

Giulia A. Disanto

Korpusbasierte Translationswissenschaft

Eine Untersuchung am Beispiel des *JRC-Acquis* Parallelkorpus Deutsch-Italienisch

Corpus-based Translation Studies. A Study of the "JRC-Acquis" Parallel Corpus German-Italian – Abstract

The aim of this paper is to describe how translation studies might benefit from the use of text corpora. The first part of the article outlines some basic theoretical questions regarding cooperation between translation studies and corpus linguistic studies. The second part is an investigation dealing with a compilation of legislative texts belonging to the European Union, the *JRC-Acquis multilingual parallel corpus*, which comprises texts listed in parallel in twenty-two languages. Using a subcorpus of translated texts in the German and Italian languages as an example, this paper provides quantitative data (word-frequency lists and collocations), while also attempting to explain them by comparing text passages in both the source and target languages. In conclusion, data regarding the frequency of Italian and German modal verbs are generated, showing the different behaviour of modal verbs in translations of legal texts.

1 Einleitung

Ein Linguist, der seine theoretischen Hypothesen am Beispiel von konkretem sprachlichem Material empirisch überprüfen und weiterentwickeln möchte, verfügt heutzutage durch das Internet über eine wachsende Menge von digitalisierten Sprachkorpora sowie über eine unbegrenzte Fülle von Texten bzw. Hypertexten, aus denen man ein eigenes zweckorientiertes Korpus aufbauen kann. Mittels spezifischer Software kann das Textkorpus dann quantitativ bzw. qualitativ analysiert und auf bestimmte sprachliche Phänomene hin untersucht werden.

Mit Korpora im engeren Sinne beschäftigt sich die Korpuslinguistik, und zwar mit der Beschreibung natürlicher (nicht formaler) Sprachen, die anhand von maschinenlesbaren Textsammlungen wissenschaftlich analysiert werden. Auf Grundlage der korpusbasierten Datenanalyse können theoretische Thesen unterschiedlichster Art formuliert bzw. widerlegt werden.

Digitalisierte Korpora erweisen sich aber unter verschiedenen Gesichtspunkten immer mehr als eine wichtige Hilfsquelle und können unterschiedlichen Zwecken dienen. Die Korpuslinguistik findet demzufolge in verschiedenen sprach- und kulturwissenschaftlichen Bereichen Anwendung und stellt auch für die Translationswissen-

schaft ein forschungsrelevantes Feld dar. An dieser Stelle soll u.a. aufgezeigt werden, wie beide Wissenschaften voneinander profitieren können.

Mit dem Einsatz von Korpusanalysen in der Translationswissenschaft sowie der Erforschung der spezifischen gemeinsamen Anwendungsbereiche von Korpuslinguistik und Translationswissenschaft befasst sich die Forschung erst seit relativ kurzer Zeit. Die Diskussion vollzieht sich hauptsächlich im englischsprachigen Raum und beschränkt sich daher meistens auf das Englische. Pionierarbeit wurde auf diesem Feld schon Anfang der neunziger Jahre von Mona Baker (1993) geleistet.

Die große Mehrheit der heute vorhandenen Arbeiten über Korpuslinguistik spricht das Thema "Translation" nicht an, und auch nur wenige behandeln das Thema "Fremdsprachenausbildung".¹ Es liegt aber klar auf der Hand, dass die Verfügbarkeit von umfangreichen Parallelkorpora in den jeweiligen Sprachenpaaren für Translationswissenschaftler, Profi-Übersetzer und Studierende der Übersetzungswissenschaft von Vorteil sein kann.

Im Gegensatz zur Korpuslinguistik im Allgemeinen muss sich die korpusbasierte Translationswissenschaft ihre Methodik für die Korpusanalyse erst noch verschaffen. Dabei sollte die Untersuchung der Daten sowohl quantitativ als auch qualitativ erfolgen. Korpusbasierte Translationswissenschaft kann nicht nur auf die Frage ausgerichtet sein: *Wie oft* wird das Segment X in der jeweiligen Ausgangssprache mit dem Segment Y in der jeweiligen Zielsprache übersetzt? Vielmehr gilt es heutzutage die Frage zu beantworten: *Wieso* wurde jenes Segment X mit dem Segment Y übersetzt oder nicht? Welche Translationsstrategien führten zur jeweiligen Übersetzung und wie kann man sie bewerten? Die Gültigkeit der Ergebnisse quantitativer Analysen sollte also durch die qualitative Interpretation sowie durch eine Diskussion anhand von Textvergleichen ergänzt werden.

Die "qualitative" Analyse basiert auf Beobachtung und Beschreibung von Daten. Das ist auch der Ansatz dieses Beitrags, in dem einerseits eines der wichtigsten mehrsprachigen Korpora – das Korpus der Rechtstexte der EU – vorgestellt und andererseits anhand von Beispielen aus einem Subkorpus den konkreten Nutzen der Anwendung von Korpora in der Translationswissenschaft dargestellt werden soll. Zu diesem Zweck wird das Subkorpus zuerst auf Wortebene analysiert und dann das Verhalten von Modalverben im deutsch-italienischen Textvergleich betrachtet. Im Grunde wird hier eine empirische Untersuchung mit Überlegungen zum Thema "Korpuslinguistik und Translationswissenschaft" verbunden, so dass eine theoretische Einleitung erforderlich wird.

Das unterschiedliche Verhältnis von Korpuslinguisten einerseits und theoretisch arbeitenden Linguisten andererseits zu Korpusdaten geht auf einen grundsätzlichen Unterschied in den erkenntnistheoretischen Grundlagen und Methoden beider Richtungen zurück. Die methodischen Grundlagen korpuslinguistischer Forschung sind *empiristisch*, die der theoretischen Linguistik *rationalistisch*. [...] Der Kern der *empiristischen* Auffassung ist die Behauptung, dass alle Erkenntnis in der sinnlichen Anschauung wurzelt. [...] Der Kern der *rationalisti-*

¹ Als umfassenden Überblick über die gleichwohl zahlreichen Bezüge zwischen Korpusanalyse und Translationswissenschaft sei an dieser Stelle auf Maeve Olohans Studie *Introducing Corpora in Translation Studies* hingewiesen (Olohan 2004).

schen Auffassung ist die Behauptung, dass Erkenntnisse durch Begriffe und Urteile gewonnen werden. (Lemnitzer/Zinsmeister 2006: 14-18)

Translationswissenschaft, die sich an Korpusanalyse orientieren möchte, sollte heute empirisch vorgehen. In Anlehnung an Elena Tognini-Bonelli, die in ihrem Buch *Corpus Linguistics at Work* zwischen "corpus-based" und "corpus-driven linguistics" unterscheidet, ist hier von korpusbasierter Translationswissenschaft die Rede.²

Den Untersuchungsgegenstand dieses Beitrags bildet eines der wichtigsten und umfangreichsten heute vorhandenen mehrsprachigen Korpora, das sich in translationswissenschaftlicher Hinsicht als besonders nützlich erweist: das *JRC-Acquis multilingual parallel corpus* der Europäischen Union (*JRC* steht für *European Commission Joint Research Center*), das zurzeit in 22 Sprachen vorliegt und die Gesamtheit der EU-Rechtstexte umfasst.³

2 Korpuslinguistik und Translationswissenschaft: allgemeine Problembestimmung

Spricht man von Korpora und Translation, so stößt man als Erstes auf Schwierigkeiten bei der Bestimmung einiger Begriffe, wie z.B. Text, Korpus, Parallelkorpus.

Obwohl der Textbegriff für die Übersetzungstheorie heute weitgehend als unumstrittener Bezugspunkt gilt, ist man sich in der Textlinguistik über seine Bestimmung noch nicht einig. Forschungen und einzelne Studien über die Multidimensionalität der Textstruktur haben bislang zu keiner allgemeingültigen Theorie geführt, die imstande wäre, das Textgebilde systematisch als Ganzes zu beschreiben.

Für eine wünschenswerte Zusammenarbeit zwischen Computer- und Textlinguistik plädiert Manfred Stede in seiner 2007 erschienenen Studie *Korpusgestützte Textanalyse – Grundzüge der Ebenen-orientierten Textlinguistik*. Obwohl das Thema Translation nicht direkt angesprochen wird, sind in dieser Studie das Grundwissen sowie interessante Überlegungen für eine text- und korpusbezogene Translationswissenschaft zu finden. Wie man bereits dem Untertitel entnehmen kann, deutet der Autor das Textgebilde als Folge der Interaktion unterschiedlicher Ebenen. Stede nennt insbesondere sechs Ebenen, die teilweise vom Texttyp abhängig sind und die Textkohärenz bestimmen: die referenzielle Struktur, die thematische Struktur, die temporale Struktur, die Illokutionsstruktur, die Argumentationsstruktur, die rhetorische Struktur.⁴ Für jede dieser Ebenen stellt der Autor spezifische Annotationswerkzeuge vor. Dabei hebt er hervor, dass die Ebenen erst einzeln annotiert werden sollten, damit man sie später zueinander in Beziehung setzen kann. Dieses Verfahren könnte sich tatsächlich

² Vgl. Tognini-Bonelli (2001). Für eine Kategorisierung der unterschiedlichen möglichen Ansätze in der Korpuslinguistik (nach den Kriterien: korpusbasiert/korpusgestützt, quantitativ/qualitativ) s. die Tabelle in Lemnitzer/Zinsmeister (2006: 33).

³ Vgl. die Webseite des *Joint Research Center* der EU: <http://langtech.jrc.it/>.

⁴ Eine weitere Ebene könnte – so schlägt Stede vor – die der Subjektivität sein: "Aus welcher Perspektive wird im Text berichtet, wem wird die Schilderung in einem Diskurssegment zugeschrieben (der Autorin oder einer im Text genannten Person), wer lässt welche Einstellungen zu Sachverhalten erkennen?" (Stede 2007: 183)

auch im Fall von Übersetzungen als fruchtbar erweisen, denn damit ließe sich nicht nur die Vielschichtigkeit des Translats erfassen, sondern auch die Voraussetzung für eine vergleichende Textanalyse anhand von Parallelkorpora schaffen.

Bei der Forschung mit digitalisierten Korpora fragt man sich nicht nur, was ein Text eigentlich ist, sondern vielmehr, welche Merkmale zu den bis heute üblichen Textdefinitionen durch das Internet, das Verfahren der Digitalisierung und die Verflechtung zwischen natürlichen und maschinellen Sprachen hinzukommen. In Zukunft werden die Einflüsse der medialen Welt auf die Arbeit mit Texten (und deren Übersetzungen) immer stärker berücksichtigt werden müssen.

In erster Linie zu überdenken ist die Gesamtsphäre der Bereitstellung, der Kategorisierung, der sprachlich/stilistischen Qualität und der Normgerechtigkeit der Texte in der Fremdsprache und Fremdkultur im Lichte verschiedener zusätzlicher Hilfsmittel, Ressourcen und Lösungsmöglichkeiten. [...] Die Pädagogik der Produktion und Rezeption fremdsprachiger Texte – und der entsprechenden Translationskompetenz – ist dringend auf eine gezielte Nutzung spezieller und allgemeiner Hilfsmittel im Internet und in den anderen neuen Medien angewiesen. (Soffritti 2006: 41)

Im dynamischen Zeitalter der medialen Kommunikation sieht also der Text – im Vergleich zu vergangenen Zeiten – ganz anders aus. Im konkreten Fall eines digitalisierten Korpus springt die Verflechtung zwischen sprachlich "traditioneller" Syntax und der Auszeichnungssprache (*markup language*), durch die der Text digital beschrieben und gestaltet wird, sofort ins Auge. Wenn man den jeweiligen Text lesen, verstehen und analysieren möchte, muss man zuerst die beiden Ebenen genau unterscheiden können.

