

Shakespeare-Datenbank: auf der Grundlage der Englisch-deutschen Studienausgabe der Dramen Shakespeares

Gabler, Hans Walter; Steppe, Wolfhard

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gabler, H. W., & Steppe, W. (1995). Shakespeare-Datenbank: auf der Grundlage der Englisch-deutschen Studienausgabe der Dramen Shakespeares. *Historical Social Research*, 20(4), 155-159. <https://doi.org/10.12759/hsr.20.1995.4.155-159>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

HUMANITIES COMPUTING

Shakespeare-Datenbank. Auf der Grundlage der Englisch-deutschen Studienausgabe der Dramen Shakespeares

*Hans Walter Gabler, Wolfhard Steppe (München)**

Die Englisch-deutsche Studienausgabe der Dramen William Shakespeares erscheint in Einzelbänden seit 1976. Unter einem deutsch-schweizer Gremium von Hauptherausgebern erarbeiten Bandherausgeber - Anglisten in Deutschland, Österreich und der Schweiz - die Dramen in Einzelausgaben. Von Shakespeares 36 Stücken sind bisher etwas über ein Drittel erschienen. Die Ausgabe versteht sich als wissenschaftlich-philologische Handreichung für Studierende und Theaterleute, und jeder Band folgt dem gleichen Grundschema. Zum englischen Originaltext tritt eine deutsche Prosaübersetzung. Diese will nicht einen spielbaren und auf der Bühne sprechbaren Text, sondern lediglich eine möglichst wörtlich akkurate Verständnisstütze geben. Was die Übersetzung selbst sprachlich nicht zu fassen erlaubt, fließt in reichliche Annotationen ein. Eine die jeweilige Forschungslage referierende Einleitung sowie ein ausführlicher diskursiver Szenenkommentar ergänzen die Zielsetzung der Studienausgabe.

Die philologische Arbeit an Übersetzung und Annotationen kumuliert Wissen und Einsicht zu Shakespeares Sprache und dramatischer Sprachkunst. Was hier von den bisherigen Bandherausgebern erarbeitet worden ist, dient künftigen Bandherausgebern als Materialrepositorium. Der Zugriff darauf soll durch eine Organisation in Datenbankform erleichtert bzw. überhaupt ermöglicht werden. Für die Anlage der Datenbank folgt daraus der primäre Einstieg über die Annotationen. Erst in zweiter Linie erlangen auch die elektronischen Möglichkeiten zur Volltextsuche im englischen und deutschen Text Bedeutung. Die Annotationen bieten einen ständig sich anreichernden Fundus an kommentiertem Vokabular, wie ein gleichfalls sich zunehmend erweiterndes Spektrum an Sinnbegriffen und -bereichen. Über die Nahzielsetzung der Datenbank als Hilfsmittel innerhalb des fortschreitenden Gesamtprojekts hinaus zeichnet

* **Protokoll des 63. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 11. Februar 1995.**

sich die weitere Perspektive ab, daß die Datenbank zu einem Werkzeug der allgemeinen und ganz grundsätzlichen Erschließung der shakespeareschen Sprache und Sprachverwendung wird.

Das Projekt der *Shakespeare-Datenbank* (das unter der Projektleitung von Wolfgang Weiß und Hans Walter Gabler von Wolfhard Steppe eingerichtet und im wesentlichen mit Hilfe von TUSTEP-Programmen durchgeführt wurde) erwuchs aus dem Wunsch, den Vorrat an Auskünften zum Vokabular Shakespeares und zu den damit bezeichneten Realien und Ideen, der sich im Fußnotenapparat der Studienausgabe angesammelt hat, unter verschiedenen Aspekten besser zugänglich zu machen. Dazu war es zunächst nötig, den relevanten Datenbestand der publizierten Bände der Ausgabe - seinerzeit 11 an der Zahl - in Dateiform zu überführen. Ein einziges Stück, dessen Druckausgabe in früheren Jahren vom Bearbeiter der Datenbank über EDV erstellt worden war, konnte aufgrund der dabei verwendeten Datei übernommen werden. Der relevante Inhalt der restlichen 10 Bände - deutsche Übersetzung und englischer Werktext (im Druck auf gegenüberstehenden Seiten synoptisch ausgebreitet), sowie der am Fuß der Seiten angebrachte Anmerkungsapparat - mußte über Scanner und Zeichenerkennungsprogramme (KDEM, OPTOPUS) eingelesen werden; die technischen Probleme, die dabei aufzutreten pflegen, haben mehr oder minder intensive Korrekturgänge erforderlich gemacht. Das Ergebnis dieses ersten Arbeitsgangs war eine sogenannte »Buchdatei«, in welcher die auf den Buchseiten der Ausgabe abgedruckte Information *paginatim* und *lineatim* exakt nachgebildet war, sie war so eingerichtet, daß die verschiedenen Bestandteile der Ausgabe - englischer Text, deutsche Übersetzung, Fußnotenapparat - grundsätzlich per Programm zu isolieren waren. Erst danach konnte die Aufbereitung der Daten für den beabsichtigten Verwendungszweck in Angriff genommen werden.

