

MANNHEIMER KORPUS UND URHEBERRECHT

Die Einbeziehung zeitgenössischer digitalisierter Texte in die computergespeicherten Korpora des IDS und ihre juristischen Grundlagen

von *Irmtraud Jüttner*

Die Idee, mit Hilfe von Computern große Textmengen nach unterschiedlichen Gesichtspunkten für wissenschaftliche – speziell sprachwissenschaftliche – Zwecke auszuwerten, ist so alt wie die Computer selbst. Das Institut für Deutsche Sprache begann bereits Mitte der Sechzigerjahre mit dem Aufbau computergestützter Textkorpora zum Gegenwartsdeutsch, um der germanistischen Wissenschaft verlässliche Hilfsmittel zur Beschreibung sprachlicher Phänomene der zeitgenössischen Gemeinsprache, bestimmter Fachsprachen, bestimmter Autoren in lexikografischem, stilistischem, sprachstatistischem Zusammenhang an die Hand zu geben und zuverlässige Aussagen über Aspekte der gegenwärtigen Sprachentwicklung zu ermöglichen. In den letzten Jahren steigen mit der weltweiten Verknüpfung von Rechnern durch das sich schnell entwickelnde Kommunikationswerkzeug Internet die wissenschaftlichen Nutzungsmöglichkeiten wie die Ansprüche an die linguistische Datenverarbeitung.

Verlage und Printmedien tragen der modernen Entwicklung zunehmend Rechnung, indem sie die Produkte ihres Wirkens – ihre Texte – auch in digitalisierter Form aufbereiten. Das IDS ist bemüht, diese digitalisierten Texte in den weiteren Ausbau, in die Aktualisierung seiner Textkorpora einzubeziehen und sie für die wissenschaftliche, nichtkommerzielle Nutzung fruchtbar zu machen. Dies erfordert eine verstärkte Zusammenarbeit der Arbeitsgruppe für Korpustechnologie (Abteilung Lexik) mit Autoren, mit Verlagen, mit Zeitungsredaktionen und die Berücksichtigung all der neuen juristischen Fundamente, die den Urheberschutz auch im Internet-Zeitalter gewährleisten sollen.

Der Umfang der Texte des Mannheimer Korpus hat sich seit Juli 1992 von ca. 28 Millionen Textwörtern auf 778 Millionen Textwörter erhöht. Diese gewaltige Textsammlung

steht allerdings nur den Mitarbeitern des IDS, seinen Gästen, die das Institut zu Forschungsaufenthalten nutzen, und seinen wissenschaftlichen Partnern für ausschließlich sprachwissenschaftliche Zwecke zur Verfügung, etwa zur automatischen Suche nach Belegen für die Verwendung bestimmter Wörter und Wortkombinationen sowie zu sprachstatistischen Analysen mit Hilfe des Computersystems COSMAS (Corpus Storage, Maintenance and Access System). Der linguistischen Fachwelt, die über das Internet – über die Internetadresse <http://www.ids-mannheim.de/> – mit dem Institut verbunden ist, können hingegen wegen urheberrechtlicher Restriktionen nur die öffentlich zugänglichen Texte des Korpus im Umfang von ca. 379 Millionen Textwörtern für die automatische Belegrecherche angeboten werden.

Welcherart sind nun die gesetzlichen Grundlagen für den Umgang mit unseren computergespeicherten Texten? Von zentraler Bedeutung ist das »Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz)« (UrhG) vom 9. September 1965, publiziert im Bundesgesetzblatt (BGBl) I, S. 1273, seither durch mehrere Veränderungen – insbesondere Ergänzungen – aktualisiert. Nach diesem Gesetz beginnt der Schutz von künstlerischen und wissenschaftlichen Leistungen, die eine gewisse Originalität und Kreativität repräsentieren, mit der Schöpfung des Werkes und endet 70 Jahre *post mortem auctoris* – nach dem Tod des Urhebers. Zu den geschützten Werken der Literatur, Wissenschaft und Kunst gehören »Sprachwerke, wie Schriftwerke, Reden und Computerprogramme« (§ 2: »Geschützte Werke«, Abs. 1, Nr. 1, UrhG in der Fassung vom 7. März 1990, BGBl I, S. 422). Gesetze, Verordnungen, amtliche Erlasse und Bekanntmachungen sowie Entscheidungen und amtlich verfasste Leitsätze zu Entscheidungen genießen keinen urheberrechtlichen Schutz. Zu den frei verwendbaren Werken zählen unter anderem Texte der als

