

...und Englisch ist doch komplexer!

Eine empirische Analyse von Nebensätzen in deutschen und englischen Fachtexten der Sprachwissenschaft

1. Einleitung¹

Deutsche Fachtexte werden im Vergleich mit englischen Fachtexten gemeinhin als schwerer verständlich empfunden. So stellt beispielsweise Siepmann (2010, 580) unter Bezug auf wissenschaftliche Texte mit einigem Recht die Frage, „[w]er hat nicht schon die Erfahrung gemacht, dass englische Texte häufig leichter zu verstehen sind als deutsche?“ Die Ursachen für diesen vermeintlichen oder wirklichen Unterschied können vielfältiger Natur sein und erstrecken sich auf lexikalische (hier vor allem: Wortbildung), syntaktische, informationsstrukturelle und textorganisatorische Phänomene.

Allerdings herrscht in der Literatur nicht immer Einigkeit darüber, was genau denn die Unterschiede sind, die die Texte aus den beiden Sprachen kennzeichnen. In Bezug auf syntaktische Strukturen gibt Pinchuck (1977, 214) folgende Einschätzung zu technischen oder wissenschaftlichen Übersetzungen: „German is more addicted to complicated sentence structure than English and in many cases an English reader would not tolerate the load carried in German“.

In diesem Artikel möchte ich mich der Komplexität von Sätzen widmen, wobei hier unter dem Begriff ‚Komplexität‘ nur die Unterordnung satzwertiger Elemente (‚clauses‘, im folgenden ‚Nebensätze‘) in größere Satzgefüge (‚sentences‘, im folgenden ‚Sätze‘) verstanden wird, und nicht etwa ganz allgemein die Anreicherung von Satzgefügen durch alle möglichen Arten von Ergänzungen und Attributen.

¹ Die hier vorgestellte Studie hat ihren Ursprung in einem studentischen Projekt, das im Rahmen meines Hauptseminars ‚Complex Clauses‘ im SS 2009 von Tabea Goerke, Juliane Willius und Anne Fritzen durchgeführt wurde. Mein besonderer Dank gilt diesen Studentinnen für die Datenerhebung und Pilotanalysen. Weitere wertvolle Hilfe verdanke ich Jennifer Schluer und Linda Hilkenbach, die an der Literaturrecherche und –auswertung sowie an der Datenkodierung maßgeblich beteiligt waren. Ich bedanke mich auch bei Sabine Arndt-Lappe, Maria Braun und Sabine Jautz für hilfreiche Kommentare zu einer früheren Version.

... und Englisch ist doch komplexer!

Was für Unterschiede können wir in diesem Bereich wohl antreffen? Nach Clyne (1987, 211; 1991, 52) sollen sich englische Argumentations- und Satzstrukturen eher durch Koordination, deutsche eher durch Subordination auszeichnen. Danach wäre beispielsweise zu erwarten, dass die Zahl der Nebensätze in englischen Fachtexten geringer ausfällt, was dazu beitragen könnte, den Eindruck der besseren Verständlichkeit zu erzeugen. Offenbar muss aber auch die Art der Nebensätze mit ins Spiel gebracht werden, denn einige Autoren treffen hierzu bestimmte Aussagen. Beier (1980, 60) schreibt zum Beispiel, dass nicht-finite Verbformen (und damit auch nicht-finite Satzgefüge, deren Kern das nicht-finite Verb bildet) ein typisches Kennzeichen der Fachsprache im Englischen seien, wobei genaue Quantifizierungen offenbar nicht vorliegen. Was die Satzlänge angeht, gibt es offenbar auch keine ganz eindeutigen Befunde. Giehl (2003) findet im Gegensatz zu einigen ihrer Vorgänger (z.B. Strevens 1977) keine signifikanten Unterschiede, sodass hier keine hinreichende Grundlage für gesicherte Schlussfolgerungen besteht.

Insgesamt ergibt sich kein sehr klares Bild, was vermutlich zwei Dingen geschuldet ist. Zum einen können Texte unterschiedlicher Fachdisziplinen durchaus in unterschiedlicher Frequenz verschiedene syntaktische Merkmale zeigen. Zum anderen weisen auch verschiedene Textsorten (z.B. Abstracts, Forschungsberichte, Rezensionen etc.) innerhalb einer Disziplin häufig Unterschiede in der Ausprägung verschiedenster Merkmale auf. In der vorliegenden Studie soll daher ein Vergleich von Texten innerhalb nur einer Disziplin, nämlich der Sprachwissenschaft, unternommen werden, beschränkt auf eine einzige, relativ eng umgrenzte Textsorte, und zwar den Forschungsartikel in einer einschlägigen wissenschaftlichen Zeitschrift.

Im Mittelpunkt stehen dabei folgende Forschungsfragen, die von der Hypothese geleitet sind, dass die (unterstellte) bessere Verständlichkeit englischer Texte auch durch eine leichter verständliche, d.h. (auf der Satzebene) weniger komplexe, Syntax zustande kommt.²

² Diese Annahme unterstellt, dass *ceteris paribus* komplexe Sätze schwerer zu verstehen sind als nicht-komplexe. Obwohl intuitiv plausibel, ist dies letztlich eine empirische Frage, die im Einzelnen schwierig zu untersuchen ist. Einschlägige Studien dazu sind mir nicht bekannt.

- Gibt es Unterschiede im Aufbau von englischen und deutschen Sätzen in Bezug auf ihre Komplexität?
- Wenn ja, welche Arten von Sätzen sind betroffen?
- Bestätigen die eventuell gefundenen Unterschiede die o.g. Hypothese?

Zur Beantwortung dieser Fragen wurden je drei Aufsätze aus jeder Sprache untersucht (Gesamttextränge: 54389 Wörter). Wie wir noch sehen werden, erlauben die angewandten statistischen Verfahren trotz der relativ kleinen Stichprobe verlässliche Aussagen über etwaige Unterschiede.

