

## Slavistik und Internet / Projekt “Kodeks”

### 1. Einführung

1996 kann als das Jahr gelten, in dem das Internet, genauer: das WWW (*world-wide web*, also die spinnenwebartige Verknüpfung von Dateien über Rechner- und Ländergrenzen hinweg), endgültig in die öffentliche Diskussion Eingang gefunden hat: alle großen Wochenzeitschriften widmeten dem Internet im Laufe des Jahres Titelgeschichten oder Serien.<sup>1</sup> Das Titelbild des SPIEGEL zum Thema wurde prämiert. Tageszeitungen widmeten sich ganz selbstverständlich den Neuigkeiten, Problemen und Kuriositäten des Internets, das somit jetzt als gesellschaftliches Phänomen etabliert ist, über das es zu berichten gilt, wobei das Grundwissen darüber, worum es sich bei dem Internet handelt, bei den Lesern schon vorausgesetzt wird. Email-Adresse und die URL (*universal resource locator*, die weltweit eindeutige Internet-Adresse) der eigenen *homepage* hielten ganz selbstverständlich Einzug auf Visitenkarten und Briefköpfen. Die – schon mal selbsternannten – Propheten des Internets wandelten sich hingegen bereits wieder vom Paulus zum Saulus und betätigten sich als Pessimisten des neuen Mediums.<sup>2</sup> Dabei mußte doch eigentlich jedem klar sein, daß viele der Versprechungen, Verlockungen und Argumente, mit denen die Computerfirmen ihre Kunden für das Internet zu gewinnen versuch(t)en, nur leere und gedankenlos nachgeplapperte Phrasen waren: regelmäßig wurde und wird hier die bloße Menge der *Informationen*, auf die man Zugriff hat, gleichgesetzt und verwechselt mit dem *Wissen*, das man besitzt. Die Euphorie, im Prinzip jederzeit auf Berge von Daten zugreifen zu können und auf der eigenen Festplatte anzusammeln, kann verglichen werden mit dem Verhalten eines Studenten, der zur Lektüre empfohlene Artikel erst einmal kopiert und dann mit sich schleppt, letzten Endes aber doch nicht durchstudiert, sondern nur oberflächlich durchliest. Auch die – im ersten Moment erklärliche – Euphorie, über das Internet wissenschaftliche Beiträge problemlos und schnell publizieren zu können,

---

1 SPIEGEL 11/96: Titelgeschichte “Die Welt online: D@s Netz”, FOCUS 35/96: Titelgeschichte und -serie “Das ABC des Internets”.

2 Eine witzige Polemik von Chr. Nürnberger fand sich in der SZ vom 30.8. 1996.

wird inzwischen bereits wegen der vielen ungeklärten Fragen und Probleme kritischer betrachtet.<sup>3</sup>

Während in anderen Philologien die Diskussion um die Möglichkeiten der Nutzung des Internets bereits konkretere Formen angenommen hat<sup>4</sup>, hinkt die Slavistik publikatorisch noch hinterher. Soll man sagen, die Slavistik hinkt der allgemeinen Entwicklung wieder einmal hinterher? Es ist also an der Zeit, sich des Gegenstandes einmal anzunehmen, zumal die Slavistik mit besonderen Problemen zu tun hat, über die sich die Vertreter der großen Disziplinen Germanistik, Anglistik und Romanistik keine Gedanken zu machen brauchen (vgl. unten). Der vorliegende Artikel kann dabei als inhaltliche Fortsetzung einer früheren Publikation (KEMPGEN 1989) verstanden werden, dessen Einleitung heute immer noch so aktuell ist wie damals. In diesem Beitrag ging es insbesondere um den Einsatz eines Hypertext-Systems (HyperCard) für Unterrichtszwecke. Abschließend wurden in dem Artikel eine ganze Reihe von konkreten Themen genannt, die von einer solchen Technik profitieren könnten, z. B. die Vermittlung der russischen Phonetik oder des Kirchenslawischen – beides inzwischen im WWW ansatzweise realisiert.

Die Grundgedanken eines Hypertext-Systems, dies zur Erinnerung, sind die nichtlineare Verknüpfung von Bild, Ton und Sprache. Die theoretische Konzeption eines Hypertext-Systems führte in einer – im Rückblick logisch erscheinenden Abfolge – zunächst zur Realisierung dieser Vorstellungen auf einem Einzelplatzrechner (d. h. einem “Personal Computer” im eigentlichen Sinne) in Gestalt des Programmes HyperCard auf dem Apple Macintosh, um dann mit der Entwicklung des Internets zu dessen WWW-Komponente zu werden: Die Verknüpfung der genannten Informationskomponenten ist nunmehr nicht mehr nur auf einen einzelnen Computer beschränkt, sondern schließt prinzipiell alle Computer ein, die einen temporären oder dauernden Zugang zum Internet haben. Die Verwandtschaft von HyperCard mit dem WWW, wie es die “Browser” genannten Programme dem Nutzer präsentieren, läßt sich auch graphisch veranschaulichen und belegen (**Abb. 1**): Grundelemente der Navigation sind die *Buttons* (symbolisiert als Richtungspfeile), mit denen man (bereits besuchte) Webseiten (HyperCard: die vorhandenen Karten) durchblättern kann, während ein weiterer Button (symbolisiert als Haus, engl. *home*) den Benutzer jederzeit zum Anfang (*Startseite, Home; HomePage*) seiner virtuellen Reise zurückbringt. Das in den Web-Adressen allgegenwärtige *http* zeigt im übrigen die angesprochene Verwandtschaft ebenfalls, denn es steht ja für *HyperText Transmission Protocol*.

---

3 R.P. PEEK/G.B. NEWBY (1996); hierzu vgl. auch die Beprechung von B. Eisenhower in der FAZ vom 31.10.96.

4 Linguistik: CÖLFEN et al. (1996), Anglistik: FELDMANN et al. (Hrsg.) (1996).

Zu der linearen Navigation (vor/zurück) kommen die *Links* hinzu, die eine nichtlineare Verknüpfung von Dokumenten erlauben: ihnen kann der Leser (bzw. Betrachter) jederzeit folgen, muß es jedoch nicht. Gerade sie sind es aber, die die Spezifik des Mediums ausmachen.