Auch der Korpusbegriff sollte möglichst eindeutig festgelegt werden. Nach der schon zitierten Studie von Lemnitzer und Zinsmeister kann man ein "Korpus" folgendermaßen definieren:

Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen. Die Daten des Korpus sind digitalisiert, d.h. auf Rechnern gespeichert und maschinenlesbar. Die Bestandteile des Korpus, die Texte oder Äußerungsfolgen, bestehen aus den Daten selbst sowie möglicherweise aus Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind. (Lemnitzer/Zinsmeister 2006: 40)

Im Allgemeinen ist ein linguistisches Korpus durch drei Merkmale gekennzeichnet:

- *Primärdaten*: Darunter sind nicht die Texte, die die Quelle des digitalisierten Textkorpus darstellen, in ihrer Originalfassung zu verstehen, sondern deren Transkription. Beim Transkribieren ist darauf zu achten, dass die Vorlagen so wenig wie möglich geändert werden. Im Allgemeinen ist damit zu rechnen, dass dabei einige Veränderungen auftreten können.
- *Metadaten*: Als Metadaten bezeichnet man die Gesamtheit von Sekundärdaten, die Informationen verschiedener Art über die Primärdaten umfassen. Zu den Metadaten gehören also zusätzliche Informationen wie Titel, Entstehungsdatum, Verfasser, Verlag, Sprache, Inhalt, Rechte, Kodierung, Vernetzung unter Texten usw. Man könnte behaupten, dass Metadaten eine Art Identität des digitalisierten Textes und damit eine wesentliche Anreicherung des Korpus darstellen. Sie werden in der Regel nach bestimmten Standards gespeichert und übertragen. Heute sind als

Datenmodelle und -formate zur Speicherung von Metadaten sehr verbreitet: *Dublin Core* (DC) oder TEI (*Text Encoding Initiative*, das Dokumentenformat der Texte im *JRC-Acquis*).

- *Linguistische Annotation*: Unter "linguistischer Annotation" versteht man die zusätzliche Ergänzung von Primär- und Sekundärdaten durch linguistische Informationen auf unterschiedlichen Ebenen (von der Wortebene bis zur Syntax, Semantik, Pragmatik usw.). Auf der Wortebene werden Korpora durch *Tagging* annotiert, wobei jede Wortform mit einer bestimmten grammatischen Kategorie (*Tag*) in Verbindung gebracht wird. Syntaktisch annotierte Korpora nennt man Baumbanken (*Treebanks*). Heutzutage findet man die verschiedensten Annotationswerkzeuge im Internet auch als *open-source* Programme. Oft sind solche Programme aber nicht wirklich benutzerfreundlich. In den meisten Fällen von bestimmten Forschergruppen zu bestimmten Zwecken entwickelt, geben sie außerdem nur in ganz bestimmten Bereichen gute Ergebnisse: für die Fragestellung, ein bestimmtes Erkenntnisinteresse, das Betriebssystem, die Sprachen usw., für die sie geschaffen wurden. Die Annotation ist in jedem Fall eine wesentliche Voraussetzung für eine erweiterte Untersuchung von Korpora. Die linguistische Annotation erweist sich als ein wichtiges Instrument auch für translationswissenschaftliche Forschungen, wir verfügen zurzeit aber über kein speziell für die Translationswissenschaft gedachtes Werkzeug, was für die Zukunft höchst wünschenswert wäre.

Für Translationswissenschaftler stehen so genannte "Parallelkorpora" im Mittelpunkt des Interesses. Im Fall eines Parallelkorpus sind einerseits die oben vorgestellten Merkmale für jede einzelne Sprache zu bestimmen und andererseits die einzelnen entsprechenden Texte im Korpus durch das sogenannte *Alignment*-Verfahren miteinander zu verknüpfen.

Es gibt unterschiedliche Arten von Parallelkorpora, wobei in der Forschung meist folgende Unterscheidung vorgenommen wird (Johansson 1998: 4):⁵

- "comparable corpora" (Vergleichskorpora): Texte, die in zwei oder mehreren Sprachen verfasst sind, werden aufgrund ihrer Vergleichbarkeit in einem Textkorpus vereint – so etwa weil sie derselben Textsorte angehören – zwischen ihnen besteht aber keine Übersetzungsrelation.
- "translation corpora" (Übersetzungskorpora): Diese bestehen aus Texten in einer bestimmten Sprache und deren Übersetzungen in eine andere oder in mehrere Sprachen. In diesem Fall sollten die Textpartien einander zugeordnet werden, was mit dem Alignment-Verfahren geschieht.⁶

⁵ Für eine Typologie von Paralleltext und Parallelkorpus vgl. Fabricius-Hansen (2004).

⁶ Stig Johansson nennt zusätzlich drei weitere Typen von Korpora: "Two additional types of corpora are of special relevance for translation studies: corpora of original and translated texts in the same language and corpora of translated texts in different languages. In addition to their usefulness in translation studies, these types are significant in providing control material when corpora of original and translated texts are used for contrastive studies. A fifth type of corpus which is relevant in contrastive studies is a learner language corpus. Different types of corpora can be combined within the same overall framework [...]" (Johansson 1998: 4-5).

Andere Wissenschaftler wollen den Ausdruck "Parallelkorpora" ausschließlich auf Übersetzungskorpora bezogen wissen (Lemnitzer/Zinsmeister 2006: 104, 198). Im vorliegenden Beitrag ist die Rede vom Parallelkorpus im Sinne von Übersetzungskorpus (bzw. "translation corpus"), denn diese Bestimmung entspricht dem Charakter des *JRC-Acquis*, einem multilingualen Parallelkorpus ("translated texts in different languages").

Die Mehrheit der Untersuchungen im Bereich der Korpuslinguistik konzentriert sich allerdings auf monolinguale Korpora, weshalb die meisten Annotationswerkzeuge auch für monolinguale Korpusanalyse konzipiert sind. Es gibt inzwischen nationale Annotationsstandards, wie das Stuttgart-Tübingen Tagset (STTS) für die morphosyntaktische Annotation deutschsprachiger Korpora, aber keine internationalen Standards, was für die korpusbasierte Translationswissenschaft von zentraler Bedeutung wäre.

Der zögerliche Einsatz von Korpora in der Translationswissenschaft hängt auch damit zusammen, dass Übersetzungen oft als etwas Zweitrangiges im Vergleich zum Original angesehen werden. Im Endeffekt sind parallele Texte (der Text und seine Übersetzung) für viele nur ein einziger Text.⁷ Damit man möglichst viele Aspekte des Übersetzungsverfahrens erfassen kann, erweist es sich natürlich als sinnvoller, die zwei parallelen Texte erst einmal als verschiedene Texte zu betrachten.

Wie bereits in der Einleitung festgestellt, stützt sich die korpusbasierte Translationswissenschaft vornehmlich auf die Beobachtung und auf die Beschreibung von Daten. Möchte man den korpusbasierten empirischen Ansatz in der Übersetzungsforschung theoretisch einordnen, dann denkt man zweifelsohne zuerst an die so genannten *Descriptive Translation Studies* (DTS). DTS sind vor allem mit dem Namen Gideon Toury verbunden, der – unter dem Einfluss der maßgeblich von Itamar Even-Zohar entwickelten *polysystem theory* (Even-Zohar 1990) – den Begriff der *assumed translation* einführte (Toury 1995): Als Untersuchungsgegenstand der Übersetzungswissenschaft kommen demnach praktisch alle Texte in Frage, die im Rahmen der Kultur der Zielsprache als solche anerkannt werden. Die Untersuchung der Ausgangstexte spielt dabei keine große Rolle. Es wird zwar postuliert, dass es einen Ausgangstext gibt, jedoch braucht man diesen Text nicht in Betracht zu ziehen.

Dieser theoretische Anhaltspunkt ist für die vorliegende Untersuchung aus zwei Gründen wichtig. Erstens basieren die ersten korpusbasierten Studien in der Translationswissenschaft der neunziger Jahre auf den Annahmen der DTS, mit denen auch die Suche nach den *universals of translation* eng zusammenhängt (Toury 2004). Zweitens entspricht der Charakter der EU-Texte eher dem Übersetzungsbegriff Tourys (Toury 1995), denn es geht um Übersetzungen, bei denen man die Ausgangssprache der Texte selten kennt. Bei Texten, die im Rahmen der EU verfasst oder übersetzt wurden, kommt dagegen vor allem der Funktion der Texte bzw. der Übersetzungen eine entscheidende Rolle zu. Die Kommunikationsabsicht, die solchen EU-Texten zugrunde liegt, ist dominant, so dass man mit der Übersetzung eine gewisse Universalität anstrebt. Die Unterscheidung zwischen Ausgangs- und Zieltext ist demzufolge nicht mehr so wichtig. Nun soll genauer betrachtet werden, zu welchen Ergebnissen die ersten

⁷ Vgl. Oakes/McEnery (2000: 1): "A bilingual parallel corpus is a corpus that contains the same text samples in each of two languages, in the sense that the samples are translations of one another".

Forschungen im Bereich der korpusbasierten Translationswissenschaft geführt haben und wie sie in die heutigen Untersuchungen einbezogen werden können.

Mitte der neunziger Jahre versuchten Mona Baker und ihre Forschungsgruppe an der Universität Manchester, spezifische und allgemeingültige Eigenschaften (so genannte "Universalien") von übersetzten Texten aufzudecken. Sie bauten ein monolinguales Korpus von Übersetzungen ins Englische aus verschiedenen Sprachen auf und verglichen es mit einem aus englischen Originaltexten bestehenden Korpus. Auf diese Weise konnte man jene Merkmale herausfinden, die die Textsorte "Übersetzung" bestimmen und die "universell" im Sinne von "unabhängig" vom Ausgangstext und von der Ausgangssprache sind. Mona Baker stellte danach einige Hypothesen auf und hob dabei folgende vier Hauptmerkmale von Übersetzungen hervor (Baker 1993):

- *Simplification*: die Tendenz des Translators, bei dem Übersetzungsverfahren, die Texte auf unterschiedliche Art – wenn auch unbewusst – zu vereinfachen.
- *Explicitation*: die Tendenz des Translators, die Sachverhalte des Ausgangstextes in der Übersetzung klarer auszudrücken, z.B., indem er Informationen hinzufügt, die im Ausgangstext nur implizit angedeutet sind.
- *Normalisation/Conservatism*: die Tendenz des Translators, auf typische Muster der Zielsprache zurückzugreifen, damit der übersetzte Text Merkmale eines in der Zielsprache verfassten Textes aufweist.
- *Levelling out*: übersetzte Texte tendieren sehr viel mehr als Originaltexte, die in der Ausgangs- oder Zielsprache verfasst wurden, dazu, Ähnlichkeiten zu zeigen.

Diese Eigenschaften der besonderen Textsorte "Übersetzung" können durch quantitative Korpusanalysen überprüft werden. Die Tendenz zur *Simplification*, die zunächst auf eine lexikalische, eine syntaktische und eine stilistische Ebene zurückgeführt wird (Laviosa-Braithwaite 1998: 288), sowie die zur *Explicitation* können durch quantitative korpusbasierte Analysen ziemlich einfach überprüft werden. Man kann z.B. annehmen, dass die beiden Tendenzen sich an der unterschiedlichen Satz- oder Textlänge von Original und Übersetzung zeigen. Auch das Verhältnis zwischen der Anzahl der Worttypes und der der Worttokens könnte ein wichtiger Parameter für die Bestimmung dieser Tendenzen sein (s. Abschnitt 4.2).⁸ Die beiden letzten Eigenschaften können gleichermaßen quantitativ bestätigt bzw. widerlegt werden, am besten durch den Einsatz von Vergleichskorpora. Auch in diesem Fall braucht aber die quantitative Analyse, damit sie wirklich zum Tragen kommt, eine sorgfältige qualitative Interpretation, die zusätzlich theoretische translationswissenschaftliche Überlegungen einbeziehen sollte. Nur durch eine Überbrückung der Kluft zwischen Korpuslinguistik und Translationswissenschaft können beide Disziplinen voneinander profitieren. Durch die Suche nach gemeinsamen Forschungsfeldern kann auch die korpusbasierte Translationswissenschaft konkrete Fortschritte erzielen. Außerdem wäre die Entwicklung von spezifischen Software-Werkzeugen, die auf die Aufgaben des Translationswissenschaftlers zuge-

⁸ Vgl. Hansen/Teich (1999).

schnitten sind, sicher ein dringendes Desiderat im Sinne eines zukünftigen Zusammenwirkens von Korpuslinguistik und Translationswissenschaft.⁹

Natürlich ist zunächst zu klären, wie die Translationswissenschaft konkret vom korpusbasierten Ansatz profitieren kann. Dazu sind interdisziplinäre Forschungsprojekte zwischen Translationswissenschaft, Korpuslinguistik und Computerlinguistik höchst wünschenswert. Jede dieser Disziplinen scheint heute ihren eigenen Weg zu gehen, nun ist es für die korpusbasierte Translationswissenschaft unerlässlich, die drei Disziplinen sehr viel näher zusammenzubringen, als dies heute der Fall ist.