Im Folgenden werde ich zunächst das methodische Vorgehen erläutern, dann in Abschnitt 3 die Ergebnisse vorstellen, deren Implikationen in Abschnitt 4 schließlich näher diskutiert werden.

2. Methode

Als Stichprobe wurde eine zufällige Auswahl von je drei Aufsätzen aus jeder der beiden Sprachen gewählt, wobei allerdings zur besseren Vergleichbarkeit alle Texte renommierten Zeitschriften entnommen wurden und sich mit morphosyntaktischen Themen befassen. Es handelt sich um die folgenden Texte:

- Allen, Cynthia L. 2003. Deflexion and the development of the genitive 's in English. *English Language and Linguistics* 7. 1-28.
- Bauer, Laurie. 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2. 65-86.
- Cornish, Francis. 2001. 'Modal' *that* as determiner and pronoun: the primacy of the cognitive-interactive dimension. *English Language and Linguistics* 5. 297-315.
- Boase-Beier, Joan & Jindrich Toman. 1986. Komposita im Text. *Zeitschrift für Literaturwissenschaft und Linguistik* 64. 61-72.
- Fuhrhop, Nanna & Rolf Thieroff. 2006. Was ist ein Attribut? *Zeitschrift für germanische Linguistik* 33. 306-342.
- Köpcke, Klaus-Michael & David Zubin. 2005. Nominalphrasen ohne lexikalischen Kopf – Zur Bedeutung des Genus für die Organisation des mentalen Lexikons am Beispiel der Autobezeichnungen im Deutschen. *Zeitschrift für Sprachwissenschaft* 24. 93-122.

Aus den ersten Seiten der Texte (knapp 5000 Wörter pro Sprache) wurden alle Satzgefüge extrahiert, die einen Nebensatz enthielten. Eine Entscheidung, welche Strukturen als Nebensätze zu klassifizieren sind und welche nicht, ist, insbesondere was nicht-finite Strukturen angeht, nicht immer einfach zu treffen. Die Problematik liegt zum einen darin, dass unterschiedliche theoretische

... und Englisch ist doch komplexer!

Annahmen zu unterschiedlichen Ergebnissen führen können, was die Satzwertigkeit einer Struktur angeht. So werden in nicht-finiten Sätzen regelmäßig Argumente des Verbs nicht auf der Oberfläche realisiert, so dass hier manche Grammatik nicht mehr von satzwertigen Konstruktionen spricht. Darüber hinaus entsteht (wie immer) beim Sprachvergleich das Problem der Vergleichbarkeit. Das Pendant dessen, was in der einen Sprache relativ klar eine satzwertige Struktur ist, mag in der anderen Sprache als nicht satzwertig angesehen werden.

Das erste Problem wurde hier methodisch so gelöst, dass Konstruktionen, die ein Verb enthielten, grundsätzlich als satzwertig eingestuft wurden. Dies geschah unter Rückgriff auf die Vorstellung, dass satzwertige Konstruktionen normalerweise Propositionen ausdrücken und Verben im Zentrum dieser Propositionen stehen. Ein spezielles Problem entsteht dabei allerdings zuweilen bei Partizipien, deren Status als Verb, insbesondere in attributiver Stellung, (z.B. *eine laufende Studie, an ongoing study*) fraglich ist. Sie werden gemeinhin als Adjektive klassifiziert und kommen demnach als Nebensätze nicht in Frage. Nehmen diese Partizipien allerdings Erweiterungen zu sich, wie in den Beispielen in (1), die alle den untersuchten deutschen Texten entnommen sind, ist die Lage nicht mehr so klar.

- (1) a. diese [für die Benennung von Autos verwendeten] Phrasen
- b. die [für die Genuszuweisung zu der Phrase verantwortliche] Instanz
- c. die [weiter unten geschilderten] textbezogenen Interpretationsmechanismen

Im Kontext dieser kontrastiven Studie sind diese Konstruktionen besonders interessant, weil sie normalerweise im Englischen ausschließlich durch nachgestellte Relativsätze ausgedrückt werden können. Das folgende Beispiel illustriert dies. (2b) ist im Englischen ungrammatisch, während seine deutsche Entsprechung vollkommen unproblematisch ist.

- (2) a. genitives [which descend directly from the old genitive case]
- b. *[descending directly from the old genitive case] genitives
- c. [direkt vom alten Genitivkasus abstammende] Genitive

Derartige erweiterte partizipiale Attribute werden in der germanistischen Literatur in aller Regel nicht als satzwertige Elemente aufgefasst, auch wenn sie alternativ durch einen nachgestellten Relativsatz ausgedrückt werden könnten und eindeutig propositionaler Natur sind. Ich orientiere mich hier aber an der

vergleichenden-typologischen Literatur, in der vorangestellte Attribute, sofern sie ein Verb oder ein prädikatives Adjektiv als Nukleus und weitere davon abhängige Ergänzungen enthalten, als Relativsätze behandelt werden (vgl. z.B. Keenan & Comrie 1977, Andrews 2007). Eine solche Entscheidung ist sinnvoll, da sie auf semantischen Kriterien (,propositionales Attribut‘) beruht und erst dadurch Vergleichbarkeit von Sprachen erzielt wird. Die deutschen Beispiele in (1) sind demnach als Relativsätze zu behandeln, und es fanden sich insgesamt 15 derartige Konstruktionen in den deutschen Daten.

Ein anderes Problem stellen verblose, propositionale Satzteile dar, so genannte *small clauses*, wie die in Klammern gesetzten Strukturen in (3) (aus Aarts 1992):

- (3) a. I consider [this man an idiot] (p. 37)
b. I find [it rather hot] (p. 38)
c. [With Peter the referee,] we might as well not play the match (p. 42)

Der Status dieser Elemente als (satzwertige) Konstituenten ist in der Literatur umstritten. In der vorliegenden Studie wurden *small clauses* wegen ihres umstrittenen Status nicht berücksichtigt, d.h. die hier untersuchten Nebensätze weisen ein an der Oberfläche realisiertes Verb auf.

