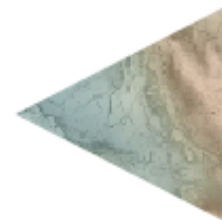


Le continuum lexique-grammaire en genre spécialisé à partir de corpus maison

9^e Colloque International de Linguistique de Corpus
Axe « Corpus, études contrastives et traduction »
31 mai 2017 - Université Paris Diderot



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

Cyril Nguyen Van, Ingénieur en analyse et structuration de données, Maison des
Sciences de l'Homme de Dijon (UBFC, USR CNRS-uB 3516)





Structure de la présentation

1. Mise en contexte : la communication de la BCE comme corpus parallèle de langue traduite institutionnelle
 2. Problématique
 3. Constitution et traitement du corpus
 4. Le continuum lexique-grammaire à l'œuvre: 3 études de cas
 5. Conclusion et perspectives
- 
- 

1. Mise en contexte

La communication de la BCE comme corpus parallèle de langue traduite institutionnelle

- Existence de travaux linguistiques sur le discours de la finance publique comme privée – dans des perspectives **essentiellement génériques** :
 - **Structuration discursive** globale du champ (Gautier, Ed., 2012; Palmieri/Rocci/Gautier, Eds, 2015)
 - Entrée par les **types de textes**, en particulier les **rapports annuels** (Jacobs 2003, Malavasi 2006, McLaren-Hankin 2008)
 - Entrée par les **fonctions discursives**, en particulier l'**argumentation** (Palmieri 2014 ; Rocci/Kudrautsava 2015)
 - Entrée par l'**ambiguïté** (Gautier/Nguyen Van 2016 ; 2017)
 - Entrée par le **niveau micro-linguistique** : métaphores (van der Yeught 2007), les schémas (Bloor/Pindi 1990), la modalité (Rocci 2012 ; Palmieri/Miecznikowski-Fuenfschilling 2016)

- Intérêt interdisciplinaire croissant pour les questions de communication des banques centrales (Allard *et al.*, 2013, Siklos 2015, Velthuis 2015, Kahveci/Odabas 2016) :
 - **Décryptées** et **analysées** par les spécialistes financiers, en particulier en termes de **tonalité** supposée « mesurable » (*certainty, optimism, realism* comme catégories d'analyse chez Kahveci/Odabas 2016)
 - Focus sur la dimension **performative** de ce discours
 - Dimension pragma-sémantique au centre des préoccupations compte tenu du caractère supposé **ambigu / implicite** de ces discours
 - ⇒ Cas de la BCE intéressant pour sa **dimension plurilingue**, à l'intersection entre **discours spécialisé** (macro-économie, finance) et discours **institutionnel**
 - ⇒ Forte tendance au **figement multi-niveaux**
 - ⇒ Enrichissement de corpus parallèles comme EU ACQUIS corpus ou UNCorpus

- Objectifs de la **Banque Centrale Européenne** (BCE)
 - Maintenir la **stabilité** des prix au sein de la zone euro
 - Sauvegarder la stabilité financière
 - Promouvoir l'intégration financière en Europe
- **Communication institutionnelle à visée performative** :
 - Outil important de la **politique monétaire européenne**
 - Mutation dans son **rôle**, devenu **majeur**, suite à la crise
 - Nouvelles **voies de communication** mise en œuvre : publications de comptes rendus, de projection économique, conférence de presse télévisées, etc.

