

# Kapitel 3

## Korpustranslatologie

Oliver Czulo<sup>a</sup> & Ekaterina Lapshinova-Koltunski<sup>b</sup>

<sup>a</sup>Institut für Translatologie gGmbH <sup>b</sup>Universität Hildesheim

Dieses Kapitel stellt die theoretischen Ansätze des Funktionalismus, der Descriptive Translation Studies und der Universalientheorie als Motivation und als theoretische Grundlage dar, sich mit Translationsdaten in Form von Korpora auseinanderzusetzen. Es wird vor allem auf die produktorientierte Perspektive in der translatologischen Korpusforschung eingegangen. Ein methodisches Kapitel führt Grundbegriffe der Korpusanalyse ein und stellt beispielhaft Top-down- und Bottom-up-Zugänge in der Korpusarbeit dar. Prozessorientierte Perspektiven – mit Blick auf kognitionsorientierte oder informationstheoretische Forschung – werden ebenfalls aufgegriffen.

### 1 Einführung

Die im späteren 20. Jahrhundert aufgekommenen funktionalen Translationstheorien (u. a. Nord 1989, Reiss & Vermeer 1984, Holz-Mänttari 1984) lösten die Dichotomie der „freien“ vs. der „nahen“ Übersetzung auf und schlugen stattdessen ausgehend vom Übersetzungsauftrag – der den Zweck und das Zielpublikum definiert – ein Spektrum an Strategien vor, die sich zwischen den beiden genannten Polen bewegen können. Die Beschaffenheit des *Translats*, d.h. des Produkts eines Translationsprozesses, hängt somit nicht allein vom Ausgangstext ab, sondern ist von verschiedenen Faktoren und den damit zusammenhängenden strategischen Erwägungen abhängig. Mit dieser Sichtweise rückt die Frage ins Blickfeld, welche Faktoren die Beschaffenheit des *Translats* wie beeinflussen, wobei inzwischen neben dem Übersetzungsauftrag zahlreiche andere situative Variablen wie z. B. das Kompetenzniveau der Translatoren, das Trägermedium des *Translats* etc. von Interesse sind.



Es sind vor allem die Ausführungen von Toury (insbes. 1995) und Baker (insbes. 1996) sowie die zunehmende Verfügbarkeit von Sprachdaten, die in den 1990er Jahren die Entstehung einer Korpustranslatologie anstoßen. Anfangs konzentrierten die Untersuchungen sich auf die Frage, welche Faktoren translations-spezifisch, oder genauer gesagt übersetzungsspezifisch sind, also was Übersetzungen von Nicht-Übersetzungen systematisch unterscheidet. Steht zu Beginn das (geschriebene) Produkt im Vordergrund, werden nach und nach Methoden – korpusbasierte wie andere – entwickelt, um auch Auswirkungen verschiedener Variablen auf den Übersetzungsprozess zu untersuchen.

Vergleichbare Untersuchungen setzen fürs Dolmetschen später als fürs Übersetzen ein. Zum einen liegt dies darin mitbegründet, dass gesprochensprachliche Daten nicht so leicht verfügbar sind wie geschriebensprachliche Daten: Erstere müssen erst transkribiert werden.<sup>1</sup> Zum anderen kann man vermuten, dass ein hinderlicher Umstand war, dass das Dolmetschen nicht in gleicher Weise in den Wissenschaftsinstitutionen verankert ist wie das Übersetzen: Dolmetschprofessuren etwa sind deutlich seltener zu finden. In jüngerer Vergangenheit weitet sich der Blick über das Geschriebene und Gesprochene hinaus auch auf multimediale Aspekte der Translation. Neben den Forschungsinteressen spielen didaktische und praktische Aspekte, d.h. der Einsatz von Korpora als Wissensquelle in der Lehre und in der Translationspraxis, früh eine Rolle in der Korpustranslatologie.

## 2 Theoretische Grundlagen der Korpustranslatologie

Die Korpustranslatologie bedient sich in vielerlei Hinsicht der Grundlagen und Erkenntnisse der Korpuslinguistik<sup>2</sup> und ist stark empirisch geprägt. Sie passt somit zur Kategorisierung der *Translation Studies* von Holmes (2000) als empirische Wissenschaft.

