



HAL
open science

Traces de subjectivité et Corpus multilingues. Introduction

Raluca Nita, Ramon Marti Solano

► **To cite this version:**

Raluca Nita, Ramon Marti Solano. Traces de subjectivité et Corpus multilingues. Introduction. Les Cahiers FoReLLIS – Formes et Représentations en Linguistique, Littérature et dans les arts de l’Image et de la Scène, 2019. halshs-02486303

HAL Id: halshs-02486303

<https://shs.hal.science/halshs-02486303v1>

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cet article est mis à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale 3.0 France. CC BY-NC 3.0 FR



Introduction

Par Raluca Nita et Ramón Martí Solano

Publication en ligne le 06 septembre 2019

Table des matières

Corpus multilingues et recherche sur la subjectivité

Présentation du numéro

— Partie 1 Quelles données contrastives pour des constructions à faible rendement ?

De la nécessité des corpus multilingues spécialisés

— Françoise Doro-Mégy : « Subjectivité et interprétation cognitive de *SEE* : comment rendre compte des traductions françaises de *be seen as vs be seen to* ? »

— Joasha Boutault et Jeanne Vigneron-Bosbach : « *Her sleeping beauty of a stepmother, sa belle-au-bois-dormant de belle-mère* : constructions *N of a N* en anglais et leurs traductions en français »

— Partie 2 Les potentialités des corpus multilingues parallèles et les enjeux de la subjectivité selon les genres

— Hélène Chuquet et Ramón Martí Solano : « Adverbes de positionnement énonciatif dans un corpus d'écrits scientifiques. Étude contrastive anglais-français-espagnol »

— Paola Artero : « Voix narrative(s), voix traduisante(s) et stéréotypes de genre en littérature de jeunesse »

— Partie 3 Les données des corpus comparables : comparaison des langues dans des genres ciblés

— Emmanuel Baumer : « Expressions référentielles et point de vue dans les nouvelles en anglais et en français »

_____ Ludivine Crible : « Emplois sous-spécifiés des marqueurs discursifs *et / and* à l'oral : stratégie (inter)subjective et variation en genre »

_____ Alexandre Dias Da Silva : “A French-English contrastive analysis of *je dis ça, je dis rien* and *(I'm) just saying* in Internet blogs and forums”

Texte intégral

Ce numéro regroupe des contributions sélectionnées [1] suite à la Journée d'étude « Traces de subjectivité et corpus multilingues » [2] qui a eu lieu à l'Université de Poitiers en mai 2015, et s'inscrit dans l'activité de constitution et d'exploitation de corpus multilingues qui a été soutenue par le PRES Poitou-Charentes-Limousin, le laboratoire FoReLLIS (Université de Poitiers) – Équipe A Linguistique « Corpus : des unités au discours » – et le laboratoire FRED (Université de Limoges) sur la période 2013-2015.

Dans cette introduction, nous présenterons la façon dont ce numéro s'articule avec les travaux de l'Équipe A du laboratoire FoReLLIS, avant une présentation détaillée des articles pour souligner leurs apports à une réflexion générale sur la place des corpus dans la recherche sur la subjectivité.

Corpus multilingues et recherche sur la subjectivité

En choisissant comme thématique les marqueurs d'inscription de l'énonciateur dans le discours analysés à travers les langues, nous avons souhaité réunir deux domaines de recherche qui animent les travaux de l'Équipe A du laboratoire FoReLLIS depuis quelques années. D'une part, la question de la subjectivité s'inscrit dans le prolongement de l'Opération « Sentiments et point de vue » menée par les membres de l'équipe de 2010 à 2017 dans les Axes « Contrastivité inter-langues » et « Didactique des langues : acquisition, contextes, outils », ce qui a donné lieu à deux manifestations scientifiques (en 2011 et 2014) et à deux publications [3]. D'autre part, la question de la collecte des données est centrale aux travaux des chercheurs de notre équipe à travers la constitution, l'analyse et l'exploitation des corpus unilingues/multilingues, écrits/oraux, diachroniques/contemporains, d'apprenants, de spécialité [4]. Plus particulièrement, nous nous situons ici dans le prolongement du travail de constitution de corpus multilingues

développé dans l'Axe « Contrastivité inter-langues » et concrétisé par le Corpus PLECI (Poitiers-Louvain Échange de Corpus Informatisés) [5].

Ces deux domaines de recherche de notre équipe se sont nourris mutuellement. Les travaux sur les « Sentiments et point de vue » ont fait mûrir l'idée d'une exploitation multilingue de la question de la subjectivité en tant que problématique de l'inscription de l'énonciateur dans l'énoncé, car il est en effet apparu que celle-ci avait été moins abordée à travers les corpus dans l'approche contrastive [6]. Cependant, en envisageant la perspective d'un traitement multilingue des traces de subjectivité, il est apparu, au fur et à mesure de nos manifestations et de nos publications, les difficultés pour les chercheurs d'accéder à des données multilingues (originaux et traductions, originaux comparables) exploitables en analyse contrastive. Ces difficultés se font encore plus marquantes lorsqu'il s'agit de langues autres que l'anglais et le français, ou de ressources discursives autres que littéraires (par exemple, journalistiques ou scientifiques), qui sont les unes et les autres moins accessibles dans la constitution des corpus bilingues parallèles et comparables (originaux traduits dans une langue ou plusieurs, originaux comparables dans deux ou plusieurs langues) [7].

