'un mot donné, il suffit de taper le mot en question. Cette concordance affiche ligne par ligne chaque occurrence d'une forme vedette au sein de son contexte d'origine :

The screenshot shows the Per-Fide search results for the word 'chat'. The interface is titled 'Per-Fide PTDC/CE/LLI/108948/2008'. The search type is set to 'monolingual' and the language is set to 'FR'. The search results are displayed in a table with the following columns: 'EuroParlV5 - Entry 1 to 30', 'EuroParlV5 - TU 30736', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150', 'EuroParlV5 - TU 31150'. The search results are as follows:

EuroParlV5 - Entry 1 to 30	EuroParlV5 - TU 30736	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150	EuroParlV5 - TU 31150
Nous devons offrir une aide constructive aux pays candidats et ne pas hésiter à appeler un chat un chat.									
Je pouvais donc me prononcer librement en faveur de la solution qui me paraît la meilleure pour les producteurs de vrai chocolat, tel que l'apprécient les gourmets qui estiment qu'il faut appeler un chat un chat et du chocolat chocolat, pour autant qu'il soit effectivement fabriqué avec du beurre de cacao.									
Je pouvais donc me prononcer librement en faveur de la solution qui me paraît la meilleure pour les producteurs de vrai chocolat, tel que l'apprécient les gourmets qui estiment qu'il faut appeler un chat et du chocolat chocolat, pour autant qu'il soit effectivement fabriqué avec du beurre de cacao.									
Nous posons la question suivante : "Savez-vous ce qu'est un chat ?" et la réponse est "Un chat, c'est un tigre qui a été auditionné par la commission des finances".									
On nous demande, effectivement, d'acheter du bûf chat en poche.									
Nous ne voulons pas acheter chat en poche, nous devrions procéder demain au vote de notre rapport et donner l'occasion à ces messieurs et dames de reconnaître leur erreur.									
Il s'agit - pour appeler un chat un chat - de l'affaire Allué et Coonan contre l'université de Venise, il s'agit de l'affaire Allué également contre l'université de Venise et il s'agit de l'affaire David Petri contre l'université de Vérone.									

Figure 2 : Concordance monolingue de *chat*

Il semblerait que le mot *chat* soit, dans le corpus *Europarl*, un excellent catalyseur d'expressions idiomatiques : *appeler un chat un chat* / *acheter chat en poche*.

2.2. Mode de recherche bilingue (simple ou double)

Une recherche bilingue devrait nous permettre de voir si l'homologue portugais de *chat* est aussi producteur d'expressions figées. La figure qui suit montre les résultats d'une requête qui permet de visualiser rapidement l'élément recherché dans différents contextes, mais aussi ses traductions en portugais, qui peuvent différer suivant les cas :

The screenshot shows the Per-Fide search interface. On the left, there are filters for 'Select Type' (set to 'bilingue'), 'Select language' (set to 'PT-FR'), and 'Select corpora' (listing various EU corpora like Comboni, Culnaria, DGT-Acquis, etc.). The main area displays 'SEARCH RESULTS' for 'chat' with 179 entries. The results are shown in a table with two columns: the Portuguese text and the French translation. The first entry shows a Portuguese sentence about helping candidate countries and its French translation. The second entry shows a Portuguese sentence about chocolate producers and its French translation. The third entry shows a Portuguese sentence about a cat and its French translation.

PT - EuroParlV5 - TU 31066	FR - EuroParlV5 - TU 31066
Nesse âmbito , temos de ajudar os países candidatos de forma crítica e construtiva , não evitando designar as coisas pelos próprios nomes .	Nous devons offrir une aide constructive aux pays candidats et ne pas hésiter à appeler un chat un chat.
PT - EuroParlV5 - TU 31489	FR - EuroParlV5 - TU 31489
Por conseguinte , podia pronunciar-me livremente a favor da solução que me parece a melhor para os produtores de verdadeiro chocolate , tal como o apreciam os bons gastrónomos , que consideram que é preciso chamar os bois pelos nomes e chocolate ao chocolate , desde que fabricado efectivamente com manteiga de cacau .	Je pouvais donc me prononcer librement en faveur de la solution qui me paraît la meilleure pour les producteurs de vrai chocolat, tel que l'apprécient les gourmets qui estiment qu'il faut appeler un chat un chat et du chocolat chocolat, pour autant qu'il soit effectivement fabriqué avec du beurre de cacao.
PT - EuroParlV5 - TU 31489	FR - EuroParlV5 - TU 31489
Por conseguinte , podia pronunciar-me livremente a favor da solução que me parece a melhor para os produtores de verdadeiro chocolate , tal como o apreciam os bons gastrónomos , que consideram que é preciso chamar os bois pelos nomes e chocolate ao chocolate , desde que fabricado efectivamente com manteiga de cacau .	Je pouvais donc me prononcer librement en faveur de la solution qui me paraît la meilleure pour les producteurs de vrai chocolat, tel que l'apprécient les gourmets qui estiment qu'il faut appeler un chat un chat et du chocolat chocolat, pour autant qu'il soit effectivement fabriqué avec du beurre de cacao.
PT - EuroParlV5 - TU 75484	FR - EuroParlV5 - TU 75484
Perguntamos : " Sabe o que é um gato ? " e a resposta é : " Um gato é um tigre que esteve em reunião de concertação com a Comissão dos Assuntos Económicos e Monetários " .	Nous posons la question suivante : " Savez-vous ce qu'est un chat ? " et la réponse est " Un chat, c'est un tigre qui a été auditionné par la commission des finances".