Biber u. a. (1998) zufolge besteht der Korpusansatz darin, Gebrauchsmuster in natürlich vorkommenden Texten zu analysieren. Grundlage ist dafür eine Sammlung von Texten, die man als *das Korpus* bezeichnet. Die Analyse bedient sich der Unterstützung von Computern und bringt Beobachtungen

- auf der *qualitativen* Ebene, also anhand von Einzelbeispielen, die meist tiefgehend analysiert werden, oft mit dem Ziel, daraus verallgemeinerbare Kategorien abzuleiten, und

---

<sup>1</sup>Jüngste Fortschritte in der automatischen Transkription können hier Abhilfe schaffen.

<sup>2</sup>Eine eingehende theoretische sowie praktische Einführung in die Korpuslinguistik findet sich in (Lemmitzer & Zinsmeister 2015).

- auf der *quantitativen* Ebene, also anhand einer Vielzahl von Einzelbeobachtungen, die anhand bestehender Kategorisierung eingeordnet, ausgezählt und interpretiert werden,

hervor. Dabei schließen diese beiden Ebenen einander nicht aus: Qualitativ hergeleitete Kategorien lassen sich quantitativ auf ihre Anwendbarkeit und Aussagekraft überprüfen; ebenso natürlich Kategorien, die quantitativ, z. B. über statistische Rechenverfahren, gewonnen wurden. Untersucht man quantitative Befunde, können Fundstellen auffallen, die nicht oder nicht so richtig in eine bestehende Kategorisierung passen; diese muss dann auf Grundlage qualitativer Analysen überarbeitet werden. Legt man sich nicht von vorneherein ausschließlich auf eine der beiden Herangehensweisen fest, wird sich so ein wechselseitiges, einander ergänzendes Verfahren ergeben.

Viele korpusbasierte Studien im Bereich der Translatologie folgen der Annahme, dass übersetzte Texte bestimmte sprachliche Gemeinsamkeiten aufweisen, die nicht oder nur in geringerem Maße in nicht übersetzten Originaltexten vorkommen. Es gibt verschiedene Namen für diese übersetzungsspezifische Phänomene. Mona Baker spricht von *Translationsuniversalia* und meint dabei, dass diese vom Einfluss der spezifischen Sprachpaare unabhängig sind (Baker 1993: 243). Toury (1995) postuliert zwei „Gesetze“ der Übersetzung, wobei er diese ausdrücklich als probabilistische Tendenzen ansieht. Das *law of growing standardization* sagt aus: „in translation, textual relations obtaining in the original are often modified [...] in favour of (more) habitual options offered by a target repertoire“ (Toury 1995: 268): Übersetzungen orientieren sich demzufolge eher als Originaltexte an etablierten Normen in Grammatik und Lexik der Zielsprache. Dem Gegenüber steht das *law of interference* (Toury 1995: 275), demzufolge Eigenschaften des Ausgangstexts tendenziell in den Zieltext übertragen werden. Diese beiden scheinbar widersprüchlichen „Gesetze“ bestimmen in einer Wechselwirkung die sprachlichen Eigenschaften des übersetzten Texts mit. Baker (1996) führt auf Grundlage von Vorläuferstudien vier Universalien auf (s. u.), die in der Folge vielfach erforscht, teils angezweifelt und später um weitere Universalien erweitert wurden. Chesterman (2004) klassifiziert die Universalien in S- und T-Universalien, d.h. Übersetzungsuniversalien, die entweder von der Ausgangs- oder Zielsprache abhängig sind. Während die S-Universalien Wechselwirkungen zwischen Übersetzungen und ihren Ausgangstexten erklären sollen, beziehen sich die T-Universalien auf Unterschiede zwischen Übersetzungen und Originalen in der Zielsprache. Die ursprünglich von Baker aufgeführten Universalien sieht Pym (2008) alle als Varianten von Tourys *law of growing standardization* an, sie würden somit nach Chesterman unter die T-Universalien fallen.

Eine weitere Bezeichnung dieser Phänomene ist *Translationese*, ein Begriff, der von Gellerstam (1986) geprägt wurde. Translationese ließe sich ins Deutsche mit „Übersetzungssprech“ oder „übersetzungstypischer Sprache“ übersetzen: Dies drückt pointiert aus, dass Übersetzungen ihren ganz eigenen Sprachstil haben, was sie, so betonen auch einige Forscher, zu einem Gegenstand echten Forschungsinteresses jenseits einer reinen „Defizitdiskussion“ – also in einfachen Worten gesagt der oft gestellten Frage, was in der Übersetzung „verloren geht“ – macht. Zu den am häufigsten erforschten Phänomenen gehören:

