

Katrin Menzel/ Ekaterina Lapshinova-Koltunski (Saarbrücken)

Kontrastive Analyse deutscher und englischer Kohäsionsmittel in verschiedenen Diskurstypen

In diesem Artikel geben wir einen Überblick über unsere Forschungsarbeiten im Zusammenhang mit der Thematik der Diskursrelationen im Sprach- und Registervergleich und dem korpuslinguistischen Forschungsprojekt „German-English contrasts in cohesion – Towards an empirically-based comparison (GECCo)“. Die Arbeiten sollen zu einer Analyse und Erklärung von Textkohäsion im Englischen und Deutschen beitragen. Beispiele für sprachvergleichende korpusbasierte Ansätze zu Fragen der englischen und deutschen Textkohäsion sind bislang selten und behandeln hauptsächlich einzelne Kohäsionsmittel, aber nicht Textualität und Diskursorganisation in einem größeren Zusammenhang. In diesem Artikel wird speziell auf die korpusbasierte Untersuchung von endophorischen Ellipsen als beispielhaft gewählte Fallstudie im Kontext des GECCo-Projektes eingegangen, wobei im Projekt die gesamte Bandbreite von Kohäsionsmitteln abgedeckt wird. Es wird gezeigt, wie sich die Subtypen von Ellipsen in englischen und deutschen Subkorpora der geschriebenen und gesprochenen Sprache verteilen und was die Präferenzen einzelner Register und Textsorten beider Sprachen in Bezug auf die Ellipsenarten sind.

Cohesion in different types of discourse—a contrastive analysis of German and English

This article provides an overview on our current research with regard to language- and register-related discourse relations and gives an outline of the corpus linguistic project “German-English contrasts in cohesion—Towards an empirically-based comparison (GECCo)“. With the design of the GECCo corpus, we intend to provide a resource allowing an intensive contrastive investigation of cohesion in English and German. Contrastive research on English and German with an empirical methodological orientation is rare and has usually only covered individual cohesive devices, but not textuality and discourse organization in a wide-ranging context. This article has a special focus on the corpus-based analysis of endophoric ellipses as a case study in the context of the GECCo project in which all cohesive devices are discussed and described. It shows some differences in the frequency distribution of ellipsis subtypes between English and German subcorpora for written and spoken data and how registers and text types in both languages differ in their preference for particular ellipsis types.

Analiza kontrastywna niemieckich i angielskich środków kohezji w różnych typach dyskursu

Niniejszy artykuł stanowi przegląd prac badawczych podjętych w ramach projektu badawczego z zakresu lingwistyki komputerowej „German-English contrasts in cohesion – Towards an empirically-based comparison (GECCo)“, obejmujących relacje dyskursywne w aspekcie porównania języków i ich rejestrów. Prace te mają stanowić wkład do analizy i wyjaśnienia zjawiska kohezji w tekstach anglo- i niemieckojęzycznych. Założenia opierające się na korpusach porównawczych, dotyczące zjawiska kohezji w tekstach anglo- i niemieckojęzycznych, stanowią obecnie rzadkość i zazwyczaj dotyczą pojedynczych środków językowych, a nie tekstualności i organizacji dyskursu w szerszym kontekście. Przedmiot niniejszego artykułu stanowi głównie oparta na badaniach korpusowych analiza elips endoforycznych będąca przykładem studium przypadków badanych w ramach projektu GECCo, obejmującego pełne spektrum środków kohezyjnych. Ukazane zostanie, w jaki sposób w poszczególnych korpusach języka mówionego i pisanego dystrybuowane są typy elips oraz jakiego rodzaju preferencje odnośnie ich występowania można zaobserwować w pojedynczych rejestrach i gatunkach tekstu w obydwu badanych językach.

1. Ziele, Motivation und Forschungsstand

In diesem Artikel geben wir einen Überblick zu unseren Forschungsarbeiten im Zusammenhang mit der Thematik der Diskursrelationen und dem DFG-geförderten Forschungsprojekt „German-English contrasts in cohesion – Towards an empirically-based comparison (GECCo)¹“, die zu einer vergleichenden Analyse und Erklärung von Textkohäsion im Englischen und Deutschen beitragen sollen. Neben kontrastiven Systemunterschieden in der Kohäsionsstiftung wollen wir auch instantielle Häufigkeiten von Kohäsionsmitteln und ihren nicht-lokalen Beziehungen in Texten ermitteln, die sich nur über eine korpusbasierte Analyse bestimmen lassen. Die Analyseergebnisse sollen auf typologischer Basis erklärt und interpretiert werden. In diesem Artikel wird dabei speziell auf die korpusbasierte Untersuchung von Ellipsen als beispielhaft gewählte Fallstudie im Kontext des Projektes eingegangen.

Im Rahmen systembasierter Arbeiten zu Kohäsionsmitteln liegen unter synchroner Perspektive bereits substantielle Erkenntnisse zum Englischen und zum Deutschen vor, allerdings meist nur mit einzelsprachlicher, nicht-empirischer Ausrichtung. Monolinguale Beschreibungen von Textkohäsion finden sich beispielsweise bei Halliday/Hasan (1976) für Englisch und De Beaugrande/Dressler (1981) für Deutsch. Kontrastive systembasierte Arbeiten sind zumeist auf spezielle lexikalische und grammatische Phänomene begrenzt, und seltener auf Textkohäsion, Textualität, Anaphorizität und Diskursorganisation fokussiert. Eine weitere Forschungslücke besteht im Hinblick auf kohäsive Relationen in der ge-

¹ www.gecco.uni-saarland.de

sprochenen Sprache. Es ist anzunehmen, dass bestimmte Kohäsionsmittel eher charakteristisch für die gesprochene Sprache und ‚konzeptionelle Mündlichkeit‘ (Koch/Oesterreicher 1994) sind, während andere typischerweise schriftsprachliche Texte auszeichnen. Zwar kann teilweise auf einzelne Untersuchungen zum Thema Kohäsion und Diskursstruktur im Zusammenhang mit gesprochener Sprache zurückgegriffen werden, allerdings handelt es sich bisher meist um Abhandlungen zu Einzelphänomenen, häufig zum anaphorischen Gebrauch von Pronomen im gesprochenen Deutsch.

Einige wenige Studien haben sich in der jüngeren Vergangenheit mit empirischen Methoden zur Untersuchung von bestimmten Kohäsionsaspekten beschäftigt. Dabei wird hauptsächlich speziellen Fragen in Bezug auf einzelne Sprachen und individuelle Kohäsionsmittel – oft zudem beschränkt auf gewisse Textsorten oder Register – nachgegangen, z. B. referenzsemantische Funktionen bestimmter Pronomen im Englischen oder im Deutschen (Gundel et al. 2007 für Personal- und Demonstrativpronomen in englischsprachigen Zeitungsartikeln, Bosch et al. 2007 für Demonstrativpronomen im Deutschen).

Beispiele für sprachvergleichende korpusbasierte Ansätze zu Fragen der englischen und deutschen Textkohäsion sind bislang äußerst selten und behandeln ebenfalls hauptsächlich einzelne Kohäsionsmittel, aber nicht Textualität und Diskursorganisation in einem größeren Zusammenhang, s. z. B. Zinsmeister et al. (2012) in Bezug auf abstrakte Anaphern oder Bührig/House (2004) für bestimmte als Kohäsionsmittel gebrauchte Konjunktionen oder Adverbien.

In den Untersuchungen im Zusammenhang mit dem GECCo-Projekt werden Kontraste in den Realisierungsmöglichkeiten für eine große Bandbreite an Kohäsionsmitteln beschrieben, wobei quantitativ und qualitativ untersucht wird, wie Textelemente im Einzelnen syntaktisch und semantisch über die Satzgrenzen hinweg miteinander in Beziehung stehen, s. z.B. Kunz/Steiner (2012, 2013), Kunz/Lapshinova (2014), Kunz/Lapshinova (erscheint). In dem vorliegenden Artikel liegt der Schwerpunkt, wie anfangs erwähnt, auf kohäsiv verwendeten Ellipsen im Englischen und im Deutschen, um die Thematik der Textkohäsion an einer ausgewählten Fallstudie zu verdeutlichen.