Le constat de la pénurie des corpus contrastifs nous a amenés ainsi à mettre en place un travail de constitution de corpus multilingues, anglais, français, espagnol, grec et roumain, qui vient compléter les ressources existantes au laboratoire d'une part par l'élargissement des ressources en contrastif fournies par le corpus PLECI dans le domaine littéraire avec l'ajout de langues peu représentées dans les corpus de traduction (l'espagnol [8], le roumain et le grec) et d'autre part, par l'enrichissement des ressources typologiques disponibles pour une recherche en contrastif avec l'ajout au PLECI littéraire et journalistique du genre scientifique avec des écrits dans le domaine de la linguistique. Le nouveau Corpus GRAFE (grec, roumain, français, anglais, espagnol) est le résultat de ce travail, et a été réalisé dans le cadre d'un projet de recherche qui a obtenu le soutien financier du PRES Poitou-Charentes-Limousin (2013-2015) et a associé le laboratoire FoReLLIS, Équipe A, de l'Université de Poitiers et le laboratoire FRED de l'Université de Limoges [9]. Ce corpus est divisé en deux sous-corpus : le Corpus GRAFE littéraire contient des extraits d'originaux en français, anglais, espagnol, roumain et leurs traductions, et des originaux en français et leurs traductions en grec ; le Corpus GRAFE académique contient des ouvrages scientifiques dans le domaine de la linguistique en français et en anglais et leurs traductions dans ces deux langues et en espagnol. Il s'agit d'un corpus bi-directionnel [10] puisqu'il permet une confrontation des données fournies par les corpus parallèles (de traduction) et les corpus comparables sur une quantité similaire de données. Ce corpus est constitué actuellement d'environ 650 000 mots dans les originaux dans les cinq langues pour la partie littéraire, et de 280 000 mots dans les originaux français et

anglais pour la partie académique, et est en cours d'alignement [11] pour une exploitation avec des concordanciers.

Présentation du numéro

C'est ainsi dans le contexte de l'activité d'enrichissement des corpus multilingues du laboratoire et du travail de l'équipe sur l'expression des sentiments et du point de vue que nous avons conçu la thématique de ce numéro. Nous avons souhaité privilégier l'approche de la subjectivité sous l'angle de l'inscription de l'énonciateur dans l'énoncé en tant que prise de position à l'égard du contenu de l'énoncé. Cela peut être mis en relation avec la définition que D. Paillard propose de la « scène énonciative » : « (...) un espace où, à partir des **agencements de formes** (y compris les marques prosodiques) constituant un énoncé, il est possible de restituer et **les positions subjectives en jeu, et le statut du dire** en tant que donnant une forme linguistique à un état de choses du monde. La scène énonciative se présente comme un espace dynamique, où, en fonction de déterminations multiples et hétérogènes, est représentée de façon infiniment variable mais régulière **la rencontre entre un (des) sujet(s), des formes linguistiques et le monde** [12]. » [13].

Les études proposées dans les articles de ce volume offrent pour analyse une large palette de marqueurs qui se dévoilent en tant que des traces de subjectivité par l'ancrage dans des typologies discursives spécifiques et par l'articulation avec d'autres marqueurs textuels. La subjectivité concerne alors la construction de la référence dans le domaine verbal (*be seen as/to*) et nominal (*her sleeping beauty of a daughter*) en anglais et en français à travers un repérage subjectif inscrit dans les marqueurs syntaxiques, lexicaux, sémantiques ; les chaînes de références nominales (pronoms vs noms propres) en anglais et en français en tant que choix énonciatifs dans le déclenchement d'origines subjectives distinctes ; la modalisation du dire (*I'm just saying ; je dis ça, je dis rien*) ; le commentaire sur son propre discours (les adverbes de phrase en anglais, français et espagnol) ; la cohésion et la cohérence discursive (*et/and*) ; le point de vue en linguistique et son pendant idéologique (*say* et les verbes introducteurs en anglais, français et italien). Sans que la réflexion sur les liens entre genres et construction de la subjectivité n'ait été une contrainte de notre thématique, la diversité des registres abordés ici (littérature générale, nouvelles, fantasy et science-fiction, littérature jeunesse, textes scientifiques, diverses formes d'échange sur internet, oral conversationnel spontané ou préparé) a fait apparaître que les phénomènes de subjectivité et leurs réalisations linguistiques sont façonnées par les genres, que ce soit du fait d'une certaine « sélection » des marqueurs opérée par les genres, ou du fait de

l'emploi spécifique des marqueurs dans le genre abordé (adverbes de phrase dans les textes scientifiques, connecteurs *et/and* à l'oral).

La spécificité des genres et des marqueurs fait que le traitement contrastif s'est vu (d'autant plus) confronté à des difficultés de collecte de données bi/multilingues, de telle façon que le terme corpus ne saurait être appréhendé ici systématiquement en tant que corpus électronique de textes sélectionnés sur des critères de représentativité et d'homogénéité, « a body of texts put together in a principled way and prepared for computer processing » [14], mais renvoie aussi au corpus en tant que base de données, corpus d'exemples, corpus d'ouvrages d'un auteur. Nous avons néanmoins choisi de garder le terme de corpus dans le titre de ce volume pour mettre en avant, à travers les articles, les défis et les limites de la constitution des corpus multilingues qui obligent, par endroits, dans certains contextes, selon l'objet d'étude, à revoir la notion de corpus en analyse contrastive. La réflexion menée dans la majorité des études sur la méthodologie de constitution et/ou d'exploitation des corpus, sur les contraintes et les obstacles dans la collecte de certaines données multilingues, sur la façon dont leur nature oriente et impacte à la fois la méthodologie et le traitement linguistique justifie que le terme corpus soit central dans le titre de ce volume.

Nous avons par conséquent choisi d'organiser les contributions de ce numéro selon leur rapport aux données multilingues, ce rapport étant influencé, au moins en partie, par la nature des marqueurs ou des constructions étudiées.

Nous partirons ainsi de deux contributions où la subjectivité est étudiée à travers des constructions que nous pourrions appeler à faible rendement étant marquées syntaxiquement, stylistiquement et discursivement. Il s'agit respectivement de *be seen to/as* et de la structure nominale N_1 of N_2 *her sleeping beauty of a mother*. La première ne semble pas avoir *a priori* des appétences pour un genre en particulier. La seconde est potentiellement confinée à un genre, la littérature, et peut dépendre d'un sous-genre (fantasy, science-fiction) ou du style de l'auteur (*her sleeping beauty of a mother*). Il s'avère que les deux constructions sont peu présentes dans des corpus bilingues déjà existants et posent ainsi un véritable défi quant à leur analyse contrastive.

Lorsque les corpus ou les données multilingues, originaux et traductions, sont disponibles elles permettent d'avoir un aperçu des enjeux de la modalisation des énoncés dans des genres spécifiques, comme par exemple les ouvrages scientifiques et la littérature de jeunesse. Ce sera notre deuxième approche des données multilingues à travers deux études portant respectivement sur la traduction des adverbes anglais en *-ly* en français et

en espagnol, et la traduction des verbes introducteurs du discours direct de l'anglais vers le français et l'italien.

