

Automatische Transformation bei der Edition und Erschließung jiddischer Literatur des 14.-18. Jahrhunderts

Rapp, Andrea; Becker, Ursula

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:

Rapp, A., & Becker, U. (1993). Automatische Transformation bei der Edition und Erschließung jiddischer Literatur des 14.-18. Jahrhunderts. *Historical Social Research*, 18(3), 99-105. <https://doi.org/10.12759/hsr.18.1993.3.99-105>

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

Automatische Transformation von Graphiesystemen bei der Edition und Erschließung jiddischer Literatur des 14.-18. Jahrhunderts

*Andrea Kapp; Ursula Becker (Trier)**

Wenn man den Stellenwert ermesen will, den die automatische Transformation von Texten aus einem in ein strukturell anderes Graphiesystem für die Erforschung des Jiddischen haben kann, tut man gut daran, sich zunächst eine Reihe von Grundtatsachen der jiddischen Sprachgeschichte zu vergegenwärtigen.¹

Das Jiddische ist eine Komponentensprache, doch innerhalb des deutschen Sprachgebiets auf der Grundlage des Deutschen entstanden. Bis heute überwiegt die deutsche Komponente mit mindestens 90% neben den hebräischen und slavischen Anteilen. Außer dem Jiddischen gibt es noch andere jüdische Gruppensprachen, wie etwa das Jüdisch-Persische in Asien und in Europa das Spaniolische oder Ladino.

Die damalige Muttersprache der im 9./10. Jahrhundert nach Deutschland einwandernden Juden waren jüdische Varianten romanischer Sprachen - des Italienischen und des Französischen. Wirtschaftliche Gründe bewogen sie, zur deutschen Sprache überzugehen, welche jedoch - zumindest beim innerjüdischen Gebrauch - durch Beibehaltung romanischen Sprachguts, ständige Inspiration am Hebräischen und Vermischung verschiedener deutsch-dialektaler Elemente einen eigenen Charakter erhielt.

Vereinzelt seit dem 12. Jahrhundert, in großen Gruppen seit den Verfolgungen 1348/49 wanderten viele dieser Juden in slavische Gebiete weiter. Obwohl sie aus Prestigegründen bewußt an ihrer Sprache festhielten, trat allmählich das slavische Element hinzu. Infolge der eigenständigen Entwicklung und langsamen Trennung des östlichen Sprachzweigs vom westlichen spricht heute die Forschung von West- und Ostjiddisch.

* Protokoll des 52. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften am 6. Juli 1991

¹ Für unsere Zwecke mag es genügen, global zu verweisen auf S. A. Birnbaum: Die jiddische Sprache. Ein kurzer Überblick und Texte aus acht Jahrhunderten. Hamburg 1974; Ders.: Yiddish- A Survey and a Grammar. Toronto 1979; M. Weinreich: History of the Yiddish Language. Aus dem Jiddischen ins Englische übersetzt v. S. Noble u. J. A. Fishman. Chicago / London 1980.

Der Untergang des Westjiddischen vollzog sich im Laufe des 19. Jahrhunderts im Gefolge der Aufklärung und der aus ihr erwachsenen Emanzipation der Juden, beide für uns verbunden mit dem Namen Moses Mendelssohn. Heute sind Überreste des Westjiddischen nur noch im Elsaß, in der Schweiz und im Burgenland zu finden, das Ostjiddische hingegen wird - in vielen Ländern der Welt - insgesamt noch von mehreren Millionen Menschen gesprochen.

Die Ursachen für die Entfernung des ‚Westjiddischen‘ vom Deutschen waren zum einen die in Deutschland immer schlechter werdenden Lebensbedingungen nach 1348/49. Die Abschließung in Ghettos lockerte die gesellschaftlichen Bindungen zur christlichen Umwelt. Freilich wirkte sich die Isolierung sprachlich nur langsam aus.

Der andere Grund war die hohe räumliche Mobilität der aschkenasischen Juden durch Ausschöpfung der wirtschaftlichen Möglichkeiten, durch auswärtige und sogar überregionale Heiraten, aber auch durch Verfolgungen und Ausweisungen aus vielen Gegenden Deutschlands, darunter aus den meisten Reichsstädten (insbesondere zwischen 1425 und 1550). Dies brachte immer neue sprachliche Konstellationen mit sich.

