Parallèles/comparables, langue originale/traduite… : s'y retrouver dans le maquis des corpus potentiels pour la traduction économique et financière

Laurent Gautier, professeur des Universités, Centre Interlangues Texte Image Langage (UBFC, EA 4182)

Structure de la présentation

1. Problématique et objectifs
2. Terminologies, corpus numériques et TAO
3. Vers une terminologie de corpus
4. Quels corpus pour quels besoins ? Etude de cas à partir de la traduction économique et financière
5. Perspectives

1. Problématique et objectifs

Terminologie(s) et traduction à l'heure numérique

- Terminologie : rôle historique à la fois dans la formation et dans la pratique des traducteurs :
  - Extraction manuelle, puis semi-automatisée
  - Rôle des experts-valideurs
  - Rôle « fétiche » du « glossaire de traduction »
- Quelles évolutions possibles à l'heure des nouveaux environnements du traducteur ?
  - Terminologie *in vivo* (vs. *in vitro*)
  - Porte d'entrée vers les contenus spécialisés (compréhension sémantique) *et* la mise en discours (choix de formulation)
  - Apports des masses de données textuelles existantes

Du corpus à la terminologie à la traduction

- Vers une saisie holistique des terminologies sur des continuums lexique-grammaire ?
- Les corpus, de référence et *ad hoc*, comme ressources premières… et uniques ?
- Vers un re-examen, par les corpus, de l'opposition langue traduite vs langue originale :

« On attend du traducteur qu'il fournisse des textes traduits dont la langue soit **naturelle** (au sens d'**idiomatique**) et qu'il demeure invisible. L'**homogénéisation** entre langue originale et langue traduite contribue selon nous à atteindre cet objectif, la **linguistique contrastive** fournissant des informations capitales à prendre en compte. » (Loock 2016, 58-59)

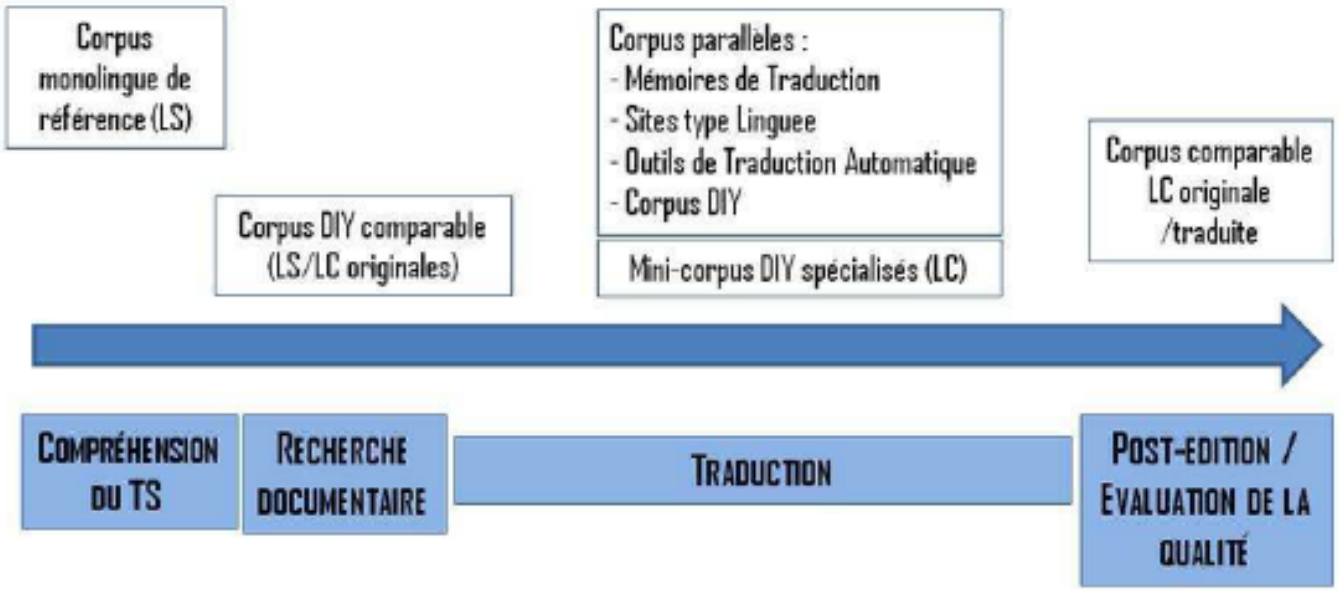
⇒ **Création, gestion et interrogation de corpus comme compétence à part entière du traducteur d'aujourd'hui**