Enfin, les corpus comparables (originaux dans plusieurs langues) permettent d'aborder des genres plus variés que les corpus traduits car la correspondance entre des originaux en plusieurs langues est plus facile à trouver. Les trois dernières contributions adoptent cette approche : à travers un corpus de nouvelles en français et en anglais est révélé le fonctionnement linguistique spécifique des marqueurs de point de vue dans chaque langue ; à travers les genres conversationnels de l'internet et les genres de l'oral on constate que des marqueurs en apparence équivalents en français et en anglais, respectivement *et/and* et *je dis ça, je dis rien/ I'm just saying*, ont un fonctionnement bien spécifique du point de vue de la construction de la subjectivité.

Partie 1 Quelles données contrastives pour des constructions à faible rendement ? De la nécessité des corpus multilingues spécialisés

Françoise Doro-Mégy : « Subjectivité et interprétation cognitive de *SEE* : comment rendre compte des traductions françaises de *be seen as vs be seen to* ? »

Françoise Doro-Mégy s'intéresse à la subjectivité à travers les constructions au passif court du verbe *see* à valeur cognitive, *be seen as* et *be seen to*. Son approche concerne à la fois les propriétés de chacune de ces deux structures et leurs traductions en français.

Dans le cadre de la Théorie des Opérations Prédicatives et Énonciatives, l'auteure montre d'une part la spécificité du passif et d'autre part la spécificité des marqueurs *as* et *to* dans la construction d'une représentation subjective dénotée par *see* cognitif. Le passif dans *X be seen as/to Y* met en évidence dans ce cas précis « la problématique de la référence, de la classification d'entités, en lien avec la représentation subjective ». C'est bien la référence du terme de départ de la relation prédicative, sujet de *be seen*, qui est en jeu et qui est construite subjectivement par rapport à une source cognitive indéterminée. Celle-ci est à l'origine du repérage du terme X en relation avec le terme Y. Selon que la mise en relation se fait par *as* ou *to*, la nature du repérage ainsi que la représentation subjective sous-tendue sont différentes. Avec *as*, il s'agit d'une identification de X par rapport au centre

organisateur du domaine notionnel représenté par Y. *As* est ainsi la trace d'une stabilité référentielle et a une valeur qualitative. *To*, généralement reconnu comme marqueur de visée, est la trace d'un repérage qui correspond à un hiatus sur le plan quantitatif (de l'occurrence spatio-temporelle) ou qualitatif (des propriétés). Il s'agit d'une instabilité référentielle de X par sa mise en relation avec Y, construit comme une prédication en attente, par le biais de *to*.

Dans l'analyse des traductions, des choix communs aux deux structures mettent en évidence « l'évaluation subjective » qui les caractérisent à travers les verbes subjectifs *considérer, percevoir, estimer, juger*, ainsi que l'indétermination de la source cognitive soit par le maintien du passif soit par l'emploi de *On*. D'autres traductions tendent néanmoins à être reliées à l'une ou l'autre des structures faisant ressortir sa spécificité. L'emploi de *comme* après *considérer* tend à s'associer à *be seen as* et confirme les hypothèses émises sur cette structure, d'autant plus que *comme* relève d'une opération similaire à *as*, celle d'une identification entre X et Y. *Sembler* apparaît davantage associé à *be seen to*, trace d'instabilité référentielle, et met ainsi au jour le hiatus dont *to* est la trace en dénotant la suspension de la validation de la relation prédicative et ainsi la modalité épistémique. Mais les traductions permettent également de s'interroger sur la relation de cette structure cognitive avec l'évidentialité du fait de la présence du passif court. L'explicitation d'une énonciation rapportée sous-tendue par un verbe cognitif au passif (*diront certains*) ou le maintien d'une dissociation énonciative entre l'énonciateur origine et la source de prise en charge de la représentation de X à travers le conditionnel journalistique (*son rôle serait avant tout défensif*) mettent en évidence des rapports avec l'évidentialité comme des pistes à explorer à l'échelle d'un corpus plus large.

L'ensemble de l'étude a été mené avec l'appui d'une combinaison de données contrastives de sources variées qui, à défaut d'avoir été quantitativement importantes pour émettre des généralités, ont permis l'identification de propriétés saillantes. La méthodologie et la problématique du recueil de données contrastives font partie de la réflexion de l'auteure dans cette étude où elle souligne la difficulté des études contrastives en raison des difficultés liées à la compilation de corpus suffisamment représentatifs du point de vue quantitatif et des genres afin d'offrir des données intéressantes notamment dans le cas de structures qui sont elles-mêmes particulières dans la langue de départ, comme c'est le cas de *be seen as/to* avec une valeur cognitive. Face à la fréquence assez faible des structures analysées dans deux corpus parallèles (PLECI-Université de Poitiers, 301 098 mots avec 46 occurrences et CODEX-Université Paris-Est Créteil, 6 040 392 mots avec 19 occurrences), l'auteure a dû se tourner vers des bases de données en ligne (*Linguee, Webitext...*) avec toutes les précautions méthodologiques nécessaires. Les hypothèses sur le contraste anglais-français restent donc à vérifier. Néanmoins les données exploitées ont permis

d'ébaucher la spécificité des structures anglaises à l'aune de la traduction et montrent l'utilité des études contrastives et ainsi la nécessité de compiler de nouveaux corpus et d'en enrichir ceux déjà disponibles.

Un objet d'étude comme celui-ci présente des exigences quant à la nature des données et l'on peut s'interroger sur la pertinence des corpus bilingues pour des objets aussi précis. Il nous semble que l'une des réponses peut être la diversification des genres représentés dans les corpus contrastifs, souvent riches en données littéraires, comme ceux utilisés ici, parfois journalistiques, mais moins, par exemple, dans le champ académiques [15]. L'autre solution, prônée par l'auteure, est celle du développement de corpus comparables. L'objectif serait en réalité la combinaison de corpus traduits et comparables et le recours à de grands corpus de référence comme cela a été fait ici (*The British National Corpus*) pour vérifier les caractéristiques de fréquence et d'usage que révèlent les corpus bilingues.