Die gefundenen Nebensätze wurden in Bezug auf ihre Länge (in Wörtern), ihre Einbettungstiefe und ihre innere Struktur klassifiziert. Bei der Länge wurde die unterordnende Instanz (Konjunktion oder Präposition) mitgezählt, da sie strukturell nicht zum übergeordneten Satz gehört, sondern mit dem eigentlichen Nebensatz eine Konstituente bildet. Die Einbettungstiefe wurde ebenfalls numerisch kodiert, mit Werten von 1 bis 5, mit 5 als Wert für die am tiefsten (d.h. fünffach) eingebettete Struktur. Bei der Kodierung der inneren Struktur ist zunächst die formale Unterscheidung in finite und nicht-finite Nebensätze relevant, wobei nicht-finite daran zu erkennen sind, dass Partizipialformen oder der Infinitiv verwendet werden und in der Regel die Realisierung von Objekt oder Subjekt nicht-kanonisch erfolgt (d.h. entweder gar nicht oder, im Falle von overten Subjekten, mit Objekt-Kasus). Weiterhin ist die Wortstellung ein Unterscheidungskriterium, da Relativsätze hier Besonderheiten aufweisen, wie die Voranstellung von Relativelementen (Relativpronomina oder entsprechenden Phrasen) und das Vorhandensein einer damit korrespondierenden Lücke im

... und Englisch ist doch komplexer!

Relativsatz (vgl. z.B. Huddleston & Pullum 2002). Der Nebensatztyp wurde daher mit den Werten finit, nicht-finit und relativ kodiert.

Durch diese Kodierung ergab sich aber das Problem der Klassifikation von eher sporadisch auftretenden nicht-finiten Relativsätzen. Die zehn nicht-finiten Relativsätze in den englischen Daten, wie in $_{NP}$ [*principles [underlying the choice]*], wurden als Relativsätze klassifiziert, und nicht als nicht-finit (in Analogie zu den in ähnlicher Anzahl vorhandenen deutschen Partizipialattributen, siehe oben).

Die Kodierungen für jeden Satz wurden in eine Tabelle eingetragen und anschließend einer statistischen Analyse unterzogen. Für die statistische Analyse wurde die frei erhältliche Software R verwendet (R Development Core Team 2007). Die Wahl quantitativer Methoden hat vor allem den Grund, dass, wie schon Hoffmann (1985, 77) schreibt, die Spezifik der Fachsprachen in erster Linie quantitativer Natur ist: Fachsprachen unterscheiden sich von der Gemeinsprache vornehmlich nicht durch die jeweiligen sprachlichen Mittel selbst, sondern vor allem durch deren Häufigkeitskonstellationen. Deshalb, so Hoffmann, „spielen statistische Methoden in der fachsprachlichen Forschung eine große Rolle“ (1985, 77). Umso mehr erstaunt es, dass die verwendeten statistischen Methoden in der Literatur oft eher einfach sind.

Die in der vorliegenden Untersuchung in Anschlag gebrachten Methoden sind komplexer und umfassen neben der statistischen Beschreibung der Daten auch das Testen von Hypothesen mittels statistischer Testverfahren und die statistische Modellierung mit Hilfe von multipler Regression. Hierzu erscheinen einige Erläuterungen angebracht, die statistisch weniger vorgebildeten Lesern und Leserinnen das Verständnis erleichtern sollen. Die anzuwendenden statistischen Tests (wie z.B. der Chi-Quadrat-Test) berechnen die Wahrscheinlichkeit p , dass eine gefundene Verteilung der Daten auf Zufall beruht. Multiple Regression hingegen ist ein Verfahren, bei dem die Wahrscheinlichkeit eines bestimmten Werts einer abhängigen Variablen mit Hilfe von mehreren unabhängigen Variablen vorhergesagt wird. Derartige multivariate Verfahren haben den Vorteil, dass mit ihrer Hilfe der Einfluss einer einzelnen Variablen unter Konstanthaltung aller anderen beteiligten Vorhersagevariablen ermittelt werden

kann. Für Einführungen in die hier verwendeten Methoden und technischen Begriffe siehe z.B. Baayen (2008), Johnson (2008), oder Gries (2009).

3. Ergebnisse

3.1. Text- und Satzlängen

Ich möchte zunächst einen Überblick über Text- und Satzlängen geben, bevor ich auf die Nebensätze im Detail eingehe. Die Text- und Satzlängen der einzelnen Texte sind in Tabelle 1 dargestellt.

Tabelle 1: Eigenschaften der Texte (Längenangaben: in Wörtern)

Text	Sprache	Gesamt- länge	Länge der Stichprobe	Sätze in Stichprobe	Nebensätze in Stichprobe
<i>Allen</i>	Englisch	11.132	1.754	62	86
<i>Bauer</i>	Englisch	10.053	1.065	38	86
<i>Cornish</i>	Englisch	8.737	2.110	73	63
<i>Fuhrhop/ Thieroff</i>	Deutsch	11.011	1.678	107	52
<i>Köpcke/ Zubin</i>	Deutsch	9.918	1.912	91	54
<i>Boase- Beier/ Toman</i>	Deutsch	3.538	1.085	50	48
Gesamt		54.389	9.604	421	389

Da wir an den Unterschieden zwischen den Sprachen und nicht an den Unterschieden zwischen den Texten interessiert sind, und wir daher die jeweiligen Texte einer Sprache zusammen behandeln wollen, ergibt sich die Frage, ob die Texte innerhalb einer Sprache überhaupt homogen genug sind, um sie zusammen behandeln zu können. Zu diesem Zweck wurden explorative Varianzanalysen zum Zusammenhang der zu untersuchenden abhängigen Variablen und dem jeweiligen Text durchgeführt, die insgesamt zeigten, dass die Texte einer Sprache in der Regel nur kleinere Unterschiede aufweisen, so dass eine vergleichende Sprachanalyse möglich ist. Darüber hinaus wird aber im Folgen-

... und Englisch ist doch komplexer!

den auch die Variation zwischen den einzelnen Texten einer Sprache dokumentiert.