Auch was das *JRC-Acquis* angeht, das offensichtlich von außerordentlichem Interesse für den Translationswissenschaftler ist, behandeln die meisten Forscher fast ausschließlich Fragen der Computerlinguistik.¹⁰ So wird z.B. darüber diskutiert, wie die Texte am besten "aligniert" werden können (d.h. wie jeder Absatz des Ausgangstextes – möglichst ohne manuelle Korrektur – mit seiner Übersetzung im Zieltexes zu verknüpfen ist), ohne sich dabei zu fragen, was diese Art der Segmentierung des Textes in Absätze vom translationswissenschaftlichen Standpunkt aus gesehen bedeutet (z.B. ob der Absatz die geeignete Einheit darstellt). Wie beeinflusst diese Art der Segmentierung die Leistungsfähigkeit von *Translation Memories*, die auf Korpora basieren?

All diese Aspekte sollte die korpusbasierte Translationswissenschaft also einbeziehen. Sie hat nicht nur neue Ansätze, Untersuchungen und Instrumente zu erproben, sondern sollte sich vor allem die Entwicklung einer spezifischen Methodik zur quantitativen und qualitativen Analyse von Parallelkorpora zum Ziel setzen, die gleichzeitig den theoretischen Überlegungen der Translationswissenschaft der letzten Jahrzehnte Rechnung trägt.

3 Das *JRC-Acquis* Parallelkorpus Deutsch-Italienisch

Das *JRC-Acquis multilingual parallel corpus* (Version 3.0; im Folgenden als *JRC-Acquis* bezeichnet) ist das mehrsprachige Korpus der Rechtstexte der Europäischen Union und liegt zurzeit in 22 Sprachen vor.¹¹ Es basiert auf dem *Acquis-Communautaire* ("Gemeinschaftlichem Besitzstand"), d.h. der gesamten Sammlung der Dokumente des Gemeinschaftsrechts und der für die Mitgliedstaaten verbindlichen Rechtsakte und umfasst dementsprechend unterschiedliche Textsorten wie z.B. Verträge, Richtlinien, Entschlüsse der Organe der EU, Rechtsvorschriften und Rechtsprechung des Europäischen Gerichtshofs, internationale sowie zwischen den Mitgliedstaaten geschlossene Abkommen, Maßnahmen und Initiativen im Rahmen der Außenpolitik usw.¹²

⁹ Für eine umfassende Einführung in die zahlreichen Bezüge zwischen Translationswissenschaft und Technologie vgl. Bowker (2002).

¹⁰ Vgl. die Ergebnisse des Workshops "Exploiting parallel corpora in up to 20 languages" (Arona, September 2005) auf der Webseite: http://langtech.jrc.it/0509_EU-Enlargement-Workshop.html.

¹¹ Bulgarisch, Dänisch, Deutsch, Englisch, Estnisch, Finnisch, Französisch, Griechisch, Italienisch, Lettisch, Litauisch, Maltesisch, Niederländisch, Polnisch, Portugiesisch, Rumänisch, Schwedisch, Slowakisch, Slowenisch, Spanisch, Tschechisch und Ungarisch.

¹² Für eine zusammenfassende Vorstellung des Korpus und der damit verbundenen *Translation Memory* vgl. John (2008).

Bevor die beitrittswilligen Staaten Mitglieder der Europäischen Union werden, müssen sie diesen Besitzstand anerkennen und übernehmen. Sie verpflichten sich damit, Gemeinschaftsrecht in nationales Recht umzuwandeln. Diese Umsetzung setzt selbstverständlich die Übersetzung der Texte des Besitzstandes in alle Sprachen der Mitgliedstaaten voraus. Das Korpus ist also durch Mehrsprachigkeit und die Einbeziehung juristischer Fachbegriffe und -ausdrücke gekennzeichnet.

Alle Dokumente des Korpus sind von zwei Webseiten der EU¹³ aus zugänglich und wurden vom *Joint Research Center* (Ispra, Italien) als HTML-Dateien heruntergeladen und in XML-Dateien konvertiert. Die sogenannte Schemasprache, die dazu benutzt wurde, ist die DTD (*Document Type Definition*). In groben Zügen könnte man sagen, dass die DTD die Grammatikregeln der digitalen Gestaltung des Textes bestimmt und dadurch die digitale Beschreibung eines Textes ermöglicht.

XML (*Extensible Markup Language*) ist eine Auszeichnungssprache für digitalisierte Dokumente, die sich in den letzten Jahren als unabdingbarer Bezugspunkt in der Computerlinguistik durchgesetzt hat. Wir verfügen heutzutage über zahlreiche Annotationswerkzeuge für Korpora, und die Tatsache, dass mit diesen Softwares XML-konforme Dateien erstellt werden, ist für deren Anwendung für verschiedene Forschungszwecke von entscheidender Bedeutung. Manchmal ist trotzdem damit zu rechnen, dass eine Konvertierung der Dateien in das Format des jeweiligen Annotationsprogramms erforderlich wird.

Mehrsprachiges Korpus bedeutet in unserem Fall, dass Dokumente in 22 Sprachen und 231 möglichen Sprachenpaaren vorliegen. Die Texte des Korpus wurden für jedes einzelne Sprachenpaar dem Alignment-Prozess unterzogen, d. h. ohne Benutzung einer Pivot-Sprache, was sicherlich zur Vermeidung etlicher Fehler bei der technischen Verknüpfung der Texte durch das Alignment beiträgt. Das bedeutet aber nicht, dass die Texte ursprünglich – und zwar vor dem Eintrag ins digitalisierte Korpus – in einer Sprache verfasst wurden, etwa auf Englisch, das nicht unbedingt zu unserem Sprachenpaar gehört. Man kann an sich davon ausgehen, dass die Texte der EU größtenteils auf Englisch verfasst wurden. Bei den Texten des *JRC-Acquis* erfolgte das Alignment durch zwei verschiedene Programme, *Vanilla* und *HunAlign*, und wurde manuell nicht überprüft.

Das Korpus, zu dem man im Internet freien Zugang hat, umfasst für jede Sprache die Dokumente aus dem Zeitraum zwischen 1958 und 2006. Die Dokumente jeder Sprache sind in getrennten Ordnern zusammengestellt und der Anwender soll selbst die Dokumente des jeweiligen Sprachenpaares alignieren. Wir haben die Dokumente des Sprachenpaares Deutsch-Italienisch aligniert und eine einzige XML-Datei (in der Größe von etwa 520 MB) hergestellt. Unsere Untersuchungen basieren sowohl auf dieser Datei als auch auf dem Vergleich der Dokumente der getrennten Ordner der zwei Sprachen. Die XML-Dateien werden mit einem normalen Browser geöffnet, anderenfalls kann man sie mit dem ausgewählten Konkordanzprogramm betrachten (fast alle Programme ermöglichen die Textansicht).

¹³ Vgl. das *Portal der Europäischen Union* (<http://europa.eu/>) und die Webseite von *CCVista Translation Database* (<http://ccvista.taie.be/>).

Durch das Alignment-Verfahren werden jeweils zwei Dokumente miteinander verknüpft, damit eine Korrespondenz zwischen den Übersetzungen hergestellt wird. Das Alignment erfolgt auf der Absatzebene (*paragraph level: <p>*), nicht auf der Satz- oder Wortebene, obwohl der Absatz oft nur aus einem einzigen – wenn auch längerem – Satz besteht.

Unter den fast 464.000 Texten des Korpus ist jede der 22 Sprachen durchschnittlich mit 23.000 Texten vertreten. Folgende Diagramme zeigen, dass die Anteile der Dokumente auf Deutsch und Italienisch im gesamten Korpus grundsätzlich vergleichbar sind und wenig über dem durchschnittlichen Wert liegen (Abb. 1). Wie insbesondere dem zweiten Diagramm (Abb. 2) zu entnehmen ist, ist der Anteil der Texte, die keine Korrespondenz im Parallelkorpus finden (d.h. Dokumente auf Deutsch bzw. Italienisch, für die keine Übersetzung auf Italienisch bzw. Deutsch angefertigt wurde), im Vergleich zu der gesamten Zahl der Dokumente äußerst gering (0.75 % der durchschnittlichen Zahl der nicht übersetzten Texte auf Deutsch und auf Italienisch). Man kann in diesem Fall nur vermuten, dass es um Texte geht, deren Wert für die jeweilige nationale Ebene begrenzt ist und für die daher keine Übersetzung benötigt wird.

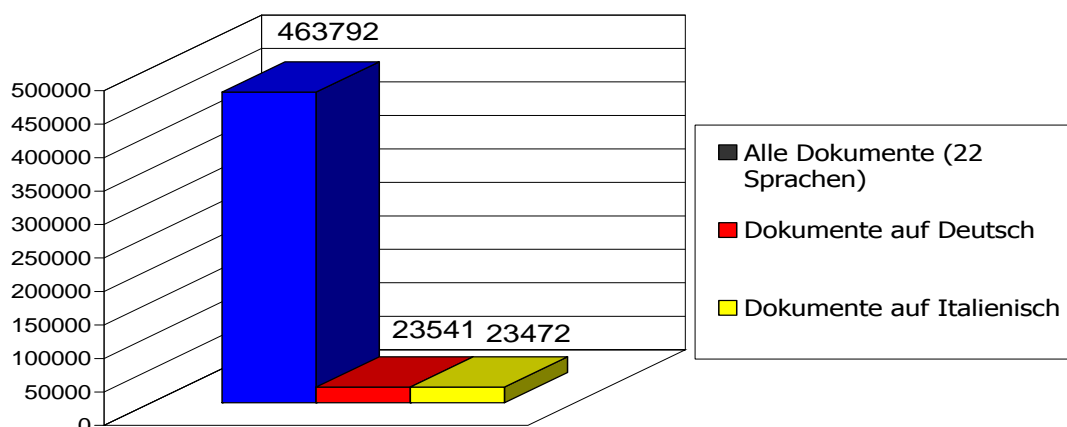


Abb. 1: Dokumente auf Deutsch und auf Italienisch im JRC-Acquis Korpus

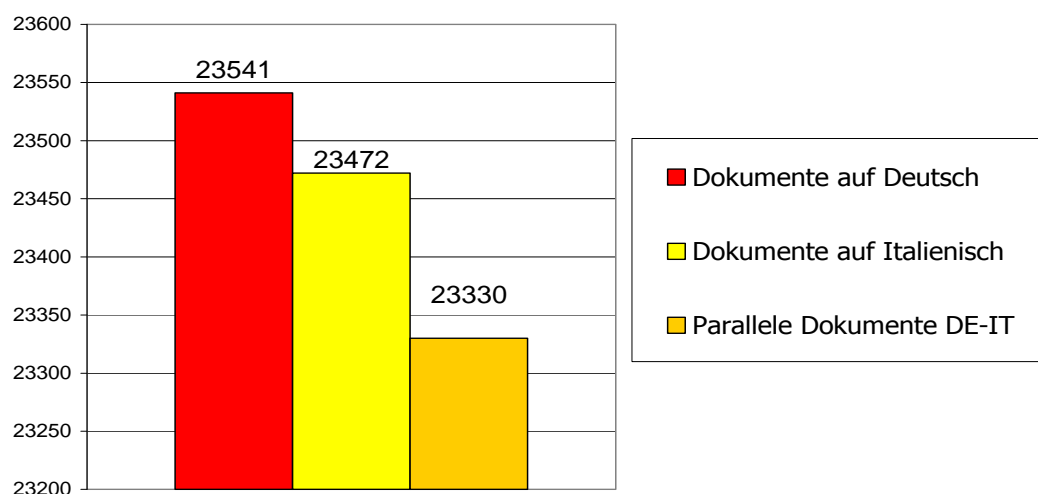


Abb. 2: JRC-Acquis Parallelkorpus DE-IT

Bevor wir zur Korpusanalyse übergehen, ist auf einige problematische Aspekte des *JRC-Acquis* hinzuweisen.