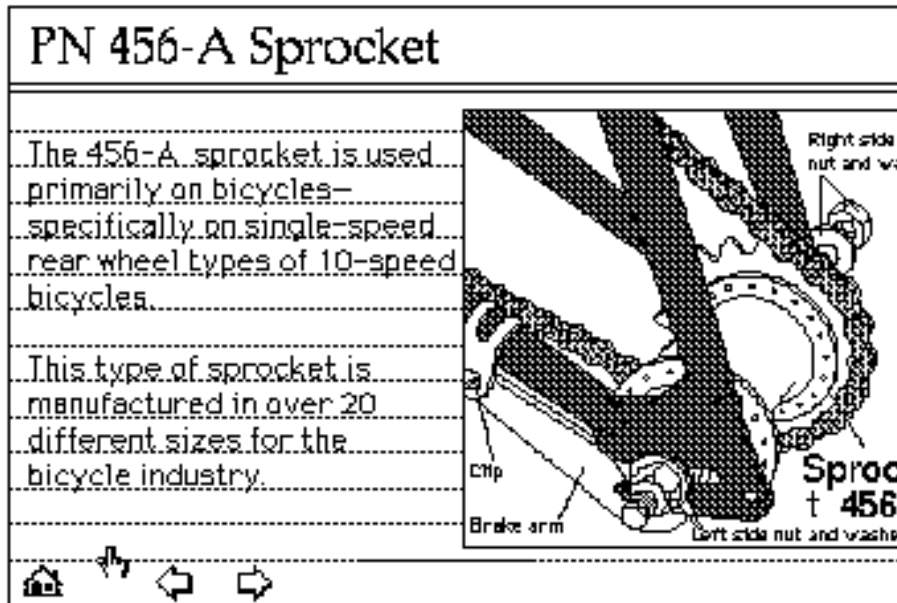


Abb. 1: HyperCard-Karte (oben) und WWW-Browser (Netscape Navigator v. 2, unten) mit Buttons zur Navigation (zurück, vor, Anfang)

## 2. Codierung von Sprache im Computer

Die eingangs erwähnten spezifischen Schwierigkeiten, mit denen ein Slavist zu tun hat, betreffen die Mensch-Maschine-Kommunikation in Verbindung mit slavischen Schriften und slavistischer Umschrift. Im Idealfall soll ja das, was “der Sender” in den Computer in den Zeichen einer natürlichen Sprache eingibt, auch genauso beim “Empfänger” wieder ankomm-

men. Um zu verstehen, wo hier die für die Slavistik bestehenden Schwierigkeiten liegen, ist ein kleiner historischer Exkurs in die Grundlagen der Codierung von Sprache im Computer notwendig, zumal wir uns derzeit mitten in einer Phase des Umbruchs befinden.<sup>5</sup>

## 2.1. Zeichencodierung

Die Entwicklung der Codierung von Sprache im Computer hat drei ganz wesentliche Etappen durchlaufen:

- 7-Bit ASCII
- 8-Bit-Codierungen
- Unicode

Die binäre Codierung alphabetischer Zeichen begann – parallel zur Entwicklung der Computer – mit dem englischen (amerikanischen) Alphabet.  $2^7 = 128$  Kombinationen reichen für alle Groß- und Kleinbuchstaben dieses lateinischen Grundalphabetes samt diverser Satzzeichen aus. Keine weitere Sprache, weder eine westliche noch gar eine osteuropäische, ist in dieser sog. ASCII-Codierung (sprich: *aski*, aufgelöst: *American Standard Code for Information Interchange*), berücksichtigt. Diese 7-Bit-Codierung ist immer noch der kleinste gemeinsame Nenner, auf dem sich alle Computer “verständigen” können, denn dies ist die einzige Codierung, die allen Computersystemen gemeinsam ist, der einzige allgemein verbindliche Standard.

Viele der “Relaisstationen” im Internet, also der Computer, die die eingehenden Informationen an den nächsten Knotenpunkt weiterreichen, verarbeiten Informationen ausschließlich im 7-Bit-Standard. Jede Information, die auf einer Codierung mit mehr Bit beruht, muß, um über das Internet transportiert zu werden, auf eine 7-Bit-Codierung gewandelt werden .

Die nächste Etappe bestand darin, 8-Bit-Codierungen zu verwenden, mit denen bereits 256 Zeichen unterschieden werden können, also doppelt so viele wie zuvor. In allen 8-Bit-Codierungen ist die 7-Bit-ASCII-Codierung enthalten; 128 weitere Zeichen können also zu dem lateinischen Ausgangsalphabet hinzukommen. Dies können z. B. die kyrillischen Buchstaben sein *oder* die lateinischen Zeichen für Westeuropa *oder* die lateinischen Zeichen für Osteuropa – aber nicht alle gleichzeitig. Eine einzelne 8-Bit-Codierung, von denen es Dutzende gibt, heißt unter DOS und Windows auch *Code-Page*. Ein Computerzeichensatz realisiert jeweils eine solche Codierungsvorschrift. Durch Verwendung mehrerer

---

5 Zum folgenden vgl. auch KEMPGEN (1995), wo die spezifischen Probleme einer Codierung des Kirchenslawischen behandelt werden.

Code-Pages, konkret: mehrerer Schriften, läßt es sich erreichen, daß in vielen Sprachen gleichzeitig (d. h. in einem Dokument) geschrieben wird.

Eine solche 8-Bit-Codierung zeigt die **Abb. 2**. Es handelt sich hierbei um die im Internet sehr gebräuchliche sog. KOI-8-Codierung (*KOI = kod obmena informaciej*) des Kyrillischen, die eigentlich nur eine Codierung der ostslawischen Zeichen ist – die besonderen serbischen und makedonischen Zeichen fehlen. An dieser Tabelle sieht man, daß jedem Zeichen eine Dezimalzahl zugeordnet ist, dem kyrillischen *a* z. B. der Dezimalwert 193, während das lateinische *a* den Wert 97 hat. Die Zeichen von dez. 128 bis 148 sind Graphikzeichen (für Rahmen, Tabellen und dergl.), die für Webseiten mit graphischer Oberfläche ohne Belang sind. Wenn man von einer Codierung spricht, meint man also immer eine solche Zuordnung von Sprachzeichen zu einem Zahlenwert; nur mit letzteren arbeitet der Computer direkt.

Leider gibt es für die 8-Bit-Codierungen keinen allgemein akzeptierten Standard<sup>6</sup> – hieraus resultieren alle Schwierigkeiten der Datenübertragung zwischen verschiedenen Systemen, die es heute gibt. Nachdem es auf PCs zunächst sogar Praxis war, daß jedes Programm seinen eigenen 8-Bit-Standard verwenden konnte, bot der Apple Macintosh zum ersten Male eine plattformweite 8-Bit-Codierung, die es ermöglichte, daß zwischen allen Programmen beliebig Text ausgetauscht werden konnte.<sup>7</sup> Mit Windows bekamen dann auch PCs eine plattformweite Codierungsgrundlage – aber eine andere als DOS oder der Macintosh. Diese Unterschiede in den Codierungen waren ursprünglich von den Hardware- und Softwareherstellern durchaus gewollt, denn sie banden die Kunden an die eigenen Produkte, weil bei jedem Versuch, Daten auszutauschen oder zu übertragen, eine Umkodierung erfolgen mußte, was oft bedeutete, daß die “Sonderzeichen” von Hand repariert werden mußten oder Daten schlicht einfach nicht austauschbar waren. Zum Glück hat sich heute die Einsicht durchgesetzt, daß eine programmspezifische Codierungsgrundlage eher hinderlich als förderlich ist, weil sonst plattformweite und erst recht plattformübergreifende Ressourcen nicht richtig genutzt werden können.

---

6 Die Betonung liegt hier auf “allgemein akzeptiert”, denn einen zur (ISO-)Norm gewordenen Standard gibt es inzwischen sehr wohl.

7 Die Macintosh-Codierungen sind dokumentiert in APPLE COMPUTER (1992, 263ff.). DOS-Codierungen finden sich in MICROSOFT CORP. (1984-95, 237ff.). Die von Windows 3.1 unterstützten CodePages sind im Windows-Handbuch nicht dokumentiert; entsprechende Unterlagen sind jedoch beim technischen Support von Microsoft für Osteuropa erhältlich. Die Unicode-Codierung ist als Buch veröffentlicht (UNICODE CONSORTIUM 1991-92).