- *Shining-through* (Teich 2003) bzw. *Durchscheinen*, also die Tendenz, die Häufigkeit von Ausgangstextmerkmalen zu reproduzieren, anstatt den Konventionen der Zielsprache zu folgen;
- *(Über-)Normalisierung* (Baker 1993, 1996), d. h. die Tendenz, sich an die typischen Merkmale der Zielsprache anzupassen und diese sogar übertrieben häufig zu verwenden;
- *Explizierung* (Baker 1996, Olohan & Baker 2000), die Tendenz, die Inhalte deutlicher auszudrücken, anstatt sie implizit zu belassen;
- *Simplifizierung* (Toury 1995), die Tendenz, die bei der Übersetzung verwendete Sprache zu vereinfachen und eindeutiger zu machen.

Nicht alle Phänomene sind als rein sprachsystembezogen zu verstehen: Phänomene wie z. B. eine Tendenz zur Simplifizierung können anderweitige Ursache haben, etwa unterschiedliche Textsortenkonventionen oder kognitive Entlastungseffekte beim Übersetzen.

Diese Übersetzungsphänomene werden im Translationsprodukt anhand von Beobachtungen über die Verteilung (*Distribution*) spezifischer sprachlicher Merkmale untersucht. Dafür werden verschiedene statistische Methoden und Techniken eingesetzt, die weiter unten in den Abschnitten 3.2 und 3.3 beschrieben werden. Anhand dieser Distributionen lassen sich übersetzte Texte von original verfassten Texten sogar automatisch unterscheiden (Volansky u. a. 2015, Baroni & Bernardini 2006).

Die Phänomene wurden in den letzten Jahren in verschiedenen Kontexten untersucht, wie z. B. in verschiedenen Textsorten (Evert & Neumann 2017, Lapshinova-Koltunski 2017), für verschiedene Modalitäten wie geschrieben und gesprochen (Lapshinova-Koltunski u. a. 2021, He u. a. 2016, Kajzer-Wietrzny 2012) für verschiedene Produzierenden wie z. B. professionelle Übersetzende, Novizen oder Laien (Bizzoni & Lapshinova-Koltunski 2021, Kunilovskaya &

Lapshinova-Koltunski 2020, Popović 2020, De Sutter u. a. 2017, Rubino u. a. 2016) und verschiedene Übersetzungsmethoden wie automatisch, händisch oder kombiniert (Popović u. a. 2023, Konovalova & Toral 2022, van der Werff u. a. 2022, Vanmassenhove u. a. 2021, Čulo u. a. 2017).

## 3 Grundbegriffe und Methoden der Korpusanalyse

### 3.1 Grundbegriffe

Ein Korpus wird von Lemnitzer & Zinsmeister (2015: 39) als „Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen“ definiert. Die Grenze zwischen einem Korpus und einer Datensammlung ziehen Lemnitzer und Zinsmeister anhand des Kriteriums, dass Korpora ganze Texte oder große Textausschnitte umfassen; allerdings betonen sie, dass diese Grenze nicht absolut definierbar sei. Sprechen wir heute über ein Korpus, meinen wir oft eine digitalisierte Sammlung. Bei der Zusammenstellung eines Korpus sind verschiedenste Kriterien denkbar, wobei einige Kriterien recht grundlegend für die wahrscheinlich meisten Korpora sind, insbesondere:

- *Repräsentativität*: Ein Korpus soll ausreichend Material enthalten, um davon verallgemeinerbare Aussagen ableiten zu können. Diese Größe ist allerdings nicht fest bestimmbar, d. h. man kann keine absoluten Zahlen angeben, ab wann ein Korpus repräsentativ ist, zumal verschiedene Phänomene unterschiedlich häufig auftreten. Hier kommen oft Erfahrungswerte zum Einsatz, oder man kann ein Korpus sukzessive erweitern und dabei überprüfen, ob Beobachtungen, die man zuvor gemacht hat, auch auf das größere Korpus zutreffen.
- *Balanciertheit* (also *Ausgewogenheit*): Ein Korpus soll Texte aus ausreichend verschiedenen Quellen umfassen, um verallgemeinerbare Aussagen machen zu können. In einem Korpus beispielsweise, das Texte von frühen Lernern einer Sprache enthält, um deren typische Fehler zu untersuchen, sollten nicht nur Texte einer oder sehr weniger Personen enthalten sein, da dies keine Verallgemeinerung auf weitere Personen erlauben würde.<sup>3</sup>

Korpora können nach Lemnitzer & Zinsmeister (2015) drei verschiedene Datentypen enthalten.

