



Fachhochschule Köln
Cologne University of Applied Sciences



The Translation Classroom: Technology, Strategies and Employability

Working with Corpora in the Translation Classroom

Ralph Krüger M.A.

PhD Student in Translation Studies

Cologne University of Applied Sciences/University of Salford



Some definitions of the term “corpus”

- „corpus now means primarily a collection of texts held in machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways ...“ (Baker 1995:225)
- “a corpus is by default assumed to be a collection of texts in electronic format which are processed and analyzed using software specifically created for linguistic research.” (Zanettin 2012:7).
- “A corpus can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria.” (Bowker and Pearson 2002:9)



Corpus typology in translation studies

- Parallel/translation corpora
 - consist of a set of texts in one language and their translations
 - may be bilingual or multilingual (e.g. translations into different languages)

- Comparable corpora
 - consist of comparable (original or translated) texts (usually the same genre, text type, register, subject matter, etc.)
 - may be monolingual, bilingual or multilingual
 - no translational relation between the corpus texts



Dichotomies in corpus linguistics/corpus-based translation studies

- general reference corpus vs. special purpose corpus
- written vs. spoken corpus
- monolingual vs. multilingual corpus
- permanent vs. disposable corpus
- synchronic vs. diachronic corpus
- open vs. closed corpus



Corpus types with direct relevance to translation teaching and practice

- Bilingual parallel/translation corpora (SL and TL texts)
 - insights into strategies employed by (professional) translators
 - provide more collocational and stylistic information than bilingual dictionaries
 - can be used to identify terminological equivalents and their collocates

- Monolingual comparable corpora (original TL texts)
 - checking terminology and collocates against original TL texts
 - searching for unknown equivalents and collocates
 - identifying genre conventions and register
 - searching for explanatory contexts (conceptual instead of linguistic information)



Situating the use of corpora in the translation process

overall translation project process

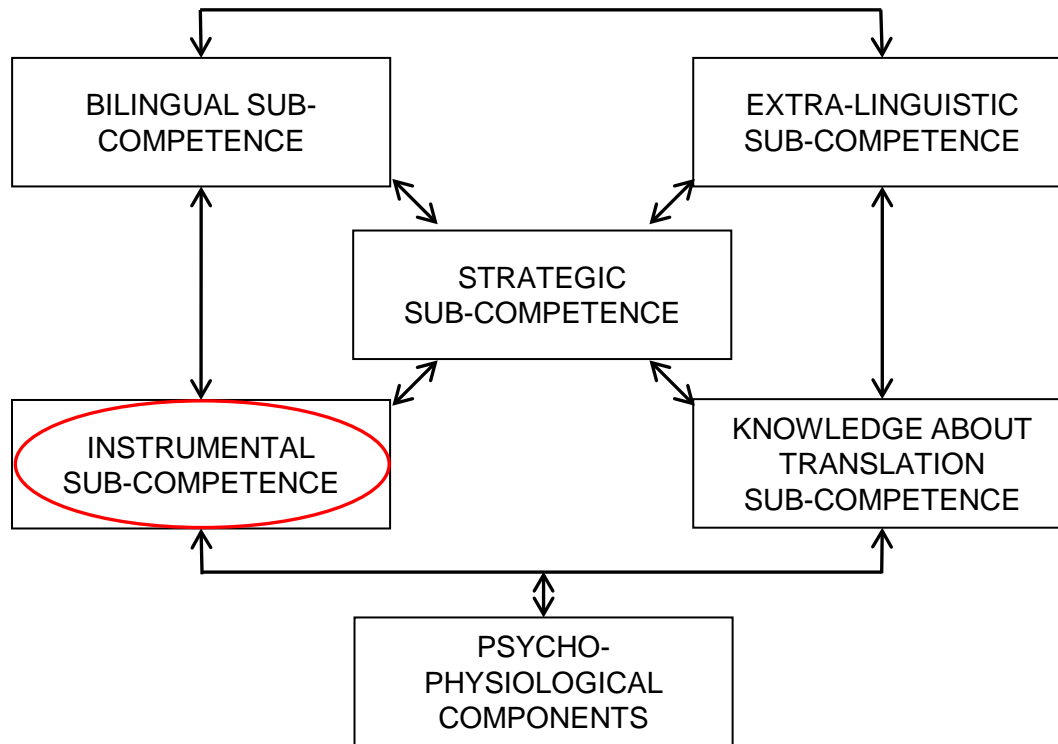
- general preparations (quotation, scheduling, etc.)
- **translation-specific preparations** (TM creation, **subject matter, register, and terminology research**, etc.)

