



FinDe

Arbeiten mit dem finnisch-deutschen Kontrastkorpus

Band 3

Leena Kolehmainen und Peter Stahl

Das zweisprachige FinDe-Korpus

FinDe

Arbeiten mit dem finnisch-deutschen Kontrastkorpus

Das FinDe-Korpus ist ein zweisprachiges finnisch-deutsches Textkorpus. Es ist bilateral angelegt, enthält deutsch- und finnischsprachige Quelltexte und ihre jeweiligen Übersetzungen in die andere Sprache. Es umfasst knapp 800.000 Wortformen. Das Korpus wurde im Rahmen des Forschungsprojekts „Wortbildung und Textanalyse im deutsch-finnischen Kontrast“ für Forschungszwecke erstellt.

Projektleiter:

Universität Würzburg: Norbert Richard Wolf, Werner Wegstein

Universität Jyväskylä: Ahti Jänntti

Universität Tampere: Marja-Leena Piitulainen

Universität Helsinki: Irma Hyvärinen

<http://www.spr.germanistik.uni-wuerzburg.de/finde-korpus>

FinDe. Arbeiten mit dem finnisch-deutschen Kontrastkorpus

Nr. 3 (Dezember 2007)

Leena Kolehmainen und Peter Stahl:

Das zweisprachige FinDe-Korpus

ISSN: 1864-8835

ISBN: 978-3-923959-37-2

© Lehrstuhl für deutsche Sprachwissenschaft

Julius-Maximilians-Universität Würzburg

Institut für deutsche Philologie

Am Hubland

97074 Würzburg

Tel.: +49 (0) 931 - 888 - 56 30

Fax: +49 (0) 931 - 888 - 46 16

<http://www.spr.germanistik.uni-wuerzburg.de>

Alle Rechte vorbehalten.

Würzburg 2007.

Dieses Dokument wird bereitgestellt durch den Online-Publikationsserver der Universität Würzburg.

Universitätsbibliothek Würzburg

Am Hubland

97074 Würzburg

Tel.: +49 (0) 931 - 888 - 59 17

Fax: +49 (0) 931 - 888 - 59 70

opus@bibliothek.uni-wuerzburg.de

<http://www.opus-bayern.de/uni-wuerzburg/>

Deckblattgestaltung: Dagmar Rußner-Blank

Das zweisprachige *FinDe*-Korpus

1 Einleitung

Der vorliegende Beitrag stellt das zweisprachige *FinDe*-Korpus vor, das im Rahmen des Forschungsprojekts »Wortbildung und Textanalyse im deutsch-finnischen Kontrast«, von 1995 bis 1997 vom DAAD und von der Akademie von Finnland finanziell unterstützt und für Forschungszwecke erstellt wurde. Das *FinDe*-Korpus enthält deutsche und finnische Quelltexte und ihre jeweiligen Übersetzungen in die andere Sprache.

2 Korpustexte

Das *FinDe*-Korpus enthält die im Folgenden aufgeführten Textpaare, die jeweils mit einer Sigle versehen sind. In den Beiträgen der elektronischen Publikationsreihe »FinDe. Arbeiten mit dem finnisch-deutschen Kontrastkorpus«¹ wird mit diesen Siglen auf die Korpustexte Bezug genommen.

Deutsche Texte und ihre finnischen Übersetzungen

- GGd Grass, Günter (1994): Unkenrufe. Eine Erzählung. Ungekürzte Ausgabe. München: Deutscher Taschenbuch Verlag (= dtv 11864). 301 Seiten. [Erstveröffentlichung: 1992]
- GGf Grass, Günter (1992): Kellosammakon huuto. Kertomus. Suomentanut Oili Suominen. Helsinki: Tammi (= Keltainen kirjasto 260). 254 Seiten.
- JH1d Habermas, Jürgen (1985): Ein Interview mit der *New Left Review*. In: Habermas, Jürgen: Die Neue Unübersichtlichkeit. Frankfurt am Main: Suhrkamp Verlag (= Kleine Politische Schriften V). S. 213–257.
- JH1f Habermas, Jürgen (1994): Haastattelu New Left Review'n kanssa. In: Habermas, Jürgen: Järki ja kommunikaatio. Tekstejä 1981–1989. 2. uudistettu painos. Valinnut ja suomentanut Jussi Kotkavirta. Tampere: Gaudeamus. S. 7–49.
- JH2d Habermas, Jürgen (1983): Die Philosophie als Platzhalter und Interpret. In: Habermas, Jürgen: Moralbewußtsein und kommunikatives Handeln. Frankfurt am Main: Suhrkamp Taschenbuch Verlag. S. 9–28.
- JH2f Habermas, Jürgen (1994): Filosofia paikanvaraajana ja tulkitsijana. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 50–67.
- JH3d Habermas, Jürgen (1984): Erläuterungen zum Begriff des kommunikativen Handelns. In: Habermas, Jürgen: Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns. Frankfurt am Main: Suhrkamp Verlag. S. 571–606.