Cyrillic KOI-8															
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
█	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
@	А	В	С	Д	Е	Ф	Г	И	Ј	К	Л	М	Н	О	
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
Р	Q	Р	Ѕ	Т	U	У	Ш	Х	У	Z	[	\	]	^	_
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
р	q	r	s	t	u	у	ш	х	у	z	{		}	~	
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
-		г	г	л	ј	ђ	ђ	т	ј	ђ	■	■	■	■	■
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
▯	▯	▯	г	л	•	√	∞	≤	≥		Ј	е	а	•	÷
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
			ё	є		і	ї								
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
	г	Г	Ё	Є		і	ї					Ў	ў	№	©
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
Ю	а	б	ц	д	е	ф	г	х	н	й	к	л	м	н	о
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
п	я	р	с	т	у	ж	в	ь	ы	э	ш	э	щ	ч	ъ
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	Н	Й	К	Л	М	Н	О
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
П	Я	Р	С	Т	У	Ж	В	ь	Ы	Э	Ш	Э	Щ	Ч	Ъ

Abb. 2: KOI-8-Codierung des Kyrillischen

Die nächste Etappe zeichnet sich ab und ist in Entwicklung. Sie heißt *Unicode* und soll die Lösung für alle Codierungsprobleme bringen. Diese Initiative vieler Computerhersteller schlägt eine 16-Bit-Codierung als gemeine Basis vor; mit ihr können praktisch alle Schriftzeichen der Welt erfaßt werden (in Bezug auf das Japanische und Chinesische gibt es aber auch hierbei noch Kompromisse). Eine ausgezeichnet Übersicht über den Hintergrund dieser Entwicklung bietet SHELDON (1991). In Version 1 des Unicode-Standards waren die slavistischen Bedürfnisse schon weitgehend berücksichtigt; Ausnahmen bildeten die Glagolica, die noch gar nicht kodiert war, sowie bestimmte Bereiche des (kyrillischen) Altkirchenslawischen, das einfach als moderne Kyrillica plus Zusatzzeichen

verstanden wurde, was diesem Schriftsystem vielleicht doch nicht ganz gerecht wird. Version 2 von Unicode bietet hier jedoch schon einen Fortschritt. Für die Philologen aller Disziplinen ist Unicode auch insofern interessant, als prinzipiell jedermann Vorschläge an das Unicode-Consortium mit dem Ziel richten kann, weitere Zeichen in diesen Standard aufzunehmen. Gegenwärtig bieten bereits Windows NT wie das Mac OS partielle Unterstützung von Unicode; sie wird in den nächsten Jahren systematisch ausgebaut werden.

7-Bit:	A - Z, a - z, 0 - 9, ( [ ! ? ] ) ... [2 <sup>7</sup> = 128] "ASCII"
8-Bit:	A - Z, a - z, 0 - 9, ( [ ! ? ] ) ... [2 <sup>8</sup> = 256] Ä Ö Ñ É Œ ä ö ß ñ æ ... "—" »•« oder A - Z, a - z, 0 - 9, ( [ ! ? ] ) ... А Б В Г Д а б в г д ... "—" »•« oder A - Z, a - z, 0 - 9, ( [ ! ? ] ) ... А Ё Ъ Ы Ь а ё ъ ы ь ... "—" »•«
Unicode:	Alle Zeichen der Welt [2 <sup>16</sup> = 65.536]

Resümieren wir diesen kurzen Überblick, so läßt sich folgendes festhalten: Die 7-Bit-Codierung ist eigentlich Vergangenheit und überholt, aber immer noch aktuell, vor allem wegen des Internets. Es kann mit Fug und Recht behauptet werden, daß das Internet in dieser Hinsicht technologisch sogar einen Rückschritt darstellt. 8-Bit-Codierungen sind der gegenwärtige Stand der Technik, aus dem jedoch viele Probleme, gerade für die Datenverarbeitung der slavischen und anderen nichtwestlichen Sprachen, resultieren. Diese Probleme werden sich erst erledigen, wenn durchgängig auf allen Computerplattformen und von allen Anwendungsprogrammen Unicode unterstützt wird. Dies ist aus heutiger Sicht Zukunftsmusik, die innerhalb der nächsten Jahre noch keine Realität werden wird, denn dazu müssen *sämtliche* Betriebssysteme und *sämtliche* Anwendungsprogramme entsprechend umgeschrieben werden.

Eine auf Systemebene angesiedelte Lösung, die beim nahtlosen Übergang auf Unicode notwendig ist, ist für den Macintosh in Entwicklung: hier soll System 8 (Veröffentlichung Mitte '97) einen *Text Encoding Converter* enthalten, der es Anwendungsprogrammen ermöglicht, vom Benutzer unbemerkt zwischen praktisch allen Codierungen aller Computerplattformen, besonders aber zwischen Unicode und allen 8-Bit-Codierungen, zu wandeln. Notwendig ist dies u. a. auch für alle Entwicklungen in *Java*, dieser plattformunabhängigen Programmiersprache, denn die legt intern Unicode zugrunde.

## 2.2. Zeichencodierung und Tastaturtreiber

An dieser Stelle sei nur kurz noch ein häufiges Mißverständnis angesprochen: von “normalen” Computernutzern werden oft *Tastaturbelegung* und *Codierung* verwechselt, was auch nicht weiter erstaunt. Dieses Mißverständnis läßt sich im Prinzip schon an den Unterschieden zwischen der amerikanischen und der deutschen Computertastatur demonstrieren, auf denen ja *z* und *y* vertauscht sind, sei hier jedoch noch einmal anhand der KOI-8-Codierung illustriert (vgl. **Abb. 3**).

Die Tastaturbelegung (der Tastaturtreiber) ist eine Softwarekomponente, die zwischen der Codierungstabelle und der eigentlichen Tastatur (der Hardware) vermittelt. Drückt der Benutzer eine Taste, so erfährt das Betriebssystem anhand eines für jede Taste spezifischen Codes zuerst einmal nur, welche Taste gedrückt wurde. Und zwar nicht in der Form, daß mitgeteilt wird, die Taste *a* sei gedrückt worden, sondern (beispielsweise) “die zweite Taste in der dritten Reihe von oben auf einer ISO-Tastatur mit 102 Tasten”. Der Computer “weiß” also nicht, was oben auf einer Taste aufgeprägt ist. Würde man beispielsweise die Tastaturkappen für *z* und *y* abhebeln und miteinander vertauschen, so würde sich die Eingabe trotzdem nicht ändern – immer noch würde die Taste rechts neben dem *t* ein *z* produzieren und die Taste neben dem *x* ein *y*.

Der jeweils aktive Tastaturtreiber sagt dem Betriebssystem im nächsten Schritt, welchem Dezimalcode der zuvor registrierte Tastendruck zugeordnet werden soll. Die Abbildung zeigt als Beispiel zwei solcher Tastaturtreiber für kyrillische Schriften in KOI-8-Codierung, einen russischen und einen nach phonetischer Zuordnung. Der Dezimalcode des kyrillischen *a*, der von der Codierung vorgegeben ist, wird – durch die beiden Tastaturtreiber vermittelt – durch unterschiedliche Tasten angesprochen.