translation process

- text reception/analysis
 - **“on-the-fly“ research (linguistic and conceptual)**
 - translation
 - quality assurance (proofreading etc.)
- follow-up measures (file clean-up, delivery, etc.)



Situating the use of corpora within a translation competence model



PACTE's Model of Translation Competence (PACTE 2003)



Different approaches to corpus use in the translation classroom

- Corpus use for learning to translate
 - corpora are selected and controlled by the teacher to provide real-life examples and exercises
 - students learn to use corpora for their actual translation task but not to compile/evaluate them

- Learning corpus use to translate
 - corpus compilation and evaluation are the students' responsibility
 - holistic approach to corpus use for translating



Do-it-yourself (DIY) corpora

- corpora of texts compiled with the sole purpose of providing information (linguistic or conceptual) for completing a specific translation task
- also called disposable or ad-hoc corpora
- useful not only as a training resource, but also as a resource for professional translators



DIY corpora

possible sources for retrieving corpus texts

- **the client**
 - lecturer can act as client
 - “corpus use for learning to translate” approach
 - students do not acquire or improve their own documentation skills

- **specialist resources**
 - databases, academic articles, etc.
 - not always available for the required subject domains

- **the internet**
 - offers vast amount of information in multiple digital formats
 - most viable alternative for documenting translation tasks
 - requires specific research and QA strategies



DIY corpora

retrieving corpus texts from the internet

- determining the characteristics of the resource used for text research
 - raising students' awareness of the difference between Surface Web and Deep Web
 - universal search engines (e.g. Google, Yahoo!, Bing) vs vertical search engines specialising in a specific field or discipline (e.g. scirius.com for scientific and technical information)
 - ranking criteria for displaying search results (e.g. Google's *PageRank* algorithm)
- specific research techniques for narrowing down the potential results
 - search within a specific website or top-level domain (e.g. *site:.edu/ac.uk*)
 - search within specific file formats (e.g. *filetype:.pdf*)
- establishing quality criteria
 - possible evaluation criteria: authorship, publishing organisation, own judgement of corpus samples (e.g. analysing translated texts for instances of *translationese*)



Excursus: the Web as corpus

- Web is used as macro-corpus
- applying specific search criteria to narrow down the search results (see previous slide)
- Shifting focus from corpus design to search strategies and interpretation of the results
- no indication of the external boundaries or the internal subdivision of the “Web subcorpus“
- Search engines do not present results in a form suitable for linguistic analysis
- Web concordancer: e.g. WebCorp Live (<http://www.webcorp.org.uk/live/>)



Excursus: the Web as corpus

listing of results in WebCorp Live

1) http://www-mdp.eng.cam.ac.uk/web/library/enginfo/aerothermal_dvd_only/aero/fprops/cvanalysis/node54.html

Text, Wordlist, text/html, ISO8859_1 (HTML source), 2005-01-01 (Copyright footer)

1: Previous: Bernoulli Equation Analysis of a **Wind Turbine** Figure 3.32: Analysis of a Wind Turbine A wind

2) http://www.ntu.ac.uk/ecoweb/carbon_elephant/wind_turbine_project/index.html

Text, Wordlist, text/html, UTF8 (HTML source), 2011-01-11 (Meta tag)

2: NTU > EcoWeb > Carbon Elephant > **Wind turbine** project EcoWeb Wind turbine project Wind turbine

3) <http://www.keele.ac.uk/geophysics/appliedseismology/wind/>

Text, Wordlist, text/html, UTF8 (HTML source), 2012-03-09 (Body near 'Last Modified')

3: Local Seismicity Microseismic Monitoring **Wind Turbine** Vibrations Small Wind Turbines Eskdalemuir

4) <http://www.inference.phy.cam.ac.uk/sustainable/energy/wind/080611/index.1.html>

Text, Wordlist, text/html, UTF8 (Content-type), 2008-06-23 (Server header)

4: | | Back to Index | David MacKay | Visit to the **wind turbine** at the organic farm in Bassingbourn, Herts..