Als erstes wenden wir uns den Satz­längen zu. Wir finden recht klare, wenn auch widersprüchliche Verhältnisse in der Verteilung. Abbildung 1 illustriert die Befunde.

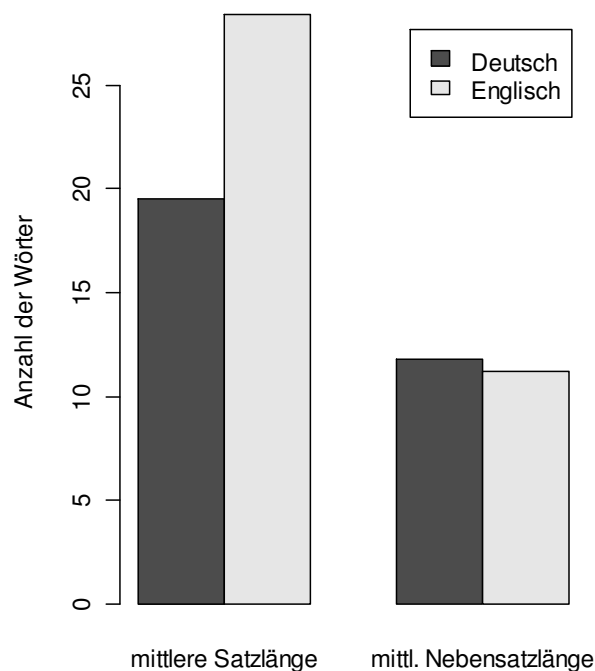


Abbildung 1: Mittlere Satz­länge und mittlere Nebensatz­länge nach Sprache

Die mittleren Satz­längen sind deutlich unterschiedlich (19,5 vs. 28,4 Wörter), wobei die englischen Sätze signifikant länger sind ($p= 0.04016$, $t(2.075)=-4.6595$). Schauen wir hingegen die Nebensätze an, sind diese praktisch identisch lang in den beiden Sprachen (11,8 Wörter im Deutschen gegenüber 11,2 im Englischen ($p=0.4722$, $t(307.682)=0.7198$)).

Eine Varianzanalyse hat ergeben, dass sich die Texte innerhalb einer Sprache kaum in Bezug auf Satz­längen unterscheiden, mit der Ausnahme des Textes von *Fuhrhop/Thieroff* im Deutschen, der sich in Bezug auf seine mittlere Satz­länge von 15,7 Wörtern noch stärker vom Englischen unterscheidet als die beiden anderen deutschen Texte es ohnehin schon tun.

3.2. Häufigkeit von Nebensätzen

Wie steht es aber mit der Frage, welche Sprache mehr Nebensätze aufweist? Hierfür stehen wenigstens drei verschiedene Maße zur Verfügung. Zum einen kann das Verhältnis von Nebensätzen zu Sätzen insgesamt als Gradmesser dienen, zum anderen kann die Zahl an Nebensätzen zur Gesamtwortzahl in Beziehung gesetzt werden. Als dritte Möglichkeit könnte man auch das Verhältnis der Anzahl der Wörter in Nebensätzen zur Anzahl der Wörter insgesamt berechnen. Allerdings ist diese Größe weniger aussagekräftig in Bezug auf die Auftretenswahrscheinlichkeit von Nebensätzen, sondern steht für die Wahrscheinlichkeit, dass ein Wort in einem Nebensatz vorkommt. Dies wiederum hängt entscheidend von der Länge der Nebensätze ab. Wir werden uns daher auf die ersten beiden Maße konzentrieren.

Mit beiden Maßen wird die Wahrscheinlichkeit des Auftretens von Nebensätzen abgebildet, nur jeweils in Bezug auf eine andere Größe. Nehmen wir die Anzahl der Sätze als Bezugsgröße, so ist die Wahrscheinlichkeit, unter einer gegebenen Anzahl deutscher Sätze solche mit Nebensätzen zu finden, wesentlich kleiner als diejenige für einen englischen Text. Im Englischen finden wir sogar mehr Nebensätze als Sätze. Dies zeigt Abbildung 2, linker Graph. Der Unterschied zwischen den Sprachen ist hoch signifikant ($p < 0,001$, $\chi^2 = 29,4167$, $df=1$, Pearson-Test mit Yates-Korrektur). Allerdings ist auch zu konstatieren, dass die Variation im Englischen breiter ist als im Deutschen. So schwankt das Verhältnis von Sätzen zu Nebensätzen in den englischen Texten zwischen 2,3 (*Bauer*) und 0,9 (*Cornish*), während es im Deutschen nur eine Spannbreite von 0,5 bis 1,1 aufweist, mit sehr ähnlichen Werten für *Fuhrhop/Thieroff* und *Köpcke/Zubin* (0,5 und 0,6) und *Boase-Beier/Toman* im unteren Bereich der englischen Werte.

Wie Abbildung 2, rechter Graph illustriert, ist die Auftretenswahrscheinlichkeit gemessen an der Wortzahl ebenfalls signifikant höher in englischen Texten (3,3 Nebensätze pro 100 Wörter im Deutschen gegenüber 4,8 im Englischen, $p = 0,03231$, Poisson-Test³). Einschränkend muss aber auch hier hinzugefügt

³ Die Auftretenswahrscheinlichkeit von Wörtern in Texten folgt einer Poisson-Verteilung (z.B. Baayen 2008). Dem hier angewandten Poisson-Test liegt ferner die Überlegung zu-

... und Englisch ist doch komplexer!

werden, dass sich die Texte innerhalb einer Sprache doch einigermaßen unterscheiden können. Dies ist in Bezug auf Nebensätze pro Wörter besonders für das Englische der Fall. Hier liegt der *Bauer*-Text weit in Führung mit 8 Nebensätzen pro 100 Wörter, während am anderen Ende der Skala der *Cornish*-Text mit 3 Nebensätzen pro 100 Wörter im Bereich des Deutschen liegt.