Tastaturbelegungen sind letzten Endes für das Problem der Datenübertragung und des Datenaustausches vollkommen irrelevant, denn sie betreffen allein die Art der Eingabe, gehören also zum *UI*, dem *User Interface*, also den Komponenten, die die Interaktion des Nutzer mit dem Gerät bestimmen. Wichtig sind für alle Probleme der Datenübertragung und des Datenaustausches jedoch allein die zugrundeliegenden Codierungen der Zeichen. Zwei Leute können also an ein und demselben Text arbeiten und dabei verschiedene Tastaturtreiber benutzen – einmal eingegeben, sieht man dem Text ja nicht mehr an, wie die Eingabe erfolgte.



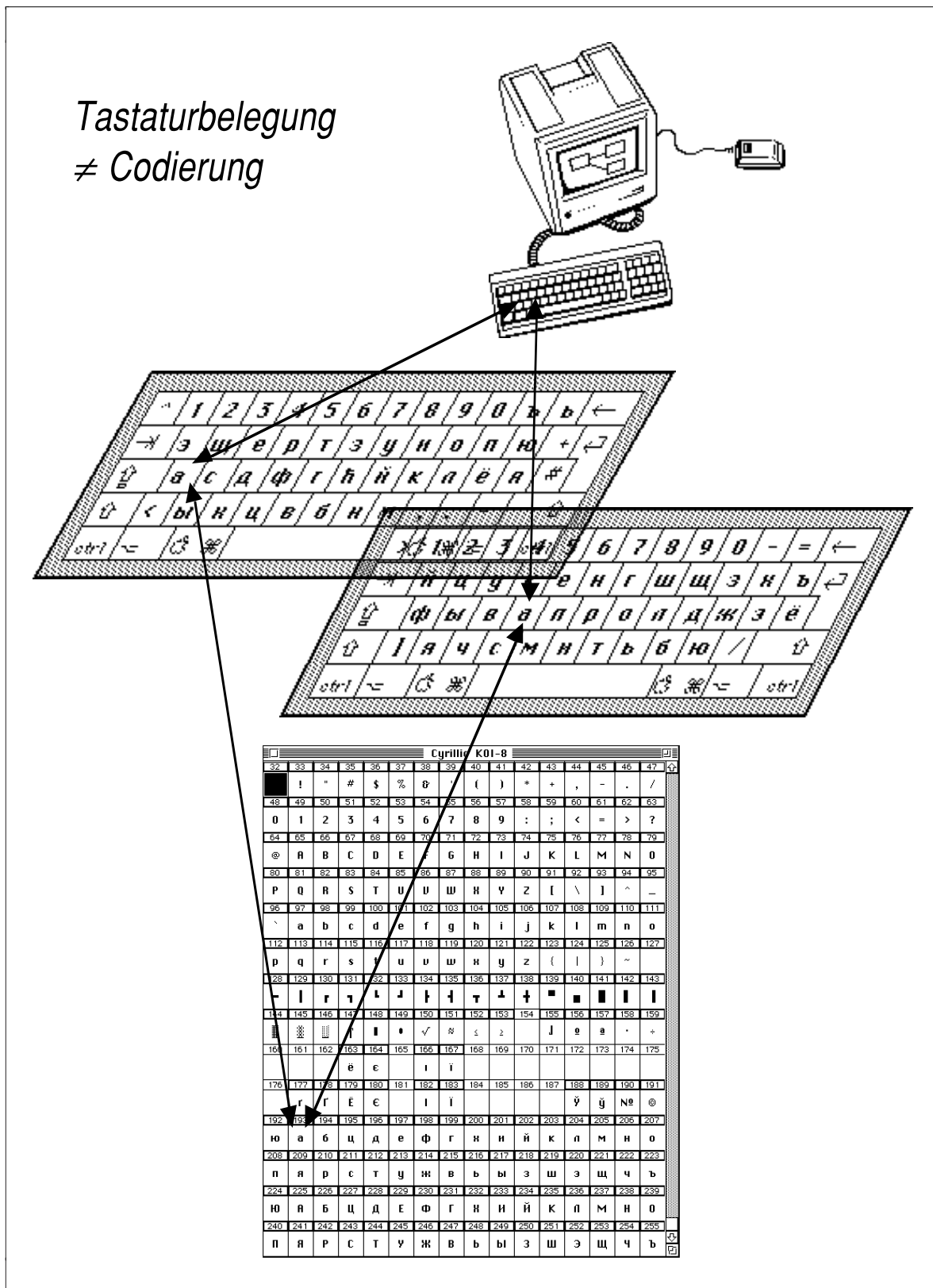


Abb. 3: Codierung, Tastatur und Tastatortreiber

### 2.3. Probleme des Datenaustausches

Von den vielen Problemen des Datenaustausches zwischen verschiedenen Computerplattformen, die es derzeit noch gibt (die aber prinzipiell lösbar sind), sei nur ein Fall demonstriert (vgl. **Abb. 4**). Derzeit ist der Stand der Programmentwicklung so, daß die “großen” Anwendungsprogramme die Standardcodierung der beiden Plattformen Windows und Macintosh kennen und zwischen ihnen automatisch vermitteln. Wird beispielsweise auf einem Macintosh das Wort “Müßiggang” geschrieben, bei dem *ü* und *ß* nicht zum 7-Bit-ASCII-Zeichensatz gehören, so entsprechen diesen beiden Zeichen die Dezimalwerte 159 und 167. Wird nun dieser Macintosh-Text auf einem PC unter Windows geöffnet, so versucht ein “intelligentes” Programm sicherzustellen, daß der Empfänger genau das sieht, was der Sender eingegeben hat. Da aber den Zeichen *ü* und *ß* im Windows-Zeichensatz die Dezimalwerte 252 und 223 zugeordnet sind, erfolgt quasi im Fluge eine Umcodierung: aus 159 wird 252, aus 167 wird 223. Dies alles geschieht automatisch zwischen dem Augenblick, in dem man auf “Öffnen...” klickt, und dem Augenblick, in dem der Text auf dem Bildschirm erscheint.

Diese Umcodierung funktioniert gut für alle westlichen Sprachen, da die Schnittmenge der sowohl auf dem Macintosh wie unter Windows unterstützten Sprachen und Alphabete fast gleich ist. Wiederum aber sind die slavistischen Bedürfnisse nicht “der Normalfall”.

Geht man aber mit der gleichen Logik nämlich an einen Text heran, der eben nicht in einer westlichen Sprache geschrieben wurde, so wird ein sinnvoller Ausgangstext vollkommen verstümmelt. Dies demonstriert die untere Hälfte der **Abb. 4**. Wird etwa unter Windows das Wort *Moskva* in der Windows-Standardcodierung des Kyrillischen geschrieben, so entsprechen diesen Zeichen die angegebenen Dezimalwerte und diesen wiederum die lateinischen Zeichen der darunterstehenden Zeile (*i*, *î* usw.). Geht das verwendete Programm – in diesem Falle fälschlich – davon aus, es handele sich um Text in einer westlichen Sprache, so liefert die automatische Umcodierung auf dem Macintosh zwar die gleichen lateinischen Buchstaben, denen aber ganz andere kyrillische Zeichen gegenüberstehen. Heraus kommt also (letzte Zeile) völliger Unsinn.