In einem ersten Schritt mußten die Fußnoten mit den Textpassagen verknüpft werden, auf die sie sich beziehen. In der Druckausgabe sind die (pro Szene jeweils neu durchgezählten) Fußnoten der *deutschen* Übersetzung zugeordnet. Besprochen werden indessen durchweg Zitate aus dem *englischen* Text; wo mehr als zwei Wörter besprochen werden, wird das in die Fußnote übernommene Lemma in der Regel durch Pünktchen abgekürzt. So war es erforderlich, diese verkürzten Informationen durch genuine Passagen aus dem vollen Werktext zu ergänzen. Die Pragmatik der Fußnoten-Kommentierung bringt es mit sich, daß einerseits einzelne Wörter, andererseits aber auch längere Bereiche des Textes zu diskutieren sind; dies führt zu Komplikationen: besprochene Einzelwörter können innerhalb besprochener Bereiche stecken; besprochene Bereiche können ineinander geschachtelt sein oder auch einander überlappen. So waren passende Markierungen in den englischen Text einzubringen, die eine saubere Extraktion und Zuordnung der betreffenden Textteile zu gewährleisten hatten. Nachdem die Anordnung der Fußnoten in der Studienausgabe durch den Textverlauf der deutschen Übersetzung bestimmt ist, kommt es öfter vor, daß

die Anordnung der Fußnoten nicht mit der Abfolge ihrer Lemmata im englischen Text übereinstimmt; die Markierungen im englischen Text mußten demnach so konzipiert werden, daß das Programm, welches jeweils den Wortlaut des Lemmas aus dem englischen Text zu extrahieren und mit dem Wortlaut der zugehörigen Fußnote zu vereinigen hatte, auch diese Komplikation erkennen und bewältigen konnte.

Zu berücksichtigen war an dieser Stelle noch ein weiterer Gesichtspunkt: Der Benutzer benötigt, um sich im Œuvre Shakespeares zurechtzufinden, ein universelles Referenzsystem, und die extrahierten Lemmata mußten selbstverständlich auch mit Referenzangaben, die sich an einem solchen System orientierten, ausgerüstet werden. Für die Studienausgabe ist vereinbart worden, daß sie sich nach der Zählung des amerikanischen *Pelican Shakespeare* zu richten hat, die einen verhältnismäßig schmalen Satzspiegel aufweist und darum die Prosa-Stellen des englischen Texts in einem anderen Zeilenfall darbietet als etwa die Studienausgabe. Zwar weichen, was die Vers-Passagen betrifft, die Zählungen der vielen miteinander konkurrierenden Shakespeare-Ausgaben kaum voneinander ab; die zahlreichen Prosa-Szenen in Shakespeares Stücken machen es jedoch nötig, sich auf eine bestimmte Zählung festzulegen. Die Festlegung auf den *Pelican Shakespeare* erfüllt in technischer Hinsicht den Zweck, daß im Rahmen der Studienausgabe mit genauer Stellenangabe auch auf solche Stücke verwiesen werden kann, die in der Studienausgabe noch gar nicht vorliegen. Für den optischen Befund der Studienausgabe ergibt sich die Konsequenz, daß die mitgedruckte Zählung in den Prosa-Passagen nicht unbedingt mit der Anzahl der tatsächlich abgedruckten Zeilen übereinstimmt: wenn fünf *Pelican-Zeilen* bereits in vier Zeilen der Studienausgabe Platz finden, dann bleibt eine Position im Zeilennummern-System unbesetzt (Für die deutsche Übersetzung ergibt sich auch das entgegengesetzte Problem: für fünf Zeilen Shakespeare-Prosa muß sie zuweilen sechs oder sieben deutsche Zeilen aufwenden - wohingegen fünf englische Verse auch einmal in drei oder vier Zeilen deutscher Übersetzungs-Prosa Platz finden.) So war es nötig, für den englischen und dann auch für den deutschen Text der Ausgabe Zählprogramme zu entwickeln, welche einerseits den tatsächlichen Textbestand sichten, andererseits aber stets von neuem auf die in der Ausgabe mitgedruckte Zählung reagieren. Da nicht gesprochener Text (Bühnenanweisungen) in der Zeilenzählung nicht berücksichtigt werden darf und da Verse, deren Vortrag auf mehrere Sprecher verteilt ist, auf eine einzige Nummer zu setzen sind, mußten zudem die betreffenden Stellen in der Buchdatei dementsprechend gekennzeichnet werden, um dem Zählprogramm den Weg zu ebnen. Um den Gesamt-Bestand des Shakespeare-Œuvres sinnvoll in einer einzigen Zählung zusammenfassen zu können, wurden (unter Ausnutzung der TUSTEP-Standard-Numerierung) fünfstellige Seitennummern und dreistellige Zeilennummern vergeben; die ersten beiden Stellen der Seitennummer bezeichnen das Stück (wobei die Stücke, in Anlehnung an die Ergebnisse des *Oxford Shakespeare* von Gary Taylor und