« Within my mandate, the ECB is ready to **do whatever it takes** to preserve the euro. And believe me, **it will be enough** » (Draghi, cité par Siklos 2015:1) => figement **certes, mais** marge d'interprétation

- Outil de communication majeur de la BCE étudié ici : la **conférence de presse**, 8 par an.
- Depuis 2015 : conférences de presse en **anglais** traduites dans les 23 langues de l'UE
- Auto-présentation du type de texte par la BCE :

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>)

2. Problématique

- Rôle fondamental des **corpus** dans **l'approche des discours spécialisés** (Gledhill/Kübler 2016 : 75) :
« 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**, which we have called **lexico-grammatical patterns**. »
- ⇒ Impacts pour notre corpus de travail :
 - Dépassement de la perspective générique pour une focalisation sur **un niveau micro-linguistique « intégratif »**
 - Dépassement de la perspective terminologique stricte
 - Mise au jour de l'articulation entre **formes morphosyntaxiques privilégiées** et **contenus sémantiques**
- Redéfinition d'une « **idiomaticité de corpus** » (corpus-based idiomaticity)

- **Omniprésence des corpus**, sous une forme plus ou moins transparente, dans tout l'**environnement** du traducteur (Loock 2016) :
 - En arrière-plan des **bases de données terminologiques** (exemples, contextes, etc.)
 - => plutôt corpus comparables
 - Sur des **sites** type *linguee.com, tradooit ou reverso context*
 - Systèmes de **TA**, en particulier SMT (*statistical machine translation*)
 - **Mémoires de traduction**
 - ⇒ Plutôt corpus parallèles
- **Apprentissage des techniques** de gestion de corpus par les futurs traducteurs pour une gestion autonome, par delà les corpus de référence à partir de corpus DIY (Varantola 2003, Mahadi *et al.* 2010)

=> Notre angle d'attaque ici.

- Les **corpus ad hoc** pour les traducteurs (Loock 2016) :
 - Utilisés comme **banque de données linguistiques** exploitables
 - Fournissent des outils sur l'usage de la langue et le processus de traduction :
 - Amélioration de la fluidité des traductions
 - Recherche d'inspiration
 - Evaluation de la qualité
 - Observation de langue source
 - Et en moins répandu : observation de la langue cible à travers l'exploitation de corpus DIY spécialisés maison en langue cible :
 - Choix **terminologiques, phraséologiques et grammaticaux des experts** du domaine



Qu'apprendre d'un mini-corpus comme celui des traductions des conférences de presse de la BCE ?




- Une approche en termes de **langue de traduction** : travail sur les trois corpus FR, DE, NL *indépendamment* du corpus-source EN
- Une **approche fonctionnelle** mettant en regard les **contenus** (domaine spécialisé), les **visées discursives** (dont la gestion de l'ambigüité) et l'utilisation du **système linguistique cible** (la « langue »)
- Dernier niveau (système linguistique) envisagé sur un **continuum lexique-grammaire**

=> **Caractérisation *sui generis*** d'un genre à envisager, dans un second temps, par rapport aux **corpus de référence**



3. Constitution et traitement du corpus

- 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
- Nombre de mots :
 - 21222 (texte original en anglais)
 - 25502 en français
 - 21143 en allemand
 - 22789 en néerlandais




- 
- Analyse **quantitative** globale, puis ciblée pour identifier les segments récurrents **et** les structures syntactico-sémantiques récurrentes
 1. Analyse **lexicométrique globale** (TXM, TreeTagger) : fréquences brutes, par classe de mots, par lemme, par n-grams (AntConc)
 2. **Alignement** des segments pertinents (LF Aligner) pour comparaison inter-langue
 - Analyse **qualitative** comparée des récurrences avec mise en regard avec les fonctions discursives
- 
- 

Alignement de segments sur un extrait contenant le terme "growth"

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.

