

## Darstellung morphologischer Zerlegungen hebräischer Wörter

*Martin Schindele – Tübingen*

### 0. Motivation

Wer Texte lesen und verstehen will, muß Wörter in ihre morphologischen Bestandteile zerlegen. Dabei versteht man unter einem Morphem zunächst einen Wortbestandteil, dem sich eine Bedeutung oder eine grammatische Funktion zuordnen läßt. Es liegt nahe, das Herstellen morphologischer Zerlegungen durch den Computer zu unterstützen. Unterstützen heißt in diesem Fall, daß die erforderlichen Programme immer interaktiv funktionieren: Die Maschine macht Zerlegungsvorschläge, der Benutzer wählt die richtigen aus. Die Vorschläge entstehen durch den Versuch, ein Wort in bereits bekannte Bausteine zu zerlegen, oder bereits akzeptierte Zerlegungen ähnlicher Wörter auf das neue Wort zu übertragen.

In diesem Artikel soll untersucht werden, wie morphologische Zerlegungen dargestellt werden können, so daß diese Darstellung einerseits der Formenlehre des Hebräischen entspricht, andererseits für Computerprogramme handhabbar ist. Daneben soll ein einfacher Kalkül zur Erzeugung binärer Zerlegungen durch Wortvergleiche vorgestellt werden und dessen Einsatz zur Darstellung ausdrucks-syntaktische Auffälligkeiten an ein paar Beispielen gezeigt werden.

Der Aufsatz entsteht vor dem Hintergrund meiner Arbeiten als wissenschaftlicher Hilfskraft bei Prof. Dr. Harald Schweizer an der Fakultät für Informatik der Universität Tübingen. Ihm möchte ich diesen Aufsatz zum fünfzigsten Geburtstag widmen.

### 1. Formale Begriffe

#### 1.1 Wort und Morphem

1.1.1 Definition.– Sei  $\Sigma$  ein Alphabet<sup>1</sup>. Unter einem Wort verstehen wir eine Folge  $W = (w_1, \dots, w_n)$  von Elementen aus  $\Sigma$ .  $n$  bezeichnet hier die Wortlänge,  $M_W$  nennen wir die Indexmenge.

<sup>1</sup>Gedacht ist hier an das Bild unseres hebräischen Alphabets in unserer Maschine, das hier durch das Transkriptionsalphabet wiedergegeben wird. Als zusätzliches Zeichen nehme ich ‚•‘ zur Wiedergabe von Dageš und Mappiq.

1.1.2 Definition.– Sei  $W = (w_1, \dots, w_n)$  ein Wort. Unter einem Morphem  $U = (w_{i_1}, \dots, w_{i_k})$  verstehen wir eine Teilfolge von  $W$ , das heißt, es gilt  $i_1 < i_2 < \dots < i_k$  und  $k \leq n$ .  $k$  bezeichnet die Länge der Teilfolge. Wir sagen auch,  $U$  komme in  $W$  vor und schreiben  $U \preceq W$  oder, falls zusätzlich  $U \neq W$  ausgedrückt wird,  $U \prec W$ .  $M_U = \{i_1, \dots, i_k\}$  ist die Indexmenge von  $U$ .<sup>2</sup>

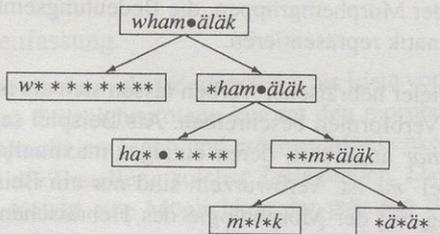
In diesem Zusammenhang kann also jede Teilzeichenfolge eines Wortes als Morphem ausgezeichnet werden. Sowohl  $ti$  als auch  $i \bullet e$  können als Morpheme von  $tim \bullet e$ <sup>3</sup> verarbeitet werden. Die Frage, welche Aufteilungen sinnvoll sind, erfordert Bedeutungswissen und damit interaktiv die Mithilfe des Benutzers.

Die Definition ist so gestaltet, daß zwei Morpheme gleich sein können, auch wenn sie verschiedene Indexmengen haben. Zur besseren Darstellung kann man mit Hilfe eines Platzhaltersymbols (hier,  $*$ ) Auskunft über die Position eines Morphems in einem Wort geben. Man schreibt also  $ti* **$  und  $*i \bullet e*$  statt  $ti$  und  $i \bullet e$ .

1.1.3 Definition.– Sei  $W = (w_1, \dots, w_n)$  ein Wort. Unter einer morphologischen Zerlegung von  $W$  verstehen wir eine Liste  $(U_1, \dots, U_m)$  von Morphemen von  $W$ , deren Indexmengen paarweise disjunkt sind und zusammen die Indexmenge von  $W$  ergeben. Gewöhnlich sollen diese Morpheme nach dem kleinsten Index aufsteigend geordnet sein.

$(t*m**', *i \bullet e*)$  ist also eine sinnvolle morphologische Zerlegung von  $tim \bullet e$ , die Zerlegung in Wurzel und das, was zur Bildung der finiten Verbform nötig ist.

Morphologische Zerlegungen können hierarchisiert werden. In diesem Fall empfiehlt sich ihre Darstellung als Baum:



<sup>2</sup>Vgl. dazu Definition 1.2.4 bei SCHINDELE 1994.

<sup>3</sup>Gen 34,5 u.a.

## 1.2 Gemeinsame Morpheme

1.2.1 Definition.– Seien  $W_1, \dots, W_n$  Wörter. Ein Morphem  $U$  heißt gemeinsames Morphem von  $W_1, \dots, W_n$ , falls  $U \preceq W_1, \dots, U \preceq W_n$  gilt. Gilt darüberhinaus für jedes Morphem  $V$ , das diese Bedingung ebenfalls erfüllt,  $U \not\prec V$ , so heißt  $U$  maximales gemeinsames Morphem von  $W_1, \dots, W_n$ . Hat jedes gemeinsame Morphem von  $W_1, \dots, W_n$  höchstens die Länge von  $U$ , so heißt  $U$  maximallanges gemeinsames Morphem von  $W_1, \dots, W_n$ .

Dabei ist ein maximallanges gemeinsames Morphem einer Wortmenge stets ein maximales gemeinsames Morphem dieser Wortmenge. Es kann aber mehrere maximallange bzw. maximale gemeinsame Morpheme einer Wortmenge geben. Diese Bedingungen erfüllen die Zeichenfolgen *bir* und *biq* für die Wörter *birqiy* und *biqrâb*<sup>4</sup>.

## 2. Anwendungen

Man kann nun die Wörter einer Wortmenge binär (d.h. in jeweils höchstens zwei Morpheme) zerlegen, indem man bei jedem Wort als ein Morphem eine maximale oder maximallange gemeinsame Zeichenfolge der Wortmenge auswählt und den Rest der Zeichen zu einem zweiten Morphem zusammenfaßt. Das Problem, ein maximales oder maximallanges gemeinsames Morphem zu einer gegebenen Wortmenge zu bestimmen, kann als algorithmisch gelöst gelten.

### 2.1 ‚Sinnvolle‘ morphologische Zerlegungen

Unter einer sinnvollen morphologischen Zerlegung verstehe ich hier Zerlegungen in Morpheme oder Morphemgruppen, die Bedeutungseinheiten im Sinne der traditionellen Grammatik repräsentieren.

2.1.1 Die Wurzeln vieler hebräischer Verben lassen sich als maximale gemeinsame Morpheme von Verbformen beschreiben. Als Beispiel seien hier die Wörter *wayyo'mâr* und *le'mor* angeführt, deren einziges maximallanges gemeinsames Morphem die Wurzel *'mr* ist. Verbwurzeln sind nur ein Beispiel dafür, daß der maschinelle Umgang mit der Morphologie des Hebräischen einen Kalkül zum Umgang mit nicht zusammenhängenden Teilzeichenfolgen von Wörtern erfordert.

---

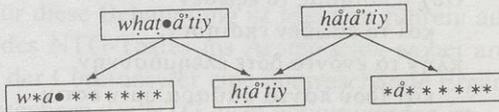
<sup>4</sup>Gen 1,15 und 2 Sam 15,5.

2.1.2 Die Wurzeln vieler hebräischer Verben lassen sich als Komplement gemeinsamer Morpheme von Verbformen beschreiben. Als Beispiel seien hier die Wörter *way•iqrd'* und *way•igdäl* angeführt, deren einziges maximallanges gemeinsames Morphem *way•iä* jeweils das Komplement der Wurzeln *qr'* bzw. *gdI* ist. Dies ist besonders dann von Nutzen, wenn zu einer Wurzel so wenig Formen belegt sind, daß eine Darstellung nach 2.1.1 nicht möglich ist.

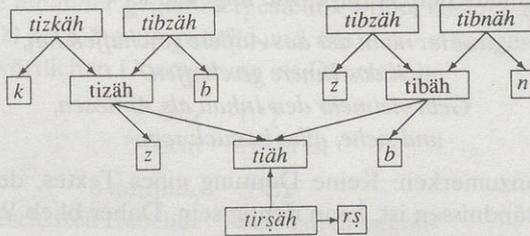
## 2.2 Beschreibung ausdrucksyntaktischer Phänomene

Binäre morphologische Zerlegungen eignen sich auch, um Wortspiele und andere Befunde der Ausdruckssyntax zu beschreiben, bei denen es nicht um die Wiederholung gleicher Wörter im Text geht.

2.2.1 In Ps 51,5f findet sich folgendes Wortspiel:



2.2.2 Etwas komplexer funktioniert das Zusammenspiel von vier Verbformen aus Ps 51,6.18ff:



## 3. Zusammenfassung

Die Darstellung morphologischer Zerlegungen als Liste von (nicht notwendig zusammenhängenden) Teilzeichenfolgen kommt den morphologischen Eigenarten der hebräischen Sprache entgegen und eröffnet Möglichkeiten der maschinellen Unterstützung bei der Erstellung solcher Zerlegungen. Darüberhinaus bietet der zugrundeliegende Kalkül eine Möglichkeit der präzisen Beschreibung von ausdrucksyntaktischen Phänomenen wie Wortspielen oder Wortähnlichkeiten.

### Literatur:

SCHINDELE, Martin; Möglichkeiten und Grenzen maschineller Befunderhebung zur Untersuchung von Formeln und geprägten Wendungen mit Beispielen aus Daniel 8; in: BADER, Winfried (Hrsg.); „Und die Wahrheit würde hinweggefegt“. Daniel 8 linguistisch interpretiert; Tübingen, Basel 1994 (THLI; Bd.9).