

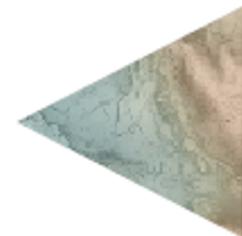
Wie vergleichbar sind vergleichbare und
parallele Fachkorpora?
Ergebnisse einer Pilotstudie zum
Sprachenpaar dt./frz. in Finanzdiskursen

IVG-Kongress 2020/2021

Sektion C11 – Kontrastive Korpuslinguistik



Università degli Studie di Palermo

Prof. Dr. Laurent Gautier, Centre Interlangues
Texte Image Langage (UBFC-Dijon, EA 4182)





Gliederung

1. Kontextualisierung
 2. Problemstellung
 3. Datenset
 4. Fallstudien zur Verwendung des Terminus
Inflation
 5. Rück- und Ausblick
- 
- 

1. Kontextualisierung

***Corpus-turn* in der (kontrastiven) Fachsprachenforschung?**

- „**Kontrastive Fachsprachenforschung**“ als Selbstverständlichkeit spätestens seit Baumann/Kalverkämper (1992):

Gerade hieraus (= aus der Selbstverständlichkeit des Begriffs und des Anliegens der k. L., LG) rechtfertigt es sich, mit dem Begriff „k. L.“ notwendige Eigenständigkeiten beim Selbstverständnis, in der Methodik, in den Analysezielen und bei der praktischen Umsetzung zu erfassen; das macht seine Attraktivität aus (...).“ (Baumann/Kalverkämper 1992, 9)

- Längere Tradition in der praktischen Terminologielehre: Textkorpora als Basis für die **Erschließung von Termini**
- Erneuerte Rolle seit der Einführung korpuslinguistischer Methoden:
 - Quantitativ: Bearbeitung **größerer** Datenmengen
 - Qualitativ:
 - Bearbeitung **neuer Texttypen**
 - Integration **neuer Dimensionen**, insbesondere zur morpho-syntaktischen Einbettung von Termini in Sätze

Korpora sind im Fachübersetzungsumfeld allgegenwärtig

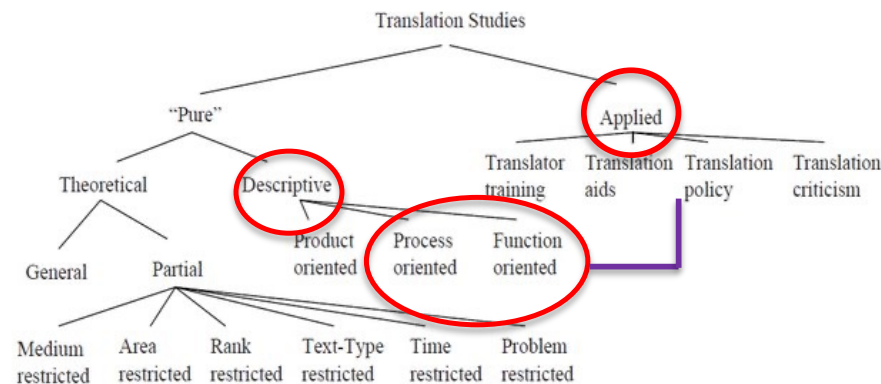
- Enge Beziehungen zwischen kontrastiver Linguistik und **Fachübersetzung**, und darüber hinaus auch eine allgemeinere „(...) zunehmende Konvergenz von Translationswissenschaft und Fachsprachenforschung“ (Ahrens *et al.* 2019, 9)
 - Somit wurden (mehrsprachige) Korpora – explizit oder implizit – ins Zentrum der Übersetzungsarbeit (Loock 2016, Krüger 2019) gerückt:
 - Im Hintergrund von **Terminologiedatenbanken** bzw. Wissensmanagementsysteme: als Ressourcen für Felder wie Definitionen, Kontext, Kollokationen, usw... => Eher **Vergleichskorpora**
 - Auf **Webseiten** wie *linguee.com*, *tradoo.it* oder *reverso context*
 - Für **MÜ-Systeme**, insbesondere SMT (*statistical machine translation*) und heutzutage immer mehr neuronale maschinelle Übersetzung
 - In **TM-Systemen**
- ⇒ Eher **Parallelkorpora**
- ⇒ Interessantes System von zu diskutierenden Relationen: kontrastive (Korpus-)Linguistik, Ü-wissenschaft, Fachübersetzung, Vergleichs- und Parallelkorpora, Äquivalenz