Figure 3: Concordance bilingue de *chat*

Si l'on compare les inventaires de structures qui correspondent en portugais à l'expression idiomatique *appeler un chat un chat* du français, on constate que la redondance sémantique du français (*un chat un chat*) se perd en portugais et il semblerait que l'on n'ait plus affaire, dans cette langue, à des *chats* mais à des *bœufs* (*bois*) ou à des *choses* (*coisas*) : *chamar os bois pelos nomes* / *designar as coisas pelos próprios nomes*. En demandant au bi-concordancier de repérer tous les couples de phrases dans lesquels *gato* apparaît du côté portugais et *chat* du côté français, le traducteur ou le linguiste peut déterminer, assez rapidement, les cas de correspondance directe entre ces deux termes :

Per-Fide
PTD/CLELL/108948/2008

Select Type
bilingual

Select language
PT-FR

Select corpora
☐ Comboni (info)
☐ Culinaría (info)
☐ DGT-Acquis (info)
☐ DGT-TM (info)
☐ ECB (info)
☐ EMEA (info)
☐ EurLex-v1 (info)
☒ EuroParlV5 (info)
☐ JRC-Acquis-V3 (info)
☐ LMD (info)
☐ PresseEU (info)
☐ SoftwarePO-2 (info)
☐ TurAccores (info)
☐ Vatican (info)

PT query
gato

FR query
chat

SEARCH RESULTS
EuroParlV5: 83 entries

EuroParlV5 – 1 to 30 entries

#1 75486	Perguntamos : " Sabe o que é um gato ? " e a resposta é : " Um gato é um tigre que esteve em reunião de concertação com a Comissão dos Assuntos Económicos e Monetários " .	Nous posons la question suivante : " Savez-vous ce qu'est un chat ? " et la réponse est " Un chat, c'est un tigre qui a été auditionné par la commission des finances " .
#2 39236	Estão efectivamente a pedir nos que compremos gato por lebre , só que na versão bovina .	On nous demande, effectivement, d'acheter du bœuf chat en poche.
#3 82743	Não devemos comprar gato por lebre , devemos votar o relatório amanhã e dar oportunidade de os senhores se informarem melhor .	Nous ne voulons pas acheter chat en poche, nous devrions procéder demain au vote de notre rapport et donner l'occasion à ces messieurs et dames de reconnaître leur erreur .
#4 159117	Não estamos interessados em comprar gato por lebre .	Nous ne voulons pas acheter chat en poche.
#5 182400	A começar pelo Comité Veterinário , que só tardiamente autorizou a vacinação , a continuar no Conselho de Ministros da Agricultura , que pareceu mais preocupado com questões de comércio externo do que com o controle da situação , e a terminar no Conselho Europeu de Estocolmo , que passou ao lado da crise como " gato por cima de brasas " , insistindo em remeter todos os custos para o quadro financeiro definido em Berlim .	À commencer par le comité vétérinaire, qui n'a autorisé que tardivement la vaccination, le Conseil des ministres de l'agriculture ensuite, qui a semblé plus préoccupé par les questions de commerce extérieur que par le contrôle de la situation, et, enfin, le Conseil européen de Stockholm, qui est passé sur la crise comme " un chat sur la braise ", en insistant pour que tous les coûts soient imputés au cadre financier défini à Berlin.

Figure 4: Concordance bilingue double de *chat-gato*

Cette recherche bilingue double montre que l'expression portugaise *comprar gato por lebre* est réellement l'équivalent de l'expression *acheter chat en poche* en français.

2.3. Recherche d'expressions multi-mots

Le corpus *Per-Fide* devient un outil encore plus précieux pour ceux qui souhaitent non seulement traduire des mots simples, mais également des mots plus complexes que l'on peut trouver dans différents contextes, ce qui est très souvent le cas dans la pratique. Si l'on tient à «contrer le trop-plein d'information, à confirmer une intuition personnelle ou encore vérifier si les équivalents proposés par les dictionnaires bilingues sont vraiment utilisés» (Langlois, 1996), on peut combiner plusieurs mots dans la syntaxe d'interrogation (simple ou double), et demander que le mot-vedette (ici, *taureau*) soit précédé ou suivi par un mot donné (par exemple *prendre/par les cornes*) :

The screenshot shows a search interface with a sidebar on the left and a main results area on the right. The sidebar includes filters for 'Select type' (set to 'bilingual'), 'Select language' (set to 'PT-FR'), and 'Select corpora' (with checkboxes for Combini, DGT-Acquis, DGT-TM, ECB, ENEA, EurLex-V1, EuroParl-V1, JRC-Acquis-V3, LIND, PresEti, SoftwarePO-2, and Vatican). The main area is titled 'SEARCH RESULTS' and shows 'PressEti: 0 entries | LIND: 0 entries | EurLex-V1: 0 entries | JRC-Acquis-V3: 0 entries | EuroParl-V1: 14 entries'. Below this, there are four search results, each with a title, a snippet of text, and a corresponding translation snippet. The results are numbered #1 through #4.