Joasha Boutault et Jeanne Vigneron-Bosbach : « *Her sleeping beauty of a stepmother, sa belle-au-bois-dormant de belle-mère* : constructions *N of a N* en anglais et leurs traductions en français »

Joasha Boutault et Jeanne Vigneron s'intéressent au fonctionnement en anglais de la structure nominale marquée N_1 of N_2 du type *her sleeping beauty of a stepmother* fondée sur la mise en relation subjective de deux notions dont l'une, correspondant à N_1 , sert à qualifier l'autre, correspondant à N_2 , à travers une comparaison sous-tendue. Les auteures identifient dans un premier temps les caractéristiques syntaxiques et sémantiques de la structure : d'une part, l'absence de contraintes quant au choix du déterminant de N_1 (démonstratif, défini, indéfini, possessif, etc.) et le caractère figé de l'emploi du déterminant *a* sur N_2 , et d'autre part, le sémantisme assez général du référent N_1 (+/- animé, +/- humain : *woman, child, barbarian, bay, smile, planet, headache, dog*). Dans un second temps, la comparaison avec les traductions en français fait ressortir le caractère spécifique de la structure en anglais, l'équivalence directe en français étant limitée au niveau du corpus (6 occurrences). Bien que les données en contrastif soient limitées, cette différence peut se justifier, au moins en partie, par les contraintes plus strictes qui régissent la structure en français : déterminants possessifs ou démonstratifs sur le N_1 , détermination zéro sur le N_2 , et sémantisme du N_2 restreint aux animés humains (*docteur, belle-mère, couturières*). En analysant la structure d'un point de vue syntaxique dans le cadre de la

grammaire générative, les auteures la mettent en relation avec l'ensemble des constructions de ce type, non-marquées, et proposent la piste d'une analyse unitaire de N_1 of N_2 en tant que constituant en faisant l'hypothèse de la portée du premier déterminant sur l'ensemble de la structure *Det* [N_1 of a N_2], hypothèse qui devra faire l'objet d'une étude plus large.

Le cœur de l'analyse des auteures se fonde sur les mécanismes qui rendent possible la mise en relation des deux noms et ainsi la construction de la subjectivité. Si la relation d'identification entre les deux noms a pu être évoquée dans la littérature (*king = bastard*) fondée sur l'idée (combattue par les auteures) d'une structure dérivée en grammaire générative justifiant la portée « inversée » des deux déterminants sur les deux noms (*that bastard of a king = that king is a bastard*), les auteures montrent avec succès que, en réalité, la structure s'appuie sur la valeur de hiatus de la préposition *of* qui permet de mettre en relation deux domaines notionnels, qu'ils soient intrinsèquement compatibles (<bastard> et <king>) ou pas (<lump> et <daughter>). A travers la préposition *of*, les propriétés notionnelles associées au N_1 servent à qualifier subjectivement le N_2 et à en offrir la propre vision de l'énonciateur. Cette analyse s'avère particulièrement pertinente lorsqu'il s'agit de distinguer entre deux valeurs de la même structure, construite avec le même N_1 : *a cold hell of a world vs a hell of a racket*. Dans le premier cas, il s'agit bien de la mise en relation de deux domaines notionnels. En revanche, dans la seconde, *a hell* opère un travail sur le seul domaine notionnel associé à N_2 exprimant le point de vue de l'énonciateur, avec une valuation positive ou négative selon le contexte, quant à l'instanciation par N_2 des propriétés du domaine notionnel : soit au plus haut degré (*a hell of a racket : un boucan terrible*) soit en deçà de ces propriétés (*a hell of a marriage : une parodie de mariage*).

L'analyse des traductions appuie les propriétés de la structure marquée et la distinction entre les deux interprétations. La qualification de N_2 par le biais d'une comparaison avec N_1 apparaît à travers la diversité des traductions : privilégier le N_1 et omettre le N_2 dans des structures attributives où N_2 est redondant par rapport aux caractéristiques du sujet (*it [Darkover]'s a cold hell of a world : c'est un enfer glacé*) ; privilégier le N_2 lorsque le N_1 est modifié par un adjectif à valeur quantitative (*a huge vat of a bath : une immense baignoire*) ; ou bien privilégier une traduction par équivalence qui montre le caractère métaphorique de la construction, formant ainsi une unité de traduction (*you have a peach of a job : vous avez la partie belle*). Dans le cas de l'évaluation des propriétés de N_2 par rapport au domaine notionnel à travers *a hell*, le N_2 ne peut être omis (à part dans des occurrences du type *a hell of a job*), car ce sont les propriétés mêmes du référent de N_2 qui sont en jeu ; *a hell of* est alors traduit par des éléments lexicaux dénotant la subjectivité de

l'évaluation des propriétés de N2 : le haut degré (*a hell of a racket : un boucan terrible*), la faible instanciation des propriétés notionnelles (*une parodie de Noël*), et une valuation positive ou négative du rapport de l'occurrence aux propriétés notionnelles (*a hell of a custom : une sacré coutume ; what a hell of a world : c'est charmant, ce monde ; you certainly picked one hell of a moment to walk in ! : Tu en a choisi un moment, pour faire ton entrée*).

Cette étude a soulevé pour les auteures un problème d'accès aux données à la fois unilingues en anglais, en raison de la spécificité de la structure, et bilingues pour la traduction en français. Les corpus parallèles disponibles et présentant une homogénéité du point de vue du genre (corpus littéraires bilingues anglais-français – PLECI et GRAFE d'environ 300 000 mots chacun) se sont avérés peu adaptés à ce type d'étude bien ciblée n'ayant pas fourni des résultats suffisants. Cela a conduit à la sélection de données de sources variées (romans de science-fiction et de fantasy, bandes dessinées, sites internet, films et séries télévisées).

La constitution et l'élargissement des corpus bilingues existants s'imposent ainsi de manière essentielle ici également. Si l'étude d'occurrences aussi spécifiques que celles envisagées dans les deux premiers articles de ce numéro peut difficilement être couverte par les corpus bilingues en raison des défis pour recueillir de données assez nombreuses quantitativement et assez variées qualitativement du point de vue des genres, l'approche contrastive à travers des corpus comparables et de traduction apparaît néanmoins comme un élément essentiel à l'approfondissement de l'analyse des structures d'une langue de départ et donc en complément des études unilingues.

