

Zur Nutzung von UNICODE in Online-Wörterbüchern

Svetlana VYVENKO

Die Vorteile elektronischer Wörterbücher sind allseits bekannt: schneller Zugang zu den Daten, Hypertextverweise zwischen Wörterbuchartikeln, die problemlos zum Auffinden der nötigen lexikographischen Informationen führen, verschiedene Suchmöglichkeiten, Volltextrecherche usw. Trotz der aufgelisteten Vorteile sehen sich die Entwickler elektronischer Wörterbücher nach wie vor mit einer Reihe von Problemen konfrontiert, die die Nutzung ihrer Produkte erschweren. Dies betrifft v.a. Wörterbücher bzw. Wörterbuchverbände von Sprachen, deren Schriftsysteme Sonderzeichen enthalten.

Im folgenden werden zunächst einige Hilfsmittel aufgelistet und demonstriert, die bei der Entwicklung elektronischer Wörterbücher Verwendung finden. Dann wird gezeigt, welche Suchmethodik in Online-Wörterbüchern benutzt wird, wobei der Schwerpunkt auf etymologischen Wörterbüchern liegen soll. Anhand einiger analysierter elektronischer Wörterbücher soll eine Antwort auf die Frage gefunden werden, wie das Problem der Eingabe von Sonderzeichen in Suchformularen zu lösen ist. Schließlich werden allgemeine Vorschläge zur Verarbeitung von UNICODE-kodierten Sonderzeichen in elektronischen Wörterbüchern gemacht.

Bei der Entwicklung elektronischer Wörterbücher gibt es verschiedene Hilfsmittel der Browser, die man abhängig von den durch sie gebotenen Datenzugriffverfahren in Gruppen unterteilen kann. Einerseits ermöglichen sie es, die gewünschten Objekte (z.B. Stichwörter der Wörterbuchartikel) mithilfe der Maus auszuwählen, andererseits erlauben sie eine Eingabe per Tastatur.

1. Bei der Auswahl der Objekte mit der Maus gibt es folgende Möglichkeiten:

1.1 Auswahl aus den Buchstaben eines “Hypertextalphabets” und Verknüpfung mit den entsprechenden Wörterbuchartikeln (vgl. das Avesta-Wörterbuch von J.H. PETERSON, Abbildung 1).

1.2. Hypertextverweise, die zur Öffnung verschiedener Fenster mit Wörterbuchartikeln führen. Dieses Vorgehen ermöglicht schnellen Vergleich zwischen verschiedenen Wörterbuchartikeln (die letzte Version von Netscape erlaubt es z.B., gleichzeitig bis zu 100 Fenster zu öffnen). Abbildung 2 demonstriert Hypertextverknüpfungen, die als Fenster

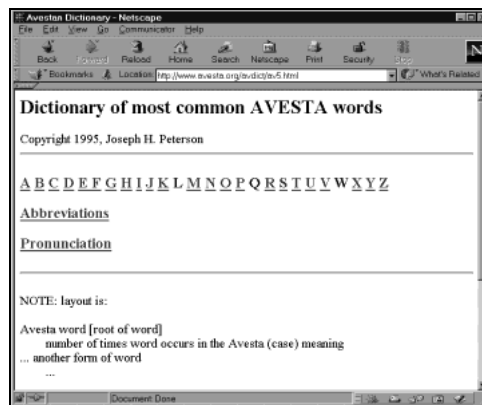


Abbildung 1: Hypertextverknüpfungen von Buchstaben des Alphabets zu entsprechenden Wörterbuchartikeln

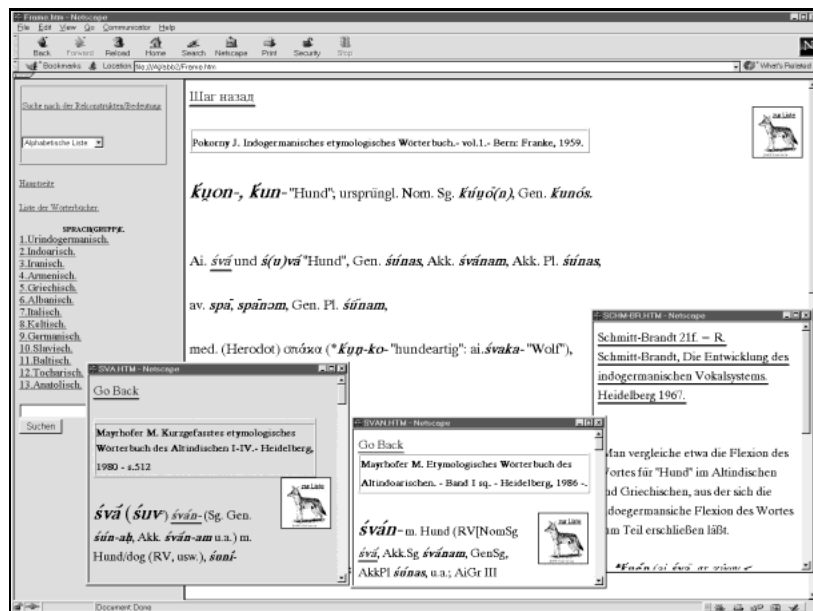


Abbildung 2: Indogermanisches etymologisches Hypertextwörterbuch

repräsentiert sind; sie enthalten Informationen aus verschiedenen indogermanistischen Standardwerken, die ich im Rahmen meines Projekts verarbeitet habe.

1.3. Hypertextverweise von Wortformen, die in einer Auswahlliste (“scrolling-list”) organisiert sind. Nach der Auswahl der Form erscheint der entsprechende Wörterbuchartikel.

2. Bei der Arbeit mit der Tastatur gibt es folgende Möglichkeiten der Auswahl lexikographischer Objekte:

2.1. Die präsentierte Auswahlliste erlaubt die Auswahl der Wortformen anhand des ersten Buchstabens. Nach der Wahl des Buchstabens per Tastatur aktiviert der Cursor das erste Wort, das mit dem gewählten Buchstaben beginnt (s. Abbildungen 3 und 4). Abbildungen 5-8 demonstrieren die Ergebnisse einer Suche von Wortformen, die mit dem Buchstaben “s” oder dessen Varianten anfangen.

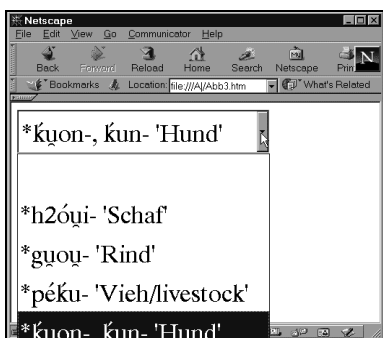


Abbildung 3: Auswahlliste mit UNICODE-kodierten Rekonstrukten



Abbildung 4: Hypertextverknüpfung mit dem entsprechenden Wörterbuchartikel

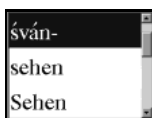


Abbildung 5

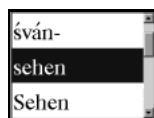


Abbildung 6

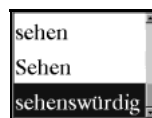


Abbildung 7

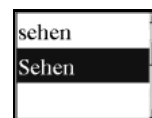


Abbildung 8

2.2. Suchformulare. Die Suche basiert auf einem per Tastatur eingegebenen Wort (vgl. Abbildung 9).

Anhand der Auswahlmöglichkeiten lexikographischer Objekte, die in elektronischen Wörterbüchern angewendet werden, ist es möglich, verschiedene Typen der Suchmethodik zu definieren. In der Arbeit von A. STORRER und K. FREESE¹ wurden bereits “Zugriffsangebote” elektronischer Wörterbücher diskutiert. Auf den Ergebnissen dieser Arbeit und weiteren Analysen von Online-Wörterbüchern beruht folgende Tabelle, die beschreibt, welche Suchmethodik in elektronischen etymologischen Wörterbüchern benutzt wird. Man kann folgende 5 Typen der Suchmethodik unterscheiden:



Abbildung 9

• A – blättrender Zugriff,
 • B – hypertextorientierter Zugriff,
 • C – Suchformulare,
 • D – Kombination der Suchformulare mit verschiedenen Mitteln zwecks Konkretisierung der Suche,
 • E – Kombination der Suchformulare mit Auswahllisten (bei der Eingabe von Wortformen mit Sonderzeichen).

	A	B	C	D	E
The Etymology of First Names ²	+				
Etymological Dictionary of Carnivorous Plants ³	+				
Onomastikon: Dictionary of Names ⁴	+				
MacBain's Gaelic Etymological Dictionary ⁵	+	+	+	+	
Prussian-German-English Dictionary (reconstructed) ⁶	+				
Proto-Romanian Etymological Dictionary ⁷	+				
Etymological Norwegian dictionary ⁸	+				
Etymologisches Indogermanisches Hypertext Wörterbuch (am Beispiel der Haustiere) ⁹	+	+	+		+

Die meisten etymologischen Wörterbücher, die momentan im Internet zugänglich sind, benutzen den sogenannten *“blätternen” Zugriff*. Das heißt, daß am Anfang des elektronischen Dokuments eine Leiste mit Alphabetbuchstaben oder Titelwörtern erscheint, die als Links formatiert sind, so daß man direkt an den Anfang der jeweils gewünschten Artikelstrecke *“springen”* kann. Es ist klar, daß einerseits solche Zugriffsangebote für umfangreiche Materialsammlungen nicht nützlich sind und andererseits das Problem der Eingabe von Wortformen mit Sonderzeichen nur umgangen wird.

Suchformulare wurden als HTML-Bestandteile für die Eingabe von Wortformen per Tastatur entwickelt. Das heißt, daß sie nur die ASCII-Zeichen, die es auf der Tastatur gibt, unterstützen. Verschiedene Wörterbücher im Internet geben Vorschläge, wie Wortformen mit Sonderzeichen, die nicht auf der Tastatur vorhanden sind, eingegeben werden können; so z.B. das Englisch-deutsche Wörterbuch der Chemnitzer Universität¹⁰, das *“virtuelle Tabellen”* benutzt, bei denen Zeichen, die nicht mit der Tastatur erzeugt werden können, durch graphische Knöpfe abrufbar sind.

Es gibt auch Wörterbücher, bei denen nach der Eingabe einer Wortform ohne Sonderzeichen eine ganze Liste von Wörtern aufgeführt wird, die alle möglichen Sonderzeichenkombinationen enthalten¹¹. Eine derartige Suchmethodik könnte man bei der Suche homonymer Formen benutzen (vgl. z.B. Abbildung 10).

Die Analyse der Suchverfahren von Wörterbüchern im Internet hat gezeigt, daß die Frage der Eingabe von Wortformen mit Sonderzeichen noch nicht zufriedenstellend gelöst ist. In einigen Wörterbüchern wird versucht, das Problem der Sonderzeichen durch eine einfachere Gestaltung der elektronischen Seiten zu umgehen. So kann

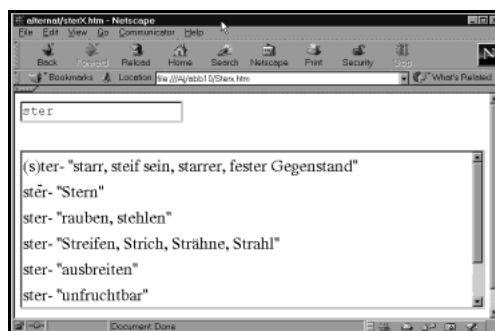


Abbildung 10: Möglichkeit der Suche nach homonymen Formen

sich z.B. die Seitengestaltung auf die Nutzung des “durchblätternen” Zugriffs beschränken, was aufgrund des umfangreichen Materials zu langen Suchzeiten führt und sich negativ auf die Bedienbarkeit auswirkt.

Um das Problem der Eingabe von Sonderzeichen zu lösen, werde ich nun zwei Suchoptionen vorschlagen, die auf der Grundlage von HTML sowie der Programmiersprache JavaScript entwickelt wurden und die UNICODE-Kodierung benutzen. Durch diese optimierten Suchfunktionen können Verfahren, die bereits in Online-Wörterbüchern Verwendung finden, ergänzt werden.

- Die erste vorgeschlagene Suchmethodik besteht aus der Kombination eines Suchformulars mit einer Auswahlliste. Die Wortform wird ohne Sonderzeichen per Tastatur eingegeben, und zwar auf der Basis eines eigens entwickelten Metasystems, in dem jedes Sonderzeichen durch ein oder mehrere ASCII-Zeichen repräsentiert wird. Im Fenster darunter wird das Ergebnis der Suche angezeigt: die kodierte Form, die Form mit Sonderzeichen und die Übersetzung (s. Abbildung 11).

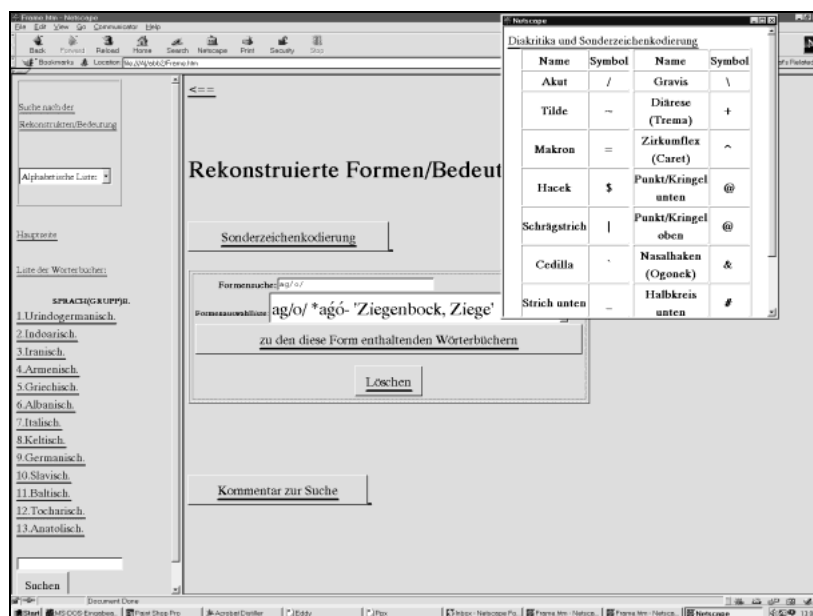


Abbildung 11: Die Suchfunktion mit erhaltenen Sonderzeichen

- Die zweite Suchfunktion wurde anhand des Wörterbuches “Linguistische Termini und ihr Umfeld Russisch-Deutsch” entwickelt. Nachdem dieses Wörterbuch als Wortliste formatiert und in UNICODE kodiert vorlag (s. <http://titus.uni-frankfurt.de/lexica/russterm.htm>), bestand die Aufgabe darin, russische Wörter mit der lateinischen Tastatur einzugeben und so eine Suche zu ermöglichen.

Zuerst wurde hierzu eine Transliterationstabelle erstellt, bei der jeder russische Buchstabe einem lateinischen Buchstaben oder einem anderen auf der Tastatur vorhandenen Zeichen entspricht. Das per Tastatur eingegebene russische Wort wird aufgrund der Transliterationstabelle mithilfe von JavaScript in UNICODE umkodiert (durch Anklicken der Taste “Umsetzen”, s. Abbildung 12). Wenn der Knopf “Suchen” aktiviert wird, erscheint das gesuchte Wort und dessen Übersetzung. Das Anklicken der Taste “Suchen” erlaubt es dann, alle weiteren Zeichenfolgen zu suchen, die das eingegebene Wort enthalten.

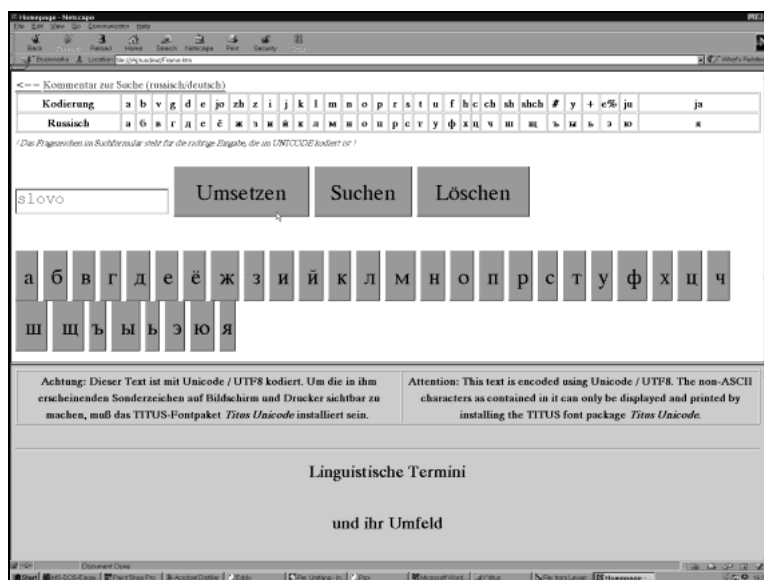


Abbildung 12: Eingabe des russ. Wortes слово mit der lateinischen Tastatur

Außer der Transliterationstabelle wurde alternativ auch eine “virtuelle Tabelle” erstellt, deren graphische Knöpfe je einem russischen Buchstaben entsprechen (s. Abbildung 13).

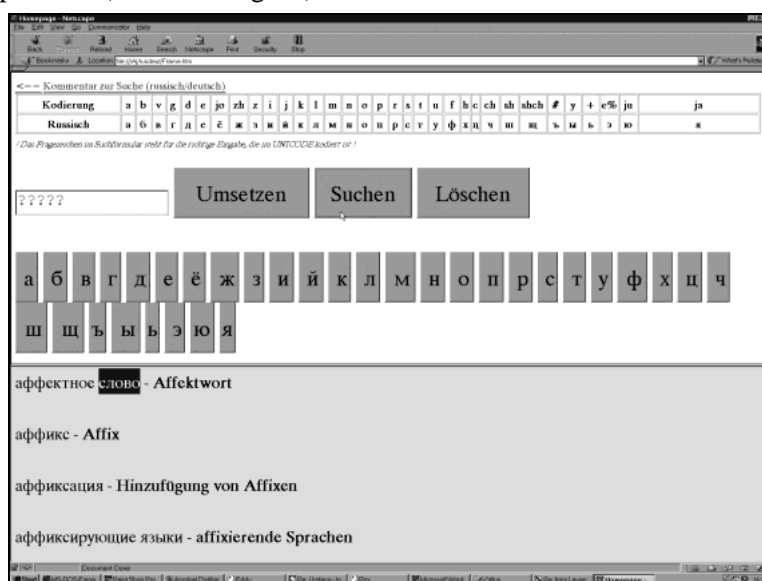


Abbildung 13: Kodierung des Wortes in UNICODE und Suche

Bei diesem Vorgehen muß man in Betracht ziehen, daß Suchformulare, die zur Worteingabe dienen, noch keine Ausgabe von UNICODE-Zeichen erlauben. Deswegen erscheinen nach der Eingabe des russischen Wortes mit der lateinischen Tastatur nur Fragezeichen, die die korrekte UNICODE-Kodierung vertreten. Sobald UNICODE-Zeichen in den Suchformularen unterstützt werden, sollte die Ausgabe russischer Buchstaben möglich sein.

Die gleiche Suchfunktion wurde auch für ein tocharisches Wörterbuch entwickelt. Dieses Wörterbuch basiert auf dem Tocharischen Elementarbuch von KRAUSE/THOMAS¹² und wurde als Textfile von N. Oettinger und S. Rössle erarbeitet. In meiner Anpassung werden die Wörter hier anhand einer Tabelle eingegeben, die eine Transliteration aller Sonderzeichen des Tocharischen enthält. Nach der Umkodierung erscheinen das

gesuchte Wort und seine Übersetzung in einem zweiten Rahmen (s. Abbildung 14).

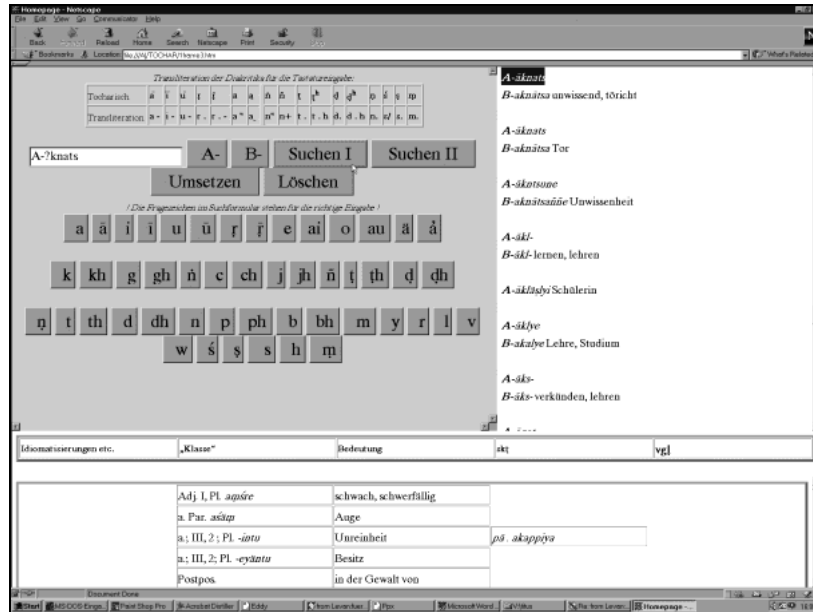


Abbildung 14: Eingabe des Wortes nach der Transliterationstabelle und Suche

Die auf JavaScript basierende Suchfunktion wurde anhand von Wörterbüchern getestet, die in einer Netscape-Umgebung benutzt werden sollen und in Form einer Liste organisiert sind. Die Suche ist dabei nicht auf das Eingabedokument beschränkt, sondern kann sich auch auf andere Dokumente beziehen. Das Beispiel des tocharischen Wörterbuches zeigt einerseits die Suche nach dem tocharischen Wort und seiner Übersetzung; andererseits kann hier die Suche aber auch ausgedehnt werden, so daß aus anderen Dokumenten Informationen über idiomatische Wendungen, etymologische Verbindungen usw. abgerufen werden (s. Abbildung 15).

Die demonstrierte Prozedur versteht sich als ein erster Vorschlag zur Eingabe von UNICODE-Zeichen in Suchmasken.

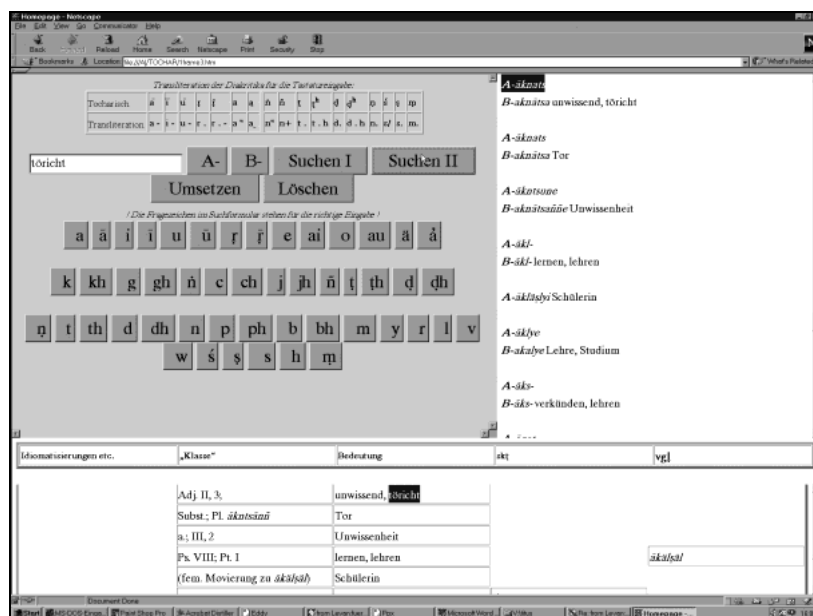


Abbildung 15: Demonstration der zwei Suchrichtungen

Noten

- ¹ STORRER, A. und FREESE, K.: Wörterbücher im Internet. In: *Deutsche Sprache* 24, 1996, 96-153.
- ² <http://www.pacificcoast.net/~muck/etym.html>
- ³ <http://www.flytrap.demon.co.uk/cpdict.htm>
- ⁴ <http://www.fairacre.demon.co.uk>
- ⁵ <http://www.ceantar.org/Dicts/MB2/index.html>
- ⁶ <http://donelaitis.vdu.lt/prussian/reconstructions.htm>
- ⁷ <http://pwl.netcom.com/~aver24/proto.html>
- ⁸ <http://www.dokpro.uio.no/ordboeker.html>
- ⁹ Das Wörterbuch ist noch in der Entwicklungsphase und noch nicht veröffentlicht.
- ¹⁰ Englisch-Deutsches Wörterbuch, Universität Chemnitz (<http://www.tu-chemnitz.de/urz/netz/forms/dict.html>)
- ¹¹ Trésor de la langue française (http://humanities.uchicago.edu/forms_unrest/TLF.html)
- ¹² KRAUSE, W. und THOMAS, W., Tocharisches Elementarbuch, Bd. II, 1964.