Das nach dem Zweiten Weltkrieg immer größer werdende Bedürfnis nach internationaler Verständigung spiegelt sich unmittelbar in der wesentlichen Rolle der Translation in allen Bereichen der EU wieder. Von Anfang an war es ein wichtiges Anliegen innerhalb der EU, den Sprachen aller Mitgliedstaaten die gleiche Bedeutung beizumessen. Mit der späteren Erweiterung war dies aber nicht mehr möglich. Zwar gilt gerade für Rechtstexte und Amtsblätter das sogenannte Vollsprachenregime (die Dokumente werden in allen EU-Amtssprachen vorgelegt), *de facto* wird intern aber mit Englisch und Französisch gearbeitet. Dadurch unterliegt das Übersetzungsverfahren innerhalb der EU zweierlei konditionierenden Einflüssen: Einerseits werden die Texte nicht für jedes Sprachenpaar übersetzt (selbst übersetzte Texte gelten als Vorlage für weitere Übersetzungen, mit einem Wort: als Original) und andererseits sind die Originale nicht immer von Muttersprachlern verfasst.¹⁴

Die Auswirkungen dieses Vorgehens auf die Qualität der Texte und deren Übersetzungen sind evident, trotzdem sind die innerhalb der EU verfassten Texte und Übersetzungen stets maßgeblich. Auf ihnen basieren *Translation Memories*, und damit gelten sie, einschließlich deren "Fehler", als Vorbild für weitere Übersetzungen. So schrieb Wolfram Wilss (1999) über das Übersetzen in der EU:

For all translators, so-called "integrated workstations" with appropriate high-tech equipment are to be created ("workbenches"). These are currently being tested. "Translation memories" are being developed, which allow all documents with recurring text segments and all administrative documents to be standardized as far as possible in accordance with

¹⁴ Vgl. Wilss (1999: 51f.).

certain patterns (standard configurations), thus accelerating the handling of translation work. The amount of "telework" is also to be increased, i.e. the use of freelance translators. There is no possibility for translators to influence or even comment on the form of the texts they work with, even though the original texts (which, as mentioned above, are frequently produced by non-native speakers) are often far from perfect. In a few especially notable cases translators have managed to refuse to translate unacceptable texts. (Wilss 1999: 57)

Auf dem *Acquis-Communautaire* basiert faktisch eine gleichfalls frei zugängliche *Translation Memory*, das *DGT-TM* (DGT = Directorate-General for Translation of the European Commission). Das *DGT-TM* umfasst vor allem fast alle Dokumente des *JRC-Acquis* sowie zusätzlich weitere Dokumente und bietet eine enorme Menge von Übersetzungseinheiten, aus denen man aber die Texte nicht rekonstruieren kann. Im Gegensatz zum *JRC-Acquis* wurde das Alignment beim *DGT-TM* manuell (d.h. mit *Euramis-Alignmenteditor*) korrigiert. Deutsch und Italienisch sind dabei mit mehr als 500.000 Übersetzungseinheiten je Sprache vertreten.¹⁵

Besonderes Augenmerk ist außerdem auf die Klassifikation der Texte im *JRC-Acquis* zu richten, die vor allem am Namen der Dateien zu erkennen ist. Der Dateiname (z.B.: "jrc31993R2454-de.xml") besteht aus: Kennzeichen "jrc" + "Celex Code" + Kennzeichen der Sprache, in der die Texte verfasst oder in die sie übersetzt sind ("de", "it"...). Folgendes Schema (Abb. 3) veranschaulicht, wie der *Celex Code* aufgebaut ist:¹⁶

3	1993	R	2454
Celex Dokumentationsbereich (Bereich 3 = abgeleitetes Gemeinschaftsrecht)	Jahr des Erlasses oder der Veröffentlichung des Rechtsakts (1993)	Rechtsform (R = Verordnung)	Nummer des Rechtsakts (Verordnung Nr. 2454 des betreffenden Jahres)

Abb. 3: Aufbau des Celex-Codes der Europäischen Gemeinschaften

Im Hinblick auf eine Klassifikation der Textsorten sind die Hinweise in der ersten und dritten Spalte bedeutsam: der "Celex-Dokumentationsbereich" und die "Rechtsform". Alle Dokumente des *JRC-Acquis* werden nach diesen zwei Kriterien kategorisiert. Juristische Textsorten sind also erstens nach einem der vier Dokumentations- bzw. Rechtsbereiche klassifiziert und zweitens nach neun Rechtsformen (Abb. 4):¹⁷

¹⁵ Vgl. folgende Webseite: <http://langtech.jrc.it/DGT-TM.html>.

¹⁶ Der Inhalt der Tabelle ist rechtlich geschützt (© Europäische Gemeinschaften 1998-2009) und abrufbar unter: <http://europa.eu/eur-lex/de/information/help/help-dir.html#codes>.

¹⁷ Die Inhalte dieser Tabelle entstammen gleichfalls der Webseite der EU: <http://europa.eu/eur-lex/de/information/help/help-dir.html#codes>.

Rechtsbereiche	
1	Verträge zur Gründung der Europäischen Gemeinschaften sowie die Verträge zu ihrer Änderung oder Ergänzung
2	Aus den Außenbeziehungen der Europäischen Gemeinschaften (oder den von den Mitgliedstaaten geschlossenen Verträgen, sofern sie die Außenbeziehungen der Gemeinschaften betreffen) hervorgegangenes Recht
3	Abgeleitetes Gemeinschaftsrecht
4	Komplementärrecht (Beschlüsse der im Rat vereinigten Vertreter der Regierungen der Mitgliedstaaten, zwischen den Mitgliedstaaten in Anwendung der Vertragsbestimmungen geschlossene Völkerrechtsabkommen usw.)
Rechtsformen	
A	Abkommen
B	Haushalt
D	Entscheidung/Beschluss
K	EGKS (<i>Europäische Gemeinschaft für Kohle und Stahl</i>)-Empfehlung
L	Richtlinie
R	Verordnung
S	Allgemeine EGKS-Entscheidung
X	Sonstige Rechtsakte
Y	Sonstige Akte (im Amtsblatt Reihe C veröffentlicht)

Abb. 4: Rechtsbereiche und Rechtsformen nach Hinweisen der Europäischen Gemeinschaften

Aus mehreren Gründen sind die Texte des *JRC-Acquis* forschungsrelevant, u.a. deshalb, weil sie durch *Translation Memories*, Terminologiedatenbanken und andere auch den externen (freelance) Übersetzern zur Verfügung gestellten Instrumente als Vorbild für weitere Übersetzungsarbeiten im juristischen Fachbereich dienen.¹⁸ All diese Ressourcen fungieren also als Hilfsmittel zum Verfassen von Texten und Übersetzungen im Rahmen internationaler sowie nationaler Rechtsbereiche und spielen daher im Allgemeinen in der Produktion und in der Auslegung juristischer Texte eine entscheidende Rolle.¹⁹ Das Verfassen von Texten bzw. Übersetzungen in mehreren Sprachen innerhalb einer internationalen Einrichtung wie der EU kann praktisch zu Veränderungen im Gebrauch der Fachsprache auch im nationalen Bereich führen.

Rechtssprache ist immer an eine Rechtsordnung oder, m.a.W., an ein Rechtssystem gebunden. [...] In gewisser Weise haben auch supranationale und internationale Rechtsordnungen Rechtssprachen hervorgebracht, die sich zumindest teilweise von den nationalen Rechtssprachen unterscheiden, gleichzeitig aber notwendigerweise eng an diese gebunden sind. (Wiesmann 2004: 19-20)

Zusammenfassend lässt sich sagen, dass das Verfassen von Texten und Übersetzungen in der EU im Sinne einer multilingualen und nicht bilingualen Auffassung von Translation zu verstehen ist. In diesem Zusammenhang wäre es interessant, der Frage nach-

¹⁸ Laut Erklärungen der *Generaldirektion Übersetzung* vergibt die EU heute ungefähr 26% ihrer gesamten Übersetzungsarbeiten freiberuflichen Übersetzern als externe Aufträge; vgl. http://ec.europa.eu/dgs/translation/index_en.htm.

¹⁹ Über die Rolle, die die Sprache (und damit die Übersetzung der Texte) für die Auslegung der Rechtstexte im jeweiligen nationalen Kontext spielt, vgl. Gréciano (2003).

zugehen, ob sich ein derartiges Verständnis von Übersetzung in der Zukunft der globalisierten Welt immer mehr durchsetzen wird und wie sie die Translation von Fachtexten in Theorie und Praxis verändern könnte.

4 Korpusanalyse

Um eine realistische Basis für eine quantitative und qualitative Korpusanalyse aus translationswissenschaftlicher Perspektive zu erhalten, haben wir den Texten des *JRC-Acquis* ein kleineres Korpus entnommen, ein sogenanntes Subkorpus.

Das *JRC-Acquis* ist zu groß für eine vollständige Analyse und tatsächlich arbeiten etliche Forschungsgruppen der Computerlinguistik noch daran, geeignete Software für die breiten Analysemöglichkeiten des *JRC-Acquis* zu entwerfen.²⁰ Jede Art Datenverarbeitung mit einem normalen Rechner wird zudem durch die enorme Datenmenge massiv erschwert. Schließlich mussten wir für die zur qualitativen Analyse herangezogenen Texte das Alignment-Verfahren manuell überprüfen.

Eine vollständige quantitative Analyse ist nicht unser Hauptanliegen, das vielmehr darin besteht, zu zeigen, in welcher Hinsicht die Korpusanalyse der Übersetzungswissenschaft dienlich sein kann, wie eine translationsorientierte Korpus- und Textanalyse konkret erfolgen könnte, welche Fragestellungen der korpusbasierten Translationswissenschaft noch offen bleiben. Dafür soll unser Korpus etwas überschaubarer werden.

Wir wollen trotzdem die quantitative Analyse aus dieser Untersuchung nicht ausschließen, denn sie stellt einen wichtigen Aspekt jeder korpusbasierten Studie dar.

Das Analyseverfahren gliedert sich in mehrere Stufen, wobei maschinelle (hauptsächlich quantitative) Untersuchungen mit manuellen (hauptsächlich qualitativen) Analysen verbunden werden. Wir werden also ein relativ breites Subkorpus vorstellen und anschließend auf Einzelbeispiele näher eingehen.

4.1 Das Subkorpus

Die Dokumente des *JRC-Acquis* verteilen sich nicht gleichmäßig auf die Zeitspanne von 1958 bis 2006. Das Korpus enthält z.B. nur 5 Dokumente (je Teilkorpus) von 1958 und mehr als 7000 von 2006 (7239 Dateien im deutschen Teilkorpus, 7128 im italienischen). Bis zum Jahre 1984 enthält jedes Teilkorpus weniger als 100 Dokumente pro Jahr. Statt Dokumente aus verschiedenen Jahren beliebig auszuwählen, haben wir uns also entschlossen, chronologisch aufeinander folgende Dokumente heranzuziehen, in der Überzeugung, dass im Laufe einer gewissen Zeitspanne Dokumente verschiedener Textsorten (d.h. verschiedener Rechtsbereiche und Rechtsformen) vertreten sind. Wir haben uns demzufolge auf die Dokumente zweier aufeinander folgender Jahrgänge beschränkt, sodass ungefähr 500 Dokumente pro Teilkorpus zur Verfügung stehen. Unser Subkorpus umfasst die Dokumente der Jahre 1991 und 1992 (s. im Detail Abb. 5). Das Parallelkorpus besteht ausschließlich aus Texten, die in beiden Sprachen

²⁰ Vgl. die Webseite: http://langtech.jrc.it/0509_EU-Enlargement-Workshop.html.

vorliegen. Wie die Tabelle veranschaulicht, wurde infolgedessen nur ein einziger Text aus dem italienischen Teilkorpus nicht berücksichtigt:²¹

	DE	IT	Parallelkorpus DE-IT
Anzahl der Dokumente	452	453	452 + 452 = 904
Dateiformat	xml	xml	xml
Dokumentengröße gesamt	11,4 MB	11,4 MB	22,8 MB
Größe der einzelnen Textdatei	4KB bis 583 KB	4KB bis 592 KB	/
Worttypes gesamt (<i>case-sensitive</i>)	53757	36632	82807
Worttypes gesamt (<i>case-insensitive</i>)	49011	31384	73156
Worttokens gesamt	1613532	1675104	3288636

Abb. 5: Daten zum untersuchten JRC-Acquis Subkorpus

Folgende Diagramme (Abb. 6 und 7) veranschaulichen die Anteile der Rechtsbereiche und Rechtsformen, wie diese im Subkorpus vertreten sind, und zeigen eine Übersicht über das Merkmal "Textsorte" im Subkorpus.²² Wie zu erwarten ist, gibt es im Subkorpus – da es die Texte aus den Jahren 1991 und 1992 enthält – keine Texte zur Gründung der Europäischen Gemeinschaften (Bereich 1). Texte zum abgeleiteten Gemeinschaftsrecht überwiegen stark (99%), nur zwei sind Texte des Komplementärrechts, 1% entfällt auf Rechtstexte mit Schwerpunkt "Außenbeziehungen der EU".