Partie 2 Les potentialités des corpus multilingues parallèles et les enjeux de la subjectivité selon les genres

Hélène Chuquet et Ramón Martí Solano : « Adverbes de positionnement énonciatif dans un corpus d'écrits scientifiques. Étude contrastive anglais-français-espagnol »

L'article d'Hélène Chuquet et Ramon Martí-Solano porte sur les adverbes de phrase marquant le positionnement énonciatif dans les écrits scientifiques. Trois langues sont analysées (anglais-français-espagnol) montrant les potentialités d'exploitation d'un corpus trilingue, le corpus GRAFE compilé à l'Université de Poitiers.

En partant de l'anglais, où les adverbes en *-ly* présentent une grande souplesse morphologique qui justifie, en partie, leur productivité et leur fréquence supérieures aux équivalents français (en *-ment*) et espagnol (en *-mente*), les auteurs étudient la nature des jugements subjectifs portés sur les contenus propositionnels en relation étroite avec le genre des écrits scientifiques. Le point de vue de l'énonciateur correspond à des valeurs modales variées qui font apparaître les contraintes rhétoriques du genre : l'évidentialité (*obviously, apparently*), la valuation (*importantly, significantly*), l'épistémique (*surely, presumably*). À ces valeurs s'ajoute une dimension intersubjective où le commentaire de l'énonciateur sur ses propos se construit par la prise en compte de la position présumée du co-énonciateur (*obviously, presumably*).

Sur le plan de la traduction, les données fournies par le corpus confirment la non-équivalence, déjà établie dans la littérature, entre adverbes en *-ly* et leurs équivalents morphologiques en français et en espagnol (*-ment, -mente*), et mettent au jour des phénomènes récurrents dans les traductions qui sont en relation à la fois avec la subjectivité spécifique du genre scientifique et avec les caractéristiques morphologiques des adverbes anglais (multiplication des suffixes et des préfixes). Dans ce cadre, les auteurs soulèvent la question de la pertinence de certaines traductions par rapport aux valeurs subjectives véhiculées par le texte scientifique (par exemple, récurrence de *manifestement* en français ou de *obvio* en espagnol comme équivalents de *obviously* alors que d'autres solutions apparaissent comme plus cohérentes pour ce registre).

En se basant sur le volet anglais-français du corpus (originaux dans les deux langues et leurs traductions), les auteurs mettent en évidence l'exploitation qui peut en être faite en linguistique (quels traits subjectifs sont-ils privilégiés dans le domaine de l'écrit scientifique ?) et en traduction (quel est l'impact de l'opération de traduction sur les formes utilisées ?). Ainsi le caractère bidirectionnel du corpus permet de vérifier pour le couple anglais-français les caractéristiques des originaux et des traductions. Les adverbes en *-ly* dans les traductions en anglais se révèlent moins nombreux mais plus variés sémantiquement que dans les originaux, alors que les constructions dans les originaux français correspondant à des traductions par les adverbes en *-ly* sont plus diversifiées syntaxiquement que dans les traductions en français.

Cet aller-retour entre traductions et originaux que permet le corpus bidirectionnel et les divergences qu'il révèle soulèvent une question plus générale, selon les auteurs, celle de la

nature de la modalisation de l'énoncé à travers les langues dans le genre de l'écrit scientifique. Selon le corpus, l'anglais privilégie la construction de l'intersubjectivité et la modalité appréciative (*significantly*), le français la modalité épistémique (*sans doute*) et le commentaire sur la place à accorder aux propos (*précisément*).

Les auteurs insistent sur les précautions à prendre quant aux résultats fournis par un corpus réduit, et interrogent l'homogénéité même du genre illustré (par rapport au public visé, au style des auteurs) bien que son unité repose sur la discipline et la situation de communication. Il n'en reste pas moins qu'un tel corpus (parallèle et comparable) permet de révéler les facettes de la subjectivité à travers les langues et leurs points de non-coïncidence, ce qui amène une réflexion sur la nature de ces divergences – sont-elles internes aux systèmes linguistiques ou bien aux pratiques discursives entre les langues ?

Paola Artero : « Voix narrative(s), voix traduisante(s) et stéréotypes de genre en littérature de jeunesse »

Paola Artero s'intéresse à la traduction des verbes introducteurs de discours direct de l'anglais vers le français et l'italien dans deux œuvres appartenant au genre de la littérature de jeunesse particulièrement connues du jeune public, *Prince Caspian* et *The Voyage of the Dawn Trader* de Clive Staples Lewis. Les verbes introducteurs sont envisagés en tant que véhicules de subjectivité, susceptibles de modifier la perception des personnages et donc la lecture des textes. Les enjeux de la traduction des verbes introducteurs deviennent d'autant plus importants lorsqu'il s'agit de littérature de jeunesse et que les personnages, féminins et masculins, peuvent transmettre à travers leur mise en scène une vision particulière des rapports filles-garçons.

Paola Artero met en relation la subjectivité avec la question du point de vue, en tant que traces linguistiques de qualification et de modalisation renvoyant à une source énonciative particulière, et avec la question de l'idéologie, en tant qu'expression dans un texte de croyances particulières. Les deux notions, utilisées dans des cadres théoriques différents, respectivement l'énonciation et la *critical analysis*, recouvrent toutes les deux la prise de position spécifique qui se manifeste dans un texte et qui correspond à la voix narrative. Or la caractéristique de la traduction est le dédoublement de cette voix par une voix traduisante. Les contraintes des systèmes linguistiques en présence et l'ancrage spatio-temporel des originaux et des traductions font qu'à l'idéologie manifestée dans le texte source peut se substituer une idéologie autre dans le texte traduit. Ce sont donc les verbes introducteurs – analysés en tant que traces de subjectivité rattachées au narrateur et au

traducteur – qui peuvent modifier la construction des personnages dans le passage d’une langue à l’autre.