---

<sup>3</sup>Dieses Kriterium gilt natürlich dann, wenn es dem Ziel der Untersuchung entspricht. Anders würde es sich verhalten, würde man beispielsweise den Schreibstil nur einer Person in einem Genre untersuchen wollen.

- *Primärdaten*: Dies ist das Material, anhand dessen bzw. über das wir Erkenntnisse gewinnen wollen, z. B. ein übersetzter Text, ein transkribiertes Gespräch oder eine Videoaufnahme.
- *Annotationen*: In unserem Fall sind dies linguistische Beschreibungen von Einheiten in den Primärdaten, z. B. das Labeln von Wortarten oder die Markierung von Sprechpausen.
- *Metadaten*: Diese beschreiben kontextuelle Aspekte der Entstehung der Primärdaten, also z. B. das Alter der Textproduzenten oder die Entstehungszeit von Texten.

Es gibt eine ganze Reihe von Korpusarten, etwa Fachsprachenkorpora, Korpora von Sprachenlernern oder multimediale Korpora. Oft lässt sich ein Korpus mehreren Typen zuordnen, z. B. ist ein Korpus mit Videoaufnahmen, in denen Grundbegriffe der Physik erklärt werden, ein multimediales Fachsprachenkorpus. In einem solchen Korpus könnte neben den zu Hilfe genommen Bildern auch interessant sein, mit welchen Gesten bestimmte Aussagen gemeinsam vorkommen und ob sich zwischen Gesten und Gesagtem ein Zusammenhang erkennen lässt. So könnte man z. B. hypothesieren, dass ein „entweder [...] oder [...]“ durch markante Handbewegungen begleitet wird, die relativ zu zwei verschiedene Positionen vor dem Körper der Sprecher – einmal links, einmal rechts der Körpermitte – erfolgen.

Für die Translatologie sind zwei Arten von Korpora von besonderem Interesse: das *Parallelkorpus* und das *Vergleichskorpus*. Der Klassifikation von Granger (2003)<sup>4</sup> folgend gelten folgende Definitionen:

- *Parallelkorpus*: Eine Sammlung von Originaltexten und deren Übersetzungen, bei denen jedem Originaltext mindestens eine Übersetzung zugewiesen ist.
- *Vergleichskorpus*: Das mehrsprachige Vergleichskorpus ist eine Sammlung von Originalen und/oder Übersetzungen in mehreren Sprachen aus demselben Genre. Das einsprachige Vergleichskorpus enthält aus derselben

---

<sup>4</sup>Die Begriffe *Parallelkorpus* und *Vergleichskorpus* wurden und werden teils noch heute uneinheitlich verwendet, so gibt es Autoren, die die Benennungen genau umgekehrt verwenden wie hier dargestellt. Die Granger'schen Definitionen haben sich inzwischen aber erfahrungsgemäß weitestgehend etabliert.

Sprache Originale und Übersetzungen oder Erstsprachlertexte<sup>5</sup> und Lernertexte. Im ersteren Fall sind die Übersetzungen nicht notwendigerweise Übersetzungen der Originaltexte.

Um Übersetzungsbeziehungen zwischen sprachlichen Einheiten oder die Effekte der Ausgangssprache (nach Chesterman S-Universalien) zu untersuchen, ist ein Parallelkorpus erforderlich. Zum Beispiel kann ein beobachteter ungewöhnlich hoher Anteil an Nomen in übersetzten Texte ein Effekt der Ausgangssprache sein, die einen höheren Anteil an Nomen verwendet als in der Zielsprache üblich ist (also ein Durchscheinen). Andererseits kann der erhöhte Anteil an Nomen auch durch Überanpassungen an Konventionen der Zielsprache verursacht sein (also eine Normalisierung, nach Chesterman in der Gruppe der T-Universalien), was wiederum mit Hilfe monolingualer vergleichbarer Korpora zu überprüfen wäre. Darüber hinaus könnte es sich auch um ein „drittes Phänomen“ handeln, also eines, das nicht von Ausgangs- oder Zielsprache abhängig ist. Dies könnte beispielsweise ein kognitiver Effekt des Übersetzungsprozesses sein, dass etwa Übersetzende tendenziell auf Nomen zurückgriffen, weil es ihnen kognitiv einfacher fiel, die Formulierungen so zu verfassen. Um letztere Annahme genauer zu beleuchten, genügen rein produktbasierte Korpora nicht, hierfür sind Prozesskorpora nötig, die Aspekte des Übersetzungsprozesses enthalten.