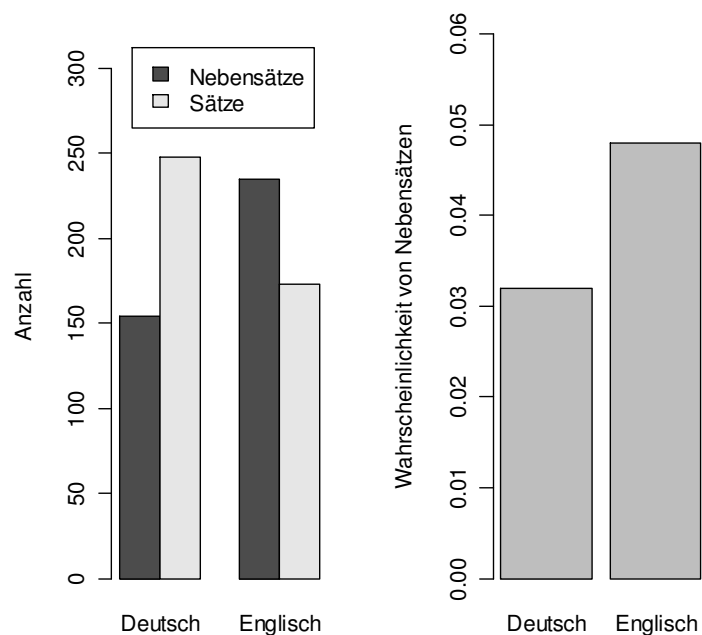


Abbildung 2: Häufigkeit von Nebensätzen nach Sprache. Linker Graph: Verhältnis von Nebensätzen (dunkel markiert) zu Sätzen (hell markiert). Rechter Graph: Wahrscheinlichkeit des Auftretens von Nebensätzen nach Wortzahl.

Auch wenn wir ausschließlich diejenigen Nebensätze anschauen, die nur einfach eingebettet sind, ergibt sich das gleiche Bild: Während im Deutschen 114 von 248 Sätzen (mindestens) einen Nebensatz haben (46 Prozent), sind es im Englischen 80 Prozent. Der Unterschied ist hoch signifikant ($p < 0,001$, $\chi^2 = 11,3065$, $df = 1$, Pearson-Test mit Yates-Korrektur).⁴

Insgesamt ist also festzuhalten, dass es in den englischen Texten sehr viel mehr Nebensätze gibt als in den deutschen, egal welches Maß wir betrachten.

grunde, dass die Auftretenswahrscheinlichkeit von Nebensätzen durch die Auftretenswahrscheinlichkeit von Nebensatzinitialen Wörtern modelliert werden kann.

⁴ Die deutschen Texte schwanken zwischen 36 und 70 Prozent Anteil an Sätzen, die einen Nebensatz aufweisen, die englischen zwischen 54 und 124 Prozent.

3.3. Nebensatztypen und Einbettungstiefe

Nachdem wir, vielleicht etwas überraschend, festgestellt haben, dass in den deutschen Texten kürzere Sätze und ein signifikant niedrigeres Auftreten von Nebensätzen zu verzeichnen ist, schauen wir uns nun die Ergebnisse für die verschiedenen Nebensatztypen an.

Eine Analyse der Verteilung von finiten, nicht-finiten und Relativsätzen zeigt, dass der einzige, aber gravierende, Unterschied zwischen den Sprachen die nicht-finiten Sätze betrifft. Der Graph in Abbildung 3 verdeutlicht dies. Auf der y-Achse ist die Zahl der Nebensätze abgetragen, die Kästchen in den Balken geben die Häufigkeit der jeweiligen Beobachtungen an. Während die Menge an finiten Nebensätzen und Relativsätzen in beiden Sprachen annähernd gleich ist, weist das Englische die siebenfache Menge an nicht-finiten Sätzen auf. Dieser Unterschied in der Verteilung ist hoch signifikant ($p < 0,001$, $\chi^2 = 39,2279$, $df = 2$, Pearson-Test).⁵

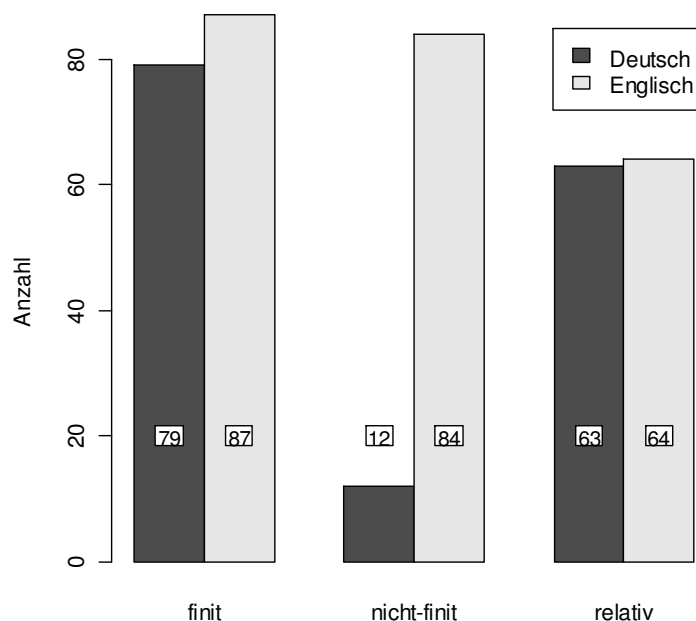


Abbildung 3: Verteilung von Nebensatztypen nach Sprache

⁵ Die Anteile der nicht-finiten Sätze im Deutschen und Englischen unterscheiden sich stark voneinander, selbst wenn man die Variation zwischen den Texten innerhalb der Sprachen mit berücksichtigt: vgl. 2, 9 und 12 Prozent in den deutschen Texten, gegenüber 24, 38 und 48 Prozent in den englischen Texten.