Veränderungen in der Syntax und Formenlehre sind erst gegen 1500 sichtbar, der Wortschatz jedoch entfernte sich wesentlich früher und stärker vom Deutschen. Mehrere Aspekte spielen hier eine Rolle:

Das Jiddische wurde nicht von den Entwicklungen der deutschen Schriftsprache beeinflusst, orientierte sich vielmehr an der gesprochenen Sprache, bewahrte aber auch altes Wortgut im Gegensatz zum Neuhochdeutschen. Zugleich vollzog sich eine Bedeutungs differenzierung vieler Wörter, gegenüber dem Deutschen unter der Wirkung der traditionellen extrem wörtlichen Übersetzung der hebräischen Bibel ins Jiddische.

Diese spezifische Bibelübersetzungssprache ist nun einer der Forschungsschwerpunkte der Trierer Jiddistik. Unter dem Dache des Faches Germanistik ist in Trier die Jiddistik seit etwa zwanzig Jahren vertreten, seit 1990 durch einen eigenen Lehrstuhl, den einzigen in Europa. Man konzentriert sich hier auf Bereiche, für deren Erforschung die Kenntnis des Deutschen und der deutschen Sprachgeschichte unumgänglich ist, vor allem auf die ältere westjiddische Sprache und Literatur, welche noch weitgehend unbekannt ist und auch für die Germanistik noch aufschlußreich sein kann.

Mit dieser Orientierung bildet die deutsche Jiddistik eine sinnvolle Ergänzung zu den Institutionen in Jerusalem, Tel Aviv, New York und neuerdings in Oxford und Paris. Dort widmet man sich vor allem den in Osteuropa, Israel, Nord- und Südamerika noch gesprochenen jiddischen Mundarten und der ostjiddischen Literatur des 19. und 20. Jahrhunderts.

Die Originalzeugnisse der westjiddischen Literatur- Handschriften und Frühdrucke - sind heute über die Bibliotheken Europas, Israels und Nordamerikas zerstreut. Der internationalen Forschung liegen sie bislang zum größten Teil nur in Mikrofilmen bzw. Mikrofiches vor. Als Basis für sprach- und

literaturwissenschaftliche Untersuchungen sind somit zunächst kritische Editionen vonnöten, und zwar zweckmäßigerweise, um eine breitere Leserschaft zu erreichen, in Form von lateinschriftlichen Transkriptionen. Das Jiddische selbst bedient sich ausschließlich des hebräischen Alphabets. Die dadurch bedingte Schriftschränke war bisher für die Germanisten, aber auch für ein breiteres Publikum ein Haupthindernis, sich mit der jiddischen Sprache und Literatur zu beschäftigen. Diese Schriftschränke durch Transkriptionen überwinden zu helfen, betrachtet die Trierer Jiddistik als eine ihrer wichtigsten Aufgaben.

Andererseits hat die hebräische Schrift für die jüdische Kultur als eine ausgesprochene Buchkultur auch eine kaum zu überschätzende ideelle Funktion: sie ist eine der wichtigsten Konstituenten des jüdischen Identitätsbewußtseins. Allerdings ist gerade bei der Anwendung dieser Schrift auf die Jiddische Sprache im Laufe der letzten sechshundert Jahre ein bemerkenswerter Wandel festzustellen. An sich ist das hebräische Alphabet ein bloßes Konsonantenalphabet, aus dem lediglich die Zeichen für *j* und *w* hilfsweise in einigem Umfang zur Andeutung von Vokalen herangezogen werden. Die jiddische Überlieferung beginnt genau auf dieser Stufe, doch dann nimmt langsam das Bestreben zu, auch die Vokale durch Buchstaben zu repräsentieren in Anlehnung an die europäische Schriftkultur. Hebräische Konsonantenzeichen werden im Jiddischen zu Vokalzeichen (כ tritt für *a* bzw. *o* ein, װ für *e*, ן für *e* bzw. *i*, ן für *o* bzw. *u*). Die Diphthonge und Umlaute werden durch Kombinationen von Waw und Jod bzw. Doppel-Jod wiedergegeben.