Wie man diese Probleme lösen kann, dürfte nach den vorangegangenen Ausführungen auch klar sein: jedem Zeichen müßte eben ein eindeutiger Wert zugeschrieben sein (wie z. B. in Unicode), oder es müßte der eingegebene Text nicht nur mit Marken für die typographischen Auszeichnungen versehen werden, sondern auch mit Markierungen für die verwendete Sprache und Codierung (von der z. B. auch Silbentrennung, Rechtschreibprüfung etc. abhängen).

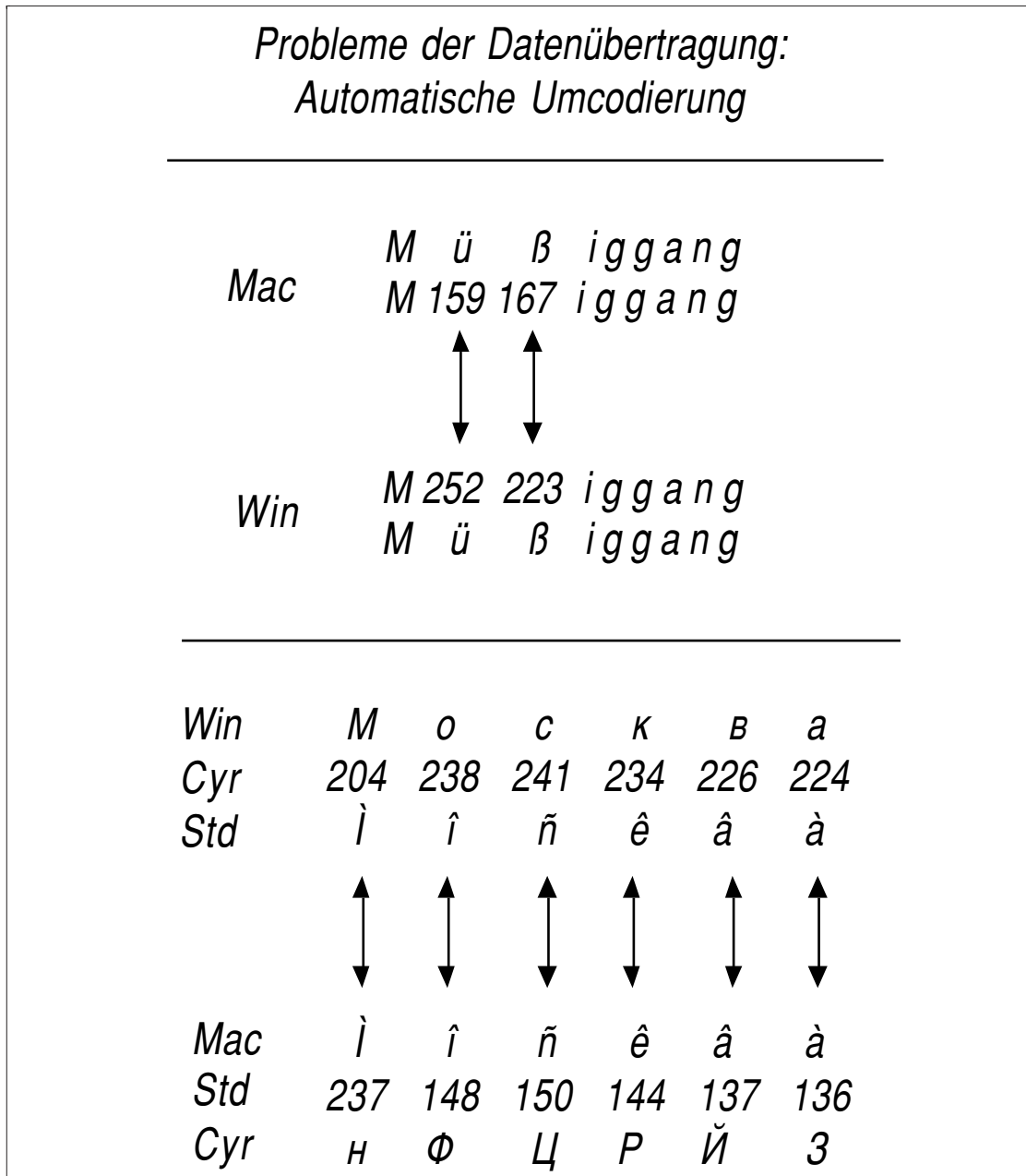


Abb. 4: Datenaustausch und Umcodierung

### 3. Zeichenübertragung im Internet

Da die 7-Bit-Codierung der kleinste gemeinsame Nenner ist, auf dem das Internet operiert, muß jede Information, die Zeichen einer erweiterten 8-Bit-Codierung nutzt, zunächst auf dem Ausgangsrechner in einen 7-Bit-Code verwandelt werden, der dann zum nächsten Gerät übertragen wird. Beim Empfänger wird dann der 7-Bit-Datenstrom rückübersetzt in eine 8-Bit-Codierung.

Als primitivstes Verfahren ist hierbei eine Datenbehandlung anzusehen, die nicht verlustfrei ist: eine Möglichkeit besteht z. B. darin, sämt-

liche Diakritika einfach wegzulassen und nur das Grundzeichen zu übertragen, z. B. anstelle eines *ä* nur ein *a*. Bei diesem Verfahren kann natürlich der Ursprungstext gar nicht mehr mechanisch richtig rekonstruiert werden, obwohl ein menschlicher Leser ihn u. U. doch einigermaßen rekonstruieren könnte. Üblicher sind glücklicherweise (mehr oder weniger) eindeutige Übertragungsmethoden; bei ihnen wird z. B. ein Zeichen aus der oberen Hälfte einer 8-Bit-Codierung<sup>8</sup> in eine *Folge* von 7-Bit-Zeichen gewandelt, die eindeutig rückübersetzt werden kann.