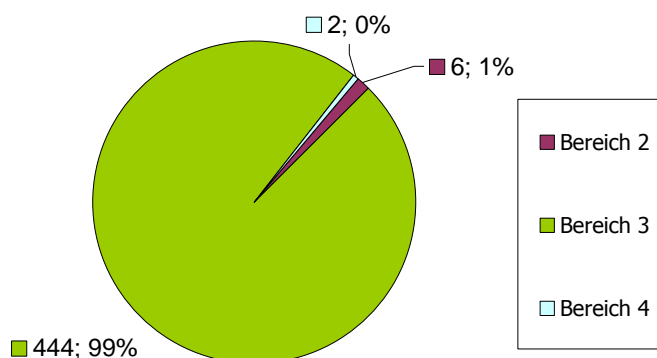


Abb. 6: Rechtsbereiche im JRC-Acquis Subkorpus

²¹ Als "Worttypes" werden Lexeme gemeint, die nur einmal gezählt wurden. Worttypes werden hier mit (*case-sensitive*) und ohne Beachtung der Groß- und Kleinschreibung (*case-insensitive*) gezählt. "Worttokens" beziehen sich hingegen auf das Vorkommen aller Wörter. Wir haben ein *open source*-Konkordanzprogramm verwendet: AntConc 3.1 (© Anthony Laurence, 2005).

²² Für die Bedeutung der Bereich-Zahlen in Abb. 6 und der Buchstaben in Abb. 7, vgl. Abb. 4 im vorliegenden Aufsatz.

Sehr relevant ist auch die Analyse der Rechtsformen, die noch mehr Informationen über die Textsorten im Korpus bietet. Das Diagramm auf Abb. 7 zeigt, dass die am meisten vertretenen Rechtsformen "Verordnungen" und "Richtlinien" sind. Beide haben relativ gut vergleichbare prozentuale Anteile. Auf die Rechtsform "Entscheidung/Beschluss" entfällt fast ein Fünftel des Gesamtkorpus. Einen äußerst geringen Anteil stellen schließlich die Formen "Abkommen" und "Sonstige Rechtsakte". Diese Daten sind nicht nur hinsichtlich der Textsorten im Korpus bedeutsam, sondern auch deswegen, weil sie uns bei der Interpretation der Ergebnisse der quantitativen Analysen behilflich sein können.

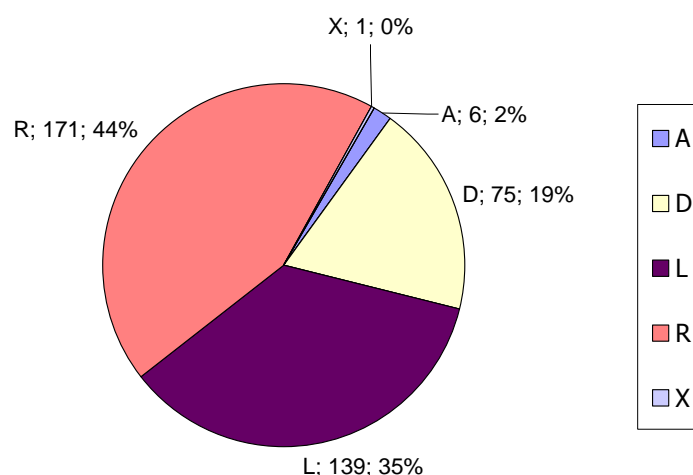


Abb. 7: Rechtsformen im JRC-Acquis Subkorpus

4.2 Untersuchungen auf der Wortebene

Auf der Wortebene können Korpora unterschiedlichen Zwecken dienen, u.a. um zu erforschen, wie Wörter in einer bestimmten Sprache gebildet, wie Wortstämme, Affixe und Flexive miteinander kombiniert werden, oder wie tief Neologismen in einer Sprache integriert sind. Auf Korpora stützt sich auch der Lexikograph, der ein Wörterbuch erstellen oder überarbeiten möchte. Korpora dienen in diesen Fällen als Belegsammlungen für die untersuchten Phänomene.

Aus translationswissenschaftlicher Sicht sind Korpora auf der Wortebene vor allem interessant, weil sie eine Wortschatz-Ressource darstellen. Das *JRC-Acquis* z.B. bietet einen riesigen Bestand an juristischem Fachwortschatz. Es lohnt sich mithin zu recherchieren, welche Wörter am häufigsten vorkommen, in welchem Kontext sie vorkommen und wie sie übersetzt werden.

Die erste übersetzungsrelevante quantitative Analyse unseres Untersuchungskorpus betrifft demzufolge die Aufstellung einer Liste von Lexemen in beiden Sprachen. Das lässt sich mit einer geeigneten Software ohne großen Aufwand machen. Hier werden beispielhaft nur die ersten zehn häufigsten Substantive im deutschen und italienischen Teilkorpus aufgelistet (Abb. 8), man könnte aber natürlich mehrere und längere solcher Listen zusammenstellen. Im Grunde basieren auch *Translation*

Memories und die damit bezogenen Terminologiedatenbanken auf derartigen Analysen und tatsächlich konnte man davon ausgehen, dass die am häufigsten vorkommenden Lexeme – je länger die Liste wird – den Grundwortschatz für Übersetzer bilden, die mit ähnlichen Textsorten arbeiten oder arbeiten werden. Die Suche erfolgte ohne Beachtung der Groß- und Kleinschreibung (*case-insensitive*), weil im Italienischen nur Eigennamen großgeschrieben werden (und Wörter wie "commissione", "articolo", etc. im Korpus sowohl als klein- als auch als großgeschrieben vorkommen).²³

DE			IT		
Lexem	Rang	Häufigkeit	Lexem	Rang	Häufigkeit
Artikel	16	10287	articolo	19	11438
EWG	21	7731	CEE	29	7638
Richtlinie	32	5920	direttiva	33	6010
Verordnung	38	4788	stati	40	4917 / Stati (3974)
Kommission	39	4425	regolamento	41	4874
Absatz	44	4121	Commissione	43	4441
Mitgliedstaaten	49	3981	membri	45	4421
Anhang	64	2841	paragrafo	47	3987
ABl [Amtsblatt]	65	2837	prodotti	55	3505
Rates	84	2175	allegato	57	3414

Abb. 8: Häufig vorkommende Substantive im deutschen und italienischen Teilkorpus

Diese Daten sind auch aus anderen Gründen relevant: Wie die Tabelle (Abb. 8) veranschaulicht, entsprechen den am häufigsten vorkommenden Substantiven aus einem Teilkorpus nicht immer deren Übersetzungen im anderen Teilkorpus (wir haben hier die häufigsten Wörter herangezogen, aber je niedriger der Rang wird, desto größer werden die Unterschiede zwischen den Teilkorpora). Dies könnte bedeuten, dass die Suche nach der "geeigneten" Übersetzung in den jeweiligen Kontexten und in den jeweiligen Sprachen zu unterschiedlichen Lexemen geführt hat. Diese Vielfalt könnte auch als Zeichen einer humanen und nicht völlig automatischen Übersetzung gelesen werden. Man könnte diese Diskrepanz aber natürlich als Zeichen bestimmter Tendenzen im Übersetzungsverfahren lesen, z.B. die Tendenz der *simplification*, von der wir im Abschnitt 2 gesprochen haben.

Wenn man die Daten der Tabelle in Abb. 5 genau betrachtet, fällt der Unterschied zwischen der Anzahl der Worttypes in den beiden Teilkorpora besonders auf: Während die Werte für die Worttokens vergleichbar sind (im Italienischen ist der Anteil sogar größer), erscheint der Anteil der Worttypes wesentlich kleiner im italienischen als im deutschen Teilkorpus. Man könnte daraus schließen, dass der im italienischen Teilkorpus benutzte Wortschatz – bei einer vergleichbaren gesamten Wörteranzahl –

²³ Folgende Tabelle beinhaltet die am häufigsten auftretenden Substantive in beiden Sprachen zusammen mit deren Häufigkeitswert und Rang. Mit "Rang" ist die vom jeweiligen Lexem in der Liste der gesamten sortierten Wörter eingenommene Position gemeint (So ist das Wort "Artikel" z.B. das am häufigsten vorkommende Substantiv, tritt aber erst an sechzehnter Stelle auf; die ersten Positionen sind von anderen Wörtern – nicht von Substantiven – besetzt).

wesentlich ärmer ist und dass dieselben Wörter viel häufiger wiederholt werden als die deutschen Wörter im entsprechenden deutschen Korpus.

Das Wort "prodotti" (Abb. 8: Lexem-Spalte, italienisch) weist einen auffällig hohen Rang als Substantiv auf (als Partizip wird es viel seltener gebraucht). Es lohnt sich, diesen Aspekt zu vertiefen, indem man z.B. die Konkordanz des Wortes in Betracht zieht. Die Konkordanz, auch *KWIC (Keyword in Context)* genannt, besteht in der Auflistung aller Kontexte, in denen ein gegebenes Wort vorkommt, und vermittelt dadurch genauere Informationen darüber, in welchem Zusammenhang jenes Wort gebraucht wird. Auf derartigen Analysen unseres parallelen Subkorpus beruhen folgende Beobachtungen.

Die häufigste deutsche Übersetzung von "prodotti" ist – wie folgendes Beispiel zeigt – "Erzeugnisse", ein Wort, das nur etwa 1200 Mal im Korpus vorkommt (dies ist nur ein Drittel im Vergleich zum Auftreten des Wortes "prodotti" in den entsprechenden italienischen Texten; wenn man Komposita mitberücksichtigt, steigt die Anzahl um 35 %).²⁴

[1, it] visto il regolamento (CEE) [...], relativo a misure specifiche in favore delle Azzorre e di Madera per taluni **prodotti** agricoli

[1, de] gestützt auf die Verordnung (EWG) [...] zum Erlaß von Sondermaßnahmen für bestimmte landwirtschaftliche **Erzeugnisse** zugunsten der Azoren und Madeiras

Die unterschiedliche Frequenz lässt sich einerseits durch eine breitere Präsenz von deutschen Übersetzungsvarianten an den gleichen Textstellen des italienischen Wortes "prodotti" (2), andererseits durch die Anwendungen von Komposita des Wortes "Erzeugnisse" (3) erklären. Viel seltener (etwa 1,5 % der Gesamtzahl; in Abb. 8 sind Komposita nicht berücksichtigt) weist auch das italienische Wort "prodotti" Komposita auf (3: "sottoprodotti").

[2, it] In attesa della conclusione formale, il verbale concordato che modifica l'accordo [...] sul commercio dei **prodotti tessili** è applicato in via provvisoria con effetto dal 1° gennaio 1991, con riserva della reciproca applicazione provvisoria da parte dell'altra parte contraente.

[2, de] Die Vereinbarte Niederschrift zur Änderung des Abkommens [...] über den Handel mit **Textilwaren** wird ab 1. Januar 1991 bis zu ihrem förmlichen Abschluß vorläufig angewandt, vorausgesetzt, daß die andere Vertragspartei sie ebenfalls vorläufig anwendet.

[3, it] Vegetali e **prodotti vegetali** [...] In attesa che vengano adottate norme tecniche comuni relative alle produzioni animali biologiche, i **sottoprodotti** dell'allevamento

²⁴ Das *JRC-Acquis multilingual parallel corpus* ist frei zugänglich. Jedoch soll man Folgendes beachten: "Only European Community legislation printed in the paper edition of the Official Journal of the European Union is deemed authentic". Für die Übersetzungen gilt folgender Hinweis: "Originally published in the official languages of the European Union in the Official Journal of the European Union by the Office for Official Publications of the European Communities. Responsibility for the translation into Italian (or German) from the original edition lies entirely with the respective translation copyright holder". Das gilt für alle in diesem Beitrag zitierten Dokumente, Abschnitte und Sätze aus dem *JRC-Acquis* Korpus. Es sei an dieser Stelle auch daran erinnert, dass weder die deutsche noch die italienische Sprache als Ausgangs- bzw. Zielsprache im traditionellen Sinne betrachtet werden kann, da dies aufgrund der multilingualen Gestaltung des *JRC-Acquis* nicht möglich ist.

possono essere utilizzati [...] deve essere inoltre tenuta una contabilità su registri o su documenti della natura, delle quantità e dei destinatari di tutti i **prodotti agricoli** venduti.

