

# Veröffentlicht im Rahmen des CoMOn-Programms, 2010:

[Http://www-ct.informatik.uni-tuebingen.de/daten/codier.pdf](http://www-ct.informatik.uni-tuebingen.de/daten/codier.pdf)

## Codierung – Computerkonkordanz – Übersetzung

Einführung in das Konkordanzsystem „Corpus Matching Online“.

Beispiel: Hebräisches Altes Testament

Harald Schweizer, Tübingen

### 1. Codierung

#### 1.1 Codiersysteme

Im Rahmen der computerunterstützten Analyse der Josefsgeschichte, war es notwendig gewesen zu klären, in welcher Codierung der Text elektronisch gespeichert werden soll. Für beides, die Speicherung in Printmedien und die damals schon aufgekommene elektronische, gab es unterschiedliche Konzepte. Sechs davon wurden in einer Begleitpublikation in Beziehung zu einander und in Beziehung zur hebräischen Klarschrift gesetzt.<sup>1</sup>

Folgende Problemstellen seien exemplarisch herausgegriffen. Immer besteht der Zwang, mit der Tastatur des Lateinischen (*ascii*) nicht-lateinische Schriftzeichen anzusteuern und wiederzugeben:

- Hebräisches **Alef** "א" kann in der Software TUSTEP<sup>2</sup> durch bloßes »a« wiedergegeben sein, verlangt dort aber zusätzlich Steuerzeichen für »Hebräisch-Anfang« und »Hebräisch-Ende«. Die komplette Codierung lautet also: »#h+a#h-«.

- Verbreitet in Printmedien, auch geltend für die ZAW, ist die Wiedergabe des Buchstabens durch » ' «. – Man kann auch – wie *Gérard Weil* vom damaligen CATAB-Institut, Lyon, mnemotechnische Hilfen einbauen und in der elektronischen Form "א" wiedergeben durch das optisch ähnliche » % «

- Die USA-Transkription<sup>3</sup> verwendet » ) «. – Wieder eine andere Anwendung (BADER): »><. – Wer mit neuronalen Netzen operiert, sollte ganz auf Ziffern umstellen: »1«.

- Komplexe Zeichen stellen komplexere Codieranforderungen. Diakritische Punkte müssen zum Grundzeichen hinzunotiert werden. In TUSTEP sähe das » װ « in der Variante *Schin* so aus: » #h+w#"8 #h-«. Andere Vorschläge für diesen Buchstaben sind: » š « (ZAW), » \$ « (USA), » W « (*Weil*), » x « (*Bader*), » 24 « (Neuronales Netz).

- Einige Buchstaben ändern ihre Form, wenn sie am Ende eines Wortes zu stehen kommen. Im Arabischen wäre dieser Punkt noch um einiges komplexer. Die Frage stellt sich: Soll man die unterschiedlichen Gestalten in der Codierung berücksichtigen? Oder nicht eher ein

1 Vgl. H. SCHWEIZER, Anhang zur Methodik, in: SCHWEIZER, H (ed.) Computerunterstützte Textinterpretation. Die Josefsgeschichte beschrieben und interpretiert im Dreischritt: Syntax-Semantik-Pragmatik. THLI 7 /i-iii. Tübingen 1995. Bd. III,16.

2 Vgl. TUSTEP. Das Handbuch. Tübinger System von Textverarbeitungsprogrammen. Ausgabe 2001. Tübingen.

3 Darunter wird verstanden die Codierung des *Claremont-Michigan-Textes* (Urheber: van DYK PARUNAK), der unserem Suchprogramm zugrunde liegt (Tradent: A. GROVES, Philadelphia).

und den selben Buchstaben immer gleich abspeichern, die unterschiedliche Gestalt anschließend durch programmierte Stellungsregeln erzeugen lassen?

Solche Vorschläge und Konzepte können nicht nach „richtig / falsch“ bewertet werden. Die Frage ist, was die jeweilige Transkription leisten soll. Ist nur eine schöne gedruckte Wiedergabe das Ziel, so darf ein "א" durch h komplexes »#h+a#h-« gespeichert sein. Sollen mit dem gespeicherten Text Sortierungen durchgeführt werden, sind einfache Codierungen nötig (*ein* Codierelement für *ein* Schriftzeichen), ebenso sollten Codierungen nicht verschiedene Zeichentypen mischen (Buchstaben, Ziffern, Sonderzeichen). Ein »+« (USA) für »ט« schafft Komplikationen, sobald man Programme schreiben soll.

## 1.2 Beispiele: Gen 37,2b\*

Wie ein und der selbe hebräische Satz in den verschiedenen erwähnten Codiersystemen wiedergegeben wird, soll hier demonstriert werden.

BHS	יוסף היה רעה את אחיו בצאן	Klarschrift	Rohtext
TUSTEP:	#h+ius^p hih ryh at axiu	Bca^n#h-	
TÜ/Tustep:	ywsp hyh r#.(h #.) #.)%.hyw	bx#.n	
TÜ:	ywsp hyh r'h 't 'hyw	bc'n	
WEIL:	YVCP HYH RJH %T %XYV	B3%N	
USA:	YWCP HYH R(H )T )XYW	BC)N	
BADER:	ywsp hyh r<h >t >xyw	bc>n	
NEUR:	9060160180050110500220170500102500108011060020160101500 (bezogen auf die obigen Tabulatorpositionen:) 90601601800 50110500 220170500 1025001080110600 20160101500		

## 1.3 Informatische Bedürfnisse

Es ist die crux bei allen nicht-lateinischen Schriften, dass mit dem traditionellen Zeichensatz (Tastatur, *ascii*) ganz andere Ergebnisse erzielt werden sollen, als man sie unmittelbar einzugeben vermag. Dafür, aber auch für die informatische Weiterverarbeitung ist die Beachtung einiger Punkte bei der Codierung wichtig:

Es gilt zu unterscheiden: Zeichen für Objektsprache und Steuerzeichen. Es hängt vom Codierverfahren ab, ob ein eingegebenes „a“ in der Ausgabe ebenfalls ein „a“ ergibt, oder ein „א“, oder ob es lediglich eine Anweisung an das ausführende Programm darstellt.

Gruppenbildung (informatisch gesprochen: *reguläre Ausdrücke*): Eine Tastatur bietet gleichmache-  
 risch viele druckbare Zeichen an. Um differenzierter arbeiten zu können, sollten Gruppen unter-  
 schieden werden: Buchstaben (groß / klein), Ziffern, Satzzeichen, »Sonstige«, z.B. § \$ % & ( ) = ?  
 @ < > +.

Folglich sollten bei Codierungen die Gruppen nicht gemischt werden. »Sonstige« dürfen nicht  
 zusammen mit »Buchstaben« oder »Ziffern« vorkommen. Inkonsistent ist demnach bei der USA-  
 Transkription das » + « als Wiedergabe von „ ʋ “, oder » ) « für „ ʀ “

Mnemotechnische Hilfen sind zwar menschenfreundlich, können aber aus genanntem Grund  
 problematisch werden, wenn etwa das System aus Lyon „ % “ für „ ʀ “ setzt, oder » 6 « für „ ʋ “.

Werden einige zusammengesetzte Zeichen verwendet - „ ħ “ oder „ š “, die anderen aber einfach  
 codiert, so ergeben sich Komplikationen bei Sortierungen: man muss beim zusammengesetzten  
 Zeichen erst den Teil definieren, der bei der Sortierung berücksichtigt werden soll.

Codierungen werden hier unter dem Primärinteresse betrachtet, dass gute Analyseergebnisse erzielt  
 werden sollen. Sekundär ist die Optik und unmittelbare Druckbarkeit. Es wird in Kauf genommen,  
 dass die optische Aufbereitung der Ergebnisse erst in einem abschließenden Schritt durchgeführt  
 werden wird.<sup>4</sup>

Zwischenergebnis: Keines der vorgestellten Transkriptionssysteme entspricht den genannten  
 Anforderungen. Für das Suchprogramm werden wir die USA-Transkription in mehrfach korrigierter  
 Form zugrunde legen:

„ ) “	→	A	für	ʀ
„ + “	→	J	für	ʋ
„ ( “	→	I	für	ʌ
„ & “	→	F	für	ʍ
„ \$ “	→	E	für	ʎ

Eine solche Verwendung von Vokalbuchstaben („F“ ausgenommen) ist möglich, weil wir die  
 lateinischen Vokalzeichen der Tastatur nicht benötigen – wir werden nur am Konsonantentext  
 arbeiten. In der korrigierten Form sieht der gewählte Beispielsatz bei uns wie folgt aus:

YWCP HYH RIH AT AXVY BCAN

## 2. Computerkonkordanz

### 2.1 Crux traditioneller Konkordanzen / Wortlisten

Konkordanzen treten mit dem Anspruch auf, nichts als Wortlisten zu sein. Das betreffende Wort  
 wird an den Fundstellen in einen kleinen Kontext eingebettet, so dass man sich ein besseres Bild  
 über den Verwendungszusammenhang machen kann. Die reine Beschränkung auf die äußere  
 Wortebene signalisiert eine Absehung von den Bedeutungen. So gesehen haftet der  
 Konkordanzarbeit immer auch eine gewisse Härte an, da man Texte, Bücher nicht wegen ihres  
 Wortinventars, sondern wegen der Bedeutungen zu lesen pflegt.<sup>5</sup>

4 Die Entzerrung von Ergebnisgewinnung und Präsentation in Klarschrift ist für Ungeübte gewöhnungsbedürftig, ist  
 man doch standardmäßig durch das WYSIWYG-Konzept („what you see is what you get“) daran gewöhnt, dass  
 etwa in einem Schreibprogramm immer bereits die Endform des gewünschten Textes präsentiert wird. Anders gehen  
 Programme/Konzepte wie TUSTEP, TEX oder inzwischen HTML vor: Datenerfassung und Präsentationsform  
 werden in zwei klar unterschiedenen Schritten vollzogen bzw. erzeugt.

5 Hier wäre der Ort – das ist aber nicht Thema des Beitrags - sich bewusst zu machen, wozu Konkordanzarbeit

Der Standard ist allerdings ein anderer: Es wird meist nach äußerlich gleichen/identischen Wortformen /-ketten gesucht, die *zugleich* inhaltlich zusammenpassen. Hat ein und die selbe äußere Wortform verschiedene Bedeutungen, werden die *Vorkommen der Wortform nach Lemmata unterschieden*. Methodisch – auch wenn man die Folgen nicht gleich überblickt – ist dies inkonsequent: neben der Beurteilung des äußeren Erscheinungsbildes wurde eine weitere Beurteilungsebene ins Spiel gebracht, die der Bedeutungen.<sup>6</sup> Auch der Begriff *Lexem* meint eine Größe, die bipolar ist: eine Wortform in Verbindung mit einer Bedeutung<sup>7</sup>

Im Fall der hebräischen Bibel entstehen Irritationen durch die masoretische Vokalisierung. Steht ein und das selbe Wort in unterschiedlichem Kontext, kann sich ein *Kamez* zu einem *Patach* verändern – mit dem Effekt, dass eine automatisierte Suche die beiden Wortformen nicht mehr als identisch erkennt.<sup>8</sup> Gravierender ist der zweite Einwand: bekanntlich ist die masoretische Vokalisierung ca. 1000 Jahre jünger als der kanonische Konsonantentext. Es bleibt als Frage, inwieweit das elaborierte Vokal- und Akzentsystem (für liturgische Zwecke entworfen) die Artikulation korrekt anzeigt, die für die Abfassungszeit der Texte angenommen werden muss. Unsere Folgerung: Für die

---

dienen kann. Eine explizite *Konkordanz-Methodik, -Didaktik* wäre in der Ausbildung angemessen. Wörter nachschlagen, gleiche grammatische Konstruktionen – das ist die zunächst nahe liegende Verwendungsweise. Wer bei der exegetischen Arbeit von der Größe „Text“ ausgeht, wird ein Interesse haben, diesen Einzeltext in Beziehung zum umgebenden Korpus zu setzen, Stichwort *Intertextualität*: in einem zunächst einfachen und gut überprüfaren Sinn: Welche Wortketten des Einzeltextes finden sich im umgebenden Korpus auch noch? Liegen die Befunde = *Suchtreffer* vor, müssen sie bewertet werden. An den Befunden selber wird nicht gezweifelt werden können – man kann sie leicht nachprüfen. Die Auswertungen und Schlussfolgerungen dagegen werden Anlass zu Debatten geben: Handelt es sich um Idiome, Gruppensprache, Anspielungen, Zitate, Verballhornungen usw. Eine breite Palette von Deutungsmöglichkeiten kann sich auftun, die entsprechend reflektiert sein müssen. - Noch eine Stufe höher, mehrere Einzeltexte und ihre intertextuellen Bezüge umfassend, kann man eine *relative Chronologie* erstellen. Dabei nähert man sich informatisch demselben Problem wie im Fall der Analyse von Gensequenzen: So wie es möglich ist, die heutige Menschheit genetisch auf 7 Urmütter in Ostafrika zurückzuführen, müsste es bezüglich der Einzeltexte des AT möglich sein, in einer Baumstruktur viele spätere von wenigen frühen Texten zu unterscheiden, basierend allein auf der Analyse von Wortketten, ohne spezielle inhaltliche oder geistesgeschichtliche Hypothesen einfließen zu lassen.

- 6 Von dem kritischen Einwand ist auch betroffen: MUTHMANN, G, Rückläufiges deutsches Wörterbuch.RGL 78. Tübingen 1988. - Deutlich reflektierter ist angelegt: DUK HO LEE, Rückläufiges Wörterbuch der deutschen Sprache. Berlin 2005.
- 7 Wer bereit ist methodisch sauber vorzugehen, sollte an dieser Stelle sehen und akzeptieren, dass es einen qualitativen Unterschied macht, ob man nur die Ebene der Ausdrücke (Wortformen und -ketten) anvisiert, oder ob man zugleich Bedeutungswissen hinzunimmt, also komplex mit zwei unterschiedlichen Kriterien gleichzeitig operiert. Im letzteren Fall möge man noch die Logik prüfen, dass beide Kriterien nicht gleichzeitig angewendet werden können. Vielmehr wird im Konfliktfall immer die Bedeutungsseite obsiegen, die Ausdrucksebene wird entsprechend übergangen oder korrigiert – was dann zur hier kritisierten Praxis führt, dass eben doch keine schlüssige und eindeutige Analyse auf Ausdrucksebene durchgeführt, eine solche allenfalls behauptet wird. Letzter Punkt: Wenn man heutzutage schon den Computer zu Analyse Zwecken nutzen kann, also zu mehr als zur Textverarbeitung oder zum Surfen und zur email, dann heißt das: nichts kann die Maschine so gut, wie Zeichenketten auf Identität / Ähnlichkeit / Häufigkeit zu überprüfen. Die Effizienz und Präzision der Maschine hierbei ist von einem menschlichen Benutzer nicht zu erreichen. Daher liegt es nahe, wohl überlegt sich dieses Hilfsmittels zu bedienen. Das verlangt jedoch auch – das versucht der Aufsatz zu erläutern -, dass man als Exeget die Notwendigkeiten informatischer Art zur Kenntnis nimmt. Es geht nicht nur darum, selbst Programmierer zu werden. Aber man kann zur Kenntnis nehmen, welche Erfordernisse entstehen, wenn ein Informatiker mitwirken soll. Die informatische Perspektive kann auf Sachzwänge verweisen, die man ohne den Computer als Analysehilfsmittel übersehen würde, ebenso die daraus resultierenden „Schäden“. Rückfragen der Informatik werden dazu führen, dass das angestammte Objekt (im Moment: hebräische Bibel) präziser und umfassender, in manchem auch anders als bisher wahrgenommen wird. Die effiziente Einbeziehung des Rechners wird nicht genau die gleichen Ergebnisse liefern wie zuvor, nur eben schneller. Vielmehr erzwingt die klar definierte Arbeitsebene des Computers (Wortformen, keine Bedeutungen) eine methodisch klarere Positionsbeschreibung des Exegeten.
- 8 Es sei denn, man verlangt bei der Suchbedingung nicht komplette Übereinstimmung, sondern lässt in gewissem Maß Variation zu (statt 100%, vielleicht nur 90% Übereinstimmung). Das hier vorgestellte Suchprogramm wird diese Option zulassen.

*online*-Konkordanz verspricht der Konsonantentext authentischere Ergebnisse. Vokale und Akzente werden ausgeschlossen.

Konkordanzen kann man zu verschiedenen Zwecken nutzen. Das hier vorzustellende Werkzeug ist auf einen Zweck zugeschnitten: einen ganzen Text komplett auf seine Verbindungen zum restlichen Korpus AT hin zu untersuchen. Das jedoch stellt an potenzielle Nutzer die Frage, ob ihnen diese Orientierung überhaupt wichtig ist. Um mit einigen Schlagworten die Richtung anzugeben: Exegese ist nicht in ihrer Breite dafür bekannt, dass sie *textlinguistisch* ausgerichtet ist. Meist wird das Stichwort *historisch-kritisch* Anspruch genommen, wobei dann die historische Fragerichtung dominiert, die literarische also nicht als eigenständige und vorrangige Perspektive gewählt wird (wir haben aber zunächst nichts als Texte, also literarische Forschungsobjekte). Mehr grammatisch ausgedrückt: Das Fahrwasser der *Satzgrammatik* dominiert die Exegese weiterhin.<sup>9</sup> Die Wende (seit ca. 3 Jahrzehnten), wonach der *Text* analytisch die entscheidende Bezugsgröße darstelle, auch eigene, textbezogene *literarische* Untersuchungen durchzuführen wären, kam in der Breite sicher noch nicht in der Exegese an.<sup>10</sup>

Das alles soll hier nicht vertieft werden, sondern nur andeuten: die vorzustellende Computerkonkordanz kann ihren Wert nur entfalten im Rahmen einer *literarisch und textorientierten* Ausrichtung. Nur so lassen sich die benötigten Reflexionen anstellen, die verhindern, dass aus Suchergebnissen naive Interpretationen abgeleitet werden.<sup>11</sup>

Die Beschränkung bisheriger Konkordanzen (Suchwortorientierung, auch bei Kontextangabe) ist – wenn der Einzeltext die entscheidende Größe ist – zu starr. Der Einzeltext selbst wird als Wortkette betrachtet.<sup>12</sup> Folglich wird benötigt: eine Orientierung der Suche an *Wortketten*. Dabei kann ein Exeget nicht im voraus wissen, für welche Wortkettenlänge Treffer zu erwarten sind. Im maximalen Fall wird ein kompletter Text irgendwo anders exakt nochmals gefunden, weil er abgeschrieben worden war. Ein solcher Befund sollte ebenso gefunden werden, wie die viel häufigeren identischen Wortketten mit Länge im Bereich von etwa 3 – 8 Wortformen. Die Unsicherheit, für welche Kettenlänge Treffer zu erwarten sind, kann informatisch aufgefangen werden. Es lässt sich algorithmisch implementieren, dass die Maschine *nach möglichst langen* Wortketten sucht – ohne dass die Länge vorab eingestellt wäre. Sobald ein nicht-

---

9 Man verzeihe den Hinweis, dass auch dies vielfach ein Euphemismus ist: Der Terminus „Satzgrammatik“ unterstellt, dass wenigstens auf diesem Feld Forschungsarbeit geleistet wird. Solche Zeiten gab es (um die Wende zum 20. Jahrhundert). Aktuell hat „Grammatik“ wohl keine Hochkonjunktur in der Exegese.

10 Interessant wäre hier ein Eingehen auf die *Formgeschichte*. Mit ihren „Kleinen Einheiten“ implizierte sie zweifellos einen Impuls in Richtung Textbeschreibung. Aber eine textorientierte Methodik, literarische Beschreibungskonzeption entwickelte sie nicht, sie war im Kern ja auch nicht *literarisch* ausgerichtet, sondern *soziologisch* („Sitz im Leben“). - Anleihen aus der Erzählforschung, die inzwischen doch eine gewissen Rezeption erfahren haben, sind gut. Solche Ansätze können aber bis ins Grammatische hinein noch wesentlich intensiviert werden. Als Vorschlag für die Schulebene vgl. <http://www.alternativ-grammatik.de>. In der Sprachdidaktik wünscht man sich seit langem einen *integrierten* Sprachunterricht, primär an Universitäten, der die Trennung in Linguistik (Grammatik) und Literaturwissenschaft aufhebt. Die Aussicht auf eine Reform in dieser Richtung sind nicht günstig: zu sehr sind die Fächer zementiert.

11 Beispiel für eine Anregung aus einer anderen Philologie: Angenommen in CoMOn wäre ein Korpus der Reden von Barack Obama und gleichzeitig ein Korpus aktuellen amerikanischen Sprachgebrauchs (z. B. ein Jahrgang Zeitungsartikel) integriert, könnte man verifizieren, was MIEDER, W., „Yes We Can“. Barack Obama's Proverbial Rhetoric. Bern 2009 plausibel nachweist: in den Reden werden in einem hohen Maß Wortketten verwendet, die die Hörer schon kennen (Obama-spezifische, Wendungen, Sprichwörter, Lebensweisheiten usw.). D. h. noch vor der inhaltlichen Prüfung, ob die betreffende Aussage passt und zutrifft, haben die Hörer unbewusst schon das Signal der Vertrautheit, der gemeinsamen Wissensbasis bekommen. Eine solche, häufig wiederholte Stilistik wird sich auch an der Wahlurne auszahlen.

12 Die von späteren Überarbeitungen befreite Josefsgeschichte hatten wir – auf Ausdrucksebene – vor uns als Objekt beginnend mit Wortform 1, endend mit Wortform 2512. Vgl. SCHWEIZER, H., Die Josefsgeschichte. Konstituierung des Textes. THLI 4. Tübingen 1991 (vgl. 3. Textversion in Band ii).

übereinstimmendes Wort folgt, bricht die Suche an der betreffenden Korpusstelle ab.

Einzustellen ist lediglich die gewünschte *Mindestlänge*. Voreingestellt ist die Länge „3“. Das lässt sich nach oben oder unten ändern. Wählt man die Länge „2“, ist erwartbar, dass die Zahl der Treffer sprunghaft ansteigt. Für manche grammatischen und/oder idiomatischen Fragen kann dies jedoch sinnvoll sein. Immer gilt: nach oben ist die Trefferlänge offen. Solange das Programm Übereinstimmungen feststellt, wird der Treffer verlängert.

Das Thema »Ähnlichkeit« ist in CoMOn vorgesehen unter dem Stichwort „Tolerance“. Standardmäßig wird unter der Bedingung „100% Ähnlichkeit = Identität“ gesucht. Das lässt sich jedoch ändern: je geringer der Prozentsatz, desto mehr Variationen werden bei den Korpus Treffern erlaubt. Es hängt von praktischen Erprobungen ab, bei welcher Einstellung interessante weitere Treffer in den Blick kommen.

Thema »Permutation« (andere Reihenfolge der Wortformen): Der Gesichtspunkt kann im »Handbetrieb« nicht zuverlässig und erschöpfend behandelt werden. Für die Maschine bedeutet das Zulassen von Umstellungen – womöglich noch kombiniert mit dem Thema „Ähnlichkeit“ - einen gewaltig vergrößerten Rechen- und Speicheraufwand. Es ist dann klug bzw. notwendig, darauf mit geringerer Suchtextlänge zu reagieren. Auch hierzu sind praktische Erprobungen notwendig. Aber es besteht die Chance, weitere interessante, nicht-identische Korpusstreffer zu finden.<sup>13</sup>

Aufwand: Aus Zeiten meiner Dissertation erinnere ich mich, dass die komplette Überprüfung eines Elischa-Erzähltextes mit gedruckter Konkordanz 14 Tage benötigte. Mit Hilfe des Computers geschieht das in ca. 1 Minute. Die Technik ermöglicht aber nicht nur einen atemberaubenden Geschwindigkeits- und Bequemlichkeitsvorteil, sondern liefert umfassendere, präzisere, klarer definierte Ergebnisse. Anders gesagt: das Objekt (hebräischer Einzeltext) wird viel genauer wahrgenommen.<sup>14</sup>

## 2.2 Computergestützte Konkordanz

Das CoMOn-Werkzeug hat eine Vorgeschichte. Sie sei knapp referiert.

Um 1990 wurden in einem sich lange hinziehenden Tübinger Colloquium die Möglichkeiten einer exklusiv auf die Ausdrucksebene bezogenen, computergestützten Recherche ausgelotet. Ergebnis waren -(1) - die in eine Dissertation mündenden *textinternen* Berechnungen, wie sie W. BADER vorgestellt hat,<sup>15</sup> (2) Erprobungen zur Wortbildung (ausdrucksformale Morphologie)<sup>16</sup>, (3) automatisierte Erkennung stilistischer Figuren wie Alliteration,<sup>17</sup> und (4) – das interessiert hier – die automatische *string*-Suche im Verhältnis „Einzeltext : Korpus“. Umgesetzt im Analysetool TUSTEP erstellte auf diese Weise M. SCHINDELE eine Beschreibung der Josefsgeschichte.<sup>18</sup>

13 Wenn es im strikten Sinn um *Formeln* geht, muss die Wortkette *identisch* sein. Aber es gibt schließlich auch *softere* Bezugnahmen, von freieren Anspielungen bis zur Einpassung von geprägten Sprachbildern in den eigenen Sprachgebrauch. „So spricht Jahwe“, die ca. 400 mal im AT vorkommende Botenformel, heißt in Gen 45,9 „so spricht [dein Sohn] J“ - soweit die Übereinstimmung. In Vollform: „so spricht dein Sohn *Josef*“. Das sieht denn doch nach Anspielung und zugleich individueller Abweichung aus. Suchtechnisch wird der Treffer gefunden, wenn man 1 Fehlwort in der Zeichenkette zulässt.

14 Im Handbetrieb wirken zwei Störfaktoren: Müdigkeit (beim Durchsuchen endlos langer Beleglisten) und die Tendenz zur Filterung aufgrund von inhaltlichen Aspekten.

15 Vgl. BADER, W, Simson bei Delila. Computerlinguistische Interpretation des Textes Ri 13-16. THLI 3. Tübingen 1991.

16 Vgl. unter Einbeziehung des Hebräischen: STENGEL, M, Strings of Natural Languages: Unsupervised Analyses and Segmentation on the Expression Level. Frankfurt/M 2008.

17 Vgl. als Beispiel: <http://www-ct.informatik.uni-tuebingen.de/daten/allit1.pdf>

18 Vgl. SCHINDELE, M, Externe Syntax: Verbindungen zu weiteren Texten. In: (1995) SCHWEIZER, Z (ed.) 42-

Untermauert war die Pioniertat durch eine formalisierend-mathematische Begleitveröffentlichung.<sup>19</sup>

Die Durchführung hieß auf damaligem Stand der Computertechnik:<sup>20</sup> Die hebräische Josefs-geschichte wurde komplett automatisch mit dem AT abgeglichen (5 Stunden auf dem Großrechner). Die Suche war auf exakte Übereinstimmung ausgerichtet, die Suchtexte schlossen noch die Vokalisierung ein.<sup>21</sup>

Etwa 10 Jahre später erstellte M. PACH (2006) in Java ein differenziertes Suchprogramm (Volltextsuche): ein Einzeltext wird im Verhältnis zu einem frei wählbaren Korpus untersucht; inklusive Permutationen, Ähnlichkeitsgrade (je einstellbar). Unter „Einzeltext = Suchtext“ wurde noch ein individuell einzugebender Textabschnitt verstanden, der vom Programm in *reguläre Ausdrücke* übersetzt wird.<sup>22</sup> Was fehlte, ging S. LAUCKNER (2008) an: ein Ausschnitt aus einem Korpus wird als Suchtext definiert und vom Programm sequenziell abgearbeitet.<sup>23</sup>

Via Studienarbeit machte S. BYKH (2009) durch ein JAVA-Applet eine abgespeckte Version des PACH-Programmes online verfügbar. Zunächst für Suchen am hebräischen AT (nur Konsonanten, in korrigierter USA-Transkription; Vokale und Akzente sind eliminiert). Das ist es, was hier unter *CoMOn* vorgestellt wird. Weitere Korpora sollen integriert werden (AT griechisch, NT, Koran – arabisch, deutsch).

Das Programm erarbeitet selbstständig, bis zu welcher Länge eine Wortkette meines Einzeltextes eine Entsprechung im restlichen AT findet (die gewünschte Länge muss man also nicht starr vorgeben). Lediglich die gewünschte Mindestlänge ist anzugeben.<sup>24</sup>

Den Such(=Einzel)-Text kann man vor der Suche noch bearbeiten, wenn es text- und literarkritische Notwendigkeiten / Erkenntnisse gibt. In diesem Fall erlischt die ursprünglich gewählte Stellenangabe für den Suchtext, denn dieser entspricht nun nicht mehr exakt der Form, wie sie im Korpus

---

88.

19 Vgl. SCHINDELE, M, Computerunterstützte Suche formelhafter Rede. Protokoll des 57. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften am 6. Februar 1993: *Literary and Linguistic Computing* 9 (1994) 169-172. - Vgl. Ders, Möglichkeiten und Grenzen maschineller Befunderhebung zur Untersuchung von Formeln und geprägten Wendungen mit Beispielen aus Dan 8, in: BADER, W (ed.), „Und die Wahrheit wurde hinweggefegt.“ Daniel 8 linguistisch interpretiert. *THLI* 9. Tübingen 1994. S. 3-16.

20 Dass die technischen Fortschritte rasant sind, ist bekannt. Verdeutlichen kann man es sich daran, dass jedes heutige Handy mehr an Rechenleistung bietet als Ende der 1960er Jahre die Mondfähre.

21 Die Befunde sind im erwähnten SCHINDELE-Beitrag (1995) aufgelistet, wie auch für die betreffende Textstelle ausgewertet und interpretiert, auch inhaltlich. Was die Befunde insgesamt für die Josefs-geschichte an Folgerungen erlauben, ist dargestellt in SCHWEIZER (1995) „Abschließende“.

22 Mit ihnen konnten Ähnlichkeitsbedingungen formuliert werden.

23 Dazu muss das Korpus mit einer Struktur vorgegeben sein – die Bibel mit ihrer traditionellen Zählung eignet sich gut dafür: Buch / Kapitel / Vers. Nur muss diese Struktur bei der Suche übergangen werden, was auch heißt: es sollen versübergreifende Wortketten gefunden werden können. Versgrenzen dürfen die Suche nicht behindern.

24 Als Standard ist „3“ eingestellt. Damit liegt man erfahrungsgemäß oberhalb häufiger belegter grammatischer Kookkurrenzen, nähert sich schon idiomatischen, geprägten Wendungen. Und natürlich sorgt die Länge „3“ dafür, dass die Suche zügig zu einem Ergebnis führt. Man kann die Mindestlänge auch verringern. Im Extremfall „1“ sucht das Programm für jedes Wort sämtliche Belege im restlichen AT. Das kann für bestimmte Fragestellungen sinnvoll sein – braucht allerdings seine Zeit und viel Speicher. Die Mindestlänge „2“ kann als Präzisierung der „3“-Suche sinnvoll sein.

aufzufinden ist. Die Suche im Korpus kann mit dem dann neu definierten Suchtext als Basis genauso gestartet werden wie auf der Basis eines unveränderten, lediglich über Buch/Kapitel/Vers definierten Einzeltextes

Es war schon beim SCHINDELE-Programm ein Schlüsselerlebnis gewesen, gleich am Beginn der ursprünglichen Josefsgeschichte mit *היה רעה את* als Dreierwortkette mit dem Befund konfrontiert zu werden, dass die Kette außer in Gen 37,2 (Suchtext) nur noch in Ex 3,1 vorkommt, sonst nirgends mehr im AT (Korpus). Mehrere Erkenntnisse kamen zusammen: (1) In einer Buchkonkordanz hätte man nach dieser Kette deswegen nicht gesucht, weil sie keinen abgeschlossenen Gedanken bietet. Mindestens ein viertes Wort hätte man noch verlangt, weil man doch bedeutungsgesteuert seine Suchbedingungen definiert. Da es dazu (4er-Kette) aber nirgendwo eine Übereinstimmung gibt, wäre der Verweis auf die Dreierverbindung auch unter den Tisch gefallen. - (2) Die Wortkette besteht aus Einzelwörtern, die für sich genommen sehr häufig in der hebräischen Bibel vorkommen. Der Schluss von der Einzelhäufigkeit der Elemente auf die dann wohl doch zu erwartende große Häufigkeit einer solchen Gesamtkette ist falsch. - (3) Eine Einmalverbindung zwischen zwei Korpusstellen, bestehend aus für sich genommen häufigen Wörtern, fällt auf. Die Kette steht folglich nicht für einen breitflächigen idiomatischen Sprachgebrauch (was auch ein interessantes Ergebnis wäre), sondern stellt einen Exklusivbezug zwischen zwei Einzeltexten her. - (4) Hat man auf Ausdrucksebene einen solchen Befund erhalten, kann man nun, aber erst als *zweiten* Schritt, auch bedeutungsmäßig der Spur folgen und fragen, ob der distributionelle Impuls auch in inhaltlicher Hinsicht interessant ist.<sup>25</sup>

Jener erste Befund traf durch den Text der Josefsgeschichte hindurch mit einer Reihe weiterer zusammen, die oft ebenfalls Exklusivbezüge darstellten. Insgesamt war der Text zwar weitgehend frei von den klassischen Formeln, wie man sie quer durch die Bücher des AT antrifft. Aber immer wieder begegneten derartige Einmalverweise auf andere Texte. Da die Veweistexte durchaus eine gewisse Streuung hatten (z. B. Exodus, David), stellte sich die Frage, wie sich Streuung und Einmalbezug deuten lassen. Die Frage der zeitlichen Ansetzung stellte sich.<sup>26</sup> Es verbot sich die Annahme, wonach die Josefsgeschichte im Kern sehr alt sei.<sup>27</sup>

Ausblick für Anschlussfragestellungen: Es liegt auf der skizzierten Linie, wenn man das AT in Zweier-Wortketten zerlegt (Wortform1 + Wortform2; Wortform2 + Wortform3 usw.) und sich separat die Fundstellen merkt. Mit diesen Zweier-Wortketten als Element lässt sich die Verteilung im AT errechnen: Wo finden sich *Cluster*, also auffallend dichte Vorkommen derartiger Idiome?<sup>28</sup>

---

25 Die Unterscheidung ist streng einzuhalten: die automatisierte Analyse der Ausdrucksseite lieferte einen Befund, bietet sozusagen eine Hypothese an: zwischen diesen beiden Texten könnte möglicherweise auch bedeutungsmäßig eine Verbindung vorliegen. Die Gleichheit der Wortketten ist nicht bloßer, nichtssagender Zufall, sondern sollte Anlass sein, in einem zweiten Schritt auch auf Bedeutungsebene zu prüfen, ob nicht auch hier der zweite Text einen Wahrnehmungshintergrund liefert, vielleicht sogar einen gewollten, somit eine gezielte Anspielung darstellt. Ein solches Urteil ist mit dem Feststellen identischer Wortketten hier und dort noch nicht gefällt. Der Befund veranlasst jedoch, auch den zweiten Schritt zu tun, also die Bedeutungen hinzuzunehmen.

26 In leicht zugänglicher und leicht lesbarer Form flossen die Befunde in die zusammenfassende Darstellung ein: <http://www-ct.informatik.uni-tuebingen.de/daten/jguebers.pdf> – Anlässlich der Übersetzung von Gen 43,29-31 wird dort auch das Thema der relativen Chronologie behandelt und dargestellt, welche Folgerungen geboten, welche jedoch unwahrscheinlich sind. Im dortigen Literaturverzeichnis werden die die Basis bildenden wissenschaftlichen Argumentationen genannt.

27 Die schon älteren Erkenntnisse, dass manches vom Sprachgebrauch der Josefsgeschichte an Wendungen aus dem Esterbuch erinnert, bekamen der Tendenz nach eine Unterstützung auf breiter Basis. Vgl. Die zusammenfassende Argumentation in SCHWEIZER, H, Nachträge zur Interpretation der Josefsgeschichte: ThZ 54 (1998) 300-324.

28 Bis hierher ist der Gedanke methodisch nicht neu, hat man doch bisher schon typische Redeformen, z. B.



Mit den Befunden kann man einen Schritt weitergehen und sich Abhängigkeiten errechnen lassen: In welchen Text(bereich)en stehen die später für die Clusterbildung typischen Ketten besonders kompakt und dicht beieinander? Es könnte sich darin um Ursprungstexte handeln, die eine sprachliche Tradition begründen. Der Befund lässt sich dann in Form einer *relativen Chronologie* auswerten. Es werden Textbereiche sichtbar, die älter sein müssen im Vergleich zu anderen, jüngeren.<sup>29</sup>

## 2.3 online – Konkordanz

Mit jedem Browser ist erreichbar: <http://www-ct.informatik.uni-tuebingen.de/Comon/www> CoMOn steht für *Corpus Matching Online*

*Technisch* ist zu beachten: Im Betriebssystem des jeweiligen Computers muss – mindestens – *java 1.6.10* installiert sein. Falls das CoMOn-Programm nicht korrekt arbeitet, oder gar abstürzt, dürfte der Grund in einer veralteten Java-Version auf dem Rechner liegen [die Rechenoperationen werden nicht auf dem Server in Tübingen, sondern auf dem lokalen Rechner durchgeführt]. In einem solchen Fall: updaten Sie, oder lassen updaten! Wenn die Java-Version nicht passt, gibt es einen roten Hinweis im CoMOn-Fenster mit weiteren Informationen.<sup>30</sup>

Erster Schritt: *Korpus* wählen. Aktuell steht das hebräische AT zur Verfügung; in absehbarer Zeit sollen das griechische AT, das griechische NT, der Koran-deutsch, der Koran-arabisch hinzukommen. – Weitere Korpora können integriert werden.

Nach *Buch/Kapitel/Vers* Anfang und Ende des Suchtextes wählen. Längenbegrenzung 70 Verse. Wer aus text- oder literarkritischen Gründen sich genötigt sieht den Suchtext zu modifizieren, kann dies tun, solange die geltenden Konventionen für die Wiedergabe der hebräischen Buchstaben beachtet werden.

Bestimmung der *Mindestlänge* der gesuchten strings. Die Voreinstellung ist „3“. Sie kann nach oben oder unten verändert werden. Je höher die Mindestlänge, desto geringer die Zahl der zu erwartenden Treffer.

Start der *Suche*.

Das *Ergebnis* wird zunächst in Tabellenform angezeigt. Eine Zeile zeigt zusammenfassend das Suchergebnis an. „matchings“ = Gesamtzahl der Treffer, „tokens“ = Länge der gefundenen Treffer (Mindestlänge + x), „avg. Length“ = durchschnittliche Länge der Ketten, „duration“ = Dauer der Suche.

„*List matchings*“: durch Anklicken des Buttons wird jeder einzelne Treffer in einer eigenen Zeile aufgeführt. Durch Anklicken der Kopfstichwörter kann die betreffende Spalte sortiert werden: aufsteigend / absteigend.

„*Show context*“: Pro Einzeltreffer (= Zeile), den man durch Klicken markiert hat, kann die Textumgebung des gefundenen Treffers angeschaut werden. Die Anzeige der Texte geschieht in unserer adaptierten Transkription.

---

*deuteronomistische* identifiziert, oder Kultermologie, typische priesterschriftliche Wendungen.

29 Die bio-informatische Genomsequenzierung hat algorithmisch viel von den (älteren) Erfahrungen zur Textanalyse gelernt (z. B. von den verschiedenen Typen einer „Schnellen Suche in Texten“). Umgekehrt kann der boomende Zweig der Bioinformatik nun auch zurückwirken auf die Textarbeit: die Suche nach *phylogenetischen Bäumen* unterscheidet sich im Kern nicht von der Erarbeitung einer *relativen Chronologie*.

30 Das CoMOn-Startfenster gibt einen Hinweis, wo man sich die aktuelle Java-Version beschaffen kann.

„generate conclusion“: Zunächst werden die Bedingungen der zurückliegenden Suche (einzelne Suchparameter) in Erinnerung gerufen. - Dann: Entlang des komplett wiedergegebenen Suchtextes (links) werden die Treffer (Kapitel/Vers) aufgelistet. Bei Schachtelungen werden die betroffenen Textpassagen wiederholt (zunächst die längste gefundene Wortkette, dann Entsprechungen für Teile davon). Wortketten, für die Treffer gefunden worden waren, stehen in einer eigenen Zeile. Das dient der Übersicht. Entsprechend wird dadurch sehr schnell klar, welche Textpassagen *keine* Entsprechung im Korpus gefunden hatten, welche also – in mehrfacher Hinsicht – *einmalig* sind.<sup>31</sup> Durch *button* (links unten) kann man hin und her schalten und dadurch bestimmen: Will man sich den Suchtext in hebräischer Klarschrift oder in Transkription anzeigen lassen?<sup>32</sup>

via »Save« können die Ergebnisse auf der lokalen Festplatte gespeichert werden. Dazu ist ein Dateiname samt der Extension „.html“ anzugeben.<sup>33</sup> Die so gespeicherte Datei kann mit einem Browser geöffnet, gedruckt, in anderer Codierung gespeichert, abgeändert und in ein anderes Dokument integriert werden.<sup>34</sup>

*Ausblick / Angebot* : Auf der Eingangsseite von CoMOn wird informiert u.a. über Korpora, die in absehbarer Zeit ebenfalls angeboten werden sollen. Wer über ein elektronisches Textkorpus verfügt, das er gern in die Online-Konkordanz integriert sähe, nehme mit uns Kontakt auf.

### 3. Verhältnis zu SESB

Zu *Stuttgart Electronic Study Bible (SESB)* bestehen signifikante Unterschiede:

Die *SESB*-Konkordanz erlaubt es, syntaktisch (im traditionellen Sinn) gleiche Strukturen zu definieren und zu suchen: Satzbau, Satzglieder sind im Fokus. Das verlangt Bedeutungsverstehen. Die Fragestellung ist natürlich legitim. Es muss nur klar sein, dass bei *CoMOn* ein völlig anderes Verständnis zugrunde liegt: die (Ausdrucks-)SYNTAX arbeitet mit der syn-tax = *Zusammenstellung* der Ausdrücke, der Wortformen allein. *SESB* verlangt nicht nur, dass man gut das Hebräische beherrscht, es bedeutungsmäßig versteht, sondern auch, dass man mit der nötigen grammatischen Terminologie vertraut ist.

Ganz anders im Fall von *CoMOn*: auch jemand, der des Hebräischen nicht kundig ist, der keine spezifischen Grammatikkenntnisse hat, kann Querverbindungen zu anderen Texten auf der Basis des Hebräischen finden, allein mit Hilfe der Stellenangaben.

*SESB* übergeht die Ebene, die zunächst jeden Leser erreicht und anspricht: die der äusseren Wortformen. Das ist linguistischer Manichäismus.<sup>35</sup>

In *SESB* ist nicht vorgesehen, dass ein ganzer Einzeltext als Suchtext, -kriterium zugrundegelegt

31 Mit einem Blick kann man damit einen Eindruck gewinnen, in welchen Bereichen der Suchtext gängigen Sprachgebrauch, Formeln, in großer Dichte aufgreift, in welchen dagegen kreative, eigene Formulierungen vorliegen.

32 Eine analoge Aufbereitung wird für Septuaginta bzw. NT, sowie für den arabischen Koran angestrebt.

33 Weitere Informationen unter: <http://www-ct.informatik.uni-tuebingen.de/daten/comon.pdf>

34 Der Rückgriff auf Ergebnisse, die mit CoMOn erzielt wurden, soll in den wissenschaftlichen Dokumentationen indiziert werden.

35 Dass Zeichen aus „Körper und Geist“ bestehen, also einem Zusammenwirken von „Ausdruck“ und „Bedeutung“, ist eine alte semiotische Metapher. Folgt man ihr und sieht, dass „Ausdrücke“ üblicherweise immer nur in Verbindung mit „Bedeutungen“ betrachtet werden, die Ausdrucksebene also nie eine eigenständige Analyse erhält, dann liegt ein Missachten der Körperseite der sprachlichen Zeichen vor, eine ständige Bevormundung durch die Bedeutungen. Der Beitrag der Sinne an der Kommunikation kann dann nicht gewürdigt werden.

wird, wobei die Ketten von Wortformen, für die Treffer gefunden werden, eine unterschiedliche Länge haben können. Der Benutzer muss nicht vorab definieren, also im Grund schon kennen, wonach gesucht werden soll. *SESB* ist nützlich für (Satz-)Grammatik-Interessierte, nicht für Textinterpreten. Die *pragmatische Wende* (seit Mitte der 1970er Jahren) wurde von *SESB* nicht vollzogen.<sup>36</sup>

Damit keine falschen Oppositionen entstehen: Das Konkordanz-Programm in *SESB* kann sinnvolle Dienste erweisen, aber auf der Ebene von »Satzbauplänen«, gleichen Inhaltsfunktionen in gleicher Abfolge. Um das Programm zu benutzen, muss man den betreffenden Satz verstanden haben.<sup>37</sup>

#### 4. Übersetzung

Das Thema soll nur angerissen werden, sofern ein Zusammenhang mit den bisherigen Punkten herzustellen ist. Am Beispiel von Gen 37,2 war erläutert worden, wie Konkordanzbefunde (im Rahmen von *CoMOn*) erzwingen, dass die jeweilige Textstelle in einem erweiterten Licht zu sehen ist. Es war oben noch nicht besprochen worden, welche *übersetzungspraktischen* Konsequenzen zu ziehen wären. Denn durch die ausführliche Korpusanalyse steht das Wissen um Querverbindungen zur Verfügung, das nicht wieder ausgeblendet werden sollte, wenn eine Übersetzung im Entstehen ist. Die schwierige Frage ist nur: Wie kann das Wissen integriert werden, ohne dass eine weitschweifige Erörterung statt einer Übersetzung/Übertragung entsteht?

Wir haben im Kontext der Josefsgeschichte eine Serie von Erfahrungen zum Thema „Übersetzung“ gemacht, auch einige Experimente durchgeführt. Das meiste davon ist veröffentlicht. Eine geraffte Darstellung findet sich in der „Einleitung“ des folgenden Manuskripts:

<http://www-ct.informatik.uni-tuebingen.de/daten/jguebers.pdf>

Die Josefsgeschichte – wie angedeutet – ist mehrfach durch solche Exklusiv-Verweise auf weitere Texte geprägt. Die wichtigsten davon sind in dem erwähnten Manuskript (im „Hauptteil“) zusammengetragen (noch präziser in den jeweiligen Ursprungspublikationen). Beim Thema „Übersetzung“ hat sich eine Dreiteilung bewährt:

eine „*Arbeitsübersetzung*“ folgt sklavisch der hebräischen Vorlage. Mit ihr macht der Übersetzer sich und anderen klar, wie er die hebräischen Wortbedeutungen und Satzkonstruktionen versteht. Er gibt sich und anderen nachprüfbar, diskutierbare Rechenschaft.

Eine „*Übertragung*“ nimmt Einsichten zu Stilistik und Begleitwissen auf und versucht, den „Ton“ zu treffen, der höchstwahrscheinlich den alten Text geprägt hatte. Zu diesem Begleitwissen gehören auch die Querverweise (Anspielungen) auf andere Texte, sofern sie sicher erhoben werden konnten.

Ein „*Essay*“ oder förmlicher Kommentar zum Text wird meist zusätzlich nötig sein, da aufgrund

36 Darin teilt das Werkzeug aber das gebrochene Verhältnis, das weite Teile der Theologie mit der Größe *Text* haben. Das Problem ist umfassend, ist nicht auf die Exegese beschränkt, tangiert vielmehr auch die Systematische Theologie, mithin auch die Kirchenstruktur. Vgl. SCHWEIZER, H, „...deine Sprache verrät dich!“ Grundkurs Religiosität. Essays zur Sprachkritik. Forum Religionskritik 1. Münster 2002.

37 Ein Zusatzproblem liegt darin – das soll nicht breit entfaltet werden –, dass zu den Inhaltsfunktionen auch eine Koppelung an bestimmte Ausdrucksmuster verlangt wird. Aber diese Orientierung ist veranlasst durch die traditionelle Grammatikauffassung, die von der Koppelung von Ausdruck und Bedeutung nicht loskommt (*Morphem* ist die kleinste bedeutungstragende Einheit – wie stereotyp betont wird). Wir trennen durch das ganze SLANG-Projekt hindurch (= „*Search for a Learning non-normative Grammar*“ vgl. SCHWEIZER (1995) „Anhang“) klar zwischen *Ausdruck* und *Bedeutung*, was im Fall des Konkordanzprogramms eine viel höhere Effizienz, Transparenz und viele, sonst meist übersehene Resultate einbringt.

der zeitlichen und kulturellen Distanz längst nicht alle notwendigen Einsichten zum Textverständnis in die „Übertragung“ gepackt werden können.

Alle drei Komponenten sind im erwähnten Manuskript - auf die jeweilige Textstelle bezogen - auf einer Doppelseite synoptisch zusammengestellt. Was ist folglich eine „Übersetzung“? Korrekt wäre es zu sagen: alle drei Elemente zusammen. „Übersetzung“ ist folglich nichts, was unmittelbar auf dem Papier steht, sondern eine Bündelung verschiedener Aspekte im Geist des Lesers:

er hat erfahren, wie der Text korrekt zu übersetzen ist (*Arbeitsübersetzung*)

er hat Zusatzinformationen bekommen (*Essay*)

er sah, wie der Exeget aus beidem versucht, einen angemessenen, gut lesbaren, den „Ton“ treffenden deutschen Text zu erstellen (*Übertragung*)

Mit diesen Komponenten ist es denkbar, dass ein Nicht-Exeget seinerseits versucht eine „Übertragung“ zu erstellen. Die notwendigen Informationen stehen ihm zur Verfügung. Keine Version von „Übertragung“ kann beanspruchen, die letztgültige zu sein. Immer neue Versuche können folgen und mittels der drei Informationsquellen diskutiert werden. Eine vierte Quelle spielt mit - und für sie ist nicht der Fachexeget der entscheidende Maßstab: es ist der gekonnte = literarisch sensible Umgang mit der Zielsprache, z.B. dem Deutschen.

Abstract: The problems of electronically encoding non-latin writings such as Hebrew were discussed. The main interest consists of a powerful search procedure in which a searchtext can be defined and where the computer automatically finds all instances of strings in a given corpus. The main characteristics of the online concordance *CoMOn* are introduced. This tool is operating exclusively on the basis of expressions/strings, without any inclusion of meaning. The results will be quasi objective, because they can easily be controlled. But a sophisticated reflection is needed to answer the questions of citation, allusion or relative chronology. Methodically the Online-concordance (working on the basis of strings alone) contrasts the wellknown printed concordances as well as the *Stuttgart Electronic Study Bible*.

Abstrait: L'article aborde les problèmes de la représentation électronique des caractères non-latins, comme l'écriture hebreux. L'interêt principal consiste dans une procédure puissante, qui d'abord défine un texte de base, et ou finalement un grand corpus sera recherché avec le but que toute chaîne de mots dans le texte de base trouvera ses correspondants dans le corpus. *Intertextualité* est le mot-clef dans un sens quasi objectif, car les correspondances trouvées peuvent être contrôlées facilement. Mais une réflexion soigneuse est nécessaire pour en tirer des conclusions convaincantes: concernant la question des citations, des allusions, de la chronologie relative. Concernant la methode le logiciel *CoMOn* se met en contrast en vue des concordances imprimées ou même de la *Stuttgart Electronic Study Bible*.