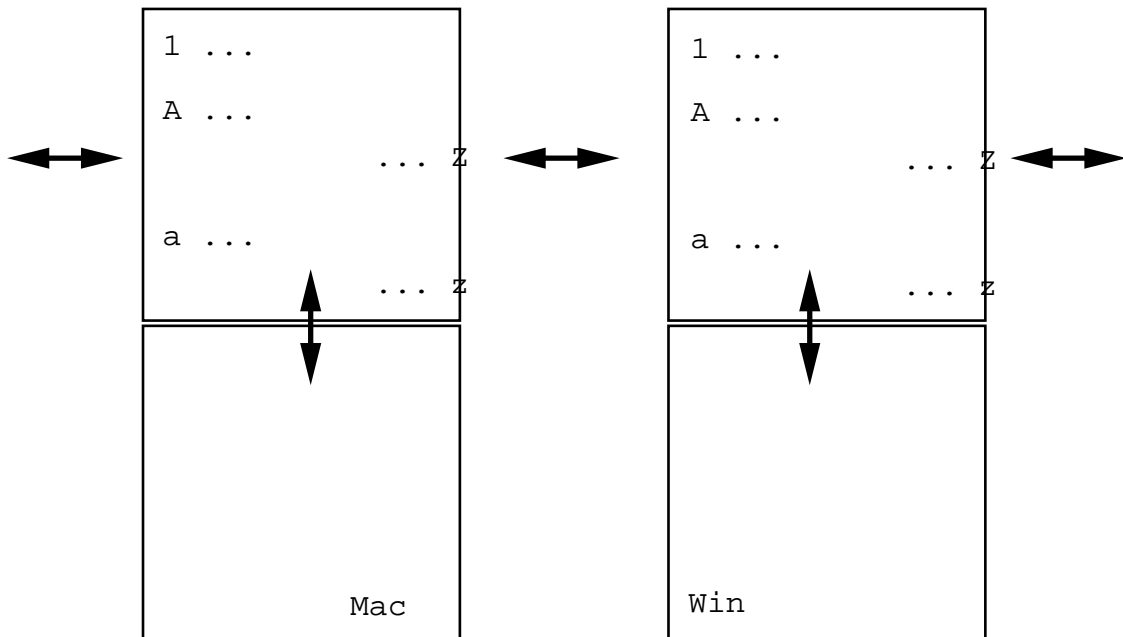


Abb. 5: Wandlung von 8-Bit-Zeichen zur Datenübertragung zwischen Geräten mit unterschiedlicher Zeichencodierung

In der illustrierenden Graphik (**Abb. 5**) stehen die beiden Blöcke für unterschiedliche Computersysteme, hier Macintosh (links) und Windows (rechts). Auf beiden Systemen liegen die deutschen Umlaute an verschiedenen Stellen der normalen Zeichentabelle. Zur Übertragung werden sie zunächst auf dem jeweiligen Gerät in 7-Bit-Code gewandelt und beim Empfänger richtig rückübersetzt. Dies alles geschieht ohne Interaktion des Benutzers.

Festzuhalten ist also: im Internet kann (derzeit jedenfalls) grundsätzlich nur das lateinische Grundalphabet übertragen werden; weder Informationen im kyrillischen Alphabet noch in einer sonstigen slavischen Sprache können direkt übertragen werden. Wenn dies doch “irgendwie”

8 In der Graphik natürlich die *untere* Hälfte; *obere* Hälfte deshalb, weil die Zeichen dieser Tabellenhälfte die höheren Dezimalwerte besitzen (man spricht deswegen auch vom *high ASCII* oder den *hohen ASCII-Werten*).

funktioniert, dann nur deshalb, weil hierzu spezielle Verfahren eingesetzt werden.

Von all dem, was sozusagen “hinter den Kulissen” vor sich geht, merkt der Benutzer erst etwas, wenn etwas einmal nicht so funktioniert, wie es sollte. Hierfür gibt es genügend Beispiele, die fast jeder Email-Nutzer kennt. Zum Beispiel kann folgendes in einer Email-Nachricht enthalten sein (echte Beispiele):

<b>m=FC=DFte</b>	(müßte)
<b>=FCbrigens</b>	(übrigens)
<b>Gr=FC=DFe</b>	(Grüße)

Hier sind die in 7-Bit-Codes gewandelten 8-Bit-Zeichen nicht richtig rückübersetzt worden; stattdessen wurden einfach die 7-Bit-Codes ausgegeben. Eine Folge wie “=FC” z. B. gibt im Hexadezimalcode die Position des Zeichens *ü* in der Codetabelle an. Glücklicherweise lassen sich solche Fälle in der alltäglichen Kommunikation durch den Kontext meist eindeutig reparieren – dank der jeder natürlichen Sprache eigenen Redundanz.

Ein anderer Fall liegt vor, wenn man etwa folgendes erhält:

<b>MenÅ</b>	(Menü)
<b>UniversitÑt</b>	(Universität)

Hier erfolgt die Ausgabe zwar richtig in 8-Bit-Zeichen, jedoch in den falschen. Solche Fehler entstehen in der Regel als Folge einer – vom Benutzer beeinflussbaren – Einstellung im verwendeten Email-Programm, die umgangssprachlich etwa folgendes besagen würde: ‘versuche einfach mal, 8-Bit-Zeichen zu übertragen und riskiere, daß dabei etwas schiefgeht’. Der konkrete Fall tritt beispielsweise auf, wenn Post von einem Windows-PC an einen Mac-Nutzer geht und die Einstellungen nicht übereinstimmen.

Web-Seiten gehen noch einen anderen Weg, um einen erweiterten Zeichensatz zu übertragen: sie benutzen sog. *Tags*, d. h. metasprachliche Marker, die alles das einklammern, was nicht als objektsprachliches 7-Bit-Zeichen zu verstehen ist. Die zulässigen Tags sind allgemeinverbindlich festgelegt, und zwar in der sog. *HTML-Norm*. HTML steht hierbei für *HyperText Markup Language*. HyperText verweist ein weiteres Mal auf die Hypertext-Grundidee des *world-wide web*, *markup language* steht für das Grundprinzip des *tagging*. Wiederum an einem konkreten Beispiel gezeigt, sieht die Übertragung eines natürlichsprachigen Wortes in einem 7-Bit-Code hierbei so aus:

**Gr&uuml;&szlig;e** (Grüße)

Das Zeichenpaar “&...;” ist einer der HTML-Tags; es besteht selbst aus 7-Bit-Zeichen, wird aber nicht objekt-, sondern metasprachlich interpretiert. Zwischen diesen Tags steht im HTML-Code für Zeichen, die nicht im ASCII-Zeichensatz enthalten sind, ein leicht entschlüsselbarer Deskriptor: “uuml” benennt zunächst das Grundzeichen *u*, dann das diakritische Zeichen *uml(aut)*, “szlig” steht für die *Ligatur* von *s* und *z*, d. h. *ß*. Solche Deskriptoren zu verwenden, ist vom Prinzip ein sehr guter Gedanke. Während nämlich ein Hexadezimalcode oder Dezimalcode, sollte er dem Benutzer zu Gesicht kommen, in der Regel auf Unverständnis stößt und jedenfalls doch geräteabhängig ist, kann ein solcher Deskriptor selbst dann verstanden werden, wenn er einmal nicht richtig übersetzt werden sollte. Im Prinzip kann man mit diesem Verfahren jedes Zeichen übertragen, wenn einige Grundprinzipien – etwa zur Benennung der Diakritika oder des Schriftsystems – festgelegt sind. Es ist von der Konzeption her nicht einmal notwendig, daß die fraglichen Zeichen auf dem Empfängergerät überhaupt darstellbar sind.<sup>9</sup>

Da nun nicht auf jedem Ausgabegerät das Vorhandensein aller erwünschten slavistischen Zeichen vorausgesetzt werden kann, müssen sich die Ersteller slavistischer Web-Seiten gegenwärtig meist irgendwie mit dem normalen West-Zeichensatz behelfen.