2. Einordnung von Ellipsen in den Zusammenhang mit anderen Kohäsionsmitteln

Kohäsion wird erzeugt durch die Gesamtheit der lexiko-grammatischen Mittel, die Sätze semantisch und syntaktisch zu einem kohärenten Text der geschriebenen oder gesprochenen Sprache verknüpfen. Während es sich bei Kohäsion um sprachlich manifestierte Textbezüge (vgl. Linke et al. 2004: 215) handelt, bezeichnet Kohärenz in der Textlinguistik den für den Leser konzep-

tionellen Zusammenhalt eines Textes in Bezug auf seine Tiefenstruktur (vgl. Crystal 2008: 85, Widdowson 1979: 312). Beide Konzepte hängen insofern miteinander zusammen, dass Kohäsionsmittel im Text eine Kohärenzrelation markieren.

Der Begriff der Textkohäsion geht zurück auf Halliday/Hasan (1976), welche das Phänomen in fünf Hauptkategorien unterteilen: Referenz, Substitution, Ellipse, Konnektoren und lexikalische Kohäsion. In unserer Studie wird zur Konzeptualisierung von Kohäsion im Wesentlichen ebenfalls auf diese Hauptklassifizierungsvorschläge von Halliday/Hasan zurückgegriffen. Dabei sowie bei der Bildung von weiteren Subkategorien werden in unserer Analyse die sprachspezifischen Besonderheiten des Englischen und Deutschen berücksichtigt. Mechanismen des Textzusammenhalts beinhalten satz- oder clause-/teilsatzübergreifende endophorische Ausdrücke, mit denen auf bereits vorher durch den Text eingeführte oder noch folgende Textelemente (Antezedens bzw. Sukzedens) Bezug genommen wird. Verschiedene sprachliche Strukturen, die als kohäsionsstiftende Mittel verwendet werden können, treten möglicherweise auch satzintern oder mit exophorischer Verweisfunktion auf. In diesem Falle spielen sie eine untergeordnete Rolle für unsere Betrachtungen zur Textkohäsion.

Des Weiteren ist zu erwähnen, dass unsere Untersuchungen von Textualitätsmerkmalen und diskursorganisatorischen Aspekten auf Erkenntnissen aus Arbeiten zum Thema Textgliederung und Einheitenbildung in der geschriebenen und gesprochenen Sprache aufbauen (u.a. Chafe/Danielwicz (1987) für Englisch oder Stein (2003) für Deutsch). Dabei wird in der Literatur oft hinterfragt, inwiefern der traditionelle Satzbegriff und der Clause/Teilsatz-Begriff in der englisch- und deutschsprachigen Literatur sich gleichermaßen auch für die Analyse gesprochener Sprache eignen. Bei der Segmentierung von Äußerungen der gesprochenen Sprache bestehen Kohäsionsrelationen eher zwischen den als geschlossenen Sinnenheiten formulierten Phrasierungseinheiten.

Tabelle 1 illustriert die Kategorien der Kohäsionsmittel mit Beispielen aus dem GECCo-Korpus (s. auch Kapitel 3).

Kohäsionsmittel	Englisch	Deutsch
(Ko-)Referenz	<p>The Chinese government funds almost all of its plant biotechnology research. It is focusing on commodities that have been mostly ignored in the laboratories of industrialized countries.</p> <p><i>(anaphorisches Pronomen)</i></p>	<p>Die chinesische Regierung finanziert fast ihre gesamte Erforschung der Pflanzenbiotechnologie. Sie konzentriert sich auf Erzeugnisse, die in den Labors der Industrieländer weitestgehend ignoriert wurden.</p>

Kohäsionsmittel	Englisch	Deutsch
Substitution	There can be no question that we have to modernize these structures . We are ready to do so and this is exactly what Agenda 2010 is all about. But we are doing so to build on our own strengths. (<i>verbale/klausale Substitution</i>)	Wir müssen diese Strukturen modernisieren . Dazu sind wir bereit, und genau das haben wir mit der Agenda 2010 begonnen. Aber wir tun das, um unsere eigenen Stärken auszubauen. (<i>verbale Substitution</i>)
Konnektoren	Russia has made significant progress on religious freedom. Therefore , the President is pursuing the removal of Russia from the application of Jackson-Vanik legislation. (<i>kausale Konjunktion</i>)	Russland hat wesentliche Fortschritte bei der Religionsfreiheit gemacht. Daher möchte der Präsident Russland von der Anwendung des Jackson-Vanik-Zusatzartikels ausschließen.
Lexikalische Kohäsion	Scalloway Castle is an impressive fortress in the old 'capital' of the islands. However, this 17th century monument to the old rulers of Shetland is not the most northerly castle in Scotland. This distinction belongs to Muness Castle on the island of Unst. Perhaps the forerunners of these fortified places are the brochs : curious round towers built for defence perhaps around 1500 years ago. (<i>Ketten von Synonymie- und Hyperonymie-Hyponymie-Relationen</i>)	Scalloway Castle ist eine eindrucksvolle Festung der alten „Hauptstadt“ der Inseln. Dieses aus dem 17. Jahrhundert stammende Monument der alten Herren von Shetland ist jedoch nicht das nördlichste Schloss Schottlands. Diese Auszeichnung gebührt Muness Castle auf der Insel Unst. Die Vorläufer dieser befestigten Anlagen waren möglicherweise die „brochs“: eigenartige Rundtürme als Verteidigungsanlagen , errichtet vielleicht vor 1500 Jahren.
kohäsiv verwendete Ellipsen	There was not one way but many [], and he followed first one [] and then the other [], choosing the straightest and stoniest []. (<i>nominale Ellipsen, teilweise im gleichen Clause</i>) We will find out which children can read and which cannot []. (<i>verbale Ellipse des lexikalischen Verbs</i>) Why in the world should an American go halfway around the world to South Korea and get wounded or killed? I'll tell you why []. (<i>klausale Ellipse, ‚shuicing‘</i>)	Nicht ein Weg , sondern eine Vielzahl von Wegen lag vor ihm, und der Jüngling folgte bald dem einen Pfad, bald dem anderen [], bemüht, sich an den geradesten und steinigsten [] zu halten. Wir werden herausfinden, welche Kinder lesen können und welche nicht []. (<i>verbale Ellipse von Modal- und lexikalischem Verb</i>) Warum sollte ein Amerikaner um die halbe Welt nach Südkorea reisen, um verwundet oder getötet zu werden? Ich werde Ihnen sagen, warum [].

Tab. 1: Beispiele der verschiedenen Kohäsionsmittel in GECCo

Bei Referenz wird eine Koreferenzbeziehung, d. h. Referenzidentität, zwischen einer Nominalphrase, einer Anapher und ihrem (außersprachlichen) Denotat durch eine Pro-Form oder ein anderes Referenzmittel hergestellt. Im Falle von Substitution wird ein Platzhalter eingesetzt, und das Phänomen beschränkt sich auf einige spezielle, semantisch ausgebleichte Ausdrücke (z. B. ‚one / do / so‘ für Nominal-, Verbal- und Klausalsubstitution im Englischen). Des Weiteren dienen zur Erzeugung von Textrelationen Konnektoren, wobei logisch-semantische Relationen zwischen Propositionen verschiedener Sätze hergestellt werden. Lexikalische Kohäsion beinhaltet z.B. Rekurrenz durch wörtliche Wiederholungen, Synonymie, Hypo-/Hyperonymie oder Generalisierungen. Dieser Zusammenhang kann auch über größere Distanzen bestehen, und die Elemente können sich als längere kohäsive Ketten durch Textabschnitte oder den gesamten Text ziehen.