L’objet de l’étude est plus particulièrement constitué par la traduction du verbe *say*, dont la répétition en anglais ne saurait être maintenue ni en français ni en italien, langues qui s’appuient sur une diversité sémantique des verbes introduisant le discours direct. Dans le contexte de ces différences discursives entre les trois langues, la question du choix du verbe introducteur pour la mise en scène de personnages susceptibles d’influencer les projections sociétales des jeunes lecteurs prend tout son sens. Statistiquement, l’auteure fait la démonstration d’une non-équivalence quantitative dans son corpus entre respectivement *say* et *dire* en français et en italien. L’étude des traductions confirme l’hypothèse d’une lecture modifiée entre les originaux et les traductions. Ce qui caractérise l’italien est l’omission fréquente de l’énoncé rapportant et donc du verbe introducteur, qui peut se justifier par des contraintes discursives (absence de spécifications contextuelles autour du verbe) et linguistiques (omission du pronom-sujet entraînant l’omission du verbe). Ce qui pourrait constituer un cadre offrant une liberté d’interprétation pour le discours direct, se rapprochant de l’effet produit par *say* en anglais, s’avère en réalité un élément perturbateur dans l’interprétation des discours directs notamment dans les échanges entre les personnages où les tours de parole ne sont plus spécifiés. La traduction amène donc des difficultés, voire des erreurs de lecture. Le français suit les caractéristiques des originaux largement vérifiées par les études de linguistique contrastive, à savoir la spécification sémantique du verbe introducteur. C’est cette spécification sémantique qui constitue le deuxième élément, essentiel, de l’étude contrastive de l’intersubjectivité dans ce corpus. L’auteure montre une tendance à la spécialisation de certains verbes introducteurs en français selon les personnages : des verbes renvoyant à une attitude psychologique de faiblesse tels *gémir, implorer, murmurer* tendent à être associés aux personnages féminins, alors que des verbes dénotant une psychologie forte, tels *commander, décider, ordonner, trancher* sont utilisés avec des personnages masculins. La subjectivité de ces éléments linguistiques n’est pas forcément en lien avec le contenu des paroles rapportées ou avec le statut des personnages et leurs traits dans le texte original, comme le souligne l’auteure au sujet des personnages féminins, Lucy et Susan, reines avec des traits de bravoure. Les choix de traduction confirment ainsi dans ces cas précis la modification de lecture que peut subir un texte traduit. Plus particulièrement, dans ce cas précis, ils construisent, selon l’auteure, une conformité aux clichés et aux stéréotypes de la société.

Il nous semble que cette étude apporte une contribution importante aux études contrastives sur les verbes introducteurs par la double approche, narratologique et traductologique, par la nature du corpus et par la perspective multilingue. La spécificité du

support d'étude - la littérature de jeunesse - amène des enjeux de construction de l'imaginaire du jeune public. La subjectivité dont sont la trace les verbes introducteurs modifie ici la caractérisation des personnages filles et garçons dans les traductions par rapport aux originaux, ce qui évoque des constructions sociales stéréotypées (faiblesse des filles, courage des garçons). La dimension trilingue de l'analyse des verbes introducteurs renforce la démonstration du rôle que ces derniers jouent dans la construction du texte. Les différences de choix sémantique entre le français et l'italien dans la traduction de *say* apportent une expérience de lecture propre selon les prises de position subjectives du traducteur, comme le souligne l'auteure. On constate alors qu'un message subjectif peut être simplement présenté comme prise de parole d'un personnage avec *say* en anglais (*Oh, do let's leave it alone, said Susan*), interprété par rapport à une prise de position dans un rapport de force entre les personnages en français (*Oh ! Laissez-le tranquille ! implore Susan*) ou situé dans le cadre d'un tour de parole en italien (*Consiglierei di lasciar perdere, per stasera – intervenne Susan*). Une telle problématique pourrait pleinement bénéficier de la constitution de corpus multilingues électroniques qui permettraient de mesurer l'impact de telles traces de subjectivité par des données plus importantes et l'association des analyses quantitatives et qualitatives.

Partie 3 Les données des corpus comparables : comparaison des langues dans des genres ciblés

Emmanuel Baumer : « Expressions référentielles et point de vue dans les nouvelles en anglais et en français »

Emanuel Baumer s'intéresse à la mise en place des chaînes de référence – noms propres et pronoms – dans les nouvelles en tant que marqueurs de la construction du point de vue. L'auteur s'appuie sur un corpus comparable de nouvelles en anglais et en français (40 nouvelles par langue) qui ont été annotées au niveau des syntagmes nominaux, notamment noms propres et pronoms, renvoyant au personnage à travers lequel le récit est filtré, à savoir l'origine subjective-point de vue. Pour faire référence à ce type de personnage, les deux langues manifestent une préférence nette pour les pronoms au détriment des noms propres (et des syntagmes nominaux). La sélection d'un sous-corpus (10 nouvelles par langue) privilégiant la « focalisation interne », démontre que la fréquence d'emploi des pronoms y est supérieure à celle dont faisait état l'étude de l'ensemble du

corpus, ce qui tend ainsi à confirmer que le pronom est un marqueur de maintien du repérage par rapport à une origine subjective distincte de l'énonciateur. L'auteur illustre ensuite ces résultats par des analyses d'extraits des deux sous-corpus et montre, sur des exemples suivis, que l'utilisation constante du pronom s'associe à d'autres marqueurs linguistiques dans le maintien de la représentation du point de vue du personnage dans le récit, et dans le passage à des formes de discours rapporté indirect libre. En revanche, le nom propre, dont le choix est la trace d'une rupture dans la structure d'un récit, correspond à l'indexation du point de vue sur l'énonciateur-narrateur à la fois dans le récit pur et dans les « intrusions narratoriales » correspondant à des commentaires sur les personnages. Le nom propre a cependant un rôle structurant dans la construction du point de vue du personnage lorsqu'il est associé à des verbes de pensée, perception, etc. Il contribue alors à « déclencher » le point de vue du personnage qui sera par la suite privilégié dans le texte et maintenu par le pronom personnel. Cet article part ainsi de données quantitatives portant sur la chaîne de référence pour montrer à travers des analyses contextuelles comment les résultats des statistiques reflètent le fonctionnement de la subjectivité associée à la question du point de vue.