Etwa um 1800 wird auf diese Weise die volle Schreibung der Vokale wie in der Lateinschrift erreicht. Auch Jiddisch-Muttersprachler haben deshalb heute Probleme beim Lesen altjiddischer Texte, nicht nur wegen des inzwischen eingetretenen beträchtlichen Laut- und sonstigen Sprachwandels, sondern auch weil sie auf Schritt und Tritt Vokalbezeichnungen vermissen. Anders ausgedrückt: die Erstellung einer historisch lautgetreuen Transkription setzt strenggenommen die Gesamtanalyse des Lautstandes dieses Textes und damit seine Situierung in die Entwicklung vom Mittelhochdeutschen zum Neujiddischen voraus.

Insgesamt ist nun aber durch die soeben beschriebene innere »Europäisierung« des jiddischen Graphiesystems die kulturelle Identifikationsfunktion der Schrift selbst keineswegs beeinträchtigt worden. Und da sie in einer Transkription zwangsläufig verlorengelht, sollte man sich der Erkenntnis nicht verschließen, daß in einer Edition neben der Transkription auch der Urtext sein Daseinsrecht hat. Bisher hätte eine Doppeledition von Original und Transkription jedoch eine zweimalige Eingabe erfordert und wäre damit fehleranfällig und zeitaufwendig gewesen. Auch in Zukunft wird eine automatische Erstellung der Transkription aus dem Urtext nicht zu realisieren sein, weil in sie sehr vielfältige lauthistorische Außeninformationen eingehen. Wohl aber ist das Umgekehrte möglich: die automatische Wiedererstellung des Originals aus der Transkription. Wir wollen sie kurz "Rückführung", ihr Ergebnis "Reoriginal" nennen.

Für den Herausgeber selbst besteht ihre Bedeutung eben darin, daß sie ihm den einen der beiden Editionsgänge einer Paralleledition abnimmt: eine *einmalige* Texteingabe kann *beide* Editionsformen generieren. Zwar wird man beide Ausdrücke Korrektur lesen müssen; aber auch dabei wird das Reoriginal die Qualität der Transkription verbessern. Man kann sich das leicht an einem Beispiel klarmachen. Im Original sei das Wort ‚Land‘ noch ohne Vokalzeichen geschrieben (לנד); der Transkriptor, der das Vokalzeichen a einsetzt (land), vergesse aber, das Diakritikum zu setzen, das den Vokal als ungeschrieben ausweist (lánd). Innerhalb der Transkription bleibt dieser Fehler unauffällig und ist deshalb schwer auszumerzen. Im Reoriginal hingegen wird der überschüssige Vollbuchstabe (לאנד) beim Vergleich mit dem Original viel eher auffallen und so die Korrektur sicherstellen.

Mehr noch als dem Herausgeber sollte aber eine so erstellte Dopeledition dem Leser zugute kommen. Mit ihr sollte man den jiddischen und hebräischen Muttersprachler ebenso wie den Muttersprachler oder Kenner des Deutschen erreichen können. Und vielleicht noch mehr: mit ihr könnte man jeden von beiden dazu bringen, allmählich auch die andere Seite der Dopeledition zu konsultieren und sich so einerseits in die jiddische Lautgeschichte, andererseits in die hebräische Schrift einzuarbeiten.

So weit unsere Rahmenvorstellungen - und nun zu den technischen Problemen.

Ansatzpunkt unserer Überlegungen ist die Transkription, weil sie eindeutig und rückführbar ist. Erika Timm formuliert in ihrer (Transkriptions-)Ausgabe der Erzählung ‚Beria und Simra‘ (1975) in gültiger Form die Anforderungen an das Transkriptionssystem, und die folgenden Prämissen orientieren sich an ihren Ausführungen:²

1. Das Transkriptionssystem soll die buchstabengetreue Rekonstruktion des Originals ermöglichen und damit den Grundanforderungen an eine wissenschaftliche Edition genügen (Rekonstruierbarkeitsprinzip).

2. Es soll über die eigentliche Transkription hinausgehend ermöglichen, den Lautstand, soweit er erschließbar ist, zu berücksichtigen und philologisch zu interpretieren (Interpretationsprinzip).

3. Es strebt bei der Auswahl der Transkriptionszeichen ein Maximum an Lesbarkeit für eine größtmögliche Leserschaft an, indem es deren orthographische Tradition berücksichtigt (Utilitätsprinzip).

4. Schließlich sollte es erlauben, mit einem Editionsengang auszukommen, um den großen Mehrfachaufwand bei einer Parallel-Edition zu ersparen (Ökonomieprinzip).