¹ WOLF, NORBERT RICHARD / WEGSTEIN, WERNER / JÄNTTI, AHTI / PIITULAINEN, MARJA-LEENA / HYVÄRINEN, IRMA. (Hg.): *FinDe. Arbeiten mit dem finnisch-deutschen Kontrastkorpus*. Würzburg (URL: www.spr.germanistik.uni-wuerzburg.de/finde-korpus).

- JH3f Habermas, Jürgen (1994): Kommunikatiivisen toiminnan käsitteen tarkastelua. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 68–97.
- JH4d Habermas, Jürgen (1983): Diskursethik – Notizen zu einem Begründungsprogramm. In: Habermas, Jürgen: Moralbewußtsein [wie JH2d]. S. 53–125.
- JH4f Habermas, Jürgen (1994): Diskurssietiikka – huomioita perustelevaan ohjelmaan. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 98–164.
- JH5d Habermas, Jürgen (1983): Die Verschlingung von Mythos und Aufklärung. Bemerkungen zur *Dialektik der Aufklärung* – nach einer erneuten Lektüre. In: Bohrer, Karl Heinz (Hg.): Mythos und Moderne. Begriff und Bild einer Rekonstruktion. Frankfurt am Main: Suhrkamp Verlag. S. 405–431.
- JH5f Habermas, Jürgen (1994): Myytin ja valistuksen yhteenkietoutuminen. Merkintöjä *Dialektik der Aufklärung* -teokseen uuden lukemisen jälkeen. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 165–189.
- JH6d Habermas, Jürgen (1985): Die Neue Unübersichtlichkeit. Die Krise des Wohlfahrtsstaates und die Erschöpfung utopischer Energien. In: Habermas, Jürgen: Die Neue Unübersichtlichkeit. Frankfurt am Main: Suhrkamp Verlag (= Kleine Politische Schriften V). S. 141–163.
- JH6f Habermas, Jürgen (1994): Uusi yleiskatsauksettomuus. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 190–209.
- JH7d Habermas, Jürgen (1989): Heidegger – Werk und Weltanschauung. In: Habermas, Jürgen: Texte und Kontexte. Frankfurt am Main: Suhrkamp Taschenbuch Verlag. S. 11–37.
- JH7f Habermas, Jürgen (1994): Martin Heidegger – tuotanto ja maailmankatsomus. In: Habermas, Jürgen: Järki ja kommunikaatio [wie JH1f]. S. 210–239.
- CHd Hein, Christoph (1994): Der Tangospieler. Erzählung. Berlin: Aufbau Taschenbuch Verlag (= AtV 1025). 206 Seiten. [Erstveröffentlichung: 1989]
- CHf Hein, Christoph (1989): Säestäjä. Suomentanut Markku Mannila. Helsinki: Otava. 167 Seiten.
- BSd Strauß, Botho (1994): Niemand anderes. 2. Auflage. München: Deutscher Taschenbuch Verlag (= dtv 11236). 221 Seiten. [Erstveröffentlichung: 1987]
- BSf Strauß, Botho (1990): Ei kukaan muu. Suomentanut Marja Kyrö. Helsinki: Otava. 216 Seiten.

Finnische Texte und ihre deutschen Übersetzungen

- PHf Haavikko, Paavo (1992): Fleurin koulusyksy. Romaani. Helsinki: Art House. 131 Seiten.
- PHd Haavikko, Paavo (1994): Fleurs mittlere Reife. Roman. Aus dem Finnischen von Gisbert Jänicke. Salzburg / Wien: Residenz Verlag. 120 Seiten.
- AIf Idström, Annika (1991): Veljeni Sebastian. Kolmas painos. Porvoo / Helsinki / Juva: WSOY. 180 Seiten. [Erstveröffentlichung: 1985]
- AId Idström, Annika (1993): Mein Bruder Sebastian. Roman. Deutsch von Gabriele Schrey-Vasara. Reinbek bei Hamburg: Rowolth Taschenbuch Verlag (= rororo 1090, neue frau 13345). 234 Seiten.
- RNf Nikula, Riitta (1993): Rakennettu maisema. Suomen arkkitehtuurin vuosisadat. Helsinki: Otava. 160 Seiten.