Extraction des 20 premières formes les plus fréquentes dans chaque langue

Anglais	F	Français	F	Allemand	F	Néerlandais	F
policy	122	euro	115	Maßnahme	70	eurogebied	93
rate	106	politique	115	Erholung	48	Groei	60
inflation	101	zone	92	Umsetzung	43	Herstel	60
area	97	mesure	84	Analyse	41	jaarbasis	58
euro	96	tau taux	84	Verbesserung	38	monetairbeleidsmaatregelen	51
price	87	croissance	81	Unternehmen	35	Raad	49
growth	80	cour cours	79	Euroraum	33	Inflatie	47
measure	68	terme	73	Haushalt	32	Vraag	45
recovery	62	inflation	65	Ankauf	31	termijn	44
loan	53	reprise	54	EZB-Rat	31	lening	43
oil	46	prêt	50	Wachstum	29	analyse	40
term	42	Conseil	49	EZB	28	huishouden	33
analysis	41	gouverneur	49	Projektion	28	tenuitvoerlegging	33
credit	40	prix	49	Entwicklung	27	olieprijs	32
Governing	38	évolution	48	Jahr	27	bedrijf	30
implementation	38	crédit	44	Inflationsrate	26	voorzicht	30
purchase	38	œuvre	43	Experte	23	hervorming	29
development	36	pétrole	43	Sektor	23	daling	26
outlook	36	mise	41	Einklang	21	ontwikkeling	26
demand	34	achat	40	Ölpreis	21	verbetering	26

- 
- **Extraction** par **classe** de mots :
 - pas suffisante pour l'analyse de la catégorie *en traduction* dans la mesure où les langues en présence ne « travaillent » pas forcément avec les mêmes classes de mots
 - Exemple sur le terme *policy* :
 - En anglais : *monetary policy*,
 - En français : *politique monétaire*
 - En allemand : adjectif *geldpolitischen*
 - En néerlandais : composé *monetairbeleidskoers*
- 
- 

4. Le continuum lexique-grammaire à l'œuvre :

3 études de cas

Une perspective intégrative du discours spécialisé autour du figement

- Perspective **top-down** en termes de **moule discursif multi-niveaux** (Gautier 2009) : intention/fonctions – contenus sémantiques – développement thématique (double linéarisation) – terminologie/phraséologie
- Figement réinvesti en termes de **réurrences** – **fossilisation => stabilité** (Feilke 1996) – **formulations préférentielles, puis exclusives**
- Saisie du moule discursif comme concaténation de **répertoires restreints** : lexique/terminologie + contraintes de linéarisation + codes grammaticaux




Terminologie et collocations conceptuelles

- « Termes » traditionnels = nœuds de structures conceptuelles organisant le domaine cognitif de référence « traduits » en langue par des mises en mots récurrentes analysables en chaînes prédicat-arguments

La syntaxe formelle limite la distribution des mots sur la base de schémas de bonne formation des expressions complexes qui ne sont pas sensibles au contenu. La syntaxe des concepts limite la distribution des mots sur **la base de procès et d'états de choses cohérents et cognitivement adéquats**, qui fournissent à la description du contenu des mots une dimension relationnelle supplémentaire. (Prandi 1998 : 36, souligné par nous)




- ~ *frame-based terminology* (Faber 2012 *et. al.*)
- 
- 

- 
- Exemple d'un terme « banal » : PIB / BIP / BBP
 - Emploi restreint à des structures de comparaison réalisant elles-mêmes un scénario de comparaison sous-jacent à tout le moule discursif :

(01 a) Ces perspectives ressortent globalement des projections macroéconomiques de mars 2016 établies par les services de la BCE pour la zone euro, qui tablent sur **une hausse du PIB annuel en volume** de 1,4 % en 2016, de 1,7 % en 2017 et de 1,8 % en 2018.

(01 b) Dieser Ausblick deckt sich weitgehend mit den von Experten der EZB erstellten gesamtwirtschaftlichen Projektionen für das Euro-Währungsgebiet vom März 2016. Den dort enthaltenen Berechnungen zufolge wird **das jährliche reale BIP** Ø 2016 um 1,4 %, Ø 2017 um 1,7 % und Ø 2018 um 1,8 % **steigen**.