5) <http://www.lancs.ac.uk/windturbine/>

Text, Wordlist, text/html, ISO8859_1 (Content-type), 2006-01-01 (Copyright footer)

5: 2012 Work has started on Lancaster University's **wind turbine** Work has started onsite to construct Lancaster



Corpus analysis software

- monolingual comparable corpora: e.g. WordSmith
(<http://www.lexically.net/wordsmith>)
- bilingual translation corpora: e.g. ParaConc
(<http://www.athel.com/para.html>)
 - texts can be prealigned with specific alignment tool (e.g. WinAlign)
 - ParaConc also offers alignment function
- texts must be converted to TXT format before analysis



Analysing a bilingual translation corpus

- looking for the terminological equivalent of “storage site”

ParaConc - [Parallel Concordance - [storage site]]

File Search Frequency Display Window Info

1 ... Except when plants are located directly above a geological [storage site](#), captured CO2 must be transported from the point of capture to a storage site. This section

2 ly above a geological storage site, captured CO2 must be transported from the point of capture to a [storage site](#). This section reviews the principal methods of CO2 transport and assesses the health, saf

1 Mit Ausnahme der Fälle, in denen sich die Anlagen unmittelbar über einer Speicherstätte befinden, ist der Transport des abgetrennten CO2 vom Ort der Abtrennung zur Speicherstätte erforderlich.

2 Mit Ausnahme der Fälle, in denen sich die Anlagen unmittelbar über einer Speicherstätte befinden, ist der Transport des abgetrennten CO2 vom Ort der Abtrennung zur Speicherstätte erforderlich.

- treatment of ST inanimate nouns + action verbs in German translation

ParaConc - [Parallel Concordance - [This section reviews]]

File Search Frequency Display Sort Window Info

1 ... gical storage site, captured CO2 must be transported from the point of capture to a storage site. [This section reviews](#) the principal methods of CO2 transport and assesses the health,

1 Dieses Kapitel beinhaltet eine Bewertung der Hauptverfahren des CO2-Transports sowie die Beurteilung der Kosten und der Gesundheits-, Sicherheits- und Umweltaspekte.



Analysing a bilingual translation corpus

- looking for German collocates of the term “CO₂”

ParaConc - [Parallel Concordance - [CO2]]

File Search Frequency Display Sort Window Info

1 ... ported from the point of capture to a storage site. This section reviews the principal methods of CO₂ transport and assesses the health, safety and environment aspects, and costs. Pipelines today oper ...
2 ... ansporting CO₂. CO₂ also can be transported as a liquid in ships, road or rail tankers that carry CO₂ in insulated tanks at a temperature well below ambient, and at much lower pressures. The first lon ...
3 ed tanks at a temperature well below ambient, and at much lower pressures. The first long-distance CO₂ pipeline came into operation in the early 1970s. In the United States, over 2,500 km of pipeline ...
4 ... nes have intermediate (booster) compressor stations. In some situations or locations, transport of CO₂ by ship may be economically more attractive, particularly when the CO₂ has to be moved over large d ...
5 ... s or locations, transport of CO₂ by ship may be economically more attractive, particularly when the CO₂ has to be moved over large distances or overseas. Liquefied petroleum gases (LPG, principally pro ...
6 G, principally propane and butane) are transported on a large commercial scale by marine tankers. CO₂ can be transported by ship in much the same way (typically at 0.7 MPa pressure), but this currently ...
7 ... his currently takes place on a small scale because of limited demand. The properties of liquefied CO₂ are similar to those of LPG, and the technology could be scaled up to large CO₂ carriers if a deman ...
8 ... erties of liquefied CO₂ are similar to those of LPG, and the technology could be scaled up to large CO₂ carriers if a demand for such systems were to materialize. Road and rail tankers also are technica ...
9 ... aterialize. Road and rail tankers also are technically feasible options. These systems transport CO₂ at a temperature of -20°C and at 2 MPa pressure. However, they are uneconomical compared to pipel