2. Terminologies, corpus numériques et TAO

Prémisses

- Omniprésence explicite ou non des corpus dans les outils de TAO et environnement du traducteur : MT, TA, BD termino, sites grand public (*linguee*)
- Définition de base : ensemble de données langagières (écrites et / ou orales) organisées, répondant à un objectif d'exploitation et remplissant un certain nombre de conditions. Ces données sont ensuite préparées pour donner lieu à des traitements automatiques plus ou moins poussés
- Critère d'**authenticité** : donné la plupart du temps, mais à traiter avec précaution quand le corpus doit être recueilli
- Critère de **représentativité**, garant de la validité des résultats

- Conditions essentielles :
  - En adéquation totale avec le **besoin de traduction**, en lien avec le flux du projet (Loock 2017 : 19):



3. Vers une terminologie de corpus

Approche intuitive

- Corpus *adhoc* français construit avec WebBootCat sous Sketchengine
- Ce qu'apprend une grammaire lexicale locale

avalanche *(noun)*
avalanches freq = 3,630 (5,233.36 per million)

verbs with "avalanche" as object		11.68
déclencher +	148	12.69
déclencher une avalanche		
provoquer	22	10.41
provoquer une avalanche		
une	12	9.64
Une avalanche		
produire	10	9.39
enfouir	7	8.92
classer	6	8.82
survenir	6	8.79
prévenir	6	8.70
éviter	6	8.57
recenser	5	8.53
risquer	5	8.53
souiller	4	8.25
dévier	4	8.25
simuler	4	8.24
ces	4	8.19
voir	6	8.13
choisir	4	8.05
représenter	4	8.02
avoir	15	7.91

verbs with "avalanche" as subject		3.55
pouvoir	46	11.09
avalanche peut		
partir	7	10.54
emporter	5	10.22
concerner	5	10.07
venir	5	9.93
constituer	4	9.82
tuer	3	9.53
couper	3	9.51
impliquer	3	9.48
évoluer	3	9.34
aller	4	9.07
faire	4	9.00
rester	3	8.98

modifiers of "avalanche"		8.43
spontané	22	10.91
avalanches spontanées		
gros	27	10.73
grosses avalanches		
accidentel	15	10.53
% des avalanches accidentelles		
naturel	17	10.22
les avalanches naturelles		
coulant	11	10.13
avalanche coulante		
dense	9	9.77
petit	16	9.56
petites avalanches peuvent se produire		
poudreux	8	9.48
mortel	7	9.39
nouveau	12	9.37
nouvelles avalanches		
préventif	8	9.31
spectaculaire	6	9.27
mixte	6	9.23
nombreux	9	9.19
possible	7	9.17

adjective predicates of "avalanche"		1.27
endémique	3	10.97
fort	4	10.79
fréquent	3	10.77
important	4	10.44

"avalanche" is a ...		0.88
masse	7	12.51
utilisation	3	11.45
technique	3	11.34

prepositional phrases		
"avalanche" de	462	12.73
"avalanche" en	63	1.74
"avalanche" à	49	1.35
"avalanche" du	22	0.61
"avalanche" au	19	0.52
"avalanche" pour	15	0.41
"avalanche" par	10	0.28
"avalanche" avec	9	0.25
"avalanche" sans	3	0.08

"avalanche" and/or ...		5.59
taille	5	9.53
dva	4	9.16
sonde	5	9.15
risque	4	8.90
deglace	3	8.90
alpug	3	8.90
schweizerische	3	8.90
cemagref	3	8.89
pida	3	8.88
donnée	3	8.88
imprimerie	3	8.87
avalanche	6	8.85
prix	3	8.85
prévision	3	8.82
secours	4	8.79
neige	5	8.79
chute	3	8.71
sécurité	3	8.45

less "avalanche" than ...		0.11
volumineux	4	13.99

- Le corpus transforme l'approche de la terminologie pour intégrer le terme dans des segments de connaissances spécialisées « prêtes à traduire » :