Result #	Source	Text Snippet	Translation Snippet
#1	EuroParl-V1 - T104612	Falei sobre este assunto nesta mesma assembleia na última sessão, mas volto a instar os Ministros da Educação da União Europeia para tomarem as rédeas da situação e certificarem-se que serão postas em prática iniciativas apropriadas e disponibilizados fundos adequados. Apelo também à senhora Comissária para que na próxima ronda sejam atribuídos fundos adequados ao Programa ERASMUS.	J'ai abordé ce sujet à l'Assemblée au cours de la session de session précédente, mais je voudrais une fois de plus inviter les ministres de l'Éducation de l'Union européenne à prendre le taureau par les cornes et à mettre en place des initiatives et des financements adéquats et je voudrais prier la commissaire de s'assurer que, lors du prochain tour, le financement du programme Erasmus sera suffisant.
#2	EuroParl-V1 - T145498	Os cidadãos do País Basco e da Catalunha, que correram riscos devido ao sobrevoo dos bombardeiros B-52, e também os cidadãos da Andaluzia, que estiveram em risco em resultado de quatro bombas que lhes caíram em cima, são cidadãos europeus e, por conseguinte, dizer que isto é um assunto interno é pura e simplesmente uma tentativa de evitar pegar o touro pelos cornos, quando na verdade é uma questão que afecta cidadãos europeus.	Les citoyens basques et catalans, qui ont été exposés à des risques en raison du survol des bombardiers B-52, ainsi que les Andalous, sur qui quatre bombes sont tombées, sont des citoyens européens. Par conséquent, répondre qu'il s'agit d'un sujet interne revient simplement à ne pas vouloir prendre le taureau par les cornes alors qu'il s'agit d'un sujet qui concerne des citoyens européens.
#3	EuroParl-V1 - T1455115	Creio que agora devíamos pôr termo a tudo isto e começar a levar as coisas a sério. Como acaba de dizer o senhor deputado Bourlanges, devíamos pegar o touro pelos cornos e respeitar os Tratados.	Je crois qu'il faut maintenant arrêter et être sérieux et, comme vient de le dire M. Bourlanges, prendre le taureau par les cornes et respecter les traités.
#4	EuroParl-V1 - T1460662	Portanto, pergunto de novo: vai a Comissão manter a sua passividade ou vai agarrar de uma vez por todas o touro pelos cornos assumindo a sua responsabilidade, de modo a remediar este real problema e a pôr ordem no tráfico selvagem que circula pelos nossos mares?	Aussi, je vous pose ma question: la Commission va-t-elle rester passive ou va-t-elle enfin prendre le taureau par les cornes et assumer ses responsabilités afin de résoudre ce problème bien réel et de rétablir l'ordre au niveau du trafic sauvage qui gouverne nos océans?

Figure 5: Concordance autour d'une expression idiomatique

Ce corpus permet donc de traduire des combinaisons de mots, des tournures idiomatiques et imagées dans leur contexte d'utilisation, en s'appuyant sur une base de données de traductions existantes, régulièrement mise à jour. L'utilisateur peut ainsi comparer les alternatives de traduction qui s'affichent et choisir celle(s) qui convien(nen)t le mieux au contexte de sa traduction. Il est intéressant de voir que l'expression *prendre le taureau par les cornes* (dont le sens propre – *se lancer dans l'arène et faire face aux coups de cornes de l'animal* – est reconnaissable sous le sens figuré – *faire face aux difficultés* – mais son usage en diffère) ne donne pas nécessairement lieu à une traduction littérale puisque l'on trouve, à côté de l'équivalent le plus direct *pegar/agarrar o touro pelos cornos*, d'autres formes de traduction (*tomar as rédeas da situação / enfrentar corajosa e rapidamente as dificuldades / encarar o problema de frente / meter mãos à obra / dar o braço a torcer, ...*) qui restituent, chacune à leur manière, le sens (à savoir *s'attaquer aux difficultés de front*) de cette expression.

2.4. Mode de recherche par les dictionnaires probabilistes de traduction

À travers l'interface de recherche du corpus *Per-Fide* l'utilisateur peut également activer le dictionnaire probabiliste de traduction (PTD) d'un mot en cliquant (non pas sur *search* mais) sur *ptd* (cf. *supra*, fig. 1). Voyez ci-dessous un exemple de recherche simple (le mot *chat*) par PTD :

Select Type
bilingual

Select language
PT-FR

Select corpora
☐ Comboni (info)
☐ DGT-Acquis (info)
☐ DGT-TM (info)
☐ ECB (info)
☐ EMEA (info)
☐ EurLex-v1 (info)
☒ EuroParlV5 (info)
☐ JRC-Acquis-V3 (info)
☐ LMD (info)
☐ PressEU (info)
☐ SoftwarePO-2 (info)
☐ Vatican (info)

PT query
 ptd→

FR query
chat ptd→
Search →
Entries per page: 20

PTD for : FR → PT

chat (229 occurrences)

38.55%	FR PT <u>gato</u> ✓	118	→
8.17%	FR PT <u>coisas</u>	9723	→
4.55%	FR PT <u>nomes</u> ✓	814	→
4.31%	FR PT <u>designadas</u> ✓	116	→
2.96%	FR PT <u>pelos</u>	26749	→
2.25%	FR PT <u>viável</u>	1310	→
2.11%	FR PT <u>postos</u>	5840	→