... und Englisch ist doch komplexer!

Auch bei der Einbettungstiefe zeichnen sich klare Unterschiede ab, wie Abbildung 4 zeigt, in der die Anteile der Nebensätze verschiedener Einbegrungsgrade innerhalb jeder Sprache dargestellt werden. Auf der y-Achse sind die jeweiligen Anteile an allen Nebensätzen abgetragen (1 = 100 Prozent). Die Balken jeder Sprache addieren sich zu 1, d.h. 100 Prozent. Die Kästchen in den Balken geben die Häufigkeiten der jeweiligen Kategorie an. Während im Deutschen über 70 Prozent der Nebensätze einfach eingebettet sind, sind es im Englischen weniger als 60 Prozent. Mit steigendem Einbegrungsgrad kehren sich die Verhältnisse ins Gegenteil. Der Anteil der zweifach eingebetteten Nebensätze ist im Englischen schon höher als im Deutschen, und mit nur 2 Prozent Anteil an allen Nebensätzen haben die deutschen Texte fast keine dreifach eingebetteten Nebensätze mehr, während unter den englischen Nebensätzen diese immerhin noch 8 Prozent ausmachen. Noch tiefer eingebettete Nebensätze kommen in den deutschen Daten gar nicht mehr vor, in den englischen wenigstens gelegentlich. Dies bedeutet, dass im Englischen die Nebensätze stärker verschachtelt sind als im Deutschen; der Unterschied ist hoch signifikant, wie die Ergebnisse der Varianzanalyse zeigen ($p < 0,001$, $df = 1$, $F = 14,029$).⁶

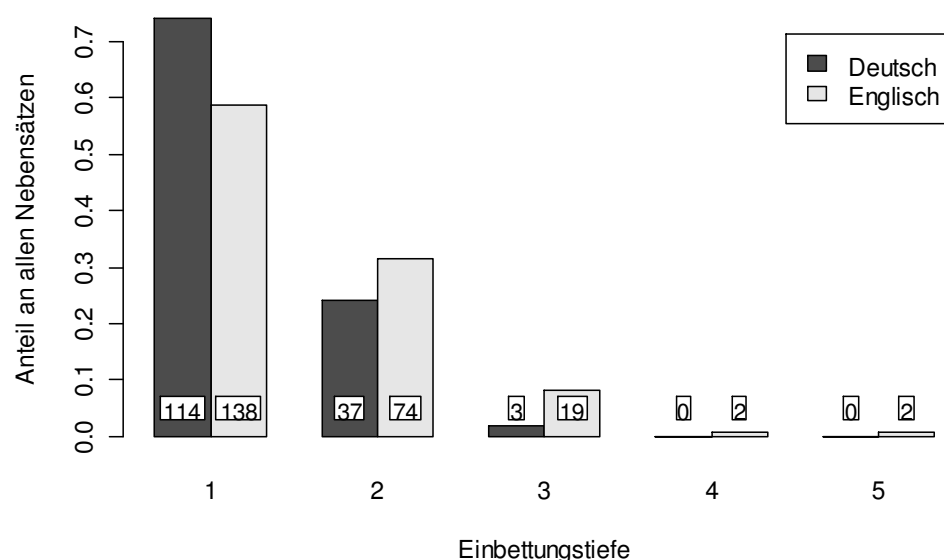


Abbildung 4: Einbettungstiefe von Nebensätzen im Deutschen und Englischen

⁶ Varianzanalysen zeigten, dass sich die Texte innerhalb einer Sprache in Bezug auf die Verteilung der Einbettungstiefen nicht voneinander unterscheiden.

Ein interessanter Punkt ist in diesem Zusammenhang die Frage, ob die Länge der Nebensätze im Zusammenhang steht mit dem Einbettungsgrad, und ob dieser Zusammenhang vielleicht je nach Sprache (oder Text) unterschiedlich ist. Man würde generell erwarten, dass tiefer eingebettete Nebensätze kürzer sind, unabhängig von Sprache oder Text. Um dies zu untersuchen, wurde eine lineare Regression durchgeführt mit WORTZAHL, SPRACHE und TEXT als Vorhersagevariablen und dem Grad der EINBETTUNG als vorherzusagende Variable. Es fanden sich ein Haupteffekt nur für WORTZAHL und keine Interaktionen. Dies bedeutet, dass weder die Verschiedenheit der Sprachen noch die Verschiedenheit der Texte einen Einfluss auf den Grad der Einbettung haben. Nach der Entfernung der nicht signifikanten Prädiktoren ergab sich das in Tabelle 2 dokumentierte Modell.

Tabelle 2: Logistisches Regressionsmodell für den Zusammenhang von Nebensatzlänge (WORTZAHL) und Einbettungstiefe.

	Koeffizient	Std. Error	t-Wert	Pr(> t)
(Intercept)	14,2524	0,8958	15,910	< 2e-16
EINBETTUNG	-1,9537	0,5655	-3,455	0,000612

Abbildung 5 zeigt die Verteilung der Beobachtungen und die Regressionslinie des Modells (durchgezogene Linie), sowie die mittlere Nebensatzlänge als gestrichelte Linie. Man sieht einen schwachen Effekt der Einbettungstiefe ($R^2 = 0,02741$) in die erwartete Richtung, d.h. mit zunehmender Einbettungstiefe werden die Nebensätze tendenziell kürzer. Im Modell wird diese Tendenz durch den Koeffizienten von EINBETTUNG (Tabelle 2, Spalte 2: -1,9537) so berechnet, dass mit jeder Erhöhung des Einbettungsgrades die Nebensatzlänge um etwa zwei Wörter abnimmt.

... und Englisch ist doch komplexer!