1 Dieses Kapitel beinhaltet eine Bewertung der Hauptverfahren des CO₂-Transports sowie die Beurteilung der Kosten und der Gesundheits-, Sicherheits- und Umweltaspekte.
2 CO₂ kann auch in flüssigem Zustand per Schiff, Lkw oder Bahn in isolierten Tanks bei sehr niedrigen Temperaturen und Drücken transportiert werden.
3 Die erste Fernpipeline wurde zu Beginn der 1970er Jahre in Betrieb genommen.
4 Je nach Gegebenheit oder Standort ist der CO₂-Transport in Tankschiffen wirtschaftlich vorteilhafter, insbesondere auf langen Transportwegen oder beim Transport nach Übersee.
5 Je nach Gegebenheit oder Standort ist der CO₂-Transport in Tankschiffen wirtschaftlich vorteilhafter, insbesondere auf langen Transportwegen oder beim Transport nach Übersee.
6 Der CO₂-Transport in Tankern läuft prinzipiell auf die gleiche Weise ab (typischerweise bei einem Druck von 0,7 MPa), aufgrund geringer Nachfrage findet er derzeit jedoch nur in begrenztem Umfang statt.
7 Die Eigenschaften verflüssigten Kohlendioxids ähneln denen von LPG, und bei einer entsprechenden Nachfrage könnte diese Technik auf große CO₂-Tanker übertragen werden.
8 Die Eigenschaften verflüssigten Kohlendioxids ähneln denen von LPG, und bei einer entsprechenden Nachfrage könnte diese Technik auf große CO₂-Tanker übertragen werden.
9 Dabei liegt das CO₂ bei einer Temperatur von 20°C und einem Druck von 2 MPa vor.



Analysing a monolingual comparable corpus

- Checking whether “CO₂-Abtrennung” or “CO₂-Abscheidung” is the more frequently used German equivalent of “CO₂ capture”