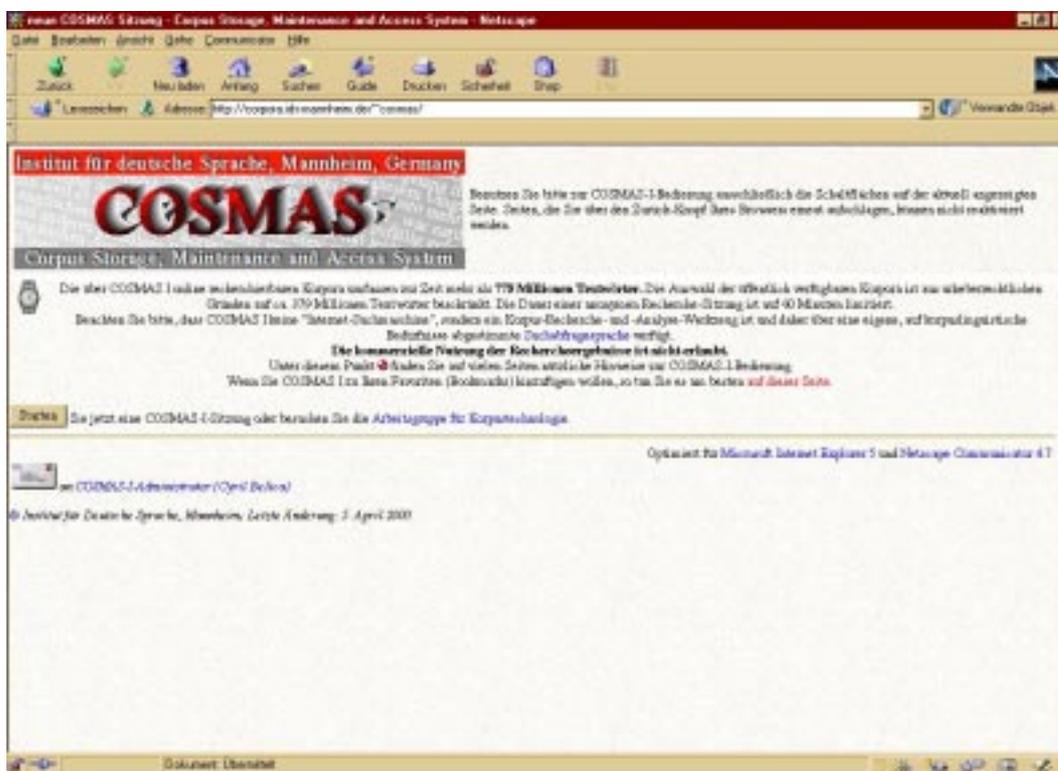
Körperschaften des öffentlichen Rechts organisierten Kirchen.

Alle Publikationen, die durch das UrhG geschützt sind, verweisen auf den Urheberschutz durch einen eigens unterhalb des Impressums eingedruckten Copyright-Vermerk mit folgendem Wortlaut: »Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verbreitung in elektronischen Systemen.«

In Bezug auf die Weitergabe von Recherche-Ergebnissen im Mannheimer Korpus ist die bisherige Gesetzgebung noch nicht eindeutig. In § 53 des UrhG (»Vervielfältigung

gesellschaft gerecht wird. Eine entsprechende Richtlinie, die der Europäische Rat im Sommer 2000 verabschiedet, soll vom Europäischen Parlament Ende dieses oder Anfang des nächsten Jahres beschlossen und dann innerhalb von 18 Monaten von den EU-Mitgliedsstaaten in nationales Recht umgesetzt werden. »Kernpunkt der Vorlage ist der Versuch, einen fairen Ausgleich zwischen den Interessen der Rechteinhaber und anderer Gruppen wie Konsumenten und Internet-Service-Provider zu schaffen.« Es ist zu hoffen, dass durch die Neuregelung des Urheberrechts auch die unsere Arbeit betreffenden Unklarheiten in der bisherigen Gesetzgebung beseitigt werden.