Ludivine Crible : « Emplois sous-spécifiés des marqueurs discursifs *et/ and* à l'oral : stratégie (inter)subjective et variation en genre »

Ludivine Crible aborde la question de la subjectivité dans un corpus oral comparable anglais-français. L'auteure s'intéresse à un trait définitoire des marqueurs discursifs, leur multifonctionnalité, qui est d'une part en relation directe avec la question de l'investissement de l'énonciateur dans son discours par le sens dont il empreigne les énoncés à travers les marqueurs discursifs, et d'autre part en relation avec la question du rapport que l'énonciateur construit avec le co-énonciateur à travers l'usage des marqueurs discursifs. Dans leur caractère multifonctionnel, les marqueurs discursifs peuvent exprimer des relations en contexte non encodées dans leur sémantisme. Il s'agit là de l'emploi « sous-spécifié » des marqueurs discursifs qui est illustré à travers *et/and*. L'auteure fait l'hypothèse que la sous-spécification est une trace de subjectivité, plutôt que d'intersubjectivité, car elle privilégie un rapport implicite entre les éléments mis en relation par les marqueurs discursifs au détriment d'une relation explicite qui tiendrait ainsi compte des besoins cognitifs/interprétatifs de l'interlocuteur. Par conséquent, certaines situations discursives peuvent mettre en place la sous-spécification. Il s'agit, par exemple,

de l'oral spontané conversationnel, alors que d'autres – l'oral préparé formel – peuvent y être réfractaires.

Cette hypothèse est testée par l'auteure pour *et/and* à partir d'un corpus oral comparable aligné au son et annoté. Sur un nombre équivalent d'occurrences de *and* et de *et*, l'auteure met en évidence la multifonctionnalité des deux marqueurs, qui peuvent exprimer des valeurs diverses en plus de leur valeur de base d'addition. Les emplois sous-spécifiés s'avèrent être pris en charge par les deux marqueurs tant dans les genres spontanés que préparés, mais néanmoins avec une préférence pour les premiers. D'un point de vue contrastif, il est intéressant de noter des fonctionnements différents dans les deux langues : le connecteur *et* privilégie l'expression de la spécification ou du changement de thème conversationnel, tandis que le connecteur *and* l'expression de la conséquence. L'analyse qualitative des emplois sous-spécifiés, notamment cause/conséquence et contraste, attire l'attention sur l'appui que prend la construction de ces sens sur le contexte. L'auteure arrive à la conclusion que dans l'expression de la conséquence, les marqueurs discursifs *et/and* activent un contenu contextuel narratif, alors que dans l'expression du contraste *et/and* jouent un rôle minimal. Ce sens est en réalité construit en contexte par une série de marqueurs, lexicaux, aspectuels, syntaxiques, etc., le rôle des marqueurs discursifs étant alors purement de liaison : « la valeur sous-spécifiée de *et/and* le rend propice à une fonction de renforcement pragmatique ou d'emphase d'éléments co-textuels ».

Alexandre Dias Da Silva : “A French-English contrastive analysis of *je dis ça, je dis rien* and *(I'm) just saying* in Internet blogs and forums”

Alexandre Dias Da Silva traite de la subjectivité dans le cadre du processus d'ajustement du dire dans la co-énonciation, opération dont sont la trace les périphrases *je dis ça, je dis rien* en français et *I'm just saying* en anglais, peu traitées dans la littérature. L'approche comparative français-anglais vise à signaler une opération commune, celle de la modalisation, mais également les valeurs distinctes amenées par les éléments morphosyntaxiques et énonciatifs propres à chacune des périphrases. L'auteur a constitué un corpus de 100 exemples dans chaque langue en s'assurant de leur comparabilité par la sélection de ressources relevant de genres similaires sur internet – groupes de discussion, commentaires d'articles, blogs, forums – et favorisant un rapprochement avec l'oral par leur dimension interlocutive, informelle et non préparée. L'auteur s'appuie dans son approche sur la Théorie des Opérations Prédicatives et Énonciatives et envisage les

opérations dont sont la trace les marqueurs linguistiques des deux périphrases pour développer son hypothèse quant à la valeur modale de désassertion.

Les deux périphrases ont une fonction métaénonciative, de commentaire sur le dire, inscrite dans le pronom *ça* en français, et l'aspect *be-ing* en anglais, les deux marqueurs ayant en outre une valeur de reprise et étant associés au verbe *dire/say*.

La valeur de modalisation (« rhetoric back-off ») au sens d'atténuation de l'assertion est nuancée dans les deux langues selon les marqueurs, respectivement les adverbes *rien* et *just*. Avec *rien*, la reprise du dire, *je dis ça*, est niée (*je ne dis rien*), ce qui conduit à la non prise en charge d'un contenu marqué par les conséquences sur l'asserteur. Avec *just*, le contenu repris (*I'm saying*) est envisagé dans le cadre d'un hiatus, d'une non-coïncidence entre ce qui a été dit et son interprétation. Ainsi l'assertion est modulée, la prise en charge est limitée (« minimal commitment »).

Le corpus permet de montrer le fonctionnement des deux périphrases dans des contextes non-assertifs pour exprimer la modalité épistémique (lorsqu'elles sont associées à des marqueurs de désassertion, *je suggère, it would seem*), la modalité appréciative (lorsqu'elles sont associées à des marqueurs de valuation, *c'est bidon, it's difficult*) ou encore la modalité radicale (lorsqu'elles sont associées à des marqueurs déontiques *pouvoir, should*). Si le corpus met en évidence un parallélisme entre les deux structures dans les emplois dits « anticipatoires », il met également au jour une différence d'emploi, celle d'ajustement, spécifique à l'anglais dans un contexte qui s'approche de la conversation en simultané et se caractérise par une désassertion des propos suite à une prise de position de l'interlocuteur à leur sujet. Cette particularité de l'anglais remet au premier plan les différences énonciatives des deux constructions, le présent français bloquant le commentaire sur une énonciation autre que celle en cours, ce que l'anglais, en revanche, peut construire avec le présent progressif.