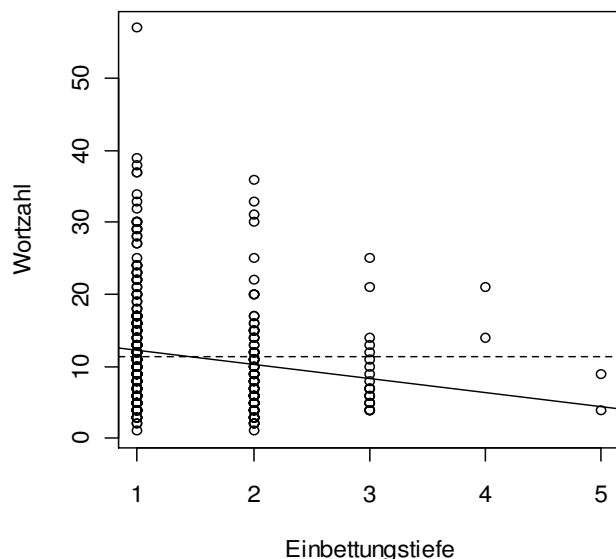


Abbildung 5: Einbettungstiefe nach Länge (Wortzahl) des Nebensatzes (alle Texte). Durchgezogene Linie: Regressionslinie, gestrichelte Linie: mittlere Nebensatzlänge.

3.4. Multivariate Analyse: Wortzahl, Nebensatztyp und Einbettungstiefe

Abschließend wurden die Daten noch einer weiteren multivariaten Analyse unterzogen, um zu prüfen, ob die Befunde aus den obigen univariaten Analysen auch unter gleichzeitiger Präsenz aller anderen Prädiktoren bestätigt werden. Hierzu wurde eine logistische Regression mit SPRACHE als abhängiger Variablen und EINBETTUNG, WORTZAHL, und Nebensatztyp (SATZTYP) als Prädiktoren durchgeführt. Mit anderen Worten, dieses Modell versucht auf der Basis der Ausprägungen von EINBETTUNG, WORTZAHL und SATZTYP vorherzusagen, ob es sich um einen deutschen oder englischen Nebensatz handelt. Als Ergebnis wurden Haupteffekte für die Einbettungstiefe und den Nebensatztyp gefunden, nicht aber für die Nebensatzlänge. Es gab keine signifikanten Interaktionen. Tabelle 3 dokumentiert das Modell (mit allen Prädiktoren, aber ohne die Interaktionen) und die dazugehörige Varianzanalyse.

Tabelle 3: Varianztabelle (Wald-Statistik) des logistischen Regressionsmodells für den Zusammenhang von EINBETTUNG, WORTZAHL, SATZTYP auf der einen und SPRACHE auf der anderen Seite.

Faktor	Chi-Quadrat	d.f.	P
EINBETTUNG	8,18	1	0,0042
WORTZAHL	0,23	1	0,6303
SATZTYP	29,46	2	<,0001
TOTAL	39,26	4	<,0001

In Abbildung 6 werden die modellierten Effekte der Variablen graphisch dargestellt. Auf der y-Achse wird die Wahrscheinlichkeit abgetragen, dass es sich um einen englischen Satz handelt. Im linken Graph steigt diese Wahrscheinlichkeit mit zunehmendem Einbettungsgrad an. Die Striche über und unter den Punkten geben das 95%- Konfidenzintervall an, ebenso wie die obere und untere gestrichelte Linie im mittleren Graph. In diesem sehen wir, dass mit steigender Wortzahl die Wahrscheinlichkeit ganz leicht ansteigt, dass es sich um einen englischen Satz handelt, aber der Effekt ist sehr schwach und erreicht keine Signifikanz (s. Tabelle 3, $p=0,6303$). Anders sieht es für die Variable SATZTYP aus, die einen klaren Effekt aufweist. Während finite Nebensätze und Relativsätze keine Tendenz aufweisen, sind nicht-finite Sätze sehr viel eher in den englischen Texten anzutreffen.

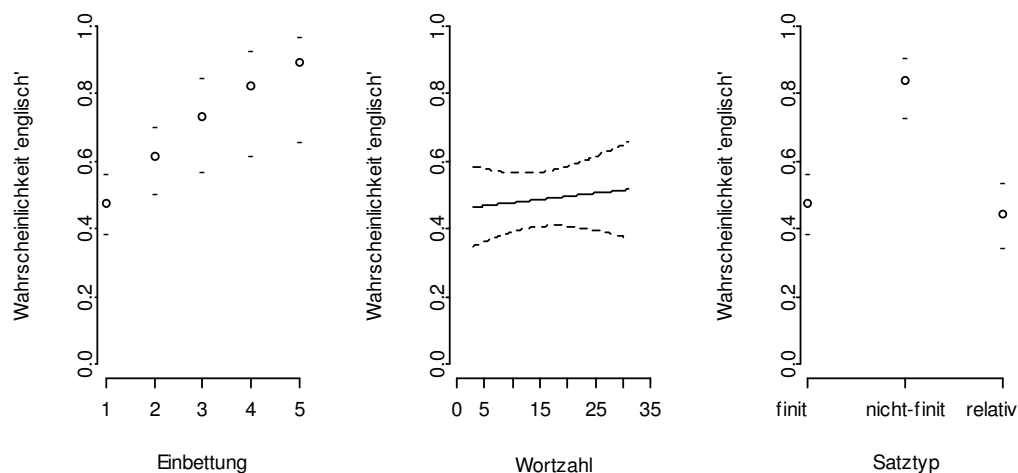


Abbildung 6: Individuelle Effekte der logistischen Regressionsanalyse

... und Englisch ist doch komplexer!

Die Ergebnisse der Regressionsanalyse bestätigen damit in vollem Umfang die Befunde aus den univariaten Analysen von Nebensatzlänge, Einbettungstiefe und Nebensatztypen. Es bestehen nicht-zufällige Zusammenhänge zwischen Einbettungstiefe, Wortzahl und Satztyp auf der einen Seite und der Sprache auf der anderen.