  « The typical linguistic features of ESP **cannot be characterised as a list of discreet items** (technical terminology, the passive, hedging, impersonal expressions, etc.), rather the most typical features of ESP texts are **chains of meaningful interlocking lexical and grammatical structures**, wich we have called **lexico-grammatical patterns**. » (Gledhill/Kübler 2016 : 75)

  =>Les grammaires de construction et la suspension de la dichotomie lexique vs. grammaire

4. Quels corpus pour quels besoins ? Etude de cas à partir de la traduction économique et financière

Une opposition tellement classique…

qu'elle s'en est fossilisée Teubert (1996)

- « A 'parallel corpus' is a bilingual or multilingual corpus that contains one set of texts in two or more languages. » (Teubert 1996 : 245)
- Textes traduits / bi-textes
  - Tertium comparationis : équivalence supposée entre les textes
  - Préparation nécessitant une opération d'alignement (semi-)automatique
  - Exemple : conférence de presse de la BCE (8/an) : TS anglais vs. TC dans les 23 langues

Anglais	Français	Allemand	Néerlandais
Given high structural unemployment and low potential output growth in the euro area, a cyclical recovery along the lines of the March ECB staff projections is no grounds for complacency.	Au vu du niveau élevé de chômage structurel et de la faible croissance potentielle dans la zone euro, une reprise conjoncturelle telle que celle ressortant des projections de mars des services de la BCE **ne permet aucun excès de confiance.**	Angesichts der hohen strukturellen Arbeitslosigkeit und des geringen Wachstums des Produktionspotenzials im Eurogebiet **gibt** eine Konjunkturerholung wie in den von Experten der EZB erstellten Projektionen vom März **keinen Anlass zur Sorglosigkeit.**	Gezien de hoge structurele werkloosheid en de lage potentiële productiegroei in het eurogebied, **is** een conjunctuurgebonden herstel zoals geschetst in de door medewerkers van de ECB opgestelde projecties van maart **geen reden om achterover te leunen.**

Quels niveaux d'**équivalence** postulés (aveuglément ?) pour la comparaison ? :

-dénotative

-conntotative

-pragmatique

-textuelle

-stylistique/esthétique (Koller 1992)

- Utilisations principales :
  - recherche d'**équivalents** terminologiques (*Weltwirtschaft/économie mondiale)* : domaine plus ou moins autonome en terminologie (ATR) / Het herstel van de binnenlandse vraag
  - recherche de **collocations** *(la croissance du PIB réel/das Wachstum des realen Bruttoinlandsprodukts/de reële bbp-groei*)
  - recherche de traits **stylistiques** ? *était supérieur à la tendance/übertraf den Trend*
- « Inconvénients »/risques :
  - tributaire de la **finesse de l'alignement**
  - tributaire de la **qualité** de la traduction
  - quelle valeur pour les résultats en dehors de l'opération de traduction ? Cf. travaux de l'UMIST (M. Baker) sur le « style du traducteur »

- « 'Comparable corpora' are corpora in two or more languages with the same or similar composition. All copora have an explicit or implicit composition. The texts they contain can be classified according to a variety of intralinguistic or extralinguistic features. » (Teubert 1996 : 245) :
  - Tertium comparationis : à définir en fonction des objectifs à atteindre car seul **garant de l'homogénéité** du corpus constitué
  - → un tc inscrit au niveau textuel / discursif
- TC possibles :
  - un type de texte
  - un domaine thématique
  - une situation énonciative
- Apports essentiels :
  - analyse de deux langues « originales »
  - travail aux niveaux textuel et discursif
  - travail possible au niveau microlinguistique (ATR par exemple) mais sans appariement automatique

- « Inconvénients »/risques :
  - précision dans la définition du tc (elle-même tributaire du niveau d'analyse)
  - part d'idiosyncrasies « culturelles » (danger pour les appariements automatisés)
  - part importante des choix initiaux et des a priori (exemple des types de texte)

- Pour notre problématique :
  - **Double comparaison** des macro-structures
  - Augmentation du degré de **granularité** de la comparaison
  - Saisie du continuum lexique-grammaire « à sa **place macro-structurelle** »