Die Korpustranslatologie geht in der Regel deskriptiv vor und beschreibt Gebrauchsmuster in Texten. Dies kann entweder, wie eben dargelegt, hypothesengeleitet (*top-down*) erfolgen oder auch datengeleitet (*bottom-up*) ablaufen. Die beiden Herangehensweisen können kombiniert werden.

### 3.2 Top-down-Analyse

Bei einer Top-down-Analyse werden ein existierendes theoretisches Modell oder Ergebnisse und Beobachtungen aus vorhergehenden Studien als Ausgangspunkt genutzt. Darauf aufbauend werden Hypothesen gebildet, die mittels Tests verworfen oder bestätigt werden können. Um sie zu testen, müssen Hypothesen *operationalisiert* werden: Es müssen beobachtbare Phänomene benannt werden, deren Erscheinungsbild als *abhängige Variable* von einer oder mehrerer *unabhängigen Variablen* beeinflusst wird. Üblicherweise formuliert man eine so genannte *Nullhypothese*, die keinen Zusammenhang zwischen unabhängigen und abhängigen Variablen sieht. Die *Alternativhypothese* geht im Gegensatz dazu von einem

<sup>5</sup>„Erstsprachler“ oder wissenschaftlich oft „L1-Sprecher“ werden umgangssprachlich als „Muttersprachler“ bezeichnet, wobei ja nicht ausgemacht ist, dass dies die Sprache einer Mutter eines Textproduzenten ist. Hier meint Erstsprache die im gegebenen Kontext stärkste Sprache von Textproduzenten, d. h. nicht notwendigerweise die Sprache, die zuerst gelernt wurde.

Zusammenhang zwischen den unabhängigen Variablen und den abhängigen Variablen aus. Weiterhin müssen die sprachlichen Merkmale definiert werden, die für ein Phänomen stehen und in Korpora gesucht werden können. Die beobachtete(n) Ausprägung(en) der sprachlichen Merkmale werden (ggf. statistisch) ausgewertet und in Bezug auf Hypothesen oder Fragestellungen interpretiert.

Zum Beispiel wird in Arbeiten im Bereich der kontrastiven Pragmatik (House 1997) ausgesagt, dass die deutsche Sprache inhaltsorientierter (unpersönlicher im Stil) als die englische Sprache ist, welche adressatenorientierter (durch persönlichere Ansprache) funktioniert. Auf diesen Beobachtungen aufbauend könnte man eine Hypothese aufstellen, dass Übersetzungen aus dem Deutschen ins Englische aufgrund von Einflüssen der Ausgangssprache inhaltsorientierter als vergleichbare Originale im Englischen sind. Hier wären die unabhängigen Variablen also die sprachlichen Konventionen der Ausgangs- und der Zielsprache sowie die Übersetzungsrichtung; die Häufigkeit der Produktion von inhalts- vs. adressatenorientierten Merkmalen wären davon abhängig. Die Nullhypothese würde besagen, dass kein Unterschied in der Häufigkeit unpersönlicher vs. persönlicher Ausdrücke zwischen übersetzten Texten und Originaltexten im Englischen bestünde; die Alternativhypothese würde unsere eben formulierte Annahme enthalten, dass die aus dem Deutschen übersetzten englischen Texte unpersönlicher gestaltet wären.

Um dies korpusbasiert zu untersuchen zu können, muss zunächst festgelegt werden, welche sprachlichen Ausdrücke im Deutschen und Englischen jeweils als inhalts- bzw. adressatenorientiert gelten. Die Inhaltsorientierung spiegelt sich z. B. in der Verwendung von Passivkonstruktionen, *zu*-Infinitiven und Reflexivverben wider. Also werden im Korpus die Häufigkeiten von Passiv- und Reflexivverben und *zu*-Infinitiven ermittelt. Passivkonstruktionen im Englischen lassen sich beispielsweise als Abfolge von einem Hilfs- oder Modalverb in finiter Form (*is, are, was, were, has, have, can* etc.), gefolgt von einem optionalen Adverb (*always, recently, etc.*), gefolgt von einem optionalen Verb *to be* im Gerundium oder als Partizip II (*been, being*) und gefolgt von Partizip II eines beliebigen Verbes definieren: *is always being done, has been asked, were written, could be clarified*.