² Beria und Simra. Eine jiddische Erzählung des 16. Jahrhunderts. Hrsg. und kommentiert v. Erika Timm. In: Literaturwissenschaftliches Jahrbuch NF 14 (1973) 1-94. Berlin 1975, hier S. 3.

Eine - hier nicht abgedruckte - Transkriptionstabelle bietet die Grundlage der wissenschaftlichen Editionen, der Texteingabe in den Computer und auch die Basis des Umcodierungsprogrammes. Sie enthält den Zeichenvorrat von »normalen« Texten des 16. Jahrhunderts, kann aber nach Bedarf erweitert werden.

Durch die Beschränkung der Diakritika auf die Vokale, die s-Laute sowie h und $\bar{\text{v}}$ -Laute wird das Lesbarkeitsprinzip erfüllt. Der Leser mit nicht-wissenschaftlichem Interesse kann sich sehr rasch daran gewöhnen, die Diakritika weniger zu beachten, aber dem philologisch Interessierten ist es möglich, die Interpretation des Lautstandes nachzuvollziehen und zu kontrollieren.

Allerdings ist die Rückführung nicht auf dem Wege einer einfachen Einzu-Eins-Ersetzungstabelle zu bewältigen. Die Schemata 1 bis 4 zeigen, daß in den meisten Fällen komplexe sprachliche Bedingungen für die jeweilige Umwandlung in die Ursprungsgraphie erfüllt sein müssen.

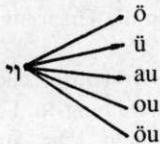
Schema 1: Umwandlemöglichkeiten vom deutschen in den hebräischen TU-STEP-Zeichensatz:

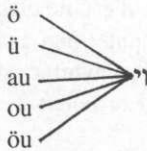
$\text{n} \rightarrow$ am Wortanfang / im Wortinnern $\rightarrow \text{n}$
 $\text{n} \rightarrow$ am Wortende $\rightarrow \hat{\text{n}}$

Schema 2: Hebräisches Ausgangszeichen \rightarrow Transkriptionszeichen \rightarrow TU-STEP-Zeichensatz:

$\text{D} \rightarrow \text{š} \rightarrow \text{s}$
 $\text{Ṛ} \rightarrow \underline{\text{s}} \rightarrow \text{z}$
 $\text{Ṛ} \rightarrow \text{s} \rightarrow$ vor Vokalen $\rightarrow \text{z}$
 $\text{W} \rightarrow \text{s} \rightarrow$ vor Konsonanten $\rightarrow \text{w}$
 $\text{W} \rightarrow \beta \rightarrow$ vor Vokalen $\rightarrow \text{w}$
 $\text{W} \rightarrow \text{sch} \rightarrow \text{w}$
 $\text{W} \rightarrow \text{š} \rightarrow$ in hebräischen Wörtern $\rightarrow \text{w}$
 $\text{W} \rightarrow \text{ś} \rightarrow$ in hebräischen Wörtern $\rightarrow \text{w}$
 $\text{T} \rightarrow \text{ss} \rightarrow$ in hebräischen Wörtern $\rightarrow \text{t}$

Schema 3: Lautliche Möglichkeiten des ם-Graphems:





Schema 4: א vor Vokalen am Wortanfang:

- i → am Wortanfang → א (ai)
- é → am Wortanfang → א (ai)
- e → am Wortanfang → ו (y)
- ó/u → am Wortanfang → א (au)
- ai/ei → am Wortanfang → א (aii)
- au/ou → am Wortanfang → א (aiu)
- a/o → am Wortanfang → א (a)

Weniger aufwendig erscheint zunächst die Unterscheidung zwischen *n* im Wortinnern und *n* am Wortende (vgl. Schema 1). Wesentlich anspruchsvoller wird die Differenzierung von vokalischem und konsonantischem Jod und Waw (vgl. Schema 4) und die der s-Laute (vgl. Schema 2). Die hebräischen s-Laute (Schema 2) ת, ש, ס, ז stehen zu ihren Transkriptionszeichen nicht in einem Eins-zu-Eins-Verhältnis und sind in der Tabelle entsprechend umgesetzt. Die Zeichenfolge א erfüllt fünf lautliche Realisierungsmöglichkeiten (vgl. Schema 3). Natürlich ist es sehr viel unkomplizierter und ökonomischer, fünf Varianten auf ein Ursprungsgraphem zurückzuführen als umgekehrt. Zudem müßte der Computer eine philologische Kompetenz besitzen, um den Lautstand korrekt zu eruieren, der zu aufwendig zu programmieren wäre.