4. Zusammenfassung und Diskussion

Insgesamt hat sich herausgestellt, dass die englischen Texte im Hinblick auf die Nebensatzstrukturen komplexer aufgebaut sind als die deutschen: Englische Sätze (inklusive der möglicherweise darin enthaltenen Nebensätze) sind länger als deutsche Sätze. Außerdem finden sich in den englischen Daten mehr Nebensätze, und diese sind auch, was weitere satzwertige Einbettungen angeht, komplexer aufgebaut als die Nebensätze in den deutschen Texten. Die englischen Nebensätze sind aber dennoch nicht länger als die deutschen. Weiterhin haben wir herausgefunden, dass englische Nebensätze zu einem sehr viel höheren Prozentsatz nicht-finit sind. Es fand sich außerdem die erwartete negative Korrelation von Einbettungstiefe und Nebensatzlänge, und zwar unabhängig von Sprache oder Text.

Die Unterschiede zwischen den Texten einer Sprache sind zwar meist eher klein, doch Aussagen über ‚das‘ Englische oder ‚das‘ Deutsche sollten nur nach eingehender statistischer Überprüfung getroffen werden. Insgesamt zeigte sich dabei, dass es häufig sehr deutliche Tendenzen gibt, die unabhängig von einzelnen Texten für eine Sprache gelten.

Was bedeuten diese Ergebnisse für die Diskussion um fachsprachliche Unterschiede Deutsch-Englisch? In Bezug auf die Frage der Neben- versus Unterordnung kann die vorliegende Studie frühere Befunde nicht bestätigen, nach denen das Englische weniger hypotaktisch organisiert sei als das Deutsche. Das genaue Gegenteil ist der Fall in den hier untersuchten Texten. Beier (1980, 57) findet in englischen chemischen Fachtexten in 50 Prozent der Sätze einen oder mehrere Nebensätze. Dies ist in etwa die Zahl, die wir in den vorliegenden deutschen Texten finden, während die englischen Texte wesentlich mehr aufweisen (80 Prozent). Ob dies der Textsorte oder der wissenschaftlichen Disziplin ge-

schuldet ist, ist unklar, da Beier die chemischen Fachtexte nicht weiter spezifiziert.

Andererseits bestätigt die vorliegende Untersuchung die von Beier (1980, 60) konstatierte Häufigkeit von nicht-finiten Verbformen im Englischen, die sich dahingehend in unseren Texten manifestiert, dass der Anteil nicht-finiten Nebensätze im Englischen den Anteil im Deutschen um das Siebenfache übersteigt. Die große Bedeutung von Relativsätzen in beiden Sprachen, die schon Hoffmann (1985, 208) empirisch belegt hat, konnte ebenfalls bestätigt werden.

In Bezug auf die Satzlänge zeigt die hier untersuchte Stichprobe, dass entgegen einiger Aussagen in der Literatur englische Sätze tatsächlich länger sind als deutsche. Insofern behält Strevens recht, der (allerdings ohne den Vergleich mit dem Deutschen) zu dem Schluss kommt, dass die englische Fachsprache gekennzeichnet sei durch „rather long sentences containing many clauses, often in complex degrees of dependency and with much embedding“ (Strevens 1977, 154).

Die oft empfundene bessere Verständlichkeit englischsprachiger Texte ist jedenfalls nicht klar aus ihrer Satzsyntax zu erklären. Vielmehr scheint es so zu sein, dass das Englische durch die stärkere Verwendung nicht-finiten Nebensätze, größerer Einbettungstiefe und längerer Sätze die Verständlichkeit eher schmälert als befördert. Auf Grund der Vielzahl von Faktoren sind verlässliche Aussagen über das sprachliche Substrat unterschiedlich guter Verständlichkeit daher wohl weiterhin schwierig.

Quellenverzeichnis

- AARTS, Bas. 1992. *Small clauses. The non-verbal types*. Berlin/New York: Mouton de Gruyter.
- ANDREWS, Avery D. 2007. Relative clauses. In: Timothy Shoben (ed.): *Language typology and linguistic description. Vol. 2: Complex constructions*. Cambridge: Cambridge University Press. 206-236.
- BAAZEN, R. Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics*. Cambridge: Cambridge University Press.
- BEIER, Rudolf. 1980. *Englische Fachsprache*. Stuttgart u.a.: Kohlhammer.
- CLYNE, Michael. 1987. Cultural differences in the organisation of academic texts. *Journal of Pragmatics*, 11.2. 211–247.
- CLYNE, Michael. 1991. The sociocultural dimension: The dilemma of the German-speaking scholar. In: Schröder, Hartmut (Hrsg.): *Subject-oriented texts: Languages for special*

... und Englisch ist doch komplexer!

- purposes and text theory*. Research in text theory / Untersuchungen zur Texttheorie, 16. Berlin/New York: de Gruyter. 49–67.
- GRIES, Stefan Th. 2009. *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck und Ruprecht.
- GIEHL, Claudia. 2003. Determinanten der Verständlichkeit englischer und deutscher Fachtexte. In: Schubert, Klaus (Hrsg.): *Übersetzen und Dolmetschen: Modelle, Methoden, Technologie*. Jahrbuch Übersetzen und Dolmetschen. Tübingen: Narr. 193–206.
- HOFFMANN, Lothar. 1985. *Kommunikationsmittel Fachsprache. Eine Einführung*. 2. Aufl. Tübingen: Narr.
- HUDDLESTON, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- JOHNSON, Keith. 2008. *Quantitative methods in linguistics*. Oxford: Blackwell.
- KEENAN, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63-99.
- PINCHUCK, Isadore. 1977. *Scientific and technical translation*. London: Deutsch.
- R DEVELOPMENT CORE TEAM. 2007. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- SIEPMANN, Dirk. 2010. Gesellschaftlich unterbewertet. Die Übersetzung deutscher wissenschaftlicher Texte. *Lehre und Forschung* 17.8. 580-582.
- STREVENS, Peter. 1977. Special-purpose language learning: A perspective. *Language Teaching* 10. 145-163.