Anhand der ermittelten Häufigkeiten dieser sprachlichen Merkmale kann untersucht werden, ob englische Übersetzungen aus dem Deutschen tatsächlich eine signifikant höhere Anzahl inhaltsorientierter Merkmale haben als die vergleichbaren Originale im Englischen. Trifft dies zu, wäre die aufgestellte Hypothese aber noch nicht bestätigt: Es muss zusätzlich noch im deutsch-englischen Parallelkorpus anhand der deutschen Originale überprüft werden, ob die unpersönlichen Ausdrücke tatsächlich aus dem Deutschen übersetzt wurden. Ist dies nicht der Fall, würde man nicht von einem Durchscheinen ausgehen, sondern



könnte u. a. annehmen, dass eine zielsprachliche Konvention beim Übersetzen ungewöhnlich oft verwendet wurde, was eher dem Phänomen der Normalisierung zuzuordnen wäre.

Anhand dieses Beispiels wird deutlich, wie ein theoriegetriebener bzw. hypothesengeleiteter Ansatz in der Korpustranslatologie eingesetzt wird: Bereits existierende Vorarbeiten und Theorien werden für die Hypothesenbildung genutzt, außerdem soweit angegeben auch als Informationsquelle für die zu untersuchenden sprachlichen Merkmale.

### 3.3 Bottom-up-Analyse

Bottom-up-Ansätze werden auch als *datengesteuert* oder *datengeleitet* bezeichnet. Bei diesen Ansätzen werden zunächst Beobachtungen festgehalten, die mit einem notwendigen Minimum an Konzepten auskommen. Für die Interpretation der Beobachtungen werden theoretische Ansätze verwendet oder gebildet.

Bottom-up-Ansätze arbeiten mit so genannten *shallow features*, d. h. oberflächlichen sprachlichen Merkmalen, die automatisch leicht zu ermitteln sind und keine besonders theorieabhängige Fundierung haben, wie etwa Wortarten.<sup>6</sup> Schon eine einfache Statistik über die Verteilung von Wortarten kann informativ sein: In einem Bottom-up-Ansatz würde man z. B. für ein Vergleichskorpus zunächst die Häufigkeitsverteilung aller Wortarten in übersetzten und Originaltexten beschreiben, um dann Abweichungen in den Verteilungen genauer unter die Lupe zu nehmen. Anders als bei einem Top-down-Ansatz würde man nicht von vorneherein bestimmte Verteilungen vermuten. Daneben wird auch mit so genannten *N-Grammen* gearbeitet, wobei *Gramme* eine kontinuierliche Abfolge von Elementen meint und *N* für eine beliebige Zahl steht. *N-Gramme* können auf der Grundlage verschiedener Größen berechnet werden, z. B. Wortformen, Wortarten oder *Tokens* (d. h. alle sprachlichen Oberflächenelemente, also Wörter, Satzzeichen, Emojis, ...). Ein Satz wie

*Ich bin ein Satz.*

würde folgende Tri-Gramme, also Abfolgen mit jeweils drei Elementen ergeben:

- Wortformen: *Ich bin ein // bin ein Satz*
- Wortarten: Pronomen–Vollverb–Artikel // Vollverb–Artikel–Nomen

---

<sup>6</sup>Streng genommen ist auch die Definition von Wortarten durchaus theorieabhängig, aber es haben sich gerade im Bereich der Korpusanalyse einige Konventionen durchgesetzt, die als quasi „theorieneutral“ gelten und aufgrund ihres häufigen Einsatzes Ergebnisse vergleichbar machen.

- Tokens: *Ich bin ein // bin ein Satz // ein Satz* .

Solche N-Gramme wurden u. a. verwendet, um Unterschiede zwischen maschinell übersetzten und menschlich übersetzten Texten in der Verteilung von Wortarten zu ermitteln (z. B. Toral u. a. 2019). Die Ergebnisse deuten darauf hin, dass diese unterschiedlich produzierten Texte tatsächlich verschiedene sprachliche Profile haben.

Im Gegensatz zum Top-down-Verfahren kommt bei einem Bottom-up-Verfahren also die Beobachtung vor der theoretischen Einordnung. In gewisser Weise kann man eine Analogie zu jenen qualitativen Verfahren ziehen, bei denen Kategorien und Hypothesen erst aus der Beobachtung (von Einzelbelegen) heraus gewonnen und nicht vorab angenommen werden.