Wie oben beschrieben erfüllen einige Zeichen vokalische und konsonantische Funktionen. In der schriftlichen Darstellung werden sie im Inlaut nicht differenziert, d.h. der Leser muß aus seiner Sprachkompetenz das Zeichen interpretieren. Im Anlaut dagegen steht immer ein Aleph vor vokalischem Jod oder Waw (vgl. Schema 3).

Hieraus wird ersichtlich, daß die philologische Arbeit bezüglich der Erstellung und Interpretation des Transkriptionssystems an erster Stelle steht. Erst wenn man die sprachlichen Bedingungen des jiddischen Schriftsystems erfaßt hat, kann man darangehen, das Programm zur Rückführung in die Ursprungsgraphie zu erstellen.

Wir wählen als ‚Werkzeug‘ für unsere Problemlösung die »Systemsprache« TUSTEP, dies aus mehreren Gründen. Die Sonderzeichenverwaltung in TUSTEP, vor allem des hebräischen Schriftsatzes ist kein Problem. Die Vorbereitung zur Edition läßt sich vollständig am PC durchführen und mit der TUSTEP-SATZ-Programm-Routine weiterbearbeiten, so daß die gesamte Kontrolle über die Edition beim Herausgeber liegt. Der Umcodierungsvorgang geschieht in mehreren komplexen Austausch-Vorgängen. Dies mit einer niedri-

gen Programmiersprache zu bewältigen, wäre zu aufwendig. Die kommerziellen Textverarbeitungsprogramme dagegen sind nicht variabel genug für unsere Zwecke, und auch die Sonderzeichenverwaltung ist kritisch. Das Tübinger System stellt quasi eine Zwischenstufe dar. Die aufwendige Programmierung der einzelnen Schritte entfällt, obwohl es variabel genug für alle Problemstellungen bleibt und den Komfort eines Textverarbeitungsprogramms bietet. Nicht zuletzt soll noch erwähnt sein, daß die TUSTEP-Programme vom Trierer Rechenzentrum hervorragend betreut werden.

Im folgenden soll noch kurz auf die »programm-internen« Anforderungen eingegangen werden. Voraussetzung für eine Automatisierung der Rückführung in die Ursprungsgraphie ist die Möglichkeit von Umgebungsdefinitionen für bestimmte sprachliche Gegebenheiten, vor allem Wortanfang - Wortinneres - Wortende oder vokalische oder konsonantische Umgebung. Ein weiterer wichtiger Punkt ist die Zusammenfassung von Zeichengruppen oder Zeichenfolgen und deren Gleichbehandlung.

Der eigentliche Inhalt des Programms ist die Rückführung der Transkription in den hebräischen TUSTEP-Zeichensatz. Zuerst werden die punktierten Vokale und Konsonanten behandelt. Danach folgen s-Laute, dann unpunktete Vokale und unproblematische Konsonanten. Dann werden die Bedingungen für Finalbuchstaben festgelegt.

Das Programm wurde in Zusammenarbeit mit dem Trierer TUSTEP-Berater Dr. Michael Trauth erstellt. Eine interaktive Benutzerführung wird eingesetzt, um es anwenderfreundlich und variabel zu gestalten und es anderen Fachkollegen zur Verfügung zu stellen, ohne daß diese die Programminterna kennen müssen.

Jeder neue Text stellt andere Anforderungen an das Programm, da sich die Texte in ihrem Lautstand im Laufe der Sprachgeschichte, wie zu Beginn beschrieben, in starkem Maße wandelten und nicht einheitlich darstellen. Jeder Text erfordert ein modifiziertes Transkriptionssystem, das der Philologe aufgrund seiner Kenntnisse der graphematischen Struktur der Texte erarbeitet - und damit natürlich auch ein modifiziertes Umcodierungsprogramm. Da das Prinzip dasselbe bleibt, ist die Erweiterung um bestimmte Definitionen und Austausch-Vorgänge relativ unproblematisch, solange ein genaues Wissen um die sprachlichen Bedingungen gegeben ist.

Jeder Editor hat andere Vorstellungen von seiner Textausgabe z. B. bezüglich der Interpunktion, der Anmerkungen, der Sonderzeichen etc. Dies läßt sich ohne größeren Aufwand variabel für beide Seiten des Ausdrucks gestalten.