Bei endophorischen Ellipsen, die zur Textverknüpfung verwendet werden können, wird der Textverweis durch Leerstellen erzeugt, wobei ein Teil einer Struktur weggelassen wird, der sich aus vorangegangenen Inhalten erschließen lässt. Aussagen zu Ellipsen in unterschiedlichen Sprachen finden sich in verschiedenen linguistischen Theorien und Modellen, besonders in der generativen und der systemisch-funktionalen Linguistik. In Ellipsentypologien werden Subkategorien häufig nach der Struktur des weggelassenen oder übriggebliebenen Fragments gebildet, nach der Art, wie die ausgelassene Information rekonstruiert werden kann (semantisch oder syntaktisch), oder in Abhängigkeit davon, in welchem Kontext die Ellipsentypen vorkommen bzw. welche Funktion sie erfüllen. Dabei mischen sich teilweise formale, funktionale und kontextuelle Faktoren. Quirk et al. (1985: 889) beschreiben Ellipsen im Englischen als graduelles Phänomen, wobei auf prototypische Fälle alle der folgenden Kriterien zutreffen: Es handelt sich um grammatikalisch unvollständige Strukturen, die aus dem textuellen Zusammenhang eindeutig verständlich sind, und durch Einfügen der fehlenden Wörter – die eine exakte Kopie des Antezedens sind – erhält man einen Satz mit unveränderter Bedeutung. Klein (1985) definierte u.a. Textsortenellipsen, Handlungsellipsen, Kontextellipsen, lexikalische Ellipsen, expressive Ellipsen als Ellipsensubkategorien im Deutschen.² Aufgrund sprachspezifischer Besonderheiten und teilweise unterschiedlicher syntaktischer und morphologischer Regeln bringt ein Vergleich von Ellipsen im Deutschen und Englischen gewisse Herausforderungen mit sich. Es ist notwendig, die zahlreichen in der englischen und deutschen Literatur beschriebenen Ellipsenkategorien unter allgemeineren Kategorien zusammenzufassen und auf den Kohäsionsaspekt zu fokussieren, um somit eine sprach- und registerübergreifende Vergleichbarkeit endophorisch verwendeter Ellipsen zu erzielen. Kontrastive Arbeiten zum Gebrauch endophori-

² S. Klein (1985: 3ff) für Details und Beispiele.

scher Ellipsen im englisch-deutschen Vergleich gibt es bisher kaum. Zwar gibt es einige Ansätze zur Entwicklung sprachübergreifender Ellipsentypologien und zur Beschreibung von elliptischen Strukturen, die in unterschiedlichen Sprachen existieren (z. B. Ellipsen im Englischen im Vergleich mit Polnisch bei Shopen/Świeczkowski (1976), im Vergleich mit Niederländisch bei Van Craenenbroeck (2004) oder mit Deutsch (Reich 2011). Häufig werden lediglich ausgewählte Einzelphänomene im Sprachvergleich betrachtet (z. B. McShane (2000) über Verbellipsen im Russischen, Polnischen und Tschechischen oder Merchant/Simpson (2012) über ‚Sluicing‘ im Englischen, Niederländischen, Türkischen, Japanischen und weiteren Sprachen). Meist liegt kein besonderer Schwerpunkt auf der möglichen satzübergreifenden und textverknüpfenden Funktion von Ellipsen, sondern eher auf der grammatische Analyse und Interpretation von elliptischen Strukturen und unvollständigen Äußerungen (außer z. B. bei Duczmal (1985) zu Ellipsen als Kohäsionsmittel im Englischen und Polnischen). Zudem gibt es bisher nur wenige Arbeiten, in denen das Phänomen ‚Ellipse‘ im tatsächlichen Sprachgebrauch und Kontext mit korpusbasierten Methoden untersucht wird. In bisherigen quantitativen Studien wurde zumindest stets mit einer sehr engen Auswahl relativ leicht abfragbarer Fälle gearbeitet, z. B. mit nominalen Ellipsen nach Adjektiven (vgl. Günther 2012).

Die Entscheidung, für die vorliegende Untersuchung über Ellipsen, wie bei den anderen Kohäsionsmitteln auch, die Hauptkategorien des Begriffes wie bei Halliday/Hasan (1976) zu verwenden und sie für den deutsch-englischen Sprachvergleich konzeptionell weiterzuentwickeln, erklärt sich daraus, dass diese sich zur Einteilung von Ellipsentypen in den untersuchten Sprachen gleichermaßen eignen, weil sie relativ breit gefasst sind. Halliday/Hasan unterscheiden zwischen nominalen, verbalen und klausalen Ellipsen. Es fallen Bestandteile einer Nominal- oder Verbalphrase bzw. ganze Konstituenten oder Teilsätze weg, die aus dem textuellen Zusammenhang erschlossen werden können.

Weitere Details zum systemischen englisch-deutschen Vergleich von verschiedenen Kohäsionsmitteln finden sich in vorherigen Publikationen, siehe z. B. Kunz/Steiner (2012) zu Referenz, Kunz/Steiner (2013) zu Substitution, Kunz/Lapshinova (2014) zu Konnektoren. In dieser Arbeit konzentrieren wir uns auf den textuellen Vergleich von Ellipsen im Deutschen und Englischen. Konzeptuelle Grundlagen dazu sind in Menzel (2014a/b) zu finden, und im Rahmen des laufenden Dissertationsprojekts der Autorin dieser Projektberichte soll eine umfassende sprach- und registervergleichende Analyse von Ellipsen erscheinen.

3. Ressourcen, Tools und Analyse

3.1 Ressourcen und Tools

Für die kontrastive Analyse wird eine empirische korpusbasierte Untersuchung gewählt, die einen Vergleich auf der Textebene ermöglicht. Gegenwärtig gibt es erst wenige Korpora, welche sich zu einer derartigen kontrastiven Analyse von Kohäsionsphänomenen eignen. Vorhandene englische und deutsche Korpora sind bisher eher monolingual, auf bestimmte Einzelregister oder bestimmte sprachliche Varietäten beschränkt, oder sie beinhalten teilweise auch Daten von Sprachlernern und Nichtmuttersprachlern. Da das Ziel darin besteht, Unterschiede hinsichtlich Form, Funktion und Häufigkeit von Kohäsionsmitteln und den semantischen Relationen in Bezug auf andere Textelemente zu identifizieren, benötigen wir ein bilinguales Korpus, das verschiedene Textsorten der geschriebenen und gesprochenen Sprache repräsentiert.

Das einzige uns bekannte Korpus, das sich für diese Fragestellung eignet, ist das bilinguale Korpus GECCo (der Name ist vom Projekttitle abgeleitet, s. Abschnitt 1), das insgesamt ca. 1,44 Millionen Tokens enthält und sowohl parallele als auch vergleichbare Subkorpora für Englisch und Deutsch umfasst. Für unsere Untersuchungen zum Sprachvergleich werden einstweilen keine Effekte berücksichtigt, die sich aus Übersetzungsaspekten ergeben, d. h. relevant sind in erster Linie die vergleichbaren Subkorpora EO (englische Originale) und GO (deutsche Originale). GECCo beinhaltet gegenwärtig 12 Register, zu denen sowohl gesprochensprachliche als auch geschriebene Textsorten gehören: ESSAY, FICTION, INSTR, POPSCI, SHARE, SPEECH, TOU, WEB, ACADEMIC, FORUM, INTERVIEW und TALKSHOW (je ca. 30.000 Tokens pro Register im jeweiligen Subkorpus). Die Eigenschaften dieser Register werden in Tabelle 2 präsentiert.