- RNd Nikula, Riitta (1993): *Bebaute Landschaft. Finnlands Architektur im Überblick*. Helsinki: Otava. 160 Seiten.
- SM1f *Sosiaalihuolto Suomessa* (1993). Helsinki: Sosiaali- ja terveystieteiden ministeriö (= Sosiaali- ja terveystieteiden ministeriön esitteitä 1993: 12). 31 Seiten.
- SM1d *Sozialwesen in Finnland* (1993). Helsinki: Sozial- und Gesundheitsministerium (= Sozial- und Gesundheitsministerium, Broschüren 1993: 12). 32 Seiten.
- SM2f *Terveydenhuolto Suomessa* (1990). Helsinki: Sosiaali- ja terveystieteiden ministeriö (= Sosiaali- ja terveystieteiden ministeriön esitteitä 5/1990). 30 Seiten.
- SM2d *Gesundheitswesen in Finnland* (1990). Helsinki: Sozial- und Gesundheitsministerium. 40 Seiten.
- JTf Tarkka, Jukka (1987): *Ei Stalin eikä Hitler. Suomen turvallisuuspolitiikka toisen maailmansodan aikana*. Helsinki: Otava. 111 Seiten.
- JTd Tarkka, Jukka (1991): *Weder Stalin noch Hitler. Finnland während des zweiten Weltkrieges*. Aus dem Finnischen von Carl-August von Willebrand. Helsinki: Otava. 112 Seiten.
- ATf Tuuri, Antti (1989): *Talvisota. Kertomus. Seitsemäs painos*. Helsinki: Otava. 242 Seiten. [Erstveröffentlichung: 1984]
- ATd Tuuri, Antti (1992): *Winterkrieg. Roman*. Aus dem Finnischen von Peter Uhlmann. Leipzig / Weimar: Gustav Kiepenhauer Verlag. 228 Seiten.

Die Korpus Texte repräsentieren die deutsche und finnische Gegenwartssprache aus der Zeit zwischen 1983 und 1994. Das Korpus ist bilateral angelegt, d. h. zu jedem Quelltext liegt eine übersetzte Entsprechung in der jeweils anderen Sprache vor. Dies erlaubt Untersuchungen in verschiedene Richtungen: Es können die Quellen zwischensprachlich oder einzelsprachlich in Opposition zu den Übersetzungen gesetzt werden. Das Korpus enthält

- belletristische Texte (GGd / GGf, CHd / CHf, BSd / Bdf, PHd / PHf, AId / AIf, ATd / ATf),
- wissenschaftliche Aufsätze (JH2d bis JH7d / JH2f bis JH7f),
- ein Interview (JH1d / JH1f)²,
- Sachprosa (RNd / RNf, JTd / JTf) und
- Informationsbroschüren (SM1d / SM1f, SM2d / SM2f).

Die belletristischen Texte und die Texte der Sachprosa bilden einen Kernbereich des Textkorpus, der in erster Linie als Grundlage für die ersten Untersuchungen in dieser Schriftenreihe dient. Die übrigen Texte gehören dem peripheren Bereich an.

Die elektronische Version des Korpus steht den Projektteilnehmerinnen und -teilnehmern in Dateien zur Verfügung, die mit Hilfe des Programms TUSTEP (Tübinger System von Textverarbeitungs-Programmen, s. dazu z. B. Stahl 1996) erstellt wurden. Die Textpaare des Korpus, die aus dem jeweiligen Original und dessen Übersetzung bestehen, wurden mit programmierten Verarbeitungsskripten automatisch synoptisiert, wobei die synoptische Darstellung auf der Ebene der Textabsätze realisiert wurde. Außer den Strukturkodierungen (für Überschriften, Absätze, Aufzählungen, Listen), die die Synoptisierung ermöglichten, enthalten die Dateien metasprachliche Informationen zu den Redewiedergaben. Auf eine Annotation zur Morphologie oder Semantik wurde zunächst verzichtet. Zum Zweck der Archivierung werden die Original-Dateien in XML-Dateien überführt; somit wird auch eine Weiterverarbeitung mittels

² Die sieben deutschen Habermas-Texte wurden in mehreren Werken veröffentlicht. Die finnischen Übersetzungen sind in einem Band zusammengefasst (s. JHf1).

anderer Programme ermöglicht. Aus Urheberrechtsgründen ist die Verwendung des elektronischen Korpus zur Zeit auf die Mitglieder des Forschungsprojekts beschränkt.