(01 c) Deze vooruitzichten komen in grote lijnen ook naar voren in de in maart 2016 door medewerkers van de ECB opgestelde macro-economische projecties voor het eurogebied, die voor 2016 **een stijging van het reële bbp op jaarbasis** voorzien van 1,4%, van 1,7% voor 2017 en van 1,8% voor 2018.

- 
- Trois instances participatives du scénario avec contraintes respectives de mise en mots :
 - **Comparé** = comparant à deux états chronologiques différents :
 - Indicateur
 - Terme brut + qualificatifs propres aux 3 langues : *reales BIP* / *reëlle BBP*
 - **Valeurs absolues** de départ et d'arrivée
 - Explicitées ou non
 - **Direction du mouvement**
 - Alternance V (*steigen*) / N (*hause, stijging*) + structure argumentale correspondante
 - **Valeur relative**
 - Préposition
 - **Période de référence**
 - Préposition
- 
- 

Emploi de N prédicatifs associés à des V supports préférentiels non prédictibles

- Certains « pseudo-termes » sont des Npred de portée « générale » nécessitant l'emploi des Vsup appropriés
- Exemple de *mesure (de politique monétaire) / (geldpolitische) Massnahmen / (monetairbeleids)maatregelen* => Combinatoire restreinte à un répertoire de Vsup + argument temporel

(2 a) Globalement, **les mesures de politique monétaire** en place depuis juin 2014 ont nettement amélioré les conditions d'emprunt pour les entreprises et les ménages ainsi que les flux de crédits dans la zone euro.

(2 b) Insgesamt haben **die seit Juni 2014 ergriffenen geldpolitischen Maßnahmen** eindeutig die Kreditbedingungen für Unternehmen und private Haushalte wie auch die Kreditströme im gesamten Euroraum verbessert.


(2 c) Over het geheel genomen hebben **de door ons sinds juni 2014 genomen monetairbeleidsmaatregelen** de kredietnemingskosten voor bedrijven en huishoudens en de kredietstromen in het eurogebied duidelijk verbeterd.

- 
- Non-traduction de verbes supports d'une langue à l'autre

(3 a) Deuxièmement, nous avons décidé de porter **les achats mensuels effectués** dans le cadre de notre programme d'achats d'actifs de 60 milliards d'euros à l'heure actuelle à 80 milliards d'euros.

(3 b) Zweitens haben wir beschlossen, **das Volumen der monatlichen Ankäufe Ø** im Rahmen unseres Programms zum Ankauf von Vermögenswerten von 60 Mrd € auf 80 Mrd € auszuweiten.

(3 c) Ten tweede hebben wij besloten **de maandelijkse aankopen 0** in het kader van ons aankoopprogramma voor activa te verhogen van €60 miljard naar €80 miljard.



Structures récurrentes en termes de routines discursives

- 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*

(4 a) Les risques pesant sur les perspectives de croissance de la zone euro **demeurent** orientés à la baisse.

(4 b) 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.

(4 c) De risico's voor de groeivoorzichten van het eurogebied **blijven** voornamelijk van neerwaartse aard.

5. Bilan et perspectives

- Un **corpus DIY** « inédit » et simple à compiler pour un (apprenti) traducteur
- Vers un **lexique-grammaire** d'un type de texte (hautement contraint) :
 - Intégrant la terminologie dans des *structures prédicat-arguments* reflets de la grammaire conceptuelle du champ ;
 - Listant les répertoires de *formes morpho-syntaxiques* préférentielles associées ;
 - Associant le lexique de liaison vu aussi en termes de fossilisation.
=> *Des constructions (CxG) spécialisées ? (Gautier 2017)*
- Degré de **validité** au delà de ce type de texte particulier pour la **traduction financière** ?
- **Comparaison** inévitable avec des corpus de **langue native de référence** : publications périodiques des banques nationales.

Merci pour votre attention !

- **Laurent Gautier** (laurent.gautier@ubfc.fr)
- **Cyril Nguyen Van** (cyril.nguyen-van@ubfc.fr)