Im IDS erfolgte im Zusammenhang mit der Fertigstellung der »Wende-Korpora« – jenes Teils des Mannheimer Korpus, der spezielle Texte zum Sprachgebrauch in der Zeit der politischen Wende 1989/90 sammelt – im Juli 1992 eine



gen zum privaten und sonstigen Gebrauch« in der Fassung vom 24. Juni 1985, BGBl I, S. 1137) wird von Ton- und Bildträgern, von Vervielfältigungen im traditionellen Sinne gesprochen, nicht aber von Datenspeicherung und Datenweitergabe, wie sie für uns wesentlich ist. Das Vervielfältigungsgesetz bezieht sich auf Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch, aber nicht auf die speziellen Erfordernisse eines wissenschaftlichen Instituts, nicht auf eine öffentlich betriebene, organisierte Textsammlung, wie sie die Textkorpora des IDS darstellen.

Laut der »Welt« vom 21. Juni 2000, S. WW2, plant die Europäische Union eine Neuregelung des Urheberrechts, die den besonderen Bedingungen der modernen Informations-

Bestandsaufnahme sämtlicher Vereinbarungen, Verträge, Regelungen mit Autoren, Verlagen und Redaktionen, die die zu diesem Zeitpunkt verfügbaren Texte des Mannheimer Korpus betrafen. Ziel der Bestandsaufnahme war 1.) die Vereinheitlichung künftiger Verträge, 2.) ihre Anpassung an modernste urheberrechtliche Standards, 3.) das Erkennen von Unklarheiten und Vertragslücken und 4.) deren Vermeidung in künftigen die Korpusarbeit betreffenden Vereinbarungen. Seit April 1994 nutzt das IDS die Unterstützung von Juristen mit Kenntnissen zum datenverarbeitungsspezifischen Urheberrecht, die dem Institut die Reichweite und die Grenzen von urheberrechtlichen Vorgaben aufzeigen. Konsultationen in Einzelfragen, juristische Vorträge, vor allem aber ein speziell für die Belange eines Forschungsinstituts und seiner Korpusarbeit

angefertigtes Rechtsgutachten zum Urheberrecht sind Formen dieser linguistisch-juristischen Zusammenarbeit.

Das IDS ist verpflichtet, bei der Aufnahme und vor allem nach erfolgter Einspeisung von digitalisierten Texten strikt die urheberrechtlichen Gesetze und juristischen Vorschriften gegenüber den Autoren der Texte, den Verlagen und sonstigen Agenturen, die über die Rechte an den Texten verfügen, einzuhalten. Für die fortlaufende Erweiterung seiner Korpora hat das Institut insbesondere in den letzten zwei Jahren folgende Vorgehensweise praktiziert:

Das Institut für Deutsche Sprache wendet sich zunächst an den Autor, den Urheber des Textes, trägt seine speziellen Wünsche vor und bittet um die Genehmigung, den computerlesbaren Text in das IDS-Rechercheprogramm aufzunehmen. Auf diese Bitte haben wir zustimmende Antworten von den Schriftstellern Martin Walser, Günter Grass, Hans Joachim Schädlich, Klaus Schlesinger, Siegfried Lenz, Christa Wolf, Stefan Heym, Bernhard Schlink und Erich Loest erhalten.

Der zweite Schritt besteht in den Verhandlungen mit den Verlagen, die ebenfalls ihr Einverständnis geben müssen. Es kann durchaus geschehen, dass zwar der Schriftsteller bereit ist, seine Werke dem IDS zur Verfügung zu stellen, der Verleger aber ablehnt.