2. Problemstellung

Forschungsfrage 1: welcher Platz für kontrastive (Korpus-)Linguistik in der Ü-wissenschaft? Eine Gretchenfrage (Gautier/Schreiber/Varga 2020)

- 4 geteilte, aber konkurrierende Achsen (Gautier 1998, 2001):
 - Sprachsystem vs. **Sprachgebrauch**: kontrastive Korpuslinguistik ist eindeutig auf der **Verwendungsebene** anzusiedeln vs. traditionelle „Translationslinguistik“ (Jäger 1975, 172ff.), die „Äquivalenzbeziehungen zw. zwei Sprachen zu beschreiben hat“ (Koller/Henjum 2020, 152)
 - Sprache als Produkt vs. **Sprache als Prozess**: neuere kognitive Ansätze sowohl in der Fachsprachenforschung (Schubert 2011, Kastberg 2019) als auch in der Ü-wiss. (Göpferich 2008, Alves/Jakobsen 2020) plädieren für einen **dynamischen Vergleichsbegriff**

Ursprüngliche Aufteilung der *Translation Studies* bei Holmes (1970)



Forschungsfrage 2a: welcher Grad von Vergleichbarkeit zw. Vergleichs- und Parallelkorpora?

- Trad. Kategorisierung **mehrsprachiger Korpora** bei **Teubert** (1996) als Ausgangspunkt (vs. Baker 1993) => Definitionen von Lemnitzer/Zinsmeister (2010, 104):
 - „*Parallelkorpora* bestehen aus Texten in einer Sprache S1 und deren **Übersetzungen** in die Sprache(n) S2...Sn. Die Textteile, z.B. Absätze oder Sätze, werden dabei einander zugeordnet (*aligniert*).“
 - „*Vergleichskorpora*, in denen Texte mehrerer Sprachen S1... Sn zu **vergleichbaren Diskursbereichen** erfasst sind, die aber keine Übersetzungen voneinander sind.“ => (extralinguistisches) *tertium comparationis* als wichtigstes Relevanzkriterium (Gläser 1992)
- => Auswirkungen auf das zu kontrastierende **Sprachmaterial**: originale (= nicht übersetzte) vs. übersetzte Sprache



Forschungsfrage 2b: welcher Grad von Vergleichbarkeit zw. beiden A- und Z-Sprachen?

- Druck der *Übersetzungsuniversalien* (Baker 1993, 1995):
„linguistic features which typically occur in translated rather than original texts and are thought to be independent of the influence of the specific language pairs involved in the process of translation“ (Baker 1993, 243)
⇒ Traditionelle Liste seit Baker (1996): Vereinfachung, Explizierung, Normalisierung und Nivellierung
- Risiko, *translationese* (Gellerstam 1986) mit dem Funktionieren der Zielsprache zu verwechseln:
„the set of ‚fingerprint‘ that one language leaves on another when a text is translated between the two“
⇒ Systematische Ergänzung von Parallel- mit Vergleichskorpora, im Sinne einer methodologischen Triangulation (Baroni/Bernardini 2005)

Forschungsfrage 3: welche linguistischen Ebenen können auf dieser Basis am effizientesten kontrastiert (und modelliert) werden?

- Fachdiskurse unterscheiden sich primär **quantitativ** von nicht fachlichen Diskursen. (Roelcke 2010)
- „Traditionelle“ Merkmale von Fachdiskursen können nicht mehr isoliert behandelt werden, sondern viel mehr als **Muster**:
« 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**. » (Gledhill/Kübler 2016 : 75)
⇒ Neudefinition einer fachspezifischen „**Korpus-Idiomatizität**“ (*corpus-based idiomaticity*)
⇒ Erstellung fachspezifischer integrativer „**Mikro-Grammatiken**“ (*integrativ = Lexik + Grammatik*)



3. Datenset

- Wahl eines hoch **spezialisierten** und **erstarnten** Textmusters (Gautier 2009) innerhalb eines **hoch spezialisierten Faches** (Roelcke 2010):
 - Horizontale Gliederung: Volkswirtschaft + Finanzen
 - Vertikale Gliederung: Expertenkommunikation
 - Grad der Lenkung (Schubert 2014, 2019): hoch bis sehr hoch

„Die Fachkommunikation ist eine Kommunikationsform, die durch Optimierung gekennzeichnet ist. Die Optimierung besteht in bewusstem lenkendem Eingreifen, das sich entweder unmittelbar auf eine konkrete kommunikative Handlung richtet oder das bei den Kommunikationsmitteln ansetzt und dadurch mittelbar auf das kommunikative Handeln einwirkt.“ (Schubert 2019, 17)
 - Textmuster: **Konjunkturbericht**
- 
- 

Parallelkorpus:

- Pressekonferenzen der EZB: Originalfassung auf Englisch + Übersetzungen in den 23 EU-Sprachen, aligniert
- 8 Pressekonferenzen, 2015-2020

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.