3 Ermittlung der Korpusstatistik

In diesem Abschnitt werden statistische Angaben über das Korpus präsentiert. Die Angabe der Korpusgröße (vgl. 3.3) wird wiedergegeben durch die Zahl der Wortformen, die die einzelnen Texte enthalten. Für die Bestimmung der Korpusgröße waren verschiedene Operationen nötig, damit die Wortzahl einheitlich berechnet werden konnte: Erstens wurde der sogenannte Textkörper definiert, der die Grundlage für alle Berechnungen in den Korpus-texten darstellt. Zweitens musste das Wort als zugrunde liegende Berechnungseinheit definiert werden.

3.1 Textkörper

Den Textkörper, in dem die Wortformen gezählt werden, bildet im Wesentlichen der Textblock selbst. Somit werden also folgende Teile der gedruckten Publikation, die vor oder nach dem Textblock stehen können, nicht dazu gerechnet: Widmungen, Copyright-Vermerke, Titelblätter, Vor- und Nachwörter, Inhaltsverzeichnisse, Register, Bildnachweise und Verlagstexte. Innerhalb der Textblöcke werden sprachliche Ausdrücke und Einheiten in Fußnoten, in Tabellen, in Erläuterungen zu Tabellen und Abbildungen sowie in Bildunterschriften bei der Bestimmung der Korpusgröße nicht berücksichtigt. Außerdem bleiben Ordnungs- bzw. Aufzählungselemente wie (1) oder (a) und die Angabe des Sprecherwechsels in JH1d / JH1f außer Acht. Textstruktureinheiten wie Überschriften und Zwischenüberschriften dagegen zählen zum Textkörper.

3.2 Wortdefinition

In der ganzseitigen Tabelle »Statistische Angaben zum *FinDe*-Korpus« auf S. 6 wird die Größe des Korpus dokumentiert. Die Angaben beziehen sich auf die Zahl der Wortformen, die in den definierten Textkörpern der einzelnen Werke auftreten. Als Wortform wird jede grafische Einheit definiert, die zwischen Spatien bzw. Satzzeichen steht und die sich nicht nur aus einem oder mehreren Satzzeichen (wie z. B. Gedankenstrich oder Auslassungspunkten) zusammensetzt. Zu solchen Einheiten zählen nicht nur Wortformen mit einer Buchstabenstruktur, sondern auch alle Zahlen in numerischer Form und Sonderzeichen wie & und %. Typografische Kennzeichen wie Großschreibung, Kursivierung und andere Schriftauszeichnungen bleiben unberücksichtigt.

Die vorgenommene Definition hat zur Folge, dass homonyme Wortformen (z. B. *Arm* und *arm*) unter einem Stichwort zusammengefasst werden. Die Definition ergibt auch, dass die verschiedenen flektierten Formen eines Lexems und Präelemente trennbarer Verben, die in einer Distanzstellung von der Basis vorkommen, in der Statistik einzeln gewertet werden. Eine weitere Folgeerscheinung ist, dass Bestandteile sogenannter Bindestrichkomposita (z. B. *Haupt- und Nebengasse*, GGd, 30.14) jeweils als eigenständige Wortform gezählt werden. Des

Weiteren werden Abkürzungen mit einem Punkt als eine Wortform gewertet. So besteht also die Abkürzung z. B. aus zwei Wortformen. Abkürzungen ohne Abkürzungspunkt werden als eine Wortform betrachtet. Enthält ein Name einen abgekürzten Bestandteil, wie beispielsweise *St. Petersburg* (RNd 69.24), gilt der ganze Name als ein Wort.