Un corpus complexe

- Conférences de presse de la BCE traduites
  - Textes compilés depuis le site de la BCE - https://www.ecb.europa.eu
  - 8 interventions en 2015 et 8 en 2016
  - Extraction, alignement et analyse des textes en anglais, français, allemand et néerlandais
- Rapports trimestriels de la Banque Nationale Suisse
  - Textes compilés depuis le site de BNS - https://www.snb.ch/fr/iabout/pub
  - 4 par ans Extraction, alignement et analyse des textes en français et allemand
- Perspectives économiques de la Banque nationale de Belgique
  - Textes compilés depuis le site de BNB - https://www.nbb.be/fr/publications-et-recherche
  - 2 par an
  - Extraction, alignement et analyse des textes en français et néerlandais
- Bulletins de la Banque de France
  - Textes compilés depuis le site de la BdF https://publications.banque-france.fr/liste-chronologique/le-bulletin-de-la-banque-de-france + anglais
  - 6 par an

Un réseau de relations à démêler

- BCE : anglais langue originale + allemand / néerlandais / français traduit

⇒ double corpus comparable nécessaire : Banque d'Angleterre, *Bundesbank*, Banque de France

- Banque Nationale Suisse : co-rédaction dans les 3 langues + anglais traduit ?

⇒ triple corpus comparable nécessaire : Banques nationales allemande, italienne et BdF

- Banque nationale de Belgique : co-rédaction dans les 2 langues + anglais traduit ?

⇒ Double corpus comparable nécessaire : Banque de France et Banque néerlandaise

- Banque de France : français langue originale + anglais traduit

=> Quel traitement du plurilinguisme ?

Un saisie d'un « environnement » réel

- Focalisation sur le « tout terminologique » peut passer sous silence d'autres récurrences préférentielles fossilisées :
  - Inclues dans les instances participatives : *par comparasion / gegenüber / ten opzichte van* vs. *par rapport à / verglichen mit / vergeleken met*
  - Comme marquage aspectuel : *nach wie vor* vs. Verbe de durativité

- Les risques pesant sur les perspectives de croissance de la zone euro **demeurent** orientés à la baisse.
- In Bezug auf die Wachstumsaussichten des Eurogebiets überwiegen **nach wie vor** die Abwärtsrisiken, was insbesondere mit der erhöhten Unsicherheit in Bezug auf die weltwirtschaftliche Entwicklung sowie den allgemeinen geopolitischen Risiken zusammenhängt.
- De risico's voor de groeivooruitzichten van het eurogebied **blijven** voornamelijk van neerwaartse aard.

Une exigence « ethnographique » en termes de méthodologie

- Conditions de rédaction des textes pour la France
- Conditions de co-rédaction pour les banques « plurilingues »
- Conditions de traduction dans les banques nationales
- Conditions de traduction à la BCE : un cas de langue traduite « institutionnelle » ?

The European Central Bank (ECB) press conference is held monthly, about 45 minutes after the Minimum Bid Rate is announced. The conference is approximately an hour long and has two parts. Firstly, a prepared statement is read (= langue écrite et préparée, pôle "distance"), then the conference is open to press questions. The press conference examines the factors which affected the ECB's interest rate decision (= dimension rétrospective, narration) and deals with the overall economic outlook and inflation (= dimension spécialisée macro-économie / finances). Most importantly, it provides clues regarding future monetary policy (= dimension prospective avec part d'incertitude). High levels of volatility can frequently be observed during the press conference as press questions lead to unscripted answers. (https://www.investing.com/economic-calendar/ecb-press-conference-396)

5. Perspectives

- Un **corpus DYI** d'un nouveau genre articulant et dépassant/transcendant l'opposition parallèle vs comparable
- Ressource pour une **appropriation croisée** des contraintes entre
  - contenus (savoirs macro-économiques et financiers) => dimension terminologique « stricte » ;
  - forme (contraintes génériques, macrostructurelles) => dimension discursive/textuelle, dont la gestion de l'ambigüité ;
  - « style » (contraintes systématiques *et* de spécialité) => dimension micro-linguistique

  ⇒ Vers un **lexique-grammaire** d'un type de texte
  Nécessité de la comparaison avec corpus de référence ?

  ⇒ Degré de **validité** au delà de ce type de texte particulier pour la **traduction financière** ?

Merci pour votre attention !

Laurent Gautier (laurent.gautier@ubfc.fr)