AntConc 3.3.0w (Windows) 2012

File Global Settings Tool Preferences About

Corpus Files

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Forschungskonze

Concordance Hits 24

Hit	KWIC
1	3.3 CO ₂ -Abscheidung und -Speicherung Vor dem Hintergrund möglicherwe
2	n entscheidender Bedeutung ist die Zeitachse. Alle Maßnahmen zur CO ₂ -Abscheidung sind nur dann sinnvoll, wenn die anschließende Ve
3	stet werden kann. Zukünftige Forschungsaktivitäten sollten daher CO ₂ -Abtrennung sowie Möglichkeiten einer nachfolgenden CO ₂ -Behand
4	iner nachfolgenden CO ₂ -Behandlung gleichrangig verfolgen. 3.3.1 CO ₂ -Abscheidung Basis der nachfolgenden Ausführungen bildet eine
5	s erforderlichen F&E-Bedarfs an. Die wichtigsten Strategien zur CO ₂ -Abscheidung können in drei grundsätzliche Kategorien eingeor
6	n etwa 13 Technologievarianten dieser Kategorie diskutiert. 3. CO ₂ -Abscheidung nach Aufkonzentration im Abgas Die Prozessführung
7	arbonisierung Erdgas Luft Gasturbine Abhitzeessel Dampfturbine CO ₂ -Abtrennung Rauchgas CO ₂ Der Gesamtwirkungsgrad der Schaltung
8	den Brennstoffzellen selbst. Der F&E-Bedarf für die eigentliche CO ₂ -Abscheidung ist nahezu identisch mit den anderen Varianten.
9	karbonisierung Kohle Luft Kessel Rauchgasreinigung Dampfturbine CO ₂ -Abtrennung Rauchgas CO ₂ Verglichen mit einem konventionellen
10	d der Technik um 6-10 Prozentpunkte niedriger als beim IGCC ohne CO ₂ -Abscheidung. Die spez. Investitionskosten liegen um 33 % 54%
11	detailliertes Flieschema für ein IGCC-Konzept mit integrierter CO ₂ -Abscheidung. Alternativ könnte Wasserstoff mittels H ₂ -Membran
12	ren [18]. Gegenüber einem fortgeschrittenen IGCC-Kraftwerk ohne CO ₂ -Abscheidung mit einem Wirkungsgrad von etwa 51% liegt er für
13	etwa 45% [16, 17]. Abbildung 20: IGCC-Prozess mit integrierter CO ₂ -Abscheidung Die Prozessroute bis zur Gasturbine entspricht d
14	nigung Dampfturbine HT-WT Entspannungsturbine Abhitzeessel Luft CO ₂ -Abtrennung CO ₂ Der Gesamtwirkungsgrad der Schaltung liegt v
15	tiven der Schaltung nicht abschließend beurteilt werden. Für die CO ₂ -Abscheidung besteht der Forschungsbedarf in der Entwicklung e
16	-Gewinnung Perowskische Oxide H ₂ -Reinigung Palladiumlegierungen CO ₂ -Abtrennung Basische Oxide R&D Themen neue Materialzusammense
17	icht wesentlich unter dem Wirkungsgrad von GuD-Kraftwerken (ohne CO ₂ -Abscheidung) liegen. Die Handhabung von Feststoffen (hier Nic
18	Luft. Das Abgas enthält kein NO _x , kein SO ₂ und keine Staub. Zur CO ₂ -Abscheidung sind zusätzliche Einrichtungen erforderlich. Gem
19	esem Fall den Wirkungsgrad des USC-Dampfkraftwerks (AD 700, ohne CO ₂ -Abscheidung) erreichen. Abbildung 25: Hydrothermaler Reaktor
20	ung von CO ₂ aus konvertiertem Synthesegas (integriertes Konzept, CO ₂ -Abscheidung vor der Verbrennung) O ₂ -Verbrennung Arbeitsproz
21	satz der Brennstoffzellentechnik gewonnen werden. Gaswischen zur CO ₂ -Abtrennung sind im Prinzip bekannt. Insbesondere im Hinblick
22	töt behandelt werden: Tabelle 14: Zukünftige F&E-Maßnahmen zur CO ₂ -Abscheidung Entwicklung von integrierten Gesamtkonzepten un
23	ge über alle technischen und wirtschaftlichen Konsequenzen einer CO ₂ -Abtrennung Klärung der Qualitätsanforderung für das zu depon
24	rbrennstoffzelle sowie entsprechenden Konzepten mit integrierter CO ₂ -Abscheidung Entwicklung eines Rauchgaskondensators 3.3.2 C



Analysing a monolingual comparable corpus

- searching for explanatory contexts: concept to be defined together with the copula “is a(n)” from the classical Aristotelian definition

5) <http://zebu.uoregon.edu/disted/ph121/l13.html>

Text, Wordlist, text/html, UTF8 (Failed), 2005-01-06 (Server header)

5: less than 1% of O₂, H₂O and CH₄ (methane). Since CO₂ is a major greenhouse gas, the radiation from the Sun is trapped

6) http://bme.usc.edu/bme403/Section_1/r9.html

Text, Wordlist, text/html, UTF8 (Failed), 1997-10-02 (Server header)

6: s what lungs do. Your tissues need O₂ to live. CO₂ is a waste product of cell metabolism, and needs to be got

7) <http://dwb.unl.edu/teacher/nsf/c09/c09links/www.casahome.org/carbondi.htm>

Text, Wordlist, text/html, UTF8 (Failed), 2001-03-22 (Server header)

7: CARBON DIOXIDE (CO₂) What is Carbon Dioxide? Carbon dioxide (CO₂) is a colourless, odourless, non-toxic gas that in its solid form is