Der dritte Verhandlungspartner ist dann die Datenträger herstellende Firma oder die Druckerei, mit der die technischen Details wie Textformat, Datenübertragungsformat, Formatierungen, Kodierungen und vor allem Fragen der Passfähigkeit zum IDS-System, aber auch die Kosten für die Herstellung und für eventuelle Bearbeitungen der Daten geklärt werden müssen.

In einer anderen Vorgehensweise wendet sich das IDS direkt an einen Verlag mit einer Anfrage nach ganz konkretem digitalisiertem Datenmaterial, das der Verlag selbst zu seinen Publikationen hergestellt hat. Hier sei der Aufbau-Verlag Berlin erwähnt, der uns die für sprachhistorische Analysen und Studien zum Sprachwandel wichtigen »Tagebücher« von Victor Klemperer (Zeitraum 1918 bis 1959) und andere Publikationen als computerlesbar gespeicherten Text überlässt.

Selbstverständlich verpflichtet sich das IDS gegenüber seinen Vertragspartnern zur Einhaltung des Urheberrechts, die die erwähnten Einschränkungen in der Arbeit mit dem Mannheimer Korpus bedingt. Dazu gehört, dass das Com-

putersystem nur die automatische Suche nach Wort- und Wortverbindungsbelegen, nur sprachstatistische Analysen, nur die Anzeige von Belegtreffern im Zeilen-, Satz- und/oder Absatzkontext erlaubt, dass die Recherche-Ergebnisse nur von einem eng begrenzten Kreis von Wissenschaftlern ausschließlich zu sprachwissenschaftlichen Zwecken benutzt werden dürfen, dass die Texte, auch ausschnittsweise, nicht im Internet zur Verfügung gestellt werden, dass die Texte nicht an Dritte weitergegeben und niemals kommerziell genutzt werden dürfen. Das IDS verpflichtet jeden Nutzer seines Recherchesystems mit einer schriftlichen Vereinbarung, die formulierten Bedingungen anzuerkennen und einzuhalten. Kein Nutzer hat einen Zugriff auf den gesamten Korpus, denn durch das Suchprogramm wird nur eine begrenzte Anzahl von Sätzen als Kontext zum gesuchten Wort zur Verfügung gestellt. Diese Maßnahmen sollen den Missbrauch von elektronischen Texten verhindern.

Der skizzierte rechtliche Hintergrund ist die Ursache dafür, dass das IDS nur ein Viertel der Texte des Mannheimer Korpus zur weltweiten Nutzung im Internet anbieten kann. So dürfen die größeren Tageszeitungen und die Zeitschriften im Korpus – »Berliner Zeitung«, »Frankfurter Allgemeine«, »die tageszeitung«, »Der Spiegel«, »Die Zeit« – nur innerhalb des IDS genutzt werden. Nur die regionale Tageszeitung »Mannheimer Morgen« steht der internationalen Forschung online in der WWW-Version von COSMAS zur Verfügung.

Das IDS bemüht sich auch weiterhin, die Mitarbeit von Autoren und Autorenverbänden, von Verlagen und Redaktionen, von Unternehmen und Stiftungen, von politischen, gemeinnützigen und anderen öffentlichen Institutionen, von Hochschul- und Forschungseinrichtungen für die Erweiterung des Mannheimer Korpus und für die Genehmigung seiner wissenschaftlichen Nutzung zu gewinnen – z.B. baut das Institut derzeit mit Unterstützung des Landes Baden-Württemberg sein Textkorpus weiter aus (Projekt Deutsches Referenzkorpus: www.ids-mannheim.de/dereko/). Ein vertrauensvolles Zusammenwirken mit ihnen allen unter Wahrung der gesetzlichen Vorschriften ist die beste Voraussetzung dafür, dass dieses Korpus von Texten der deutschen Gegenwartssprache als Schatzkammer unserer Sprachwirklichkeit immer reicher, als Instrument der germanistisch-linguistischen Forschung immer fruchtbarer wird.

Die Autorin ist wissenschaftliche Mitarbeiterin am Institut für Deutsche Sprache in Mannheim.