Stanley Wells, in chronologischer Abfolge gezählt werden); die dritte Stelle bezeichnet den Akt, die beiden letzten die Szene. Die Ergebnisse der so durchgeführten Zählung konnten sodann den Zitaten, die aus dem englischen Text zu extrahieren und als Lemmata an ihre jeweiligen Fußnoten heranzuführen waren, beigegeben werden.

Die Fußnoten und ihre vervollständigten und mit Referenzen ausgerüsteten Lemmata wurden somit jeweils zu einem Datenbankeintrag vereinigt und in einer eigenen Datei gesammelt, die als eigentliche »Datenbankdatei« (im Gegensatz zur »Buchdatei«) zur Materialbasis für ein Abfrage-Programm auszubauen war, das in Abstimmung mit der Aufbereitung des Datenmaterials in Gestalt eines TUSTEP-Makros (5429 Sätze) entwickelt wurde. Da gewünscht wurde, daß man beim Studium eines Datenbankeintrags über die zitierten (und nunmehr im Wortlaut vervollständigten) Lemmata hinaus noch einen etwas größeren Kontext der besprochenen Stelle sowohl im Original wie in der Übersetzung einsehen können sollte, wurden die Datenbankeinträge mit einer dem jeweiligen Fall angepaßten Reihe von Platzhaltern ausgerüstet, die bei der Datenbankabfrage die Option ermöglichen, wahlweise einen kürzeren oder längeren Kontextabschnitt aus dem englischen und/oder deutschen Text heranzuholen. Diese Platzhalter wurden aus der Referenz des Lemmas extrapoliert; sie erfassen die numerischen Positionen in den Textdateien, die durch das Zitat selbst abgedeckt sind, sowie noch etliche Zusatz-Positionen, die den Kontext erweitern. Sie wurden so gestaltet, daß sie als »Kürzel« im Sinne des TUSTEP-Kommandos »EINFUEGE« dienen konnten. Es erschien zweckmäßig, dieses Programm einzusetzen, weil es die Möglichkeit bietet, in *einem* Programmlauf das gleiche Kürzel beliebig oft auszuwerten; da bei der Abfrage der *Shakespeare-Datenbank* jederzeit damit zu rechnen ist, daß die zu verschiedenen Datenbankeinträgen angeforderten Kontexte einander überschneiden, kann auf diese Möglichkeit nicht verzichtet werden. - Um diese Kürzel aktivieren zu können, mußten zwei Dateien mit dem vollen englischen und deutschen Text bereitgestellt werden, in denen die einzelnen Textzeilen mit den gleichen Kürzeln markiert waren. Formal sind die hierfür verwendeten Kürzel nichts anderes als die am Anfang jeder Zeile in den Text eingetragenen Seiten-Zeilennummern der einzelnen Datensätze.