[3, de] Pflanzen und **Pflanzenerzeugnisse** [...] Im Hinblick auf die Genehmigung gemeinsamer technischer Vorschriften für die biologische tierische Produktion können **Nebenerzeugnisse** der Tierhaltung verwendet werden [...] ferner ist eine durch Eintragung oder Unterlagen belegte Betriebsbuchführung über Art, Menge und Abnehmer aller verkauften **Agrarerzeugnisse** Vorschrift.

Auch in dieser Hinsicht scheint es also im italienischen Teilkorpus eine größere Tendenz zur lexikalischen *simplification* zu geben, obwohl sich die quantitativen Ergebnisse zum Teil auch durch eine stärkere lexikalische Verdichtung des deutschen Wortes "Erzeugnisse" im Vergleich zum italienischen Wort "prodotti" erklären lassen könnten.

Der im Deutschen verbreitetere Gebrauch von Komposita begründet – mindestens teilweise – den oben erwähnten Unterschied zwischen der Anzahl der Worttypes im deutschen und im italienischen Teilkorpus. Lässt sich aber die Tatsache, dass die Anzahl der Worttokens im Italienischen sogar größer ist als im entsprechenden deutschen Korpus, allein durch den häufigeren Gebrauch von Komposita im Deutschen erklären?

Eine weitere mögliche Erklärung für die unterschiedliche Wortfrequenz und die unterschiedliche Anzahl der Worttypes ergibt sich aus der verschiedenartigen Mikrostruktur der Fachtexte in beiden Sprachen. Qualitative Analysen zeigen, dass bei der Übersetzung der Texte auch ihr syntaktischer Aufbau stark verändert wird, da auch sprachen- und kulturspezifische Merkmale des Fachtextes berücksichtigt werden. Sehr oft wird in den italienischen Texten etwa bei der Anführung der Grundlagen für eine Richtlinie das Gerundium gebraucht, um eine Reihe von Nebensätzen einzuführen. In der entsprechenden deutschen Version hingegen führt man eingangs einen neuen Satz ein und verzichtet im Folgenden auf Wiederholungen. Bei den folgenden Sätzen handelt es sich – dank diesem stilistischen Eingriff – meist um Hauptsätze, so dass der deutsche Text mehr an Verständlichkeit gewinnt. Folgende Tabelle (Abb. 9) bietet dafür ein Beispiel, wie quantitative Analysen auf der Wortebene durch qualitative Untersuchungen, in diesem Fall durch die vergleichende Textanalyse ergänzt werden können. Sie liefert insbesondere ein syntaktisch-stilistisches Muster, das praktisch alle Texte der Sorte "Richtlinie" der jeweiligen Teilkorpora durchzieht. Der einleitende Teil wird durch die Großbuchstaben des Hauptsatzes ("DER RAT DER EUROPÄISCHEN GEMEINSCHAFTEN / IL CONSIGLIO DELLE COMUNITÀ EUROPEE [...] HAT FOLGENDE RICHTLINIE ERLASSEN / HA ADOTTATO LA PRESENTE DIRETTIVA") markiert. Innerhalb des ganzen Teiles zeigen die Texte ganz unterschiedliche Gliederungssignale (de: "in Erwägung nachstehender Gründe" + Reihe von Aussagesätzen; it: statt eines einmaligen einleitenden Ausdruckes die Wiederholung des Gerundiums "considerando" + Anreihung von Nebensätzen durch "che"), hinter denen zwei kulturspezifische Denkweisen stecken.

<p>de: DER RAT DER EUROPÄISCHEN GEMEINSCHAFTEN – gestützt auf [...] auf Vorschlag [...] nach Stellungnahme [...] nach Stellungnahme [...] in Erwägung nachstehender Gründe: In der Richtlinie [...] über [...] sind gemeinsame Regeln zu [...] erlassen worden. Die Entschließung des Rates [...] sowie das Aktionsprogramm der Europäischen Gemeinschaften [...], das Gegenstand der Entschließung des Rates der Europäischen Gemeinschaften [...] über [...] ist, sehen gemeinschaftliche Maßnahmen zu [...] vor. Die allgemeinen Regeln für [...] Die ordnungsgemäße Bewirtschaftung [...] Es muß sichergestellt werden, daß [...] Die Anpassung der Vorschriften der vorliegenden Richtlinie an [...] HAT FOLGENDE RICHTLINIE ERLASSEN: [...]</p>	<p>it: IL CONSIGLIO DELLE COMUNITÀ EUROPEE, visto [...] vista la proposta [...] visto il parere [...] visto il parere [...] --- considerando che la direttiva [...], relativa a [...], ha stabilito norme comunitarie su [...] considerando che la risoluzione del Consiglio [...] ed il programma d'azione delle Comunità europee [...], il quale è oggetto della risoluzione del Consiglio delle Comunità europee [...], concernente [...], prevedono misure comunitarie per [...] considerando che la normativa generale su [...] considerando che una corretta gestione [...] considerando che è necessario provvedere affinché [...] considerando che deve essere possibile adattare rapidamente le disposizioni della presente direttiva a [...] HA ADOTTATO LA PRESENTE DIRETTIVA: [...]</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Abb. 9: Beispiel für qualitative Analyse anhand des Textvergleichs

Hat man auch nur die kurzen Abschnitte der Texte in Abb. 9 durchgelesen, dann wird sofort klar, wie aufschlussreich korpusbasierte Analysen für terminologische Untersuchungen (und darüber hinaus) sein können. Dazu gehört auch die Ermittlung juristischer Formeln (Ausdrücke wie "gestützt auf", "in Erwägung nachstehender Gründe" usw.) durch das Untersuchungsinstrument, das nach Nachbarwörtern sucht. Untersuchungen auf der Wortebene erweisen sich im Fall des juristischen Textes als besonders nützlich, weil dieser Fachtext stark durch lexikalische Dichte gekennzeichnet ist.

In diesem Abschnitt haben wir einen kurzen Querschnitt durch die Möglichkeiten der Korpusanalyse auf der Wortebene gegeben. Mit Wörterlisten, KWIC, *Collocates* kann man die Recherchen fortsetzen, quantitative Überblicke herstellen und qualitativ auf Einzelbeispiele eingehen. Analysen auf der Wortebene stellen die Vorstufe vieler Korpusanfragen dar, die man natürlich – besonders mittels der Annotation – noch unterspezifizieren könnte.

4.3 Modalverben im Übersetzungsvergleich

Wie wir am Beispiel der Texte in Abb. 9 gesehen haben, ist die Rolle der Textsorte (Abb. 9 zeigt ein Beispiel für die Textsorte "Richtlinie") sowie der spezifischen Merkmale des Fachtextes bei der Auswertung der Daten sehr wichtig. Betrachtungen zum Fachtext und zur Textsorte können zudem zu bestimmten Korpusanfragen anregen. Dies ist z.B. bei Modalverben der Fall, denn sie sind im juristischen Fachtext besonders häufig vertreten.²⁵

Wir werden in diesem Abschnitt versuchen, anhand von quantitativen und qualitativen Analysen die Verwendungsweise der Modalverben in den Übersetzungen unseres Subkorpus zu erörtern. Die Untersuchung betrifft die traditionellen Modalverben – für das Deutsche: "dürfen", "können", "mögen", "müssen", "sollen", "wollen", für das Italienische: "potere", "dovere", "volere". Modalausdrücke unterschiedlicher Natur finden im qualitativen Teil der Analyse Berücksichtigung, sofern sie dazu beitragen, für die Unterschiede im Gebrauch der Modalverben in den beiden Teilkorpora nähere Erklärungen zu bringen. Es ist nicht nur das Italienische, das aufgrund der geringen Zahl der Modalverben auf Umschreibungen mit modaler Bedeutung zurückgreift; auch im deutschen Korpus findet man besondere Merkmale des modalen Ausdrucks. Beim Übersetzungsvergleich zeigt sich ein fachtextspezifischer Gebrauch von Modalverben.

In Abb. 10 und Abb. 11 ist jeweils die Häufigkeit der italienischen und deutschen Modalverben im Subkorpus angegeben. In beiden Fällen werden nur Tempora und Modi miteinbezogen, die im Korpus eine gewisse Relevanz aufzeigen, das gleiche Kriterium gilt für die Personenangaben (die 1. sowie die 2. Person Singular und Plural sind in juristischen Fachtexten kaum vertreten).

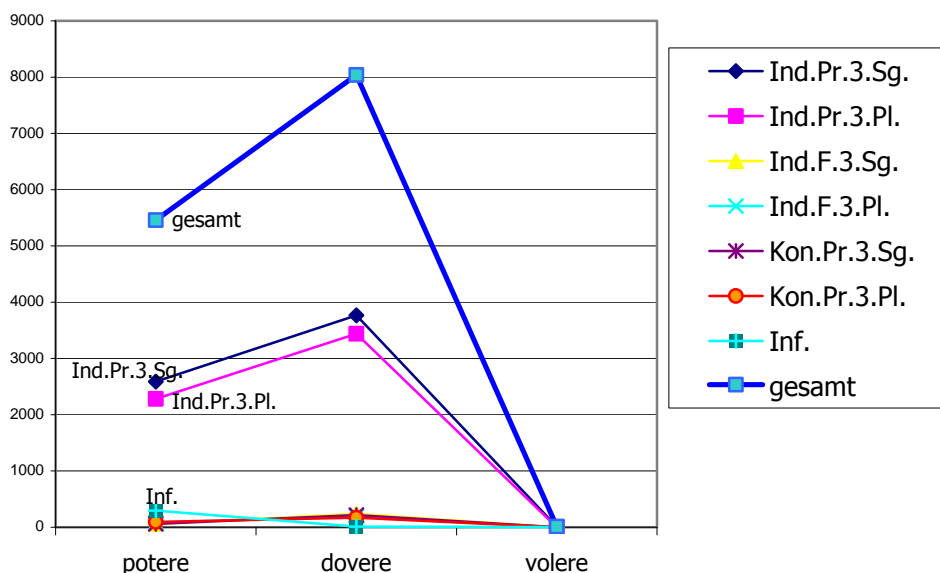


Abb. 10: Italienische Modalverben im Subkorpus

²⁵ Zu den Besonderheiten im Gebrauch von Modalverben im Fachtexten vgl. Heller (1999).

Abb. 10²⁶ lassen sich folgende Beobachtungen entnehmen:

- Das Verb "dovere" erscheint dominant (es macht etwa 59% der Gesamtzahl der Modalverben aus);
- Das Verb "potere" ist relativ häufig (etwa 40%);
- Formen des Verbs "volere" sind praktisch nicht vertreten (etwa 0,1%);
- Was die Modi angeht, zeigt sich die absolute Dominanz des Indikativ Präsens; Futur Indikativ und Konjunktiv sind gleichmäßig vertreten; ziemlich gut vertreten ist auch der Infinitiv von "potere".