- Umfang:

<u>Sprache</u>	<u>Jahr</u>	<u>Tokens</u>
EN	2015	11326
EN	2016	10665
EN	2017	9085
EN	2018	9413
EN	2019	10243
EN	2020	14483
EN	total	65215

<u>Sprache</u>	<u>Jahr</u>	<u>Tokens</u>
DE	2015	11364
DE	2016	10559
DE	2017	9169
DE	2018	9510
DE	2019	10229
DE	2020	14414
DE	total	65245

<u>Sprache</u>	<u>Jahr</u>	<u>Tokens</u>
FR	2015	13715
FR	2016	12693
FR	2017	10801
FR	2018	11189
FR	2019	12408
FR	2020	17203
FR	total	78009

Vergleichskorpora (als Kontrollkorpora):

- DE: **Monatsberichte** der Bundesbank
 - Gesamtumfang: 1949-2019
 - Ausgewertetes Teilkorpus: 252 Textexemplare => 20 Mio. Tokens
- FR: **Rapports mensuels** der *Banque de France*
 - Gesamtumfang: 1994-2018: 300 Textexemplare => 5,5 Mio. Tokens
- Tertia comparationis:
 - **Serielle** Publikation mit stark **erstarrender Makrostruktur**
 - Schnittstelle mehrerer „traditioneller“ **Fächer**: VWL, Statistiken, Finanzwissenschaften, usw.
 - Vergleichbare **Position** im gesamten Finanz-**Diskursnetzwerk**:
 - Technische Studien der Bundesbank
 - Wissenschaftliche Publikationen
 - Medienberichterstattung
 - Regierungserklärungen
 - Parteipolitische Stellungnahmen...

4. Fallstudien zur Verwendung des Terminus *Inflation*

- **Termini** als Schlüsselstein von Fachtexten können nicht mehr isoliert betrachtet werden => keine Frage der reinen Äquivalenz, sondern eine Frage von **Gebrauchsmustern**
- **Termini** heute als Eckpunkte konzeptueller Strukturen aufgefasst, um die das Fachwissen organisiert ist

Inflation: Schulbeispiel eines volkswirtschaftlichen Terminus

Term: Inflation

Term reference: EZB->Lehrmaterial <https://www.ecb.europa.eu/ecb/educational/hicp/html/index.de.html> (4.7.16)

Term type: term

Reliability: ★★★

Note: ANDERS: monetäre Inflation (= bewusst gesteuerte Ausweitung der Geldmenge durch Zentralbanken), zu deren Folgen die Preisinflation (s. Term 2) gehört

Contexts:

Term in context: Inflation ist für die klassische Denkschule der Ökonomie, der sich die meisten der führenden Wirtschaftswissenschaftler zurechnen, nichts als ein Anstieg der allgemeinen Konsumentenpreise einer Volkswirtschaft als klassische Schulen definiert die österreichische Schule der Nationalökonomie Inflation als ungezielte Ausweitung der Geldmenge, also als eine *monetäre Inflation*

Context reference: Hayek Institut <http://www.hayek-institut.at/you-said-inflation-ein-notwendiger-terminologischer-exkurs-inflationpolitik-der-ezb-besser-verstehen-und-kritisieren-zu-können.html> (4.7.16)

Owner: COM

Definition: augmentation générale et durable des prix des biens et des services, d'où une perte du pouvoir d'achat de la monnaie

Definition reference: Conseil-FR d'après :
- Eurostat, <http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Inflation/fr>
- Institut national de la statistique et des études économiques, <http://www.insee.fr/fr/methodes/default.asp?page=definitions/inflation.htm> [30.6.2016]

Note: À ne pas confondre avec *inflation monétaire* [IATE:839411]. Le terme "inflation" signifie alors augmentation de la masse monétaire; selon le courant monétariste, c'est l'augmentation de la masse monétaire qui est la cause unique de la hausse des prix.
Aujourd'hui, cependant, on entend par inflation hausse généralisée du niveau des prix.
Voir également:

- désinflation [IATE:1104508]
- déflation [IATE:1104509]
- stagflation [IATE:1071621]
- érosion monétaire [IATE:784083]