Some disadvantages of using special corpus-analysis software

- conversion of corpus files into txt format and file alignment/preparation for automatic alignment may be quite time consuming
- loss of formatting information (boldface, italics, etc.)
- loss of image information (no study of image-text interaction possible)
- focus on micro level, i.e. search words and their immediate linguistic context (no study of macro-level structures possible)

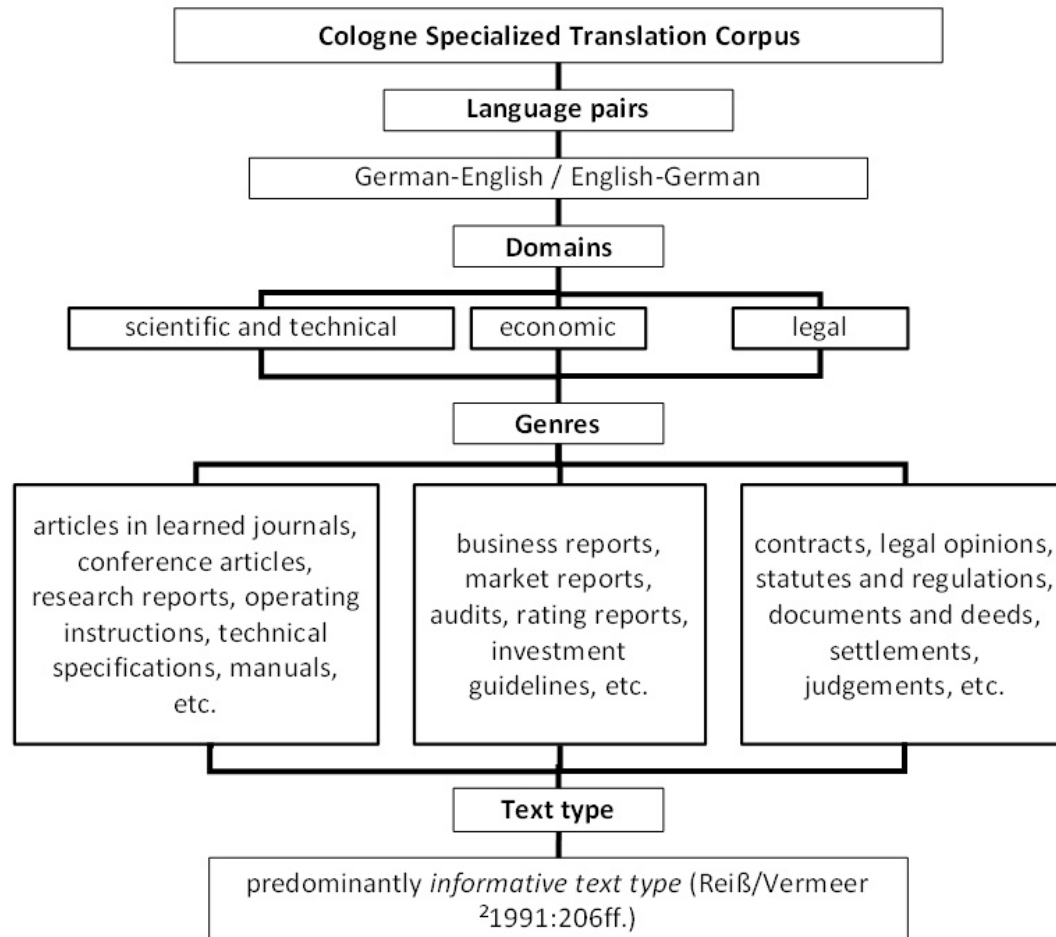


The Cologne Specialised Translation Corpus

- high-quality bilingual specialised translation corpus (English and German)
- expert communication in different domains and different genres
- used for translator training as well as for research projects (MA and PhD theses)
- Example of corpus-based PhD research: investigation of explicitation and implicitation in scientific and technical translation (language directions EN-DE and DE-EN, texts differ in degrees of technicality)