Figure 6: Dictionnaire probabiliste de traduction de *chat*

L'affichage en temps réel d'un PTD lorsque l'utilisateur passe la souris sur n'importe quel mot d'une concordance est un véritable atout de ce corpus. Ce PTD est capable d'identifier (grâce aux méthodes statistiques) les correspondances traductionnelles de l'expression recherchée. Dans le sous-corpus *Europarl*, *gato* est le meilleur candidat à la traduction du mot *chat* mais il est intéressant de voir que, dans le réseau des autres équivalents de ce mot, on trouve également, bien que dans une moindre mesure, les mots *coisas* et *nomes* (qui figurent, on le rappelle, dans les expressions idiomatiques *chamar os bois pelos nomes*/*chamar as coisas pelos próprios nomes* signalées plus haut). Plutôt que d'avoir à parcourir de longues listes de concordances pour examiner les équivalences traductionnelles d'une unité, l'utilisateur peut donc très rapidement activer le PTD d'un terme pour en déterminer, en un seul clic, les correspondants les plus directs. Cette approche du lexique est quantitative, mais ne doit, en aucun cas, négliger le retour à la concordance, qui est la seule à pouvoir offrir

une représentation fidèle et claire du voisinage d'une forme vedette dans une langue et des différentes manières de la rendre dans une autre langue. Pour assurer ce retour à la concordance, il suffit de cliquer sur la flèche qui figure à droite de chaque candidat à la traduction (cf. *supra*, fig. 6) et on a aussitôt accès à la concordance double *chat-gato*, *chat-coisas*, *chat-nomes*, etc., ce qui nous permet donc de visualiser ces termes en contexte et de voir qu'ils s'inscrivent, pour la plupart, dans des groupes de mots reliés par de fortes relations sémantiques et syntaxiques. Pour simplifier notre appréhension, le PTD opère donc le découpage initial du texte en unités, mais il est clair que ces unités ne sont pas seulement des mots isolés mais, également des unités complexes, plus larges, qu'il est nécessaire de retrouver et d'expliciter par le contexte. Nous verrons, notamment avec la méthode des cooccurrences du programme *CooCS*, comment le repérage de ces formes complexes peut se faire automatiquement.

Il faut préciser, par ailleurs, que ces PTDs fonctionnent selon un principe de circularité dans la mesure où ils affichent non seulement les correspondances traductionnelles du terme recherché, mais également les traductions de chacune de ces correspondances, ce qui permet de revenir à la langue de départ de notre requête initiale :

PTD for : FR → PT

chat (229 occurrences)

38.55%	<u>gato</u> ✓	118	→
8.17%	<u>coisas</u>	9723	→
4.55%	<u>nomes</u> ✓	814	→
4.31%	<u>designadas</u> ✓	116	→
2.96%	<u>pelos</u>	26749	→
2.25%	<u>viável</u>	1310	→
2.11%	<u>postos</u>	5840	→

PTD for : FR → PT

chat (229 occurrences)

	<u>gato</u> ✓	118	→
61.63%	<u>chat</u>	229	→
5.37%	<u>cat</u>	3	→
3.89%	<u>on</u>	45221	→
3.04%	<u>seulement</u>	26143	→
1.92%	<u>réclamait</u>	76	→
1.69%	<u>extraordinaires</u>	310	→
1.62%	<u>piste</u>	191	→
1.19%	<u>passer</u>	5795	→
0.81%	<u>bath</u>	1	→
0.38%	<u>proverbe</u>	234	→
0.37%	<u>sable</u>	122	→

Figure 7: Circularité des dictionnaires probabilistes de traduction

En effet, en cliquant sur les flèches centrifuges qui se trouvent à gauche de chaque alternative de traduction, l'utilisateur gère le PTD de chacune de ces alternatives, ce qui lui permet, par exemple, de conclure, de manière très simple et rapide, que le pourcentage de correspondance de *chat-gato* est nettement plus élevé du PT au FR que l'inverse (→ 38,55% contre 61,63%). On remarquera, en effet, que le PTD est traversé par un gradient chromatique qui mesure le degré de probabilité d'un terme d'être traduit par un autre. Le rouge pur s'impose lorsque le degré de correspondance entre deux termes tend vers les 0%, le vert pur lorsque ce degré se rapproche, au contraire, des 100% et le jaune pur s'assume comme une couleur intermédiaire qui marque un degré de correspondance de l'ordre des 50%. Il est à noter que plus ce degré de correspondance baisse, plus l'on aura affaire à des stratégies de traduction indirecte. En effet, comme on l'a signalé plus haut, d'autres termes (*coisas*, *bois*) entrent en concurrence avec le terme *gato* dans la traduction des expressions idiomatiques françaises comportant le nom de cet animal et donc la correspondance *chat-gato* se situe nécessairement dans un pourcentage inférieur à 50 % (d'où la couleur orange des 38,55%).

Il est intéressant de remarquer qu'un même terme peut donner lieu à des PTDs différents en fonction du corpus dans lequel on effectue la requête. En effet, si l'on calcule, par exemple, le PTD de *chat* dans un corpus plus technique lié à l'informatique, on tombe alors (non plus sur le terme *chat* mais) sur celui de *conversa* qui nous fait entrer, de plein fouet, dans le monde des messages SMS, des *chats*, des courriers électroniques très prisés par les jeunes.