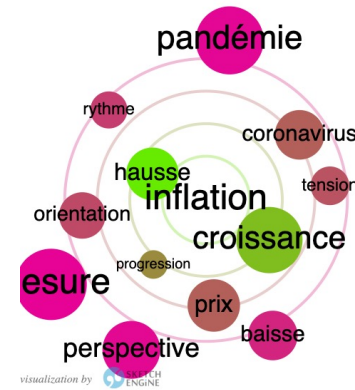
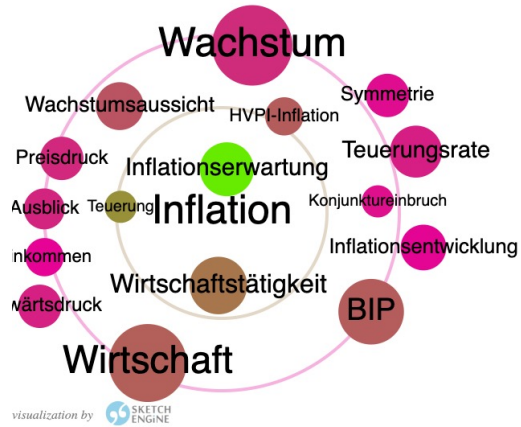
Owner: COM

- Hypothese 1: **terminologische Struktur** bleibt in beiden Korpusarten **gleich**
- ⇒ Erklärung: stabile **ontologische Architektur** des Fachs + **normierte** internationale Terminologie
- Aufgedeckte Tendenzen:
 - Hypothese bestätigt für stark terminologisierte Bezeichnungen: *Inflationsrate/taux d'inflation*
 - Größeres Repertoire an Formen in den Originalkorpora, mit systembasierten Unterschieden: Komposita im dt. vs. NPs im frz., dazu mit zwei Mustern: N Präp *inflation* // *inflation* Adj

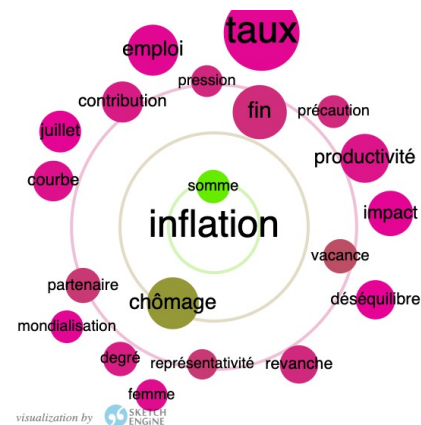
Inflationsrate, - erwartungen, -ziel, - unterschied, -risiko, - risikoprämie		
	mesure, anticipation, niveau, taux d'inflation	
		Inflation cumulée, annuelle, trimestrielle,

– Thesaurus-Funktion unterstreicht in den vier Fällen **starke Divergenzen**

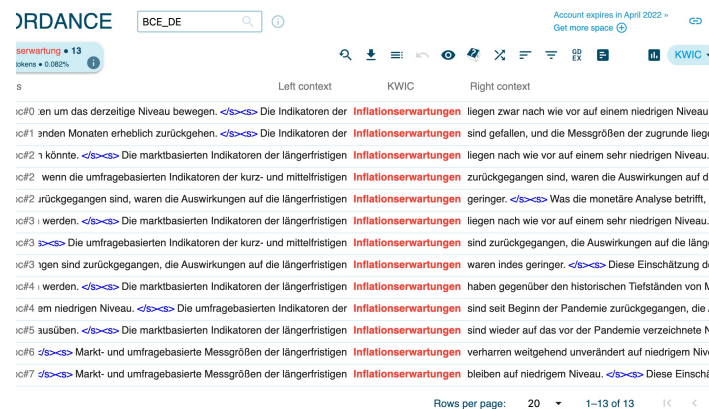
- Parallelkorpus



- Vergleichskorpora



- Aufgedeckte Tendenzen:
 - Einige Idiosynkrasien im Parallelkorpus im Vergleich zu den Originaltexten, stärker repräsentiert im Deutschen
 - *Inflationsschwäche* (im Vergleichskorpus nicht verwendet)
 - *Inflationsentwicklung* (einmalig verwendet im Vergleichskorpus)
 - *Inflationserwartungen* vs. *-aussichten* (im Vergleichskorpus nicht verwendet), mit unterschiedlicher Distribution : *I-erwartungen*: 0,082 vs. *I-aussichten*: 0,094