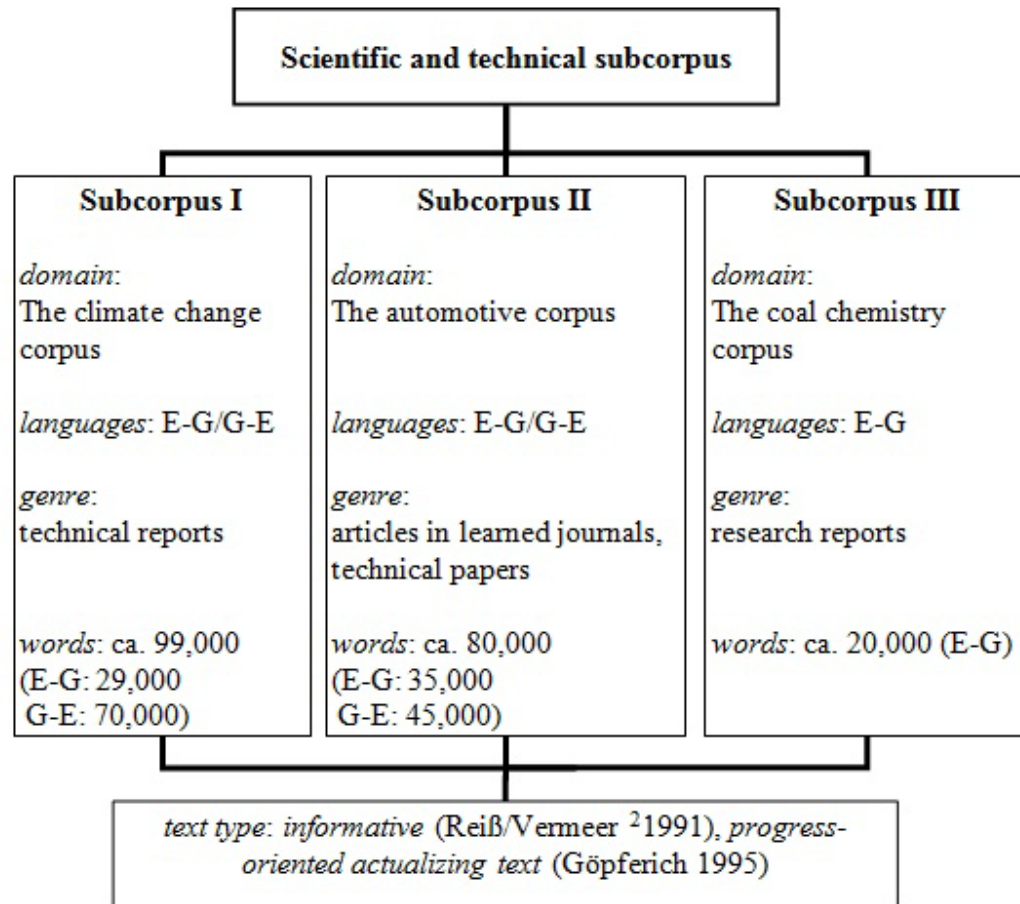


The Cologne Specialised Translation Corpus





The Cologne Specialised Translation Corpus





References

- Baker, Mona (1995): *Corpora in Translation Studies. An Overview and Some Suggestions for Future Research*. *Target*, 7:2, 223-243.
- Beeby, Allison; et al. (2009): *Corpus Use and Translating*. Amsterdam/Philadelphia: Benjamins.
- Bowker, Lynne; Pearson, Jennifer (2002): *Working with Specialized Language. A practical guide to using corpora*. London/New York: Routledge.
- Krein-Kühle, Monika (2011): "Register Shifts in Scientific and Technical Translation. A Corpus-in-Context Study", in: Olohan, Maeve; Salama-Carr, Myriam (eds): *Science in Translation*, Special Issue of the *Translator* 17(2). Manchester: St. Jerome, 391-413
- Laviosa, Sara (2002): *Corpus-based Translation Studies*. Amsterdam/New York: Rodopi.
- Olohan, Maeve (2004): *Introducing Corpora in Translation Studies*. London/New York: Routledge.
- Zanettin, Federico; et al. (2003): *Corpora in Translator Education*. Manchester: St. Jerome.
- Zanettin, Federico (2012): *Translation-driven Corpora*. Manchester: St. Jerome.