	Register	Textkategorien
geschrieben	ESSAY	politische Aufsätze
	FICTION	Belletristik
	INSTR	Bedienungsanleitungen
	POPSCI	populärwissenschaftliche Texte
	SHARE	Aktionärsbriefe
	SPEECH	vorformulierte politische Reden
	TOU	Tourismusprospekte
	WEB	Webseiten von Firmen und Organisationen

	Register	Textkategorien
gesprochen	ACADEMIC	akademische Vorträge
	FORUM	mündlich konzipierte Dialoge aus Internetforen
	INTERVIEW	Interviews über Beruf und Alltagsleben
	TALKSHOW	Fernsehtalkshows

Tab. 2: Register in GECCo

Die Texte der geschriebenen Register sind aus dem CroCo-Korpus übernommen worden, vgl. Hansen-Schirra et al. (2012). Register gesprochener Sprache wurden im Rahmen des GECCo-Projekts gesammelt und annotiert. Probleme der Korpuskompilierung, Zusammenstellung und Annotation dieser Register wurden in Lapshinova et al. (2012) beschrieben. Beschaffung des gesprochensprachlichen Materials ist kompliziert, weil es viel manuellen Aufwand (Transkription) mit sich bringt. Zu beachten ist dabei, dass das Sprachmaterial in beiden Sprachen gesammelt wird und vergleichbar sein soll. Aus diesem Grund ist der geschriebene Teil von GECCo bisher größer als der gesprochene. Dennoch ermöglicht schon diese Zusammenstellung die Durchführung kontrastiver Analysen. Der gesprochene Teil wird jedoch kontinuierlich erweitert.

Das Korpus ist auf verschiedenen Ebenen annotiert und enthält Informationen über Tokens, Lemmata, morpho-syntaktische Eigenschaften (z. B. Genus, Numerus, Kasus usw.) und Wortarten. Syntaktische Annotationen schließen Phrasen-(chunks) und Satzgrenzen ein. Phrasen sind zusätzlich mit ihren grammatischen Funktionen markiert. Diese Annotationen der geschriebenen Korpusregister wurden aus CroCo importiert. Für das gesprochene englische Subkorpus verwenden wir den Stanford POS Tagger (vgl. Toutanova et al. 2003) und den Stanford Parser (vgl. Klein/Manning 2003). Für das gesprochene Deutsch wurde MATE (vgl. Bohnet 2010) verwendet. Das Gesamtkorpus ist im CWB-Format gespeichert (CWB, 2010), welches Abfragen mit dem Corpus Query Processor (CQP; Evert 2005) ermöglicht. GECCo hat eine webbasierte Version, die über CLARIN-D gewährt wurde und mit einem eingeschränkten Zugang (auf Anfrage) der wissenschaftlichen Öffentlichkeit zur Verfügung steht.³

Auf den vorhandenen Informationen basiert das Verfahren der Annotation von Kohäsionsmitteln. In dem automatischen Annotationsschritt greifen wir auf die Tags mit Informationen über Phrasen- und Satzgrenzen zu. In einigen Fällen, z. B. bei der Annotation von demonstrativer Referenz und konjunktionalen Relationen, wird auch von der Information über ihre Funktionen Gebrauch gemacht. Das semi-automatische Annotationsverfahren für Koreferenz, Konnektoren und

³ <https://fedora.clarin-d.uni-saarland.de/cqpweb/>

Substitution ist in Lapshinova/Kunz (2014) detailliert beschrieben. Darüber hinaus liefern diese Annotationsebenen zusätzliche Angaben über morpho-syntaktische Präferenzen der Kohäsionsmittel selbst: ihre Position im Satz oder Phrase, die Kettenlänge, Funktionen der Antezedenzen, ihre Parallelität mit den entsprechenden Anaphern, usw. Die Annotation von Ellipsen erfolgt zurzeit ausschließlich manuell. Dabei wird das Tool MMAX2 (vgl. Müller/Strube 2006) verwendet, das eine Visualisierung von Annotationskategorien ermöglicht. Das Schema und die Kategorien sind in detaillierten projektinternen Richtlinien dokumentiert (vgl. Menzel 2014c). Auf der Basis der annotierten Ellipsen, ihrer Häufigkeiten, syntaktischen Umgebungen sowie Eigenschaften der Antezedenzen sollen im Rahmen des Projektes bessere automatische Annotationsmöglichkeiten entwickelt werden.

3.2 Korpusextraktionen

Wie oben beschrieben, ist GECCo in CWB-Format encodiert und kann mit CQP (offline oder online über CQP-Web) abgefragt werden. Die Kategorien der Kohäsionsmittel (Tab. 1 aus Abschnitt 2) können mithilfe von CQP-Queries abgefragt werden. Wir extrahieren die Instanzierungen dieser Kategorien aus dem Korpus, sortieren sie nach Sprache (Englisch vs. Deutsch), Sprachmodus (geschrieben vs. gesprochen) und Register (ESSAY vs. FICTION, usw.). Insgesamt stehen 24 Subsets zum Vergleich, z. B. EO-ACADEMIC, GO-ACADEMIC, EO-FICTION, GO-FICTION usw.

In Tabelle 3 sind Beispiele von Queries aufgelistet, die für die Extraktion von Subtypen von Ellipsen verwendet werden.

Query		Erläuterung	Beispiele
1	[_ellipsis_type="nominal"]	nominale Ellipse	<i>Können wir das Mikro bisschen lauter machen? Hab ich das falsche genommen?</i>
2	[_ellipsis_type="verbal"]	verbale Ellipse	<i>Can we show it? Yes we can.</i>
3	[_ellipsis_type="clausal"]	klausale Ellipse	<i>Macht er euch Geschenke? Welche?</i>
4	[_ellipsis_problematic="yes"]	problematische Fälle	<i>This agreement applies to the civil aviation industry. The EU itself has explicitly agreed that it does.</i> (ähnelt verbaler Substitution)

Tab. 3: Beispielqueries für die Extraktion einzelner Kohäsionsmittel

Zum Beispiel extrahieren wir Nominalellipsen mit Query 1, verbale Ellipsen mit Query 2, und klausale Ellipsen mit Query 3. Darüber hinaus können wir Fälle extrahieren, die von Annotatoren als ‚problematisch‘ markiert sind. Meistens sind es mehrdeutige Fälle, die entweder Grenzfälle zur Substitution oder aus anderen Gründen schwer zu klassifizieren sind. Solche Fälle kann man leicht mit Query 4 aus dem Korpus extrahieren und separat untersuchen. Die extrahierten numerischen Ergebnisse werden sortiert und in Tabellen und Datenmatrizen gespeichert. Diese werden für Interpretationen und weitere Evaluationen verwendet, z. B. für multivariate statistische Analysen in R (Venables/Smith 2010).

3.3 Analysemethoden

Unser Forschungsvorhaben umfasst mehrere Fragestellungen, deren Beantwortung unterschiedliche Analysen erfordert. Wir analysieren Kontraste nicht nur zwischen Sprachen, sondern auch zwischen Sprachmodi und Textregistern. Aus diesem Grund werden Antworten auf verschiedene Fragen gesucht: Welche Kohäsionsphänomene kommen in unseren Daten vor und was sind ihre Eigenschaften? Welche kohäsiven Phänomene tragen zur Sprachvariation in unseren Daten bei? Wo ist die größte Variation zu beobachten?