Left context	KWIC	Right context
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
<S> In Anbetracht der nach wie vor verhaltenen	Inflationsaussichten	ist es zugleich erforderlich, weiterhin für einen längeren i
sbedingungen im Einklang mit Änderungen der	Inflationsaussichten	entwickeln. <S> In jedem Fall ist der EZB-Rat nach v
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
it erhebliche Abwärtsrisiken für die kurzfristigen	Inflationsaussichten	. <S> Was die monetäre Analyse betrifft, schwächte
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
zwärtskorrektur der Wirtschaftstätigkeit und der	Inflationsaussichten	gerechnet, wobei das Basisszenario mit einem außerger
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
lick auf ihre Implikationen für die mittelfristigen	Inflationsaussichten	. <S> Der EZB-Rat ist nach wie vor bereit, alle seine
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
leiben werden, bis wir feststellen, dass sich die	Inflationsaussichten	in unserem Projektionszeitraum deutlich einem Niveau a
möglichen Auswirkungen auf die mittelfristigen	Inflationsaussichten	auch weiterhin beobachten. <S> Wir sind daher nach
Projektionen vom September 2020 wurden die	Inflationsaussichten	für die Jahre 2020 und 2022 nach unten korrigiert. <S>



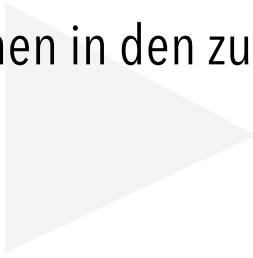
=> Über die Termini(-äquivalente) hinaus: Ü-Segmente (siehe 2. Erklärung)

- Potentielle Erklärungen:
 - Unterschiede in den repräsentierten Wissenssegmenten, aber angesichts des chronologischen Umfangs der Vergleichskorpora nicht sicher
 - Lenkungsrichtlinien bei der institutionellen Übersetzung der ZEB

- Hypothese 2: Stärkere **Lenkung** der Redaktion in den übersetzten Texten spiegelt sich in **längeren fossilisierten Segmenten** wider
- ⇒ Erklärung: Druck von **Richtlinien** + **Übersetzungstools**, die ihrerseits auch korpusbasiert sind und zu **Instrumenten der Lenkung** werden
- Aufgedeckte Tendenzen:
 - N-Gramme mit blockierter Position für *Inflation* sind viel aussagekräftiger im Parallelkorpus
 - Segmente als Ü-Einheiten, welche über den Terminus hinaus als integrierte Patterns fungieren:

Word
1 Gesamtinflation in den kommenden Monaten
2 die Gesamtinflation in den kommenden
3 dürfte die Gesamtinflation in den
4 in Deutschland dürfte die Gesamtinflation
5 Öl dürfte die Gesamtinflation in
6 für Öl dürfte die Gesamtinflation
7 sich die Gesamtinflation in den
8 dürfte die Gesamtinflation bis Anfang
9 dürfte sich die Gesamtinflation in
10 Deutschland dürfte die Gesamtinflation bis

- * Teil eines Vergleichsframes (*cf. Infra*)
- * gelenkte Enkodierung aller Frame-Elemente
- * textmuster-spezifische Enkodierung bestimmter Aspektualitäts- und Modalitätsdimensionen (hier: epistemische Modalität (Laifa/Gautier/Cruz 2021a, 2021b))

- 
- Drei beteiligte Instanzen (FEs), die alle bestimmten Restriktionen unterliegen:
 - **Vergleichsbasis** = Komparandum zu 2 unterschiedlichen Zeitpunkten:
 - Indikator (BIP)
 - Terminus mit Restriktionen in den zugelassenen Adjektiven: *reales BIP*
 - **Ausgangs- und Zielwerte**
 - Nicht immer erwähnt
 - **Bewegungsrichtung**
 - Restriktionen in der Abwechslung von V (*steigen*) + entsprechende Argumentstruktur
 - **Relativer Wert**
 - Restriktionen in den zugelassenen Präpositionen
 - **Vergleichsperiode**
 - Restriktionen in den zugelassenen Präpositionen
- 
- 

5. Rück- und Ausblick

- Wie kann kontrastive Korpuslinguistik zur Erstellung integrativer fachspezifischer Repertoires führen?
 - Durch die systematische **Triangulation** beider **Korpus**typen;
 - Durch die Assoziation aller **zugelassenen und fossilisierten morpho-syntaktischen Ausdrucksformen**, die global als **lenkungsbedingte Restriktionen** aufgefasst werden können;
 - Durch einen **holistischen** Ansatz, der **Termini** zwar berücksichtigt, aber „nur“ als **Bestandteil von Patterns**.
- Einsetzbarkeit:
 - **Tools** für Fachübersetzung und technische Redaktion
 - Didaktische Aufbereitung für den **Fremd(fach)sprachenunterricht**



Vielen Dank für die Aufmerksamkeit!

Laurent Gautier (laurent.gautier@ubfc.fr)