Figure 8: Technicité des corpus (chat = conversa ≠ gato)

C'est justement parce que nous avons conscience de l'impact direct du degré de technicité des corpus sur les résultats obtenus que nous avons cherché à diversifier au maximum la typologie de textes qui intègre le corpus *Per-Fide*.

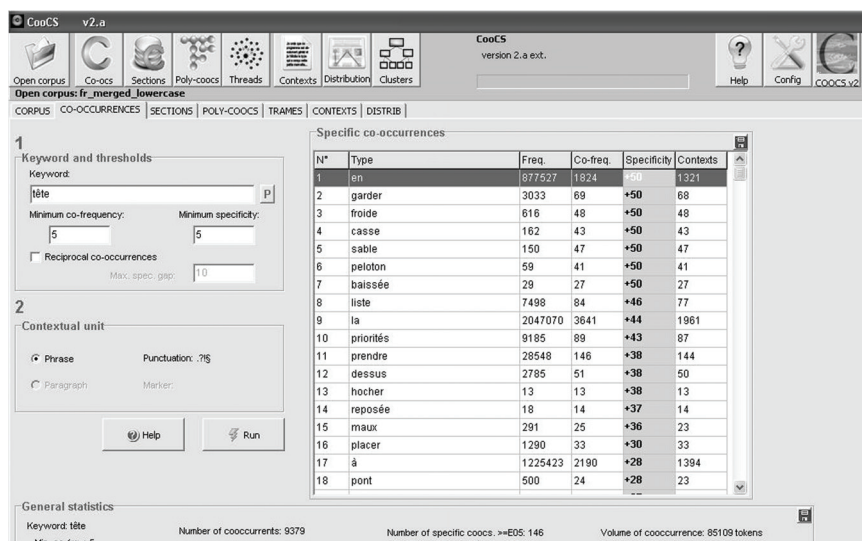


Figure 9 : Cooccurrences positifs du pôle tête

On calculera ainsi que, dans ses 2505 phrases d'occurrence, le pôle tête apparaît plus souvent en co-fréquence avec le nom peloton qu'avec le nom maux (co-fréq. 41 contre 25). La forme peloton se distingue par un indice de spécificité maximal (+50) qui rend compte de ses surprenantes coïncidences avec le pôle tête (41 sur 59, soit deux tiers de ses occurrences contre 25 sur 291 pour *maux*). Comme on peut le constater, en dressant un inventaire hiérarchisé des associations répétées autour d'un pôle donné, la méthode cooccurentielle définit ce que l'on peut appeler sa valence lexicale, c'est-à-dire sa capacité à attirer d'autres formes de manière récurrente en contexte (Martinez, 2003).

3.2. Au-delà de la cooccurrence binaire : les poly-cooccurrences

En examinant les systèmes d'attraction lexicale au-delà du lien de cooccurrence binaire, on peut recenser des attractions multiples et simultanées qui s'opèrent autour d'une forme pivot et d'identifier, par conséquent, des expressions semi-figées ou idiomatiques dont le sens ne peut être déduit à partir des mots qui les constituent (Martinez, 2012). On trouve ci-dessous un extrait du graphe qui reflète une partie du complexe réseau cooccurentiel élaboré à partir du pôle tête :

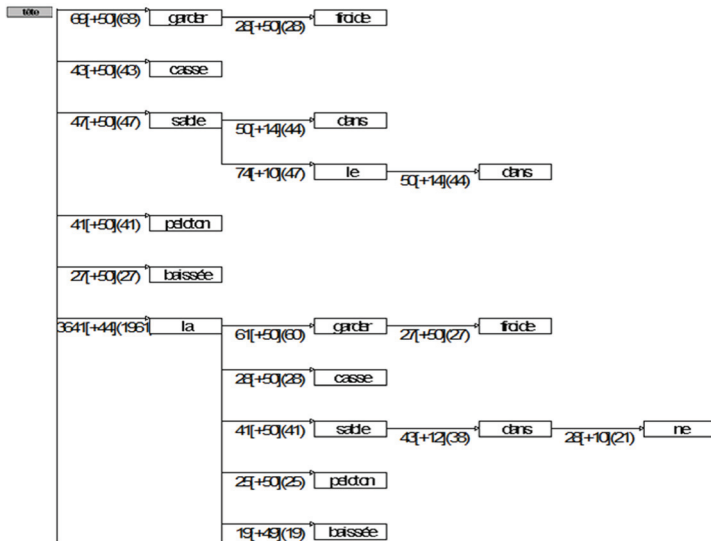


Figure 10: Réseau de cooccurrence du pôle tête

En explorant, avec ce niveau de précision, l'environnement contextuel du pôle tête, on découvre à ce dernier un certain nombre de cooccurents précédemment mis en avant par la méthode de cooccurrence initiale. Articulés autour des verbes (*casser, garder, placer, figurer,...*), qui restent les cooccurents les plus spécifiques de tête, apparaissent des prédicats nominaux (*sable, peloton, liste, etc.*) et adjectivaux (*froide, haute, baissée, etc.*). Les cartouches d'information qui jalonnent ce graphe nous renseignent sur la réalité statistique de ces attractions multiples :

xx [yy] (zz) = co-fréquence [indice de spécificité] (contextes de rencontre)

L'arbre donne donc pour chaque nœud le nombre exact de rencontres et le nombre de contextes où ces coïncidences ont lieu. Considérons le chemin de cooccurrence tête + sable. Le diagnostic de la dernière cooccurrence avec la forme sable, 47[+50](47), nous informe sur l'existence de 47 contextes de réalisation pour cette poly-cooccurrence. En poursuivant la lecture à droite du chemin, on voit que celui-ci se prolonge en deux réalisations plus spécifiques intégrant les formes le et dans pour lesquels on trouvera respectivement 47 et 44 réalisations en contexte. La concordance bilingue qui suit restitue une partie des contextes de réalisation de cette poly-cooccurrence :

FR	PT
(1a) Nous ne pouvons nous <i>enfoncer la tête dans le sable</i> et nier que c'est déjà le cas.	(1b) Não podemos <i>enterrar a cabeça na areia</i> e negar a realidade do que está a acontecer.
(2a) Il ne sert à rien de <i>s'enfouir la tête dans le sable</i> et d'affirmer sans cesse le contraire selon la devise [...]	(2b) Não vale a pena <i>enterrar a cabeça na areia</i> e afirmar continuamente o contrário, adoptando o lema [...]
(3a) Certains <i>mettent leur tête dans le sable</i> comme des autruches et pensent que tout ira bien.	(3b) Alguns <i>enfiar a cabeça na areia</i> e pensam que tudo acabará bem.
(4a) Vous pouvez continuer à <i>vous cacher la tête dans le sable</i> si bon vous semble.	(4b) Podem continuar a <i>enterrar a cabeça na areia</i> , se preferirem.
(5a) L'autruche, alors qu'elle est capable de <i>fourrer sa tête dans le sable</i> dans certaines situations, est aussi l'animal qui court le plus vite sur le sol.	(5b) A avestruz, embora possa em determinado momento <i>enfiar a cabeça na areia</i> , é também o corredor mais veloz em terra.

Tableau 1 : Concordance bilingue de la poly-cooccurrence verbe
+ *la tête dans le sable* (*Europarl/Per-Fide*)

Il est intéressant de voir, en effet, que l'association *tête dans le sable* apparaît bel et bien dans le sous-corpus *Europarl* du corpus *Per-Fide* et on ne peut s'empêcher de constater les variations du verbe qui affectent cette poly-cooccurrence. En effet, on trouve aussi bien le verbe *s'enfoncer* que les verbes *se cacher*, *s'enfouir*, *s'enterrer*, *fourrer*, *mettre*, ... *la tête dans le sable*. Certains auteurs parlent de classes d'équivalences (Gross, 1988) ; de possibilité de paradigme (Gross, 1996) ; de variantes lexicales (García, 2011). C'est cette même variation verbale que l'on observe sur la séquence *tête + baissée* (cf. *supra*, figure 10) :

FR	PT
(6a) Je demande donc à la Commission si elle veut <i>se jeter tête baissée</i> dans ce problème?	(6b) Pergunto aqui à Comissão: quer ela <i>meter-se às cegas</i> neste problema?
(7a) Au lieu de soutenir Israël en ce sens, d'importants États membres de l'UE <i>se précipitent tête baissée</i> sur l'État juif.	(7b) Em vez de apoiar Israel nesta questão, proeminentes Estados-Membros da UE <i>caem irreflectidamente</i> sobre o Estado judaico.
(8a) Il est regrettable que la Commission ait préféré <i>foncer tête baissée</i> , en s'en tenant à une proposition idéologique qui ne tient pas compte des opinions de la population.	(8b) É lamentável que a Comissão tenha optado, em vez disso, por <i>avançar de rom-pante</i> com a sua proposta ideológica que não presta atenção às opiniões dos cidadãos.
(9a) La Russie <i>plonge tête baissée</i> dans le piège et soutient indirectement et sans le savoir les forces en Géorgie qui pourraient vouloir utiliser le conflit pour satisfaire leurs propres fins politiques.	(9b) A Rússia <i>está a cair exactamente nesta mesma armadilha</i> e apoia mesmo, de modo indirecto e involuntário, as forças que, na Geórgia, poderão querer usar o conflito para obtenção dos seus próprios fins políticos.

FR	PT
(10a) J'espère donc que le Conseil entendra la voix de la sagesse et que les ministres ne <i>poursuivront pas tête baissée</i> le scénario de l'unification, qui serait sans issue.	(10b) Espero, portanto, que o Conselho oiça a voz da sensatez e que os Ministros não prosigam, <i>cabisbaixos</i> , o cenário da unificação, que não teria saída.

Tableau 2 : Concordance bilingue de verbe + *tête baissée* (Europarl/Per-Fide)

Dans le contexte plus immédiat à gauche de *tête baissée*, on trouve également une grande panoplie de verbes : *se jeter*, *se précipiter*, *foncer*, *plonger*, *poursuivre*... *tête baissée*. Outre cette variation d'ordre purement lexical, on ne peut s'empêcher de souligner la très grande diversité de moyens linguistiques mis en œuvre pour rendre, en portugais, cette expression idiomatique : *meter-se às cegas*, *avançar de rompante*, *cair (irreflectidamente)*, *cabisbaixos*, ... C'est, au contraire, une traduction littérale que l'on trouve en portugais (*enterrar a cabeça na areia*) pour l'expression française *s'enfoncer la tête dans le sable* mentionnée auparavant. Il semblerait que la traduction portugaise de l'expression idiomatique *jeter de l'huile sur le feu* subisse, en revanche, d'immenses variations dans notre corpus :

FR	PT
(11a) Les institutions européennes auraient dû y penser en 1991-1992, lorsqu'elles ont reconnu, contrairement aux règles en vigueur en matière de protection des minorités, la Slovénie d'abord et la Croatie ensuite, et ainsi <i>jeté de l'huile sur le feu</i> et déclenché la catastrophe des Balkans.	(11b) As instituições da UE deveriam ter pensado neste aspecto em 1991 1992 quando, contra as regras em vigor relativas à proteção das minorias, se reconheceu primeiro a Eslovénia e mais tarde a Croácia, <i>tendo assim posto achas na fogueira</i> que, mais tarde, se transformou na catástrofe dos Balcãs.
(12a) Une opération militaire de grande ampleur dans le nord de l'Iraq ne ferait qu' <i>ajouter de l'huile sur le feu</i> dans la région et pourrait également causer de sérieux problèmes en Turquie.	(12b) Uma intervenção militar em grande escala no norte do Iraque seria apenas <i>lançar mais achas para a fogueira</i> naquele país, além de que poderia igualmente causar graves problemas à Turquia.
(13a) Par ailleurs, il faut également faire comprendre au Président palestinien, M. Arafat, que son attitude ne favorise pas non plus le processus de paix. Il ne doit pas <i>mettre de l'huile sur le feu</i> en brandissant la menace d'une déclaration d'indépendance unilatérale de l'Etat palestinien.	(13b) De resto, deve ser dito que a atitude do presidente da Palestina, Arafat, não ajuda em nada o processo de paz, não devendo este <i>deitar mais achas na fogueira</i> ao ameaçar com a proclamação unilateral da independência da Palestina.
(14a) En présentant les villes et villages arabes comme un «cinquième front », les dirigeants des partis de droite <i>ont versé de l'huile sur le feu</i> .	(14b) Apresentando as cidades e as aldeias árabes como uma "quinta frente", os dirigentes dos partidos de direita <i>atiraram lenha para a fogueira</i> .

Tableau 3 : Concordance bilingue de verbe + *de l'huile sur le feu* (Europarl/Per-Fide)

En effet, alors qu'en français, l'expression ne manifeste, une fois encore, qu'une simple variation du verbe (*jeter/mettre/verser de l'huile sur le feu*), les alternatives de traduction proposées, en portugais, pour cette expression française opèrent une variation aussi bien au niveau du verbe qu'au niveau du nom et même de la préposition : *pôr/atirar/lançar/deitar* (verbes) *achas/lenha* (noms) *na/para a* (prépositions) *fogueira*. Une recherche sur corpus présente donc l'intérêt d'offrir une représentation fidèle des différentes mutations morphosyntaxiques et sémantiques qu'une expression idiomatique peut subir en contexte. Évidemment, cette variation est rarement répertoriée dans les dictionnaires. Par exemple, si l'on considère l'expression idiomatique dont il vient d'être question, on s'aperçoit qu'elle apparaît sous une forme bien plus pétrifiée dans les dictionnaires portugais *Infopédia* (<http://www.infopedia.pt/>) et *Priberam* (<http://www.priberam.pt/dlpo/>) puisque elle y est, respectivement, fixée comme suit : *deitar achas/lenha na fogueira* ; *deitar lenha na fogueira*. Mais un survol de notre corpus montre que les traductions de cette expression peuvent diverger considérablement de cette version dictionnarisée. En effet, en contexte, une expression de ce type peut ne mettre en avant que son sens propre (*lançar petróleo/gasolina na fogueira*) (cf. (15a)-(16b)) ou donner, au contraire, à son sens figuré une grande liberté d'expression (*evitar inflamar o debate, atear o fogo de um sentimento, alimentar o próprio monstro, ...*) (cf. (17a)-(19b)) :

FR	PT
(15a) Saddam <i>jette</i> en effet constamment <i>de l'huile sur le feu</i> - qui n'a rien de sacré - du conflit israélo-palestinien.	(15b) Com efeito, Saddam Hussein lança constantemente <i>petróleo para a fogueira</i> profana do conflito israelo-palestiniano.
(16a) Plusieurs pays de la région peuvent y prendre part directement car ils <i>jettent</i> sciemment <i>de l'huile sur le feu</i> terroriste palestinien et libanais.	(16b) Esse facto pode preocupar diversos países da região, porquanto <i>lançam</i> , a torto e a direito, <i>gasolina na fogueira</i> terrorista da Palestina e do Líbano.
(17a) Nous devons tous éviter de jeter de l'huile sur le feu en soulevant des questions qui pourraient être jugées menaçantes dans nombre de capitales nationales.	(17b) Todos nós devemos <i>evitar inflamar o debate</i> levantando questões que possam ser interpretadas como uma ameaça em muitas capitais nacionais.
(18a) [...], des déclarations qui contribuent à <i>jeter de l'huile sur le feu</i> d'un sentiment anti-sémite qui se répand et qui est déjà trop violent en Europe.	(18b) [...] declarações essas que contribuem para <i>atear o fogo</i> de um sentimento anti-semita que grassa e é já muito violento na Europa.
(19a) Si nous n'évitons pas les réponses extrémistes, nous pourrions bien <i>jeter de l'huile sur le feu</i> que nous essayons d'éteindre.	(19b) Se não evitarmos reacções extremistas, podemos acabar por vir a <i>alimentar o próprio monstro</i> que pretendemos destruir.

Tableau 4: Concordance bilingue de *jeter de l'huile sur le feu* (Europarl/Per-Fide)

Ces expressions caractérisées par le figement constituent un phénomène extrêmement complexe du point de vue formel et sémantique (Lamiroy, 2008 ; Mejri, 2008) qui présente un intérêt certain sur le plan contrastif, puisque, comme nous venons d'en rendre compte, la plupart d'entre elles échappent à une traduction directe lors de la recherche d'équivalents (figés ou non) dans une autre langue.

4. Considérations finales

Il est donc important de motiver l'intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction (outre celui de la linguistique et des langues) car le recours à ce type de ressources améliore la qualité des traductions par rapport à la seule utilisation de dictionnaires, notamment au niveau du choix des termes et de l'utilisation d'expressions idiomatiques (Frérot, 2010). Comme le note, à juste titre, Kraif (2006), les outils d'alignement et de concordance bilingue sont encore largement sous-exploités. Familiariser les utilisateurs – traducteurs, étudiants ou linguistes – avec les principaux concepts-clés du domaine et avec les techniques utilisées pour exploiter des corpus (Araújo *et al.*, 2011) se révèle donc essentiel. Nous avons tenu à montrer que le corpus multilingue *Per-Fide* met gratuitement à disposition une interface qui permet différents modes de recherche (monolingue ou bilingue simple et double) ; dictionnaire probabiliste de traduction (PTD) dans des langues et des domaines de spécialité variés. Les corpus parallèles, qui y sont gratuitement consultables en ligne, permettent d'étudier deux langues, en les confrontant et les éclairant réciproquement. Les PTDs sont tout aussi précieux, par exemple pour la traduction statistique automatique, mais aussi pour assister le travail des lexicographes et des terminologues en réalisant une partie du travail préliminaire de reconnaissance. Les textes sont, en effet, des objets complexes qui se construisent sur des relations d'attraction lexicale que les PTDs ne parviennent qu'à suggérer. Pour mettre à jour ces phénomènes de cooccurrence binaire ou multiple (notamment les expressions figées idiosyncrasiques et imprévisibles d'une langue donnée), il est possible, comme on l'a vu, d'utiliser toutes ces ressources en complémentarité avec d'autres outils pour mener des études plus textométriques.

Université du Minho

Université du Minho

Institut de Linguistique Théorique et Computationnelle

Sílvia ARAÚJO

Ana CORREIA

William MARTINEZ

Références bibliographiques

- Araújo, Sílvia / Almeida, José-João / Dias, Idalete / Simões, Alberto, 2010. «Apresentação do projecto Per-Fide: Paralelizando o Português com seis outras línguas», *Linguamática* 2, 71-74.
- Araújo, Sílvia / Oliveira, Ana / Dias, Idalete, 2011. «Como pesquisar em corpora», *Workshop présenté aux I Journées Internationales Corpora & Traduction*, 22 et 23 novembre 2011, Université du Minho, Braga - Portugal <http://perfide.ilch.uminho.pt/site.pl/workshop.pt>.
- Frérot, Cécile, 2010. «Outils d'aide à la traduction : pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs», *Les Cahiers du GEPE*, 2. Outils de traduction - outils du traducteur?
- Gross, Maurice, 1988. «Les limites de la phrase figée», *Langages*, 90, 7-22.
- Gross, Gaston, 1996. *Les expressions figées en français : noms composés et autres locutions*, Paris, Ophrys.
- García-Page, Mario, 2011. «Collocations complexes (application à l'espagnol)», *Linguisticae Investigationes*, 34, 68-111.
- Kraif, Olivier, 2006. «Qu'attendre de l'alignement de corpus multilingues?», *Traduire. 4^e Journée de la traduction professionnelle* 210, 17-37.
- Lamiroy, Béatrice, 2008. «Les expressions figées: à la recherche d'une définition», in: Blumenthal, Peter & Mejri, Salah (ed.), *Les séquences figées : entre langue et discours*. Zeitschrift für Französische Sprache und Literatur, Beihefte 36, 85-98.
- Langlois, Lucie, 1996. *Bitexte, bi-concordance et collocation*, Thèse pour l'obtention de la Maîtrise en Traduction, Université d'Ottawa, Canada.
- Martinez, William, 2003. *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*, Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne Nouvelle - Paris 3, France.
- Martinez, William, 2012. «Au-delà de la cooccurrence binaire : poly-cooccurrences et trames de cooccurrence», *Corpus* 11.
- Mejri, Salah, 2008. «Figement et traduction: problématique générale», *Meta : journal des traducteurs* 53, 244-252.
- Williams, Geoffrey (dir.), 2005. *La Linguistique de Corpus*, Rennes, Presses Universitaires de Rennes.