## 4 Ausblick

Dieser sehr komprimierte Einblick in die Grundlagen der Korpustranslatologie wird der Breite des Feldes eigentlich kaum gerecht. Nicht nur werden inzwischen deutlich elaboriertere statistische Modelle verwendet, um Produktdaten Einsichten in Übersetzungsphänomene zu entlocken, sondern es werden hierfür zunehmend Prozessdaten – also z. B. Blickbewegungsdaten oder Tastaturanschlagsprotokolle – ausgewertet. Nur angerissen wurde außerdem die Relevanz kognitiver Faktoren, deren Untersuchung Einblicke in die Verarbeitungsprozesse bei der Übersetzung geben, oder multimedialer Faktoren, die durch neue digitale Möglichkeiten an Wichtigkeit zunehmen; und mit diesen Faktoren hören die Möglichkeiten der Beschreibung translatorischer Kontexte nicht auf. Die korpusbasierte Erforschung von Dolmetschdaten konnte ebenfalls nicht ausgeführt werden. Neben Untersuchungen auf all diesen Feldern bietet die korpustranslatologische Literatur zahlreiche korpuspraktische und -didaktische Überlegungen, die ebenfalls gewinnbringend sind.

## Literatur

Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (Hrsg.), *Text and technology: In honour of John Sinclair*, 233–250. Amsterdam & Philadelphia: John Benjamins.

- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (Hrsg.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager*, 175–186. Amsterdam: John Benjamins.
- Baroni, Marco & Silvia Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21(3). 259–274. DOI: 10.1093/llc/fqi039.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bizzoni, Yuri & Ekaterina Lapshinova-Koltunski. 2021. Measuring translationese across levels of expertise: Are professionals more surprising than students? In *Proceedings of the 23rd nordic conference on computational linguistics (NoDaLiDa)*, 53–63. Reykjavik, Iceland: Linköping University Electronic Press, Sweden. <https://aclanthology.org/2021.nodalida-main.6>.
- Chesterman, Andrew. 2004. Beyond the particular. In Anna Mauranen & Pekka Kujamäki (Hrsg.), *Translation Universals: Do they exist?*, 33–49. Amsterdam: John Benjamins.
- Čulo, Oliver, Silvia Hansen-Schirra & Jean Nitzke. 2017. Contrasting terminological variation in post-editing and human translation of texts from the technical and medical domain. In Gert De Sutter, Isabelle Delaere & Marie-Aude Lefer (Hrsg.), *Empirical translation studies: New theoretical and methodological traditions* (Trends in Linguistics Studies and Monographs 300). Berlin; Boston: De Gruyter Mouton.
- De Sutter, Gert, Bert Cappelle, Orphée De Clercq, Rudy Looock & Koen Plevoets. 2017. Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations. *Linguistica Antverpiensia, New Series—Themes in Translation Studies* 16.
- Evert, Stefan & Stella Neumann. 2017. The impact of translation direction on characteristics of translated texts: A multivariate analysis for english and german. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (Hrsg.), *Empirical translation studies: New theoretical and methodological traditions*, 47–80. Berlin; Boston: De Gruyter Mouton.
- Gellerstam, Martin. 1986. Translationese in Swedish novels translated from English. In Lars Wollin & Hans Lindquist (Hrsg.), *Translation Studies in Scandinavia*, 88–95. Lund: CWK Gleerup.
- Granger, Sylviane. 2003. The corpus approach: A common way forward for Contrastive Linguistics and Translation Studies? In Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (Hrsg.), *Corpus-based approaches to Contrastive Linguistics and Translation Studies*, 17–30. Amsterdam: Rodopi.