Notes

[1] Membres du comité scientifique : Joasha Boutault (Université de Poitiers), Hélène Chuquet (Université de Poitiers), Jean Chuquet (Université de Poitiers), Jean-Charles Khalifa (Université de Poitiers), Ramon Marti-Solano (Université de Limoges), Raluca Nita (Université de Poitiers), Michel Paillard (Université de Poitiers), Jeanne Vigneron-Bosbach (Université de Caen). Les coordonnateurs de ce numéro tiennent à remercier leurs

collègues membres du comité scientifique de leur relecture attentive de l'ensemble des contributions.

[2] Nous remercions Sylviane Granger (Université Catholique de Louvain) de la conférence plénière, *Recherches sur corpus multilingues : Acquis, défis et perspectives*, donnée en ouverture de cette journée d'étude.

[3] Nita, R. et F. Valetopoulos, (dir.), 2018, *L'expression des sentiments : de l'analyse linguistique aux applications*. Rennes : Presses Universitaires.

[4] Une présentation des productions en matière de corpus de l'Équipe A est disponible sur le site du laboratoire FoReLLIS : <http://forellis.labo.univ-poitiers.fr/vie-du-laboratoire/corpus/>.

[5] Corpus littéraire et journalistique d'originaux et traductions (anglais, français allemand) compilé conjointement par l'Équipe A du laboratoire FoReLLIS et le CECL (*Centre for English Corpus Linguistics*) de l'Université Catholique de Louvain-la-Neuve, Belgique, sous la direction d'Hélène Chuquet et de Sylviane Granger.

[6] On peut néanmoins citer deux projets centrés sur le positionnement énonciatif mais dans un genre spécifique, l'écrit scientifique : le *Projet Scientext* (LIDILEM Grenoble 3, Licorn UBS-Lorient, LLS Chambéry) portant sur la constitution et l'analyse de corpus d'articles scientifiques en français et anglais et d'écrits universitaires en français langue étrangère relevant de disciplines variées ; Fløttum, K., Dahl, T. & Kinn, T. (dir.), 2006, *Academic Voices. Across Languages and Disciplines*, Amsterdam/Philadelphia, John Benjamin.

[7] Au sujet de la pénurie des données en traduction pour des langues autres que l'anglais et le français voir, par exemple, Johansson, S., 1998, « On the role of corpora in cross-linguistic research » in S. Johansson & S. Oksefjell (dir.) *Corpora and Cross-Linguistic Research: Theory, Method and Case Studies*. Rodopi, pp. 3-24.

[8] Méritent mention les corpus parallèles et comparables anglais-espagnol ACTRES de l'Université de Leon, Espagne et les corpus parallèles anglais-espagnol MULTINOT de l'Université Complutense de Madrid, Espagne.

[9] Avec l'aide technique de la Plateforme Numérique de la MSHS, Poitiers (Responsable, David Chesnet). Responsables scientifiques : Raluca Nita, Université de Poitiers, Laboratoire FoReLLIS E.A. 3816 et Ramon Marti Solano, Université de Limoges, Laboratoire FRED E.A. 6311. Membres du projet (2013-2015) : Université de Poitiers, E.A. 3816 FoReLL : Joasha Boutault (Maître de conférences), Hélène Chuquet (Professeur émérite), Raluca Nita

(Maître de conférences), Freiderikos Valetopoulos (Professeur), Jeanne Vigneron (Doctorante) ; Université de Limoges, E.A. 6311 FRED : Ramon Marti Solano (Maître de conférences) ; Université de Tours, E.A. 6297 ICD : Manuel Torrellas (Maître de conférences) ; Université de Chypre, Chypre : Efi Lamprou (Maître de conférences).

[10] « ...a comparable corpus and a translation corpus are combined within the same overall framework, and each type can be used to control and supplement the other » (Johansson, S., 2007, *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*, John Benjamins Publishing p. 11)

[11] Alignement du GRAFE académique dans le cadre du projet CRISTAL (2017-2018) – « Humanité numérique », soutenu par la MSHS Poitiers, porteurs : Laboratoire FoReLLIS et Laboratoire MIMOC – Université de Poitiers. Nous remercions Michaël Nauge, ingénieur d'étude au laboratoire FoReLLIS, de son aide technique et de sa collaboration à la réflexion autour du traitement du corpus.

[12] Notre soulignement.

[13] Paillard, D., 2009, « Prise en charge, commitment ou scène énonciative » in *Langue Française*, 2, pp. 109-208.

[14] Johansson, S., 1998, « On the role of corpora in cross-linguistic research » in S. Johansson & S. Oksefjell (éds.) *Corpora and Cross-Linguistic Research: Theory, Method and Case Studies*. Rodopi, pp. 3-24.

[15] Il est intéressant de noter que le corpus GRAFE dans son volet de textes académiques en anglais (environ 145 000 mots) fournit 13 résultats de *be seen as/to* à valeur cognitive. Nous donnons ici comme exemple l'une des traductions du corpus GRAFE qui n'est pas répertoriée dans le corpus de l'auteure : "These have on the face of them the look—or at least the grammatical make-up—of 'statements'; but nevertheless they are seen, when more closely inspected, to be, quite plainly, not utterances which could be 'true' or 'false'" (J. L. Austin, *How To Do Things With Words*, Oxford University Press, 1975 (1962)) vs « Ces énonciations ont l'air, à première vue, d'« affirmations' — ou du moins en portent-elles le maquillage grammatical. On remarque toutefois, lorsqu'on les examine de plus près, qu'elles ne sont manifestement pas des énonciations susceptibles d'être 'vraies' ou 'fausses'. » (J. L. Austin, *Quand dire c'est faire*, Traduction G. Lane, Seuil, 1991)

Pour citer ce document

Par Raluca Nita et Ramón Martí Solano, «Introduction», *Cahiers FoReLLIS - Formes et Représentations en Linguistique, Littérature et dans les arts de l'Image et de la Scène* [En ligne], Traces de subjectivité et corpus multilingues. URL : <https://cahiersforell.edel.univ-poitiers.fr/index.php?id=711>

Quelques mots à propos de : [Raluca Nita](#)

Université de Poitiers, FoReLLIS EA 3816

Quelques mots à propos de : [Ramón Martí Solano](#)

Université de Limoges, CeReS, EA 3648

Cet article est mis à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale 3.0 France. CC BY-NC 3.0 FR