Die allereinfachste und vielfach verwendete Lösung besteht darin, alle speziellen slavistischen Diakritika schlicht wegzulassen und nur das Grundzeichen zu benutzen.<sup>10</sup> Eine andere Möglichkeit ergibt sich durch kreativen Ad hoc-Gebrauch der im Standard-Zeichensatz vorhandenen Komponenten. So kann man z. B. ersatzweise auf eine Zeichenkombination zurückgreifen, wenn eine bestimmte Zeichenkombination nicht zur Verfügung steht. So kann man für *š* beispielsweise *^s* oder *s^* schreiben. Oder man verwendet, vor allem im englischsprachigen Raum, die dort übliche Transkription des Russischen, also *kh*, *cz*, *sh*, *zh* usw. für die Zischlaute; im französischsprachigen Raum ist *è* als Transliteration für das *e oborotnoe* in Gebrauch, was als Ersatz für normalerweise fehlendes *e mit Punkt* genommen werden kann. Wiederum andere Autoren verdoppeln einfach das Grundzeichen, um das Vorhandensein eines Diakritikons

---

9 Hierin zeigt sich die enge Verwandtschaft von HTML zu SGML, was soviel heißt wie *Structuralized Generalized Markup Language*. Dies ist ein sehr viel ambitiöseres Codierungssystem, das im Prinzip einer Seitenbeschreibungssprache entspricht. Zu Grundlagen und philologischen Anwendungen vgl. auch BIRNBAUM et al (1995).

10 Aus diesem Grund ist schon vermutet worden, daß es langfristig eine Rückkopplung von diesen Beschränkungen des Internets auf die orthographischen Normen einzelner Sprachen geben könnte. Digraphen wie *sz*, *cz* oder *rz*, die aus Zeichen des 7-Bit-Zeichensatzes bestehen, sind nämlich von Haus aus “Internet-tauglich”, während es die Diakritika nicht sind.

anzudeuten und überlassen dem Leser ihrer Web-Seiten die richtige Rekonstruktion. So findet man dann z. B. Artefakte wie *Ccakavski, mozze se, Sveuccilissna naklada* auf kroatischen Webseiten. Wer kroatisch kann, wird sich in einem solcherart verfremdeten Text durchaus zurechtfinden können. Daß dies ebenfalls nur eine Ad hoc-Lösung ist, die nicht wirklich befriedigen kann, ist jedoch nicht zu leugnen.

#### 4. Slavistische Quellen im Internet

Natürlich kann es nicht Sinn und Zweck dieses kurzen Beitrages sein, einen Überblick über die wichtigsten slavistischen Quellen im Internet auch nur zu versuchen. Bei dem sich ständig wandelnden Gehalt des Internets wäre dies von vornherein ein aussichtsloses Unterfangen. Stattdessen sei nur kurz auf einige wichtige Quellen zur allgemeinen Orientierung verwiesen, wobei nur akademische, nichtkommerzielle *Web-Sites*<sup>11</sup> berücksichtigt wurden.

Die HomePages der einzelnen slavischen Seminare stellen für den individuellen Web-Surfer naturgemäß einen naheliegenden Ausgangspunkt dar. Diese Seiten sind im Grunde elektronische "Schwarze Bretter", geben aktuelle Informationen zu den Veranstaltungen des Instituts, zu Gastvorträgen, zu Studiengängen, Studien- und Prüfungsordnungen, d. h. ergänzen die mündliche Studienberatung, werben für ein Seminar, sind also auch auf prospektive Studenten (Wechsler und Anfänger), auf "Auslenkung", gerichtet. Eine Seite mit Links zu allen Seminaren ist innerhalb einer systematischen Fächerübersicht verfügbar:

■ <http://www.rz.uni-karlsruhe.de/Outerspace/VirtualLibrary/491.de.html>

Wer in dieser Liste mit seinem Institut vertreten sein will, muß sich allerdings aktiv um die Aufnahme seines Links bemühen.

Slavistische Links bieten mehrere große Web-Sites. Allen voran sei das *REES-Web* genannt (*Russian and East European Studies*), eine vorzügliche, gut gepflegte Site der University of Pittsburgh, die nicht nur Links zu allen möglichen Quellen versammelt, sondern auch kommentiert, so daß gleich eine erste Abschätzung der Relevanz möglich ist, bevor man einem der angegebenen Links folgt:

■ <http://www.pit.edu/~cjp/rslang.html>

---

11 Web-Site: Allgemein eine zusammenhängende Gruppe von Web-Seiten innerhalb einer *Domäne*, wobei jede Seite ein einzelnes, auf einem Computer gespeichertes Dokument darstellt.

Slavistisches Material selbst ist in Form von belletristischen Texten, Zeitungen, landeskundlichem Material, Unterrichtshilfen, Programmen für den computergestützten Unterricht usw. vorhanden. Als Text-Quelle sei beispielsweise genannt:

■ <http://koi.www.online.ru/sp/rel/russian/>

Die *Публичная Электронная Библиотека*, die sich hinter diesem Link verbirgt, bietet kürzere und längere Texte russischen Klassiker, die auf dem Bildschirm gelesen und frei weiterverwendet werden können. Eine Besonderheit hier: alle Texte liegen in mehreren Codierungen vor (Windows, KOI-8, Mac). Diese Texte werden teilweise auf dem Kodeks-Server (s.u.) gespiegelt und können auch von dort heruntergeladen werden.

Von der Slavistik der Universität Potsdam wird ein Zeitungsarchiv mit aktuellen Ausgaben wichtiger Tages- und Wochenzeitungen gepflegt:

■ <http://www.uni-potsdam.de/u/slavistik>

*George Fowler* von der University of Indiana hat ein Korpus diverser russischer Texte gesammelt, die von dem folgenden Server frei heruntergeladen werden können:

■ [http://infomeister.osc.edu/pub/central\\_eastern\\_europe/russian/corpora](http://infomeister.osc.edu/pub/central_eastern_europe/russian/corpora)

Diese Texte sollen künftig ebenfalls auf dem Kodeks-Server gespiegelt werden.

Als deutsches slavistisches Projekt sei an dieser Stelle auch die *Li-Lab*-Site genannt, das von Chr. Sappok geleitete *Linguistische Laboratorium* am Seminar für Slavistik der Universität Bochum:

■ <http://www.ruhr-uni-bochum.de/lilab/>

Hier geht es um phonetisches Material, Lautbeispiele zu Publikationen usw. Mit diesem Projekt gewinnen Arbeiten zur Phonetik eine völlig neue Dimension: Material, das bislang höchstens Lehrbüchern in Form von Tonbändern, Kassetten, Disketten oder CDs beigegeben war, kann jetzt auch wissenschaftliche Publikationen ergänzen.

Die folgende Webseite versammelt eine große Zahl an Links zu ‘Russian Media’:

■ <http://www.hibo.no/stud/sh4/media.html>

## 5. Projekt “Kodeks”

Das Projekt “Kodeks” entsprang inhaltlich der wohl unbestreitbaren Einsicht, daß der Zustand der im deutschsprachigen Raum verwendeten Ein-



fürungen in die älteren slavischen Sprachstufen durchaus nicht so ist, daß hier nicht Raum für Verbesserungen vorhanden wäre. Für das Altkirchenslawische beispielsweise ergibt sich die bekannte Problematik der vorhandenen Lehrwerke daraus, daß sie oft zu vieles gleichzeitig sein wollen: Grammatik, Lesebuch, Lexikon, Kulturgeschichte und didaktisches Lehrbuch. Das LESKIENSche Handbuch beispielsweise ist als Lesebuch immer noch gut zu gebrauchen, in seiner Darstellung von Grammatik jedoch überholt. Bei Benutzung des vielverwendeten Lehrbuches von TRUNTE sind Umfang, Reihenfolge und Aufbau der zu behandelnden Themen und Texte durch die einzelnen Lektionen vorgegeben; als grammatisches Nachschlagewerk eignet sich das Buch hingegen weniger gut, da der Stoff in viele kleine Portionen auseinandergerissen ist.<sup>12</sup> Ein deutsches Lehrbuch zum Altrussischen fehlt ganz – eine bemerkenswerte Lücke in der deutschen Slavistik.