Von dem erläuterten Prinzip der Wortdefinition musste in folgendem Punkt abgewichen werden, der eine besondere Handhabung forderte: Im Finnischen sind ein Doppelpunkt und ein Apostroph in den Fällen nicht als Satzzeichen zu verstehen, in denen unmittelbar nach ihnen ein Flexionsmorphem folgt. Vgl.:

DDR:ssä (GGf, 50.8)
[DDR.INESSIVENDUNG]
'in der DDR'

Foucault'sta (JH1f, 13.1)
[Foucault.ELATIVENDUNG]
'von Foucault'

Derartige Einheiten stellen also jeweils eine Wortform dar.

3.3 Statistische Angaben

Die Tabelle »Statistische Angaben zum *FinDe*-Korpus« vermittelt einen Überblick über die Größe der Originaltexte und deren Übersetzungen im Einzelnen sowie über das *FinDe*-Korpus insgesamt. Zu jedem Werk werden paarweise die Angaben zunächst für die deutsche Version, darunter die für die finnische aufgeführt. Zur leichteren Orientierung ist das jeweilige Original links neben der Werksigle mit einem Punkt markiert. Die Tabelle gliedert sich vertikal in zwei Bereiche. Im oberen Teil stehen die Zahlen zu den deutschen Originaltexten und ihren finnischen Übersetzungen, im unteren die Zahlen für die finnischen Quellen und deren Übersetzungen ins Deutsche. Jeder Teilbereich wird abgeschlossen durch die Angabe der Summen der jeweiligen Wortformen und der anderen damit zusammenhängenden Zahlen. Am Tabellenende befinden sich die Summen der Wortformen aller deutschen Texte (447.098) sowie aller finnischen (344.073). Somit besteht das *FinDe*-Korpus insgesamt aus 791.171 Wortformen.

Die in Spalte (1) aufgeführten Zahlen geben die Anzahl der Wortformen wieder, die als *Tokens* bezeichnet werden. Daraus geht beispielsweise hervor, dass unter allen deutschsprachigen Quelltexten GGd mit 60.441 der längste, JH2d mit 5.702 der kürzeste Text ist. Unter den finnischen Übersetzungen finden sich dieselben Größenrelationen wieder: GGf stellt den längsten, JH2f den kürzesten Einzeltext dar. Anhand der numerischen Daten wird ersichtlich, dass die finnischen Texte durchweg weniger *Tokens* aufweisen als die deutschen. Diese zwischensprachliche Differenz wird in Spalte (6) prozentual angegeben. Die Daten dieser Spalte sind folgendermaßen zu lesen: Die 40.907 Wortformen in BSf entsprechen 81,0 % von 50.518 in BSd. Insgesamt liegt der Anteil der Wortformen in den finnischen Übersetzungen im Vergleich zu den deutschen Quelltexten zwischen 72,2 % und 83,6 %. Den 232.471 Wortformen der deutschen Originaltexte stehen 185.113 in den finnischen Übersetzungen gegenüber, was 79,6 % entspricht.

Statistische Angaben zum <i>FinDe</i> -Korpus						
	(1) <i>Tokens</i>	(2) <i>Types</i>	(3) <i>Singles</i>	(4) <i>Tokens / Types- Relation</i>	(5) <i>Singles- Anteil</i>	(6) <i>Fin- / De- Anteil</i>
· GGd	60.411	13.320	8.603	4,5 ×	64,6 %	
· GGf	50.481	18.674	13.685	2,7 ×	73,3 %	83,6 %
· JH1d	13.765	4.271	2.944	3,2 ×	68,9 %	
· JH1f	10.469	5.455	4.176	1,9 ×	76,6 %	76,1 %
· JH2d	5.702	1.962	1.348	2,9 ×	68,7 %	
· JH2f	4.216	2.399	1.873	1,8 ×	78,1 %	74,0 %
· JH3d	9.491	2.488	1.489	3,8 ×	59,8 %	
· JH3f	7.243	3.200	2.219	2,3 ×	69,3 %	76,3 %
· JH4d	20.706	4.599	2.650	4,5 ×	57,6 %	
· JH4f	15.510	5.869	3.937	2,6 ×	67,1 %	74,9 %
· JH5d	7.826	2.558	1.792	3,1 ×	70,1 %	
· JH5f	5.910	3.110	2.429	1,9 ×	78,1 %	75,6 %
· JH6d	6.182	2.371	1.720	2,6 ×	72,5 %	
· JH6f	4.614	2.797	2.228	1,6 ×	79,7 %	74,7 %
· JH7d	8.945	2.908	1.972	3,1 ×	67,8 %	
· JH7f	6.457	3.493	2.720	1,8 ×	77,9 %	72,2 %
· CHd	48.925	7.012	4.139	7,0 ×	59,0 %	
· CHf	39.306	10.859	7.268	3,6 ×	66,9 %	80,3 %
· BSd	50.518	11.289	7.741	4,5 ×	68,6 %	
· BSf	40.907	15.578	11.764	2,6 ×	75,5 %	81,0 %
· De-Originale	232.471	32.783	19.602	7,1 ×	59,8 %	
· Übers. ins Fin	185.113	49.580	33.616	3,7 ×	67,8 %	79,6 %
· PHd	31.047	6.226	4.080	5,0 ×	65,5 %	136,4 %
· PHf	22.774	8.300	6.187	2,7 ×	74,5 %	
· AId	47.892	7.940	4.758	6,0 ×	59,9 %	138,7 %
· AIf	34.526	11.402	7.753	3,0 ×	68,0 %	
· RNd	43.971	10.144	6.343	4,3 ×	62,5 %	134,1 %
· RNf	32.784	14.722	10.581	2,2 ×	71,9 %	
· SM1d	8.113	2.107	1.277	3,9 ×	60,6 %	151,6 %
· SM1f	5.353	2.712	2.004	2,0 ×	73,9 %	
· SM2d	8.607	2.305	1.381	3,7 ×	59,9 %	171,5 %
· SM2f	5.022	2.633	1.878	1,9 ×	71,3 %	
· JTd	21.733	4.881	3.011	4,5 ×	61,7 %	141,5 %
· JTf	15.362	6.294	4.493	2,4 ×	71,4 %	
· ATd	53.264	6.038	3.133	8,8 ×	51,9 %	123,5 %
· ATf	43.139	9.845	6.194	4,4 ×	62,9 %	
· Übers. ins De	214.627	26.565	15.191	8,1 ×	57,2 %	135,0 %
· Fin-Originale	158.960	42.510	28.195	3,7 ×	66,3 %	
· Gesamt De	447.098	49.453	28.341	9,0 ×	57,3 %	
· Gesamt Fin	344.073	79.432	52.013	4,3 ×	65,5 %	
· Gesamt	791.171					

Anhand der Daten (s. den unteren Tabellenteil) lässt sich auch beobachten, dass sich dieselben *Token*-Verhältnisse in der umgekehrten Richtung wiederholen: Die deutschsprachigen Übersetzungen weisen ausnahmslos mehr *Tokens* auf als ihre finnischen Originale, wobei dieselben proportionellen Relationen weitgehend beibehalten werden. Der untere Tabellenteil verdeutlicht, dass ATf der längste finnische Quelltext ist; auch dessen Übersetzung ist unter den deutschen Wiedergaben der umfangreichste. Es fällt allerdings auf, dass dem kürzesten Text von SM2f nicht die kürzeste Übersetzung gegenübersteht. Der Grund dafür liegt in den zusätzlichen erklärenden Texteschüben für die deutschen Leserinnen und Leser. Deshalb weist die Übersetzung in Bezug auf die Anzahl der Wortformen 171,5 % des finnischen Originals auf.

Die durchgehende Differenz in der *Token*-Zahl ist auf sprachtypologische Unterschiede zurückzuführen. Im Vergleich zum Deutschen stellt das Finnische eine agglutinierende Sprache mit einem synthetischen Charakter dar (vgl. dazu z. B. Karlsson 1984 und Fromm 1982), was die geringere Anzahl von *Tokens* in den finnischen Texten bedingt. Zur Veranschaulichung seien hier einige Gegebenheiten aufgeführt: Erstens ist das Finnische eine artikellose Sprache (vgl. *talo* 'ein Haus, das Haus'). Zweitens werden viele grammatische Relationen, die im Deutschen durch frei auftretende Morpheme bezeichnet werden, im Finnischen mit gebundenen Morphemen ausgedrückt. Beispielsweise verfügt es über fünfzehn Kasus, den deutschen Possessivpronomina entsprechen finnische Possessivsuffixe (*talo-ni* 'mein Haus'), klitische Partikeln sind eine frequente Erscheinung (*talo-ni-kin* 'auch mein Haus') und die Morphologie der infiniten Verbformen weist einen Reichtum mit bis zu fünf Infinitivklassen und deren verschiedenen Kasusformen auf. Des Weiteren werden das Passiv, das Reflexiv und die Modi durch Endungen signalisiert, vgl.:

Talo myy-dä-än.