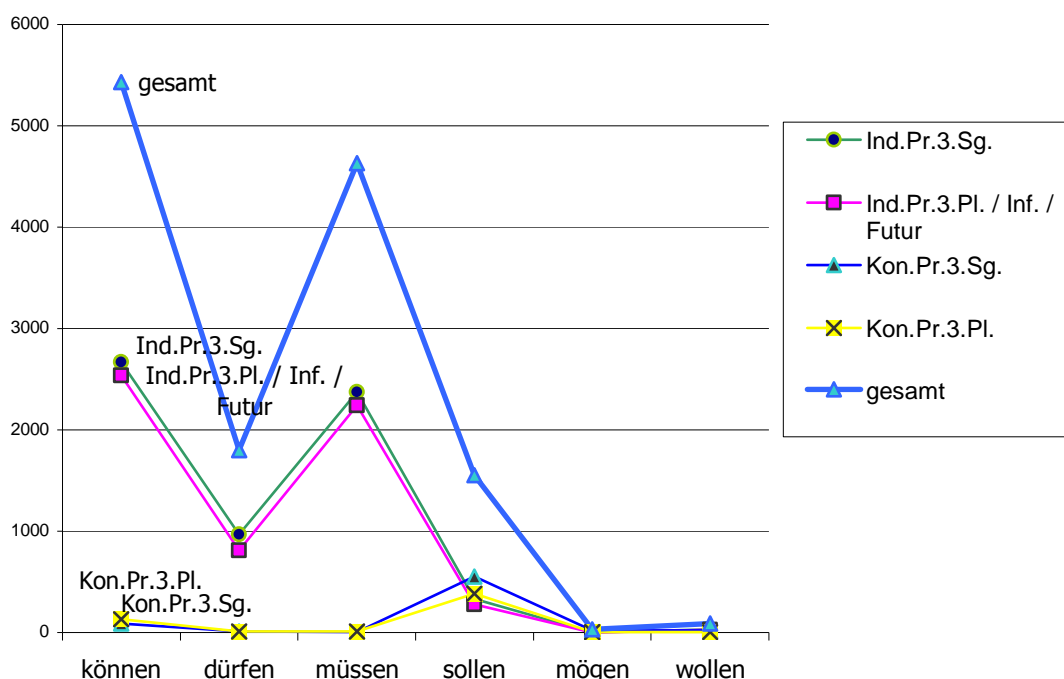


Abb. 11: Deutsche Modalverben im Subkorpus

²⁶ Wie die Abkürzungen der rechten Spalte andeuten, wurden folgende Modi/Tempora berücksichtigt: Indikativ Präsens: 3. Person Singular und Plural (può, deve, vuole; posso, devo, voglio); Indikativ Futur ["futuro semplice indicativo"]: 3. Person Singular und Plural (potrà, dovrà, vorrà; potranno, dovranno, vorranno); Konjunktiv Präsens ["congiuntivo presente"]: 3. Person Singular und Plural (potrebbe, dovrebbe, vorrebbe; potrebbero, dovrebbero, vorrebbero); Infinitiv Präsens (potere, dovere, volere). Besondere Schreibweisen wurden mitberücksichtigt ("poter" statt "potere" usw.). Es wurde natürlich auch auf die Disambiguierung zwischen dem Substantiv ("il potere", "il dovere"...) und den Modalverben selbst geachtet; dabei wurden die Substantive nicht in Betracht gezogen.

Abb. 11²⁷ – hingegen – veranschaulicht folgende Merkmale deutscher Modalverben im Subkorpus:

- Die Dominanz des Verbs "können" (etwa 40%; "können" und "dürfen" zusammen bilden etwa 53%);
- Das Verb "müssen" ist stark vertreten (etwa 34%; zusammen mit "sollen" etwa 46%);
- Das Verb "wollen" ist sehr selten vertreten, auf jeden Fall etwas mehr als im entsprechenden italienischen Teilkorpus; "wollen" und "mögen" zusammen bilden 1%);
- Was die Modi angeht, so zeigt sich, wie im Italienischen, die Dominanz des Indikativ Präsens (mit Ausnahme vom Verb "sollen", das vielmehr im Modus Konjunktiv vertreten ist), wobei die Form des Indikativ Präsens in der 3. Person Plural mit möglichen Formen des Infinitivs oder Futurs zusammenfällt.

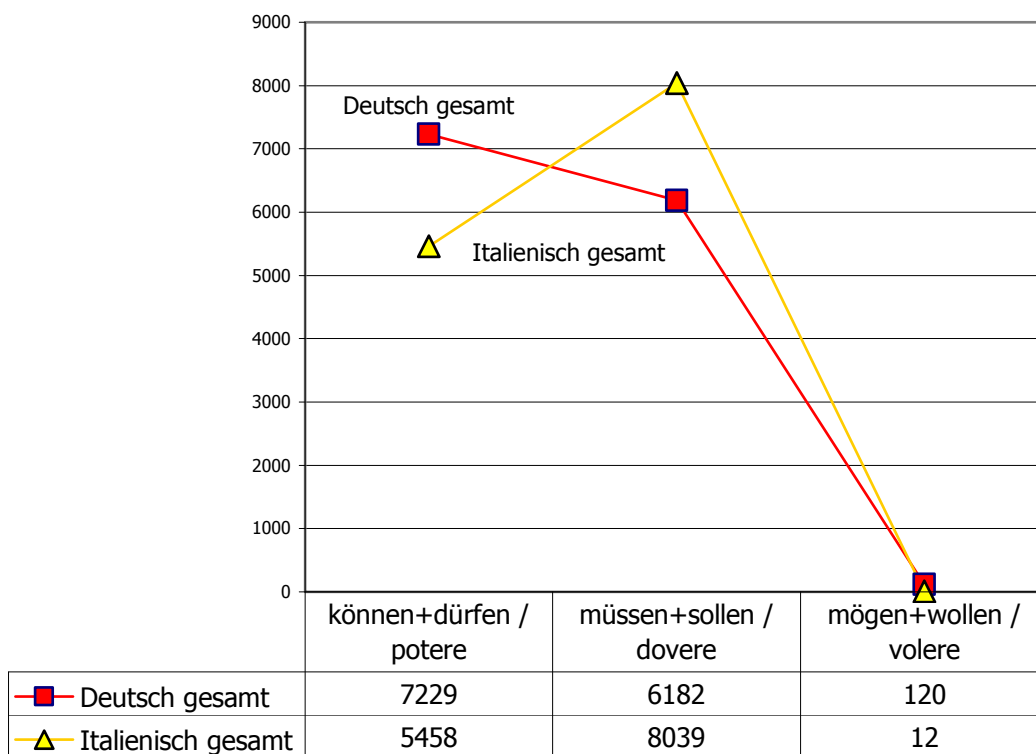


Abb. 12: Italienische und deutsche Modalverben im Vergleich

²⁷ Folgende Modi/Tempora wurden berücksichtigt: Indikativ Präsens: 3. Person Singular und Plural (kann, darf, muss...; können, dürfen, müssen...); Konjunktiv Präsens: 3. Person Singular und Plural (könnte, dürfte...; könnten, dürften...). Durch die Anwendung von sogenannten regulären Ausdrücken wurde auf Besonderheiten der Rechtschreibung geachtet.

Aus Abb. 12 lässt sich entnehmen, dass die Häufigkeitswerte von "potere" und "dovere" sich im Vergleich zu denen von "können"/"dürfen" und "müssen"/"sollen" genau umgekehrt verhalten. Das ist der markanteste Unterschied zwischen den zwei Teilkorpora.²⁸

Versuchen wir nun, die vorliegenden Ergebnisse der quantitativen Analyse zu interpretieren, indem wir auf einzelne Übersetzungsbeispiele eingehen.²⁹ Dabei gehen wir davon aus, dass die quantitativen Unterschiede durch Abweichungen von der "Standard"-Übersetzung (Übersetzung von "können"/"dürfen" mit "potere" und umgekehrt, von "müssen"/"sollen" mit "dovere" und umgekehrt) erklärbar sind. In allen Fällen, in denen "dovere" in einem italienischen Text vorkommt und im entsprechenden deutschen Text an der gleichen Stellen weder "müssen" noch "sollen" vorkommen, hat man im Deutschen das Modalverb einfach anders ausgedrückt (das Gleiche gilt für das Italienische). Folgende Tabelle (Abb. 13) veranschaulicht beispielhaft eine Reihe von Textabschnitten mit Übersetzungen von Modalverben. Die ersten Beispiele (Gruppe I) beziehen sich auf die Übersetzungen, die wir als "Standard" bezeichnet haben, die folgenden hingegen stellen jeweils alternative Ausdrücke im Deutschen (Gruppe II) und im Italienischen (Gruppe III) dar.

	de	it
	GRUPPE I	
1	Sie können zu [...] eingesetzt werden	[Essi] possono essere utilizzati per [...]
2	<i>Jede</i> in Absatz 1 genannte <i>Regierung kann</i> [...]	<i>I governi</i> di cui al paragrafo 1 del presente articolo possono [...]
3	Alternative Transaktionswerte dürfen nach Buchstabe b) nicht festgesetzt werden	Non possono essere stabiliti valori sostitutivi ai sensi della predetta lettera b)
4	Auf Österreichischem Territorium dürfen in der Gemeinschaft zugelassene Lastkraftwagen, Sattelanhänger und Anhänger verkehren	Possono circolare sul territorio austriaco gli autocarri, i semirimorchi e i rimorchi autorizzati nella Comunità
5	[...] müssen die Mitgliedstaaten die Betriebserlaubnis mit nationaler Geltung für einen Kraftfahrzeugtyp verweigern	gli Stati membri [...] devono rifiutare l'omologazione di portata nazionale di un tipo di veicolo a motore
6	Diese Anreize müssen im Einklang mit dem Vertrag stehen und [X] darüber hinaus folgende Bedingungen erfüllen	Detti incentivi devono essere conformi alle disposizioni del trattato e devono inoltre soddisfare le seguenti condizioni
7	Die Kommission soll einen Vorschlag für eine Richtlinie über Maßnahmen zu [...] vorlegen	considerando che la Commissione deve presentare una proposta di direttiva che prevede [...]
8	[...], die vor allem auf gemeinsamen Überwachungsindikatoren beruhen sollte	[...], il quale dovrà segnatamente basarsi su indicatori di sorveglianza comuni

²⁸ Das Verb "dürfen" besitzt einen besonderen Status, denn es wird in der negativen Form oft mit dem Italienischen "non dovere" übersetzt. In dieser quantitativen Untersuchung haben wir uns trotzdem an die "Standard"-Übersetzung gehalten (s. folgenden Abschnitt im Text), denn eine präzisere Untersuchung könnte erst nach einer Disambiguierung erfolgen. Dazu müsste das Korpus annotiert sein, was bei dem *JRC-Acquis* nicht der Fall ist.

²⁹ Aufgrund seiner extremen Unterzahl wird das Verb "volere" in die folgenden Analysen nicht einbezogen.

GRUPPE II		
9	[...] die nicht in der Lage sind , eine Änderung anzunehmen	[...] che non possono accettare un emendamento
10	Teilstimmen sind nicht zulässig	I voti non possono essere frazionati
11	Die Teilnahme im Ausschuß steht allen Mitgliedern offen	Tutti i membri possono partecipare al Comitato
12	Diese Dauerhaltbarkeitsprüfung ist auf 80 000 gefahrene Kilometer anzulegen	il controllo della durata deve essere eseguito a 80 000 km
13	Mit Ausnahme der Fahrzeuge nach Abschnitt 8.1 werden Fahrzeuge mit Fremdzündungsmotor den folgenden Prüfungen unterzogen	Fatta eccezione dei veicoli di cui al punto 8.1, i veicoli muniti di motore ad accensione comandata devono essere sottoposti alle seguenti prove
14	nach den vom Rat [...] festzulegenden Hygienevorschriften	conformemente a norme igieniche che dovranno essere stabilite dal Consiglio
GRUPPE III		
15	Unternehmen, die das Direktversicherungsgeschäft betreiben, können in der Regel auch Rückversicherungen übernehmen	le imprese che esercitano l'assicurazione diretta praticano generalmente anche la riassicurazione
16	<i>Wenn</i> Diethylether <i>nicht benutzt werden kann</i> [...]	<i>se non è possibile</i> usare l'etere etilico [...]
17	<i>Dadurch kann</i> die Laufzeit um ein Drittel verkürzt [...] werden	<i>questo accorgimento permette</i> di ridurre di circa 1/3 il tempo di sviluppo [...]
18	jedoch darf die Gesamtzahl der Grundstimmen 150 nicht übersteigen	a condizione che il totale dei voti di base non sia superiore a 150
19	Bei der Wahrnehmung ihrer Aufgaben dürfen der Exekutivdirektor und das sonstige Personal von keinem Mitglied und von keiner Stelle ausserhalb der Organisation Weisungen einholen oder entgegennehmen	Nell'adempimento delle loro mansioni, il direttore esecutivo ed il personale non sollecitano né accettano istruzioni di alcun membro, né da alcuna autorità esterna all'Organizzazione
20	Der nach Absatz 1 ermittelte Zollwert darf nicht zur Grundlage haben [...]	Il valore in dogana ai sensi del paragrafo 1 non si basa [...]
21	[...] darf der nächste Buchstabe in der in diesem Absatz festgelegten Reihenfolge herangezogen werden	[...] è consentito applicare la lettera immediatamente successiva nell'ordine stabilito dal presente paragrafo
22	[...] müssen die strengeren europäischen Normen auf der Grundlage einer vollständigen Harmonisierung eingeführt werden	[...] occorre applicare norme europee più rigorose basate su un'armonizzazione totale
23	Sie müssen einer Gesundheitskontrolle gemäß Kapitel V des Anhangs unterzogen worden sein . [...] Sie müssen gemäß Kapitel VI des Anhangs vorschriftsmässig verpackt worden sein	sono stati sottoposti ad un controllo sanitario, nell'osservanza del capitolo V dell'allegato; [...] sono stati confezionati nel modo appropriato, nell'osservanza del capitolo VI dell'allegato
24	sie müssen die allgemeine Wettbewerbsfähigkeit oder die Marktaussichten von [...] verbessern können	essere in grado di migliorare la capacità concorrenziale globale oppure le prospettive del mercato di [...]