Die letzten zwei Fragen lassen sich am besten mit einer statistischen Analyse beantworten. In unseren vorherigen Analysen verwendeten wir Methoden des unüberwachten und überwachten Lernens. In der Studie zu Referenz in Kunz/Lapshinova (erscheint) wird beispielsweise eine Korrespondenzanalyse (vgl. Nenadic/Greenacre 2007, Baayen 2008 und Venables/Smith 2010) benutzt, um zu sehen, welche Register und Sprachen sich gruppieren und welche Subtypen der Referenz dazu beitragen. In dem Artikel Degaetano-Ortlieb et al. (eingereicht) analysieren wir ein breiteres Spektrum der Kohäsionsmittel einschließlich Ketten solcher Mittel und verwenden dafür Klassifikation mit Support Vector Machines (vgl. Vapnik et al. 1974, Joachims 1998). Dieses Verfahren erlaubt die Bestimmung der bedeutendsten Kohäsionsmittel, die unsere Subkorpora voneinander unterscheiden.

In dem vorliegenden Artikel konzentrieren wir uns auf die deskriptive Analyse, die auf den Distributionen einzelner Phänomene basiert. Wir analysieren hier das Vorkommen von Ellipsen sowie ihrer Subtypen in verschiedenen Registern beider Sprachen. Dabei berücksichtigen wir folgende Fragen: 1) Wie verteilen sich die Subtypen von Ellipsen in englischen und deutschen Subkorpora? 2) Wo kommen mehr Ellipsen vor: in geschriebenen oder gesprochenen Registern? 3) Was sind die Präferenzen einzelner Register beider Sprachen in Bezug auf die Ellipsenarten? Die quantitativen Ergebnisse sowie ihre Interpretationen präsentieren wir im nächsten Abschnitt.

4. Ergebnisse

4.1 Ellipsen im Deutschen und Englischen

Im ersten Schritt vergleichen wir die Distribution von Ellipsen im Gesamtkorpus.⁴ Tabelle 4 demonstriert die Verteilung der drei zu untersuchenden Typen in GECCo.

abs.		nominal		verbal		clausal		TOTAL	
		in %	abs.	in %	abs.	in %	abs.	in %	abs.
EO	SPOKEN	154	40,31	87	22,77	141	36,91	382	100,00
	WRITTEN	385	55,08	148	21,17	166	23,75	699	100,00
	TOTAL	539	49,86	235	21,74	307	28,40	1081	100,00
GO	SPOKEN	192	57,83	24	7,23	116	34,94	332	100,00
	WRITTEN	282	53,82	68	12,98	174	33,21	524	100,00
	TOTAL	474	55,37	92	10,75	290	33,88	856	100,00
TOTAL		1013	52,30	327	16,88	597	30,82	1937	100,00

Tab. 4: Ellipsen in GECCo (in %)

Allgemein gesehen finden wir mehr Vorkommen von Ellipsen im Englischen als im Deutschen. In beiden Sprachen dominieren Nominalellipsen, gefolgt von Klausal- und dann erst Verbalellipsen, wobei die letzten viel häufiger im Englischen als im Deutschen sind. Häufiger als verbale Ellipsen findet man in deutschen Subkorpora nominale und klausale Ellipsen. Bestimmte Sachverhalte im Englischen werden teilweise durch Verbalellipsen ausgedrückt, wo das Deutsche prinzipiell eher die Verwendung einer Verbalphrase vermeidet, durch klausale Ellipsen auf noch reduziertere Strukturen als das Englische oder auf unterschiedliche Weise auf Auslassungen zurückgreift, oder lexikalisches Material wiederaufnimmt. Die englische Verbalphrase ist häufig auch länger durch die Verwendung von ‚do‘ als Operator oder einer Form von ‚be‘ bei Verlaufsformen, die es im Deutschen nicht gibt, und bietet daher mehr Auslassungsmöglichkeiten (s. Tabelle 5).

⁴ Vorerst nur ACADEMIC und INTERVIEW für den gesprochen sprachlichen Teil, da die Daten für die anderen Subkorpora noch ausgewertet werden. Die gesamten Annotationen werden zudem noch in einem weiteren Schritt überprüft. Für diese Analyse wurde das Corpus-Release vom Jahr 2013 verwendet.

Englisch	Deutsch
<i>It was a singular object that somebody treasured almost the way they would [] an oil painting.</i>	<i>Es war ein einzigartiges Objekt, das fast als genauso wertvoll wie ein Ölgemälde betrachtet wurde.</i>
<i>They talked about quitting but they didn't [].</i>	<i>Sie haben davon gesprochen, dass sie kündigen wollen, aber sie haben's nicht gemacht.</i>
<i>And finally one day Greywing begins to cry, even more bitterly than she has [] so often.</i>	<i>Bis eines Tages Greywing in Tränen ausbricht und noch bitterlicher weint als so oft schon.</i>
<i>The quacks tell the patient that they're expecting a daughter, even when they aren't [].</i>	<i>Dann erzählen die Quacksalber den Patientinnen, dass sie eine Tochter erwarten, obwohl das gar nicht der Fall ist.</i>
<i>Has the Islamicization of the Orient reached, or passed, its zenith? Abbas thinks it has [].</i>	<i>Hat die Islamisierung des Orients ihren Zenit erreicht bzw. überschritten? Abbas glaubt ja.</i>
<i>You're blushing! - I am not [].</i>	<i>Du wirst ja noch rot! - [] Gar nicht.</i>

Tab. 5: Korpusbeispiele englischer Verbalellipsen, denen im Deutschen andere Konstruktionen entsprechen

Wenn wir die Unterschiede zwischen gesprochener und geschriebener Sprache in EO und GO in Betracht ziehen, sehen wir, dass die Verteilung von Ellipsentypen unterschiedlich ist. Zwar dominieren Nominalellipsen in allen Subkorpora, und verbale Ellipsen kommen am seltensten vor, dennoch sind die verbalen Ellipsen am wenigsten in beiden deutschen Subkorpora zu finden. Hier ist auch der Unterschied zwischen gesprochenem und geschriebenem Deutsch zu betonen: im gesprochenen Teil von GECCo kommen weniger verbale Ellipsen vor als im geschriebenen. Für die englischen Verbalellipsen ist diese Tendenz nicht festzustellen. Der Unterschied liegt eher in der Distribution nominaler und kausaler Ellipsen, s. Abbildung 1.

Interessant ist auch das Verhältnis von Nominalellipsen in deutschen und englischen Subkorpora. Im EO dominiert dieser Typ in geschriebenen Registern, wobei in GO eine höhere Anzahl solcher Ellipsen im gesprochenen Teil zu beobachten ist. Die untersuchten deutschen akademischen Texte wiesen eine besonders hohe Anzahl von nominalen Ellipsen nach Adjektiven und Numeralien auf (Bsp. 1).

1 a) *Soweit zu den neuen Dingen, die wir vor uns haben. Kommen wir zu den alten [].*

1 b) *Ich entwickle jetzt die perfekte Lösung für zwanzig Punkte. Achtzehn [] kriegt man sehr gut.*

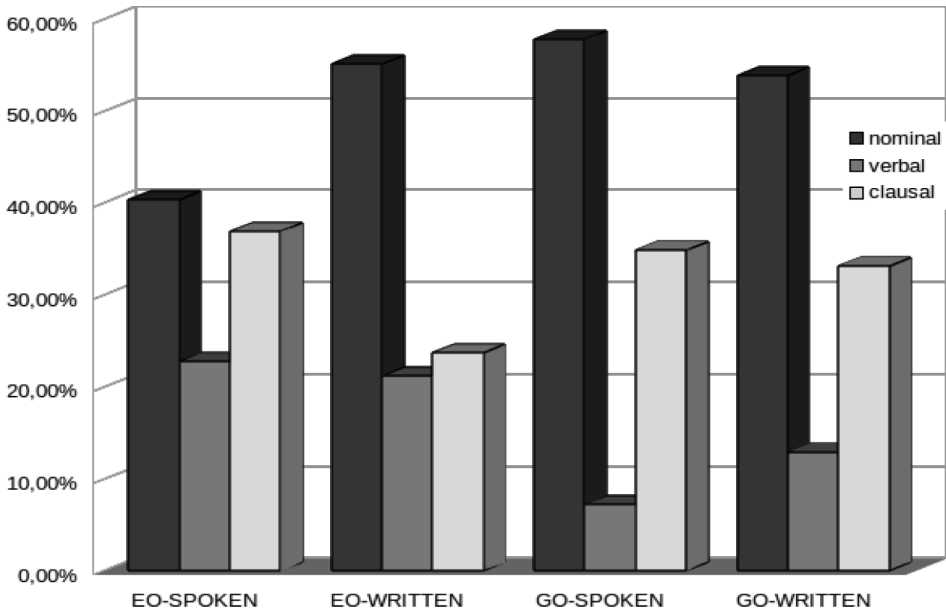


Abb. 1: Ellipsen in gesprochenen und geschriebenen Registern in GECCo

Nominale Ellipsen sind prinzipiell kennzeichnend für Fachkommunikation mit nominalem Stil und zahlreichen Möglichkeiten, Kopfnomen in komplexen Nominalphrasen auszulassen.