[Haus.NOMINATIV verkaufen.PASSIV]

'Das Haus wird verkauft.'

Minä pese-ydy-n.

[Ich.NOMINATIV waschen.REFLEXIV.1SG]

'Ich wasche mich.'

Jo-isi-n mielelläni kahvia.

[trinken.KONDITIONALIS.1SG gern Kaffee.PARTITIV]

'Ich würde gern Kaffee trinken.'

Die geringere Anzahl der Wortformen im finnischen Korpusteil könnte den Eindruck erwecken, dass die Texte weniger bedruckten Raum einnehmen als die deutschen. Dies ist jedoch nicht der Fall, weil die synthetische Formenbildung längere Wortformen zur Folge hat. So besteht z. B. Günter Grass' »Unkenrufe« aus insgesamt 366.285 Buchstaben und Satzzeichen. Die Übersetzung »Kellosammokonhuuto« enthält 381.182. In diesen Texten bestehen somit die deutschen grafischen Einheiten aus durchschnittlich 6,06 und die finnischen aus 7,55 Elementen.

Spalte (2) informiert über die Anzahl der unterschiedlichen Wortformen, der *Types*. Obwohl die finnischen Texte eine geringere Anzahl an *Tokens* aufweisen, enthalten sie eine höhere Anzahl an *Types*. Dies ist u.a. durch die bereits erläuterte Formenvielfalt bedingt, die durch die gebundenen Morpheme hervorgerufen wird.

Tokens und *Types* werden in Spalte (4) zueinander in Relation gesetzt. Diese Spalte gibt an, wie häufig eine Wortform durchschnittlich in dem jeweiligen Text verwendet wird. Der kleinere Quotient für die finnischen Texte verdeutlicht die Vielfalt der flektierten Formen.

Spalte (3) verzeichnet die Anzahl der Wortformen, die nur ein einziges Mal belegt sind. Wird die Zahl der *Singles* mit den *Tokens* verglichen, ergibt sich, dass die *Tokens* allein keinen Rückschluss auf die Anzahl der *Singles* erlauben: Beispielsweise sind die Texte BSd und CHd mit 50.518 bzw. 48.925 ungefähr von gleichem Umfang. Die Zahl der *Singles* differiert aber stark: BSd enthält insgesamt 7.741 *Singles*, CHd jedoch nur gut die Hälfte davon, nämlich 4.139. Die Anzahl der *Singles* muss vielmehr in Relation zu den *Types* gesehen werden. Von den 11.289 *Types* in BSd sind 68,6 % Einzelbelege einer Wortform, von 7.012 *Types* in CHd sind es 59,0 %.

Der Prozentanteil an *Singles* nimmt allerdings nicht signifikant ab, wenn immer größere Textmengen analysiert werden. Denn alle deutschen Quelltexte, die 232.471 *Tokens* 32.783 *Types* aufweisen, enthalten 19.602 *Singles*, was 59,8 % der *Types* entspricht. Werden die Übersetzungen ins Deutsche hinzugenommen, weist die fast doppelte Zahl der *Tokens*, nämlich 447.098, nur 49.453 *Types* auf, von denen immer noch 57,3 %, also 28.341, *Singles* sind.

4 Literaturverzeichnis

FROMM, HANS (1982): Finnische Grammatik. Heidelberg.

KARLSSON, FRED (1984): Finnische Grammatik. Autorisierte Übertragung aus dem Finnischen von KARL-HEINZ RABE. Hamburg.

STAHL, PETER (1996): TUSTEP für Einsteiger. Eine Einführung in das »Tübinger System von Textverarbeitungs-Programmen«. Würzburg.

WOLF, NORBERT RICHARD / WEGSTEIN, WERNER / JÄNTTI, AHTI / PIITULAINEN, MARJA-LEENA / HYVÄRINEN, IRMA. (Hg.): FinDe. Arbeiten mit dem finnisch-deutschen Kontrastkorpus. Würzburg (URL: www.spr.germanistik.uni-wuerzburg.de/finde-korpus).

Hinweis: Der vorliegende Beitrag war bereits im Jahr 1998 abgeschlossen.