25	[...] in diesem Fall soll beim Einblasen mit geringstmöglichem Luftdruck und -volumen gearbeitet werden	In questo caso occorre ridurre al minimo la pressione e il volume dell'aria
26	[...] so kann dies nicht bedeuten, dass damit ein [...] unabhängiges Normenwerk geschaffen wird	[...] ciò non deve condurre all'elaborazione di una serie di norme dissociate [...]

Abb. 13: Deutsche und italienische Modalverben im Übersetzungsvergleich

Anhand der quantitativen Ergebnisse und der Beispiele der Tabelle in Abb. 13 lässt sich vermuten, dass der unterschiedliche Verlauf der Kurven in Abb. 12 durch Übersetzungen der Art nach Gruppe II bzw. III zu erklären ist. Wären die Beispiele der Gruppe I zahlreicher gewesen, dann müssten die zwei Kurven in Abb. 12 näher beieinander liegen. Sehen wir uns also die Beispiele der Gruppen II und III genauer an. In beiden Sprachen werden als Alternative zum Modalverb periphrastische Bildungen (Abb. 13: de: 9, 10, 11; it: 17, 18, 24) verwendet. Im Italienischen ist der Rückgriff auf unpersönliche Modalausdrücke (16, 21, 22, 25) häufig zu beobachten. Das Gleiche gilt für die Passivform des Deutschen "sein+zu+Infinitiv" (statt "müssen"), für die es im Italienischen keine wörtliche Korrespondenz gibt (12). Diese Form ist sehr verbreitet und daher weist "dovere" teilweise einen größeren Anteil im Vergleich zu "müssen"/"sollen" auf. All diese Beispiele beziehen sich auf grammatische Unterschiede, die auch in der Allgemeinsprache sehr verbreitet sind.

Die restlichen Beispiele hingegen sind stark von den kulturspezifischen Merkmalen der Textsorten "Verordnung"/"Richtlinie" geprägt. Wie man aus einer genaueren Lektüre der Texte schließen kann, neigt man im Italienischen mehr dazu, den normativen Charakter solcher Textsorten ausdrücklicher zu markieren: Das geschieht vor allem durch die Verwendung der Formen des Verbs "dovere" und dessen Wiederholung (6), durch die fachsprachliche Anwendung des Indikativs des Hauptverbs und durch die damit verbundene Aufhebung des Modals (19, 20, 23), durch die tendenzielle Vermeidung von periphrastischen Bildungen, wenn das Verb "potere" negiert wird (9, 10). Sehr bedeutsam in dieser Hinsicht ist das Beispiel 26: Der deutsche Satz mit "nicht können" wird ins Italienische zu einem Satz mit "non dovere". Das italienische Verb "dovere" taucht überall auf, wo eine Art Verbot zum Ausdruck kommen muss.

In all diesen Fällen scheint das Deutsche durch seine sechs Modalverben breitere Möglichkeiten zu haben, um den normativen Charakter der Texte zu unterstreichen. Für das Verbot bedient sich das Deutsche tatsächlich des Verbs "dürfen".

Aus der Analyse kann man schließlich folgende Rückschlüsse ziehen: In den EU-Rechtstexten zeigen sich fachtextspezifische Merkmale der jeweiligen Sprachen, Kulturen und nationalen juristischen Traditionen. Das Verfassen von Übersetzungen im Kontext der EU ist aber Teil des Prozesses, der zur Herausbildung und Konsolidierung einer internationalen Institution führt. Es entwickelt sich eine neue Rechtssprache, zu deren Genese multilinguale Übersetzungen entscheidend beitragen. Dies könnte zukünftig u.a. dazu führen, dass sich sprachen- und kulturspezifische Merkmale, wie

sie heute im Textvergleich festgestellt werden können, immer stärker ausgleichen, wodurch sich die Auffassung von Fachübersetzung gründlich verändern könnte.

5 Zusammenfassung und Ausblick

Das Ziel der vorliegenden Untersuchung bestand darin, zu zeigen, wie die Translationswissenschaft von der digitalen Korpusanalyse profitieren kann. Dabei wurden zuerst theoretische Fragen angegangen und dann konkrete quantitative und qualitative Analysen durchgeführt. Gegenstand der Studie war das EU-Rechtstextekörpus (*JRC-Acquis*) in den Sprachen Deutsch und Italienisch.

Texte eines aus dem *JRC-Acquis* gewonnenen Subkorpus (DE-IT) wurden zunächst der Analyse auf der Wortebene unterzogen. Dann wurde versucht, textsortenspezifische Merkmale anhand der Übersetzungen von Modalverben zu bestimmen. Dadurch konnte diese Arbeit neues Licht auf die zahlreichen möglichen Entwicklungen der korpusbasierten Translationswissenschaft werfen.

Vieles muss für die Zusammenarbeit von Korpuslinguistik und Translationswissenschaft noch gemacht werden. Die Tatsache, dass das wichtigste heute existierende multilinguale Korpus – das *JRC-Acquis multilingual parallel corpus* der Europäischen Union – für translatorische Zwecke bis jetzt kaum untersucht wurde, ist dafür ein klarer Beweis. Forschungsdesiderat bleibt die Entwicklung eines für die Translationswissenschaft gedachten Annotationssystems, das man auch für das *JRC-Acquis* anwenden könnte.

Literatur

- Baker, Mona (1993): "Corpus Linguistics and Translation Studies. Implications and Applications." Mona Baker (Hg.): *Text and Technology. In Honour of John Sinclair*. Amsterdam etc.: Benjamins, 233-251
- Bowker, Lynne (2002): *Computer-Aided Translation Technology. A Practical Introduction*. Ottawa: University of Ottawa Press
- Even-Zohar, Itamar (1990): *Polysystem Studies*. Tel Aviv: The Porter Institut for Poetics and Semiotics
- Fabricius-Hansen, Cathrine (2004): "Paralleltext und Übersetzung in sprachwissenschaftlicher Sicht." Harald Kittel, Armin Paul Frank, Norbert Greiner, Theo Hermans, Werner Koller, José Lambert mit Juliane House, Brigitte Schultze (Hg.): *Übersetzung / Translation / Traduction*. Teilband 1. (Handbücher zur Sprach- und Kommunikationswissenschaft 26.1.) Berlin/New York: de Gruyter, 322-329
- Gréciano, Gertrud (2003): "Übersetzung und (über)einzel sprachliche Textsortenadäquatheit im Europakorpus." Heidrun Gerzymisch-Arbogast, Eva Hajičová, Petr Sgall, Zuzana Jettmarová, Anneli Rothkegel, Dorothee Rothfuß-Bastian (Hg.): *Textologie und Translation*. (Jahrbuch Übersetzen und Dolmetschen 4/II.) Tübingen: Narr, 85-106 – http://www.translationconcepts.org/pdf/jahrbuch_4.2.pdf (29.12.2008)
- Hansen, Silvia; Elke Teich (1999): "Kontrastive Analyse von Übersetzungskorpora: ein funktionales Modell." Jost E. Gippert (Hg.): *Multilinguale Corpora. Codierung, Strukturierung, Analyse*. Prague: Enigma, 311-322

- Heller, Dorothee (1999): "Das Problem der Modalisierung in fachlichen Texten." Heidrun Gerzymisch-Arbogast, Daniel Gile, Juliane House, Annely Rothkegel mit Silke Buhl (Hg.): *Wege der Übersetzungs- und Dolmetschforschung*. (Jahrbuch Übersetzen und Dolmetschen 1.) Tübingen: Narr, 143-163 – http://www.translationconcepts.org/pdf/jahrbuch_1.pdf (18.11.2008)
- Johansson, Stig (1998): "On the Role of Corpora in the Cross-linguistic Research." Stig Johansson, Signe Oksefjell (Hg.): *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*. Amsterdam etc.: Rodopi, 3-24
- John, Elisabeth (2008), "Mehrsprachiger EU-Rechtstextekorpus im TMX-Format frei verfügbar." *Forum. ATICOM Verbandszeitschrift* 2: 17-23
- Laviosa-Braithwaite, Sara (1998): "Universals of Translation." Mona Baker (Hg.): *Routledge Encyclopedia of Translation Studies*. London: Routledge, 288-291
- Lemnitzer, Lothar; Heike Zinsmeister (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr
- Oakes, Michael; Anthony M. McEnery (2000): "Bilingual Text Alignment – an Overview." Simon P. Botley, Anthony M. McEnery, Andrew Wilson (Hg.): *Multilingual Corpora in Teaching and Research*. Amsterdam etc.: Rodopi, 1-37
- Olohan, Maeve (2004): *Introducing Corpora in Translation Studies*. London etc.: Routledge
- Soffritti, Marcello (2006): "Textlinguistik und Texte: Was bestimmt das Tempo der Entwicklung?" Marina Foschi Albert, Marianne Hepp, Eva Neuland (Hg.): *Texte in Sprachforschung und Sprachunterricht. Pisaner Fachtagung 2004 zu neuen Wegen der italienisch-deutschen Kooperation*. München: Iudicium, 34-47
- Stede, Manfred (2007): *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik*. Tübingen: Narr
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam etc.: Benjamins
- Toury, Gideon (2004): "Probabilistic Explanations in Translation Studies: Welcome as They Are, Would They Qualify as Universals?" Anna Mauranen, Pekka Kujamäki (Hg.): *Translation Universals: Do They Exist?* Amsterdam etc.: Benjamins, 15-32
- Toury, Gideon (1995): *Descriptive Translation Studies and Beyond*. Amsterdam etc.: Benjamins

trans-kom

ISSN 1867-4844

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der trans-kom-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh
Artesis Hogeschool Antwerpen
Vertalers en Tolken
Schilderstraat 41
B-2000 Antwerpen
Belgien
leona.vanvaerenbergh@scarlet.be

Klaus Schubert
Universität Hildesheim
Institut für Übersetzungswissenschaft
und Fachkommunikation
Marienburger Platz 22
D-31141 Hildesheim
Deutschland
klaus.schubert@uni-hildesheim.de

Wiesmann, Eva (2004): *Rechtsübersetzung und Hilfsmittel zur Translation. Wissenschaftliche Grundlagen und computergestützte Umsetzung eines lexikographischen Konzepts*. Tübingen: Narr

Wilss, Wolfram (1999): *Translation and Interpreting in the 20th Century: Focus on German*. Amsterdam etc.: Benjamins

Webseiten

CCVista Translation Database: <http://ccvista.taieex.be/>

Directorate-General for Translation of the European Commission (DGT):
http://ec.europa.eu/dgs/translation/index_en.htm

DGT-TM: <http://langtech.jrc.it/DGT-TM.html>

European Commission Joint Research Center: <http://langtech.jrc.it/>

Portal der EU: <http://europa.eu/>

Rechtsbereiche und Rechtsformen im *JRC-Acquis* Korpus:

<http://europa.eu/eur-lex/de/information/help/help-dir.html#codes>

Workshop zum *JRC-Acquis* Korpus ("Exploiting parallel corpora in up to 20 languages", Arona, September 2005): http://langtech.jrc.it/0509_EU-Enlargement-Workshop.html

Autorin

Giulia A. Disanto studierte Germanistik und Romanistik an der Universität Bari, an der Universität Münster und an der Freien Universität Berlin. 2006 Promotion in Germanistik. 2007 DAAD-Stipendiatin an der FU-Berlin. Seit 2007 Post-Doc-Stipendiatin am Lehrstuhl für Deutsche Sprache und Literatur der Universität Bari. Forschungsgebiete: Translationswissenschaft, Übersetzungstheorie. Freiberufliche Übersetzerin.

E-mail: giuliadisanto@gmail.com