Um eine zielgerichtete Datenbankabfrage zu ermöglichen, wurde versucht, ein Angebot zu erarbeiten, das die in den Anmerkungen akkumulierten Informationen dem Benutzer in sinnvoller Weise aufschließt. Dabei wurde ein doppelter Weg beschritten: einerseits wurde das in einer Fußnote besprochene *Vokabelgut* erfaßt, und zum zweiten wurde versucht, in stichwortartigen *Begriffen* festzuhalten, welche Themen die Fußnote anschneidet. Daß es auch Einträge gibt, die sich so punktuell zu einer Textstelle äußern, daß sie keine weiterführenden Auskünfte abzuwerfen scheinen, wurde in Kauf genommen. Grundsätzlich aber war es möglich, die Mehrzahl der Datenbankeinträge soweit auszurüsten, daß das Abfrage-Makro entweder über bestimmte *Vokabeln* und/oder

über bestimmte *Begriffe* auf sie zugreifen kann. Dies kann so vor sich gehen, daß der Benutzer zunächst über eine Angebots-Liste der *Vokabeln* oder der *Begriffe*, zu denen die Datenbank triftige Auskünfte enthält, in die Abfrage einsteigt; mittels eines im Makro einkomponierten EDITOR-Laufes wird ihm die gewünschte Liste (Vokabeln: 8450; Stichwörter 1344) auf den Bildschirm gebracht; durch »Ankreuzen« einzelner Vokabeln (oder Begriffe) kann er sodann seine Wünsche präzisieren, worauf die volle Information (wahlweise mit englischem und/oder deutschem Kontext) herangeschafft wird. Dies geschieht in einem weiteren EDITOR-Lauf mit Hilfe der Suchanweisungen für strukturierte Daten, wobei der Benutzer wiederum die Möglichkeit hat, einzelne Einträge für eine anschließende Ausgabe (auf Papier, in eine Datei) auszuwählen. Es wird auch die Möglichkeit angeboten, Vokabeln und/oder Stichwörter in beliebiger Kombination (als Zusatzbedingung; als Ausschlußbedingung; Einstieg über Vokabeln, Umsteigen auf die mit solchen Vokabeln gekoppelten Begriffe) ohne Rückgriff auf die vorbereiteten Listen direkt einzugeben, doch hat es sich gezeigt, daß diese technischen Möglichkeiten angesichts der Eigenart der vorhandenen Informationen nicht immer zu besonders sinnvollen Ergebnissen führen. Es ist versucht worden, dem nicht an TUSTEP gewohnten Benutzer durch Hilfstexte den Weg zu ebnen und auf eventuelle Fehlbedienungen mit informativen Meldungen zu antworten.

Eine neue Dimension hat das Projekt dadurch gewonnen, daß sich zusätzlich noch die Möglichkeit eröffnet hat, den erarbeiteten Datenvorrat in eine CD-ROM-Version zu überführen, die sich die Strukturen der *hypertext-Technik* zunutze macht. Es war möglich, die unter TUSTEP erstellten und für das TUSTEP-Makro entworfenen Dateien (Datenbank: 125860 Sätze; englischer Text: 30401 Sätze; deutscher Text: 30440 Sätze) mit geringfügigen Adaptationen an die Mitarbeiter der Firma MAKROLOG (Wiesbaden) zu übergeben, die hierfür Partner des Projekts ist. Eine Test-Version konnte von Herrn Dr. Mohn in Tübingen bereits vorgestellt werden, wobei einige der erweiterten Möglichkeiten dieses Mediums sichtbar wurden: das bequeme Blättern in den Listen; die Möglichkeit, über *hypertext-Marken* von einem Eintrag aus über andere dort verankerte *Vokabeln* oder *Stichwörter* zu weiteren Einträgen vorzudringen; die Möglichkeit, jederzeit die von einem Eintrag berührten Stellen im englischen (oder deutschen) Text zuzuschalten (und eventuell dann dort weiterzumachen); die Möglichkeit, englischen und deutschen Text parallel zu lesen; die Möglichkeit, rasch zu erfahren, ob man zu einem beliebigen Wort im englischen Text, zu dem es keine eigene Fußnote gibt, dennoch anderweitig im Corpus der Studienausgabe einschlägige Informationen bekommen kann etc. Die Arbeiten von MAKROLOG sind noch nicht abgeschlossen.