- He, He, Jordan Boyd-Graber & Hal Daumé III. 2016. Interpretese vs. translationese: The uniqueness of human strategies in simultaneous interpretation. In Kevin Knight, Ani Nenkova & Owen Rambow (Hrsg.), *Proceedings of the 2016 conference of the north American chapter of the association for computational linguistics: Human language technologies*, 971–976. San Diego, California: Association for Computational Linguistics. DOI: 10.18653/v1/N16-1111. <https://aclanthology.org/N16-1111>.
- Holmes, James S. 2000. The name and nature of Translation Studies. In Lawrence Venuti (Hrsg.), *The Translation Studies reader*, 172–185. London: Routledge.
- Holz-Mänttääri, Justa. 1984. *Translatorisches Handeln: Theorie und Methode*. Helsinki: Suomalainen tiedeakatemia.
- House, Juliane. 1997. *Translation Quality Assessment: A model revisited* (Tübinger Beiträge zur Linguistik 410). Tübingen: Gunter Narr.
- Kajzer-Wietrzny, Marta. 2012. *Interpreting universals and interpreting style*. Adam Mickiewicz University, Poznan. (Dissertation).
- Konovalova, Aleksandra & Antonio Toral. 2022. Man vs. machine: Extracting character networks from human and machine translations. In *Proceedings of LateCHCLFL*, 75–82. Gyeongju, Republic of Korea: COLING. <https://aclanthology.org/2022.latechclfl-1.10>.
- Kunilovskaya, Maria & Ekaterina Lapshinova-Koltunski. 2020. Lexicogrammatic translationese across two targets and competence levels. In *Proceedings of LREC-2020*, 4102–4112. Marseille, France.
- Lapshinova-Koltunski, Ekaterina. 2017. Exploratory analysis of dimensions influencing variation in translation: The case of text register and translation method. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (Hrsg.), *Empirical translation studies: New methodological and theoretical traditions* (Trends in Linguistics Studies and Monographs 300), 207–234. Berlin; Boston: De Gruyter Mouton.
- Lapshinova-Koltunski, Ekaterina, Yuri Bizzoni, Heike Przybyl & Elke Teich. 2021. Found in translation/interpreting: Combining data-driven and supervised methods to analyse cross-linguistically mediated communication. In Yuri Bizzoni, Elke Teich, Cristina España-Bonet & Josef van Genabith (Hrsg.), *Proceedings for the first workshop on modelling translation: Translatology in the digital age*, 82–90. Association for Computational Linguistics. <https://aclanthology.org/2021.motra-1.9>.
- Lemmitzer, Lothar & Heike Zinsmeister. 2015. *Korpuslinguistik: eine Einführung*. 3., überarbeitete und erweiterte Auflage (Narr Studienbücher). Tübingen: Narr Francke Attempto.

- Nord, Christiane. 1989. Loyalität statt Treue. Vorschläge zu einer funktionalen Übersetzungstypologie. *Lebende Sprachen* 3. 100–105.
- Olohan, Maeve & Mona Baker. 2000. Reporting ‘that’ in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1(2). 141–158. DOI: 10.1556/acr.1.2000.2.1.
- Popović, Maja. 2020. On the differences between human translations. In *Proceedings of the EAMT-2020*, 365–374. Lissabon: European Association for Machine Translation.
- Popović, Maja, Ekaterina Lapshinova-Koltunski & Maarit Koponen. 2023. Computational analysis of different translations: By professionals, students and machines. In Mary Nurminen, Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov, Frederike Schierl, Tharindu Ranasinghe, Eva Vanmassenhove, Sergi Alvarez Vidal, Nora Aranberri, Mara Nunziatini, Carla Parra Escartín, Mikel Forcada, Maja Popovic, Carolina Scarton & Helena Moniz (Hrsg.), *Proceedings of the 24th annual conference of the european association for machine translation*, 365–374. Tampere, Finland: European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.36>.
- Pym, Anthony. 2008. On Toury’s laws of how translators translate. In Anthony Pym, Miriam Shlesinger & Daniel Simeoni (Hrsg.), *Descriptive Translation Studies and beyond*, 311–328. Amsterdam: Benjamins.
- Reiss, Katharina & Hans J. Vermeer. 1984. *Grundlegung einer allgemeinen Translationstheorie* (Linguistische Arbeiten 147). Tübingen: M. Niemeyer.
- Rubino, Raphael, Ekaterina Lapshinova-Koltunski & Josef van Genabith. 2016. Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of NAACL-HLT-2016*, 960–970. San Diego, California.
- Teich, Elke. 2003. *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin & New York: Mouton de Gruyter.
- Toral, Antonio, Mikel Forcada & Andy Way. 2019. Post-editese: An exacerbated translationese. In *Proceedings of MT Summit XVII*, 273–281. Dublin. <http://arxiv.org/abs/1907.00900>.
- Toury, Gideon. 1995. *Descriptive Translation Studies – and beyond* (Benjamin Translation Library 4). Amsterdam: John Benjamins.
- van der Werff, Tobias, Rik van Noord & Antonio Toral. 2022. Automatic discrimination of human and neural machine translation: A study with multiple pre-trained models and longer context. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, 161–170. Ghent,

- Belgium: European Association for Machine Translation. <https://aclanthology.org/2022.eamt-1.19>.
- Vanmassenhove, Eva, Dimitar Shterionov & Matthew Gwilliam. 2021. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In *Proceedings of EACL-2021*, 2203–2213. ACL. DOI: 10.18653/v1/2021.eacl-main.188.
- Volansky, Vered, Noam Ordan & Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30(1). 98–118. DOI: 10.1093/llc/fqt031.