4.2 Ellipsen in spezifischen deutschen und englischen Registern

Im nächsten Schritt vergleichen wir die Distribution der drei Ellipsen in verschiedenen Textregistern beider Sprachen. Diese ist in Abbildung 2 dargestellt.

Gut zu sehen ist hier die ungleiche Verteilung über die verschiedenen Register im Englischen und Deutschen. In beiden Sprachen ist zwar zu bemerken, dass ACADEMIC, FICTION und INTERVIEW die meisten Fälle aufweisen⁵, aber es lassen sich größere Unterschiede zwischen den deutschen Registern als zwischen den englischen beobachten. Das bestätigt unsere Beobachtungen aus anderen Studien zu anderen Kohäsionsmitteln, s. z. B. Kunz/Lapshinova (2014, erscheint). Der Grund dafür könnte die Tendenz des Englischen zur Verwendung von gesprochensprachlichen Elementen auch im geschriebenen Kontext sein (z. B. Studien

⁵ Im Deutschen haben ACADEMIC und FICTION die höchste Anzahl, im Englischen INTERVIEW und FICTION.

von Mair 2004 und 2006), allerdings haben beispielsweise die deutschen äußerst schriftsprachlich konzipierten Register ESSAY und POPSCI auch eine vergleichsweise hohe Anzahl von nominalen und teilweise auch von klausalen Ellipsen.

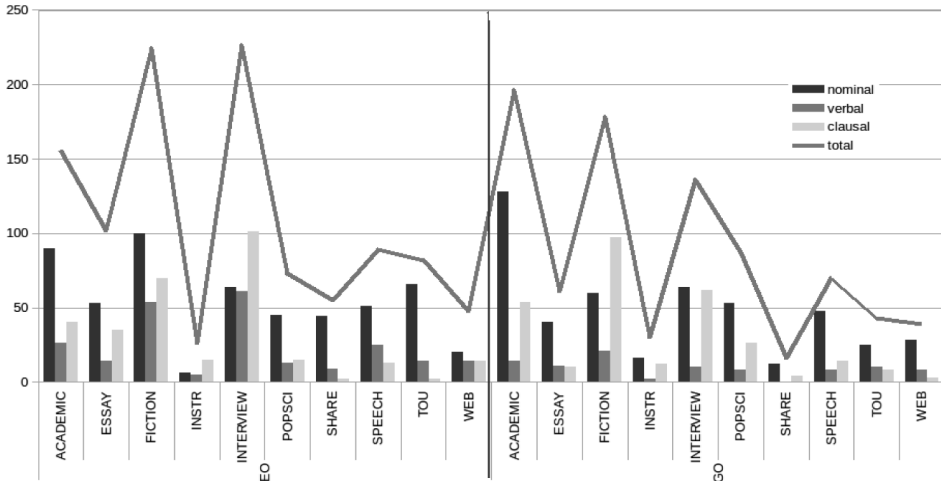


Abb. 2: Ellipsen in deutschen und englischen Textregistern in GECCo

McCarthy (1998: 76) und Schwitalla (2012: 101) hatten vermutet, dass in geschriebener Sprache im Englischen wie im Deutschen insgesamt weniger Ellipsen vorkommen als in gesprochener Sprache, und bezogen sich dabei auf die Gesamtheit aller Ellipsenarten. Man muss dabei aber noch genauer differenzieren, inwiefern bestimmte Ellipsensubtypen eher ein Merkmal der geschriebenen oder der gesprochenen Sprache bzw. bestimmter Kommunikationsszenarien (symmetrische vs. asymmetrische / monologische vs. dialogische Kommunikation etc.) sind. Dass klausale Ellipsen z. B. kennzeichnend für dialogische Diskurstypen sind, zeigt sich u.a. daran, dass diese Ellipsen häufiger in Interviewdialogen vertreten waren als in den eher monologischen Texten aus akademischen Veranstaltungen. In beiden Sprachen sind weitere Unterschiede zwischen Registern in ihren Präferenzen für bestimmte Art der Ellipse zu beobachten. In englischen Interviews und Gebrauchsanweisungen sowie deutschen fiktionalen Texten findet man mehr Klausal- als Nominalellipsen. Diese Ellipsenart ist auch in deutschen Interviews sehr häufig und nur geringfügig weniger vertreten als Nominalellipsen. So werden gewisse Ellipsentypen teilweise bewusst als rhetorisches Stilmittel in formeller Sprache und in geschriebenen Texten verwendet, während andere eher mit gesprochener Sprache oder schlechtem Stil assoziiert werden. Milroy/Milroy (1999: 54) stellen beispielsweise fest, dass mit Vagheit und Ambiguität verbundene Ellipsen in geschriebenen englischen Texten weniger toleriert werden als in

gesprochener Sprache. Deutsch tendiert scheinbar noch stärker dazu, explizitere Strukturen in der Schriftsprache zu verwenden. Dies kann ein Grund dafür sein, warum insgesamt weniger Ellipsen in geschriebenen Registern gefunden werden.

4.3 Mehrdeutigkeiten und Problemfälle

Manche Ellipsenfälle mögen auf den ersten Blick nicht leicht zu identifizieren oder zu kategorisieren sein, weil wir es manchmal mit Grenzfällen zwischen Ellipse und Substitution oder mit problematischen Fällen an der Grenze zwischen Satzspaltung oder syntaktischer Umstellung und elliptischer Struktur zu tun haben bzw. zwischen kohäsiver und nicht-kohäsiv gebrauchter Ellipse. Die Annotationen in unserem Korpus ermöglichen eine automatische Extraktion solcher Fälle, die dann getrennt analysiert werden können bzw. über deren endgültige Kategorisierung noch entschieden werden kann.⁶ In Abbildung 3 präsentieren wir die Distribution von Fällen, die bei der Annotation im Englischen und Deutschen als problematisch markiert wurden.

Der englische Teil von GECCo scheint dabei mehr problematische Ellipsenfälle zu enthalten. Dies kann an der höheren Ambiguität von häufig auftretenden Strukturen in dieser Sprache liegen. In beiden Sprachen ist man mehr mit Problemen bei den Nominaellipsen konfrontiert, die auch am häufigsten in fast allen Subkorpora vorkommen, s. 4.1 und 4.2. Im Deutschen verschaffen aber auch klausale Ellipsen den Annotatoren Probleme. Die Anzahl solcher Fälle ist allerdings etwas geringer als die der Nominaellipsen. Bei den typischen Auslösern von nominalen Ellipsen haben einige Formen ein Flexionsmorphem, das keine Ergänzung durch Nomen zulässt (Deutsch: *eine, keine, meine* [] / *eins, keins, meins* [/]; Englisch: *the other* [] / *the others* [/], *his* [] / *mine* [/]). Bei einigen Fällen nominaler Ellipsen war fraglich, ob sie tatsächlich kohäsionsstiftend sind oder sich eher durch syntaktische Prinzipien erklären lassen. Inkonsistente Schreibweisen bei Nominalisierungen oder gar fälschliche Nominalisierung von Adjektiven oder Zahlworten im Korpus fallen ebenfalls unter die Fälle von nominalen Ellipsen, die auf den ersten Blick nicht eindeutig sind (z. B. „*Nicht weniger beeindruckend sind die weit über tausend Linien, die die Hochebene überziehen. Die Längste erstreckt sich über ca. 9 Kilometer.*“ Die korrekte Schreibweise wäre hier: „*Die längste* []...“, und es handelt sich um Adjektiv + nominale Ellipse). Manchmal bleibt es ambig oder ist nicht sofort ersichtlich, ob es sich um eine exophorische Ellipse oder um einen Verweis auf ein spezifisches Textelement handelt (z. B. „...*irakische Professoren, die*

⁶ Darauf aufbauend kann auch das Annotationsschema weiter verfeinert werden.

ihre Posten nicht dazu missbraucht haben, andere [Personen/Iraker/Professoren?] einzuschüchtern.“ Einige Fälle liegen an der Grenze zwischen einer willkürlicher Satzspaltung (z. B. durch kreative Satzzeichenverwendung) und kohäsiver Ellipse, da meist theoretisch noch beliebig Material hinzugefügt werden kann – auch Material, was die vorherige Struktur wiederaufnimmt (z. B. „*Sie werden geprägt. In der Familie, der Schule, im Wohnviertel.*“). In mehreren Fällen handelt es sich aber nur durch die vom üblichen Gebrauch abweichende Satzzeichenverwendung zum Anzeigen einer Sprecherpause, wodurch ein Fragment entsteht.

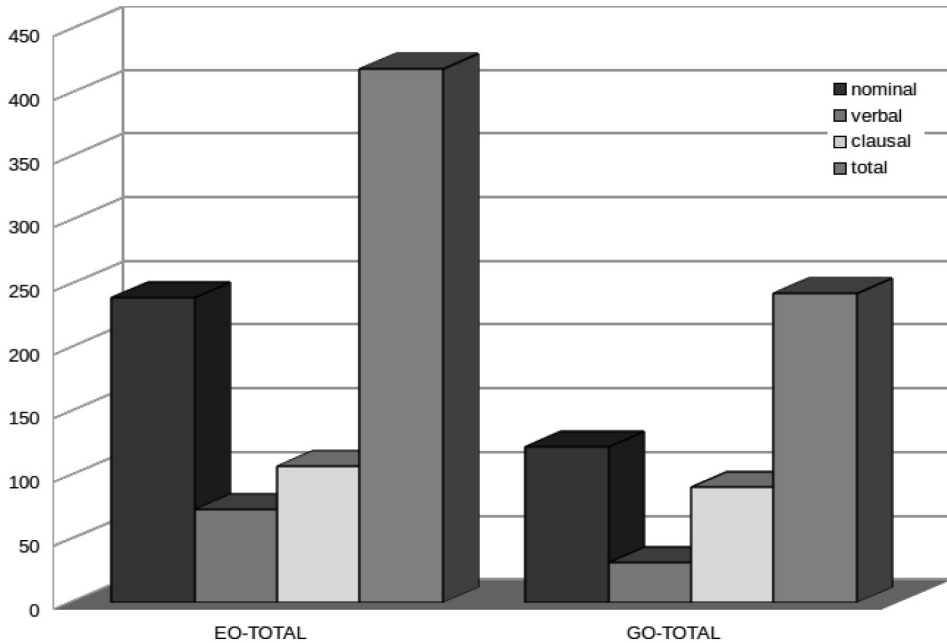


Abb. 3: Problematische Fälle bei der Annotation von Ellipsen

5. Zusammenfassung und Ausblick

In diesem Artikel konnte anhand einer Fallstudie im Zusammenhang mit der Thematik Kohäsionsstiftung im kontrastiven Vergleich gezeigt werden, wie sich die Subtypen von Ellipsen in englischen und deutschen Subkorpora der geschriebenen und gesprochenen Sprache verteilen und was die Präferenzen einzelner Register beider Sprachen in Bezug auf die Ellipsenarten sind. Insgesamt gibt es vermutlich weniger systemische Unterschiede im Hinblick auf Ellipsen zwischen Englisch und Deutsch als zwischen den Registern und Textsorten

innerhalb der Sprachen, da beide – wenn man die Kategorien relativ breit fasst – ähnliche elliptische Strukturen ermöglichen, die je nach Register und Ellipsensubtyp vergleichbare Funktionen übernehmen. Diese dienen zur Raffung, Kennzeichnung von konzeptioneller Mündlichkeit, zur Textverknüpfung oder als syntaktische Stilfigur. Es bleibt abzuwarten, welchen Einfluss weitere neue gesprochen sprachliche Korpusregister auf die weiteren Ergebnisse haben werden und auch welche Unterschiede im Ellipsengebrauch es zwischen Originaltexten und Übersetzungen im Englischen und Deutschen gibt. Dafür benötigen wir eine statistische Auswertung der Daten, z. B. eine Signifikanzanalyse, die zeigen soll, dass die Unterschiede keinen zufälligen Charakter haben. Diese werden wir in den nachfolgenden Studien durchführen. Vermutlich lässt sich eine Tendenz im Übersetzerverhalten feststellen, expliziter als im Original zu formulieren und weniger Ellipsen als im Original zu verwenden. Übersetzer fügen in solchen Fällen tendenziell etwas hinzu, greifen vorher Genanntes wieder auf und fügen auch anderswo im Text kaum neue Ellipsen ein, da solche Strukturen stilistisch markiert sind.

Literatur

- Baayen, Harald (2008): *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge.
- Bohnet, Bernd (2010): Very High Accuracy and Fast Dependency Parsing is not a Contradiction. The 23rd International Conference on Computational Linguistics (COLING 2010). Beijing, China.
- CWB (2010): The IMS Open Corpus Workbench. <http://www.cwb.sourceforge.net> (27.07.2014)
- Bosch, Peter/ Katz, Graham/ Umbach, Carla (2007): The Non-Subject Bias of German Demonstrative Pronouns. In: Monika Schwarz-Friesel/ Manfred Consten/ Mareile Knees (Hrsg.): *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference.*, Amsterdam/Philadelphia, S. 145–164.
- Bührrig, Kristin/ House, Juliane (2004): Connectivity in translation: Transitions from orality to literacy. In: House, Juliane/ Rehbein, Jochen (Hrsg.) *Multilingual Communication*. Amsterdam, S. 87–114.
- Chafe, Wallace L./ Danielwicz, Jane (1987): Properties of spoken and written language. In: Horowitz, Rosalind/ Samuels, S. Jay (Hrsg.): *Comprehending oral and written language*. New York, 83–113.
- Van Craenenbroeck, Jeroen (2004): *Ellipsis in Dutch Dialects*. Utrecht.
- Crystal, David (2008): *A Dictionary of Linguistics and Phonetics*. 6. Ausgabe. Oxford.
- De Beaugrande, Robert-Alain/ Dressler, Wolfgang (1981): *Einführung in die Textlinguistik*. Tübingen.
- Degaetano-Ortlieb, Stefania/ Kunz, Kerstin/ Lapshinova-Koltunski, Ekaterina/ Menzel, Katrin/ Steiner, Erich (eingereicht): GECCo - an empirically-based comparison of English-German cohesion. In: De Sutter, Gert/ Delaere, Isabelle and Lefer, Marie-Aude (Hrsg.): *New Ways of Analysing Translational Behaviour in Corpus-Based Translation Studies*. TILSM series. Berlin/New York.

- Duczmal, Stanislaw (1985): Ellipsis as a cohesive device in English and Polish: a contrastive analysis. In: Poznan Studies in Contemporary Linguistics, Vol. 20, S. 115–128.
- Evert, Stefan (2005): The CQP Query Language Tutorial. Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, April. CWB version 2.2.b90.
- Gundel, Jeanette/ Hedberg, Nancy/ Zacharski, Ron (2007): Directly and Indirectly Anaphoric Demonstrative and Personal Pronouns in Newspaper Articles. Proceedings of the Sixth Annual Discourse Anaphora and Anaphora Resolution Colloquium.
- Günther, Christine (2012): The Elliptical Noun Phrase in English: Structure and Use. New York.
- Halliday, Michael A.K./ Hasan, Ruqaiya (1976): Cohesion in English. London/New York.
- Hansen-Schirra, Silvia/ Neumann, Stella/ Steiner, Erich (2012): Cross-linguistic Corpora for the Study of Translations. Insights from the language pair English – German. Series Text, Translation, Computational Processing. Berlin/New York.
- Joachims, Thorsten (1998): Text categorization with support vector machines: Learning with many relevant features. Machine Learning: ECML98, S. 137–142.
- Klein, Wolfgang (1985): Ellipse, Fokusgliederung und thematischer Stand. In: Meyer-Herrmann, Reinhard/ Hannes Rieser (Hrsg.): Ellipsen und fragmentarische Ausdrücke Bd.1.31. Tübingen, S. 1–24.
- Klein, Dan/ Manning, Christopher (2003): Accurate Unlexicalized Parsing. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics – Volume 1, ACL '03, Stroudsburg, PA, USA. Association for Computational Linguistics, S. 423–430.
- Koch, Peter/ Oesterreicher, Wulf. (1994): Schriftlichkeit und Sprache. In: Günther, Hartmut/ Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung. Berlin/New York, S. 587–604.
- Kunz, Kerstin/ Steiner, Erich (2012): Towards a comparison of cohesive reference in English and German: System and text. In: Taboada, Maite/ Doval Suárez, Susana/ González Álvarez, Elsa (Hrsg.): Contrastive Discourse Analysis. Functional and Corpus Perspectives. London.
- Kunz, Kerstin/ Steiner, Erich (2013): Cohesive substitution in English and German: a contrastive and corpus-based perspective. In: Aijmer, Karin/ Altenberg, Bengt (Hrsg.): Advances in corpus-based contrastive linguistics. Studies in honour of Stig Johansson. Amsterdam, S. 201–231.
- Kunz, Kerstin/ Lapshinova-Koltunski, Ekaterina (2014): Cohesive conjunctions in English and German: Systemic contrasts and textual differences. In: Lieven, Vandelanotte/ Davidse, Kristin/ Gentens, Caroline/ Kimps, Ditte (Hrsg.): Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora. Amsterdam/New York, S. 229–262.
- Kunz, Kerstin/ Lapshinova-Koltunski, Ekaterina (erscheint): Cross-linguistic analysis of discourse variation across registers. In: Aijmer, Karin/ Hassegard, Hilde (Hrsg.): Special Issue of Nordic Journal of English Studies.
- Lapshinova-Koltunski, Ekaterina/ Kunz, Kerstin/ Amoia, Marilisa (2012): Compiling a Multilingual Corpus. In: Mello, Heliana/ Pettorino, Massimo/ Raso, Tommaso (Hrsg.): Proceedings of the VIIth GSCP-2012 International Conference: Speech and Corpora. Firenze. S. 29–34.
- Lapshinova-Koltunski, Ekaterina/ Kunz, Kerstin (2014): Annotating Cohesion for Multilingual Analysis. In: Proceedings of the 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, Reykjavik, May 26, 2014.
- Linke, Angelika/ Nussbaumer, Markus/ Portmann, Paul R. (2004): Studienbuch Linguistik. Tübingen.
- Mair, Christian (2004). Corpus linguistics and grammaticalisation theory: Statistics, frequencies, and beyond. In: Lindquist, Hans/ Mair, Christian (Hrsg.): Corpus Approaches to Grammaticalization in English. Amsterdam/Philadelphia, S. 121–150.
- Mair, Christian (2006): Twentieth-Century English. History, variation and standardization. Cambridge.

- McCarthy, Michael (1998): *Spoken Language and Applied Linguistics*. Cambridge.
- McShane, Marjorie J. (2000): Verbal ellipsis in Russian, Polish and Czech. *Slavic and East European Journal* 44, S. 195–233.
- Menzel, Katrin (2014a): Ellipsen als Stil- und Kohäsionsmittel in deutschen und englischen politischen Reden. In: Leuschner, Torsten/ Koliopoulou, Maria (Hrsg.): *Germanistische Mitteilungen*. Zeitschrift für Deutsche Sprache, Literatur und Kultur, 40.1: Deutsch kontrastiv. Brüssel. S. 31–50.
- Menzel, Katrin (2014b): Types and Functions of Ellipses in an English-German Corpus of Written and Spoken Language, DFG-Projektbericht.
- Menzel, Katrin (2014c): Guidelines for the manual annotation of ellipses, DFG-Projektbericht.
- Merchant, Jason/ Andrew Simpson (Hrsg.) (2012): *Sluicing: Cross-linguistic perspectives*. Oxford.
- Milroy, James/Milroy, Lesley (1999): *Authority in Language. Investigating Standard English*. London/New York.
- Müller, Christoph/ Strube, Michael (2006): Multi-Level Annotation of Linguistic Data with MMAX2. In: Braun/ Sabine, Kohn, Kurt/ Mukherjee, Joybrato (Hrsg.): *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods (English Corpus Linguistics, Vol.3)*. Frankfurt am Main, S. 197–214.
- Nenadic, Oleg/ Greenacre, Michael (2007): Correspondence analysis in R, with two-and three-dimensional graphics: The ca package. *Journal of Statistical Software* 20(3): S. 1–13.
- Quirk, Randolph/ Greenbaum, Sidney/ Svartvik, Jan/ Leech, Geoffrey (1985): *A Comprehensive Grammar of the English Language*. London.
- Reich, Ingo (2011): Ellipsis. In: Maienborn, Claudia/ von Heusinger, Klaus/ Portner, Paul (Hrsg.): *Semantics: an International Handbook of Natural Language Meaning*. Berlin/New York, S. 1849–1874.
- Schwitalla, Johannes (2012): *Gesprochenes Deutsch. Eine Einführung*. Berlin.
- Shopen, Tim/ Świczkowski, Walerian (1981): Some remarks on ellipsis in Polish and English. In: *Poznań Studies in Contemporary Linguistics*, Vol. 20, S. 111–126.
- Stein, Stephan (2003): *Textgliederung. Einheitenbildung im geschriebenen und gesprochenen Deutsch: Theorie und Empirie*. Berlin/New York.
- Toutanova, Kristina/ Klein, Dan/ Manning, Christopher D./ Singer, Yoram (2003): Feature-rich part-of-speech tagging with a cyclic dependency network. In: *NAACL'03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (Vol. 1)*, S. 173–180.
- Venables, William N./ Smith, David M. (2010): *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*. Electronic edition.
- Vapnik, Vladimir N./ Chervonenkis, Alexey J. (1974): *Theory of Pattern Recognition*. Moskau.
- Widdowson, Henry (1979): *Explorations in applied linguistics*. Oxford: OUP.
- Zinsmeister, Heike/ Dipper, Stefanie/ Seiss, Melanie (2012): Abstract pronominal anaphors and label nouns in German and English: Selected case studies and quantitative investigations. In: *TC3. Translation: Computation, Corpora, Cognition*. (2) 1, S. 47–80.

Katrin Menzel / Dr. Ekaterina Lapshinova-Koltunski
Universität des Saarlandes
Institut für Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen
D- 66123 Saarbrücken
E-Mail: e.lapshinova@mx.uni-saarland.de
E-Mail: k.menzel@mx.uni-saarland.de