Ermöglicht wurde das Projekt “Kodeks” durch eine Zusammenarbeit von Ulrich Schweier (Universität München) mit dem Autor des vorliegenden Artikels. Technisch verbirgt sich hinter “Kodeks” ein Macintosh Internet Server mit entsprechender Software<sup>13</sup>, der seit dem 1. August ’96 läuft und vom Autor betreut wird. Wählt man mit seinem Internet-Browser die zugehörige Web-Adresse,

■ <http://kodeks.uni-bamberg.de>

so sieht man die Startseite oder HomePage des Servers (**Abb. 6**). Inhalt, Aufbau, Gestaltung und Zielsetzung des Servers haben sich im Laufe der ersten Monate naturgemäß gewandelt, erweitert, präzisiert, jedoch inzwischen soweit stabilisiert, so daß die nachfolgenden Erläuterungen nicht gleich wieder überholt sein werden.

Die älteren Sprachstufen wurden auch deshalb als Schwerpunkt gewählt, weil den Autoren klar ist, daß sie nicht immer eine besonders beliebte Studienkomponente darstellen, aber auch deshalb, weil diese Bereiche im Ruf stehen, ein besonderes Expertenwissen zu erfordern, bevor man sich mit ihnen aktiv beschäftigen könne. Die Benutzung neuester

---

12 Problematisch sind an diesem Lehrbuch viele weitere Dinge, wie z.B. die offensichtliche Manie, möglichst viele “exotische” Schriften im Original zu benutzen, jedoch in einem so kleinen Schriftgrad, daß keinerlei Einsicht in die Struktur fremder Grapheme möglich ist.

13 Konkret: Power Macintosh 7250/120 unter Mac OS 7.5.5. mit 64 MB Arbeitsspeicher, auf dem die folgende Software zum Einsatz kommt: *WebStar* als HTTP-Server, *AppleSearch* für Volltext-Indizierung (Verbindung zu WebStar über *AppleSearch.acgi*), *Apple NetFinder.acgi* für die Download-Funktion. Die Web-Seiten werden mit *Adobe PageMill 2.0* direkt auf dem Server erstellt. Außerdem läuft auf dem Server *Claris FileMaker Pro Server*, ein Datenbankserver. Das System hat sich im Betrieb als außerordentlich stabil erwiesen.

Technik soll also nicht zuletzt einen gewissen Anreiz und Einstieg bieten, sich mit "alten" Themen zu beschäftigen.

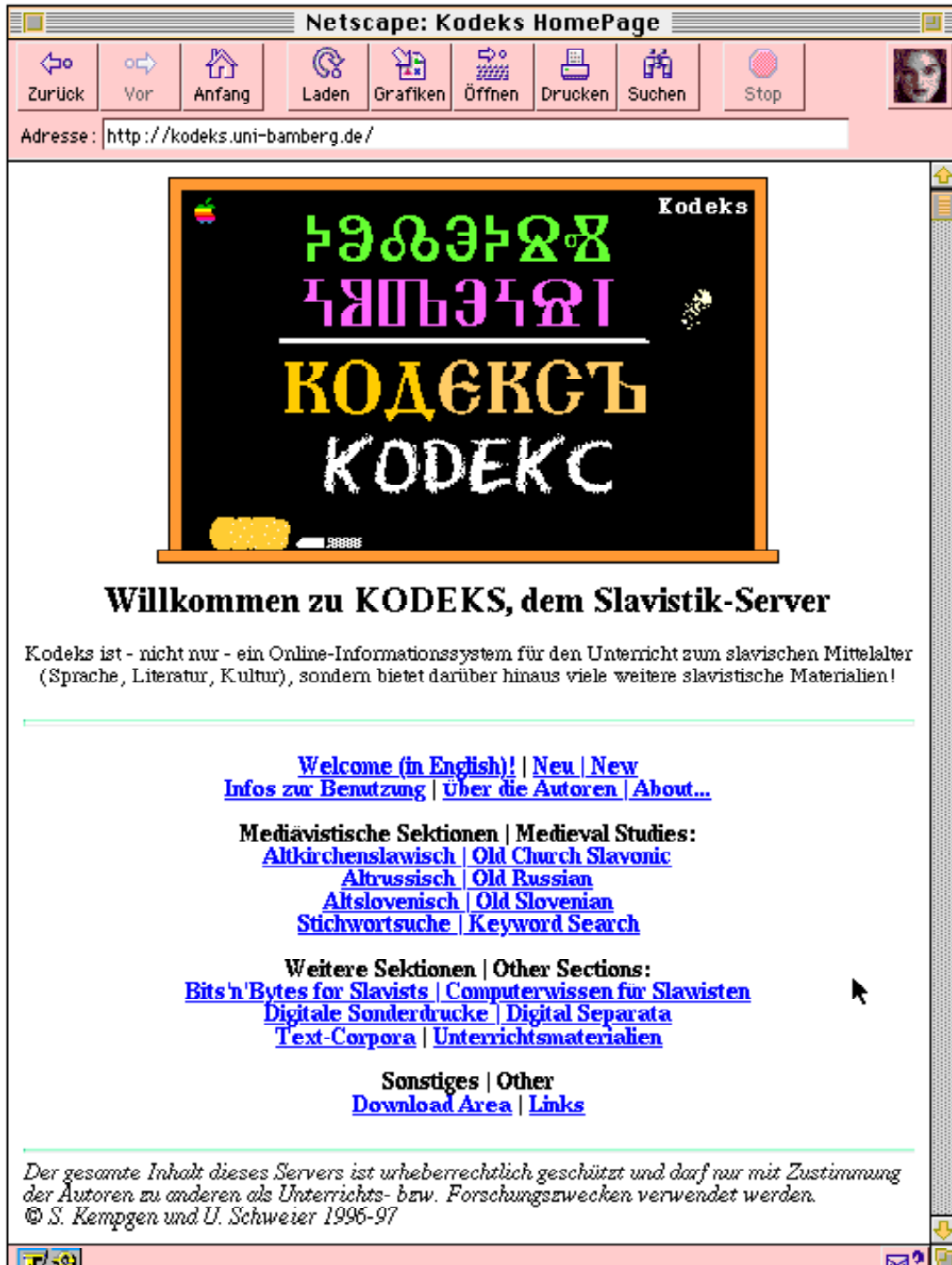


Abb. 6: Startseite des Kodeks-Servers (April '97)

## 5.1. Prinzipielle Überlegungen

Der Server richtet sich an die deutschsprachige Slavistik, und zwar an Studierende wie Lehrende des Faches. Er soll ihnen u.a. Materialien an die Hand geben, die die Beschäftigung mit den älteren Sprachstufen des Slavischen erleichtern, sei es zum Selbststudium oder auch für den Unterricht. Der Kodeks-Server soll dabei weder den Unterricht ersetzen noch überflüssig machen, sondern dem Nutzer vielmehr einzelne, immer wieder verwendbare Bausteine zur Verfügung stellen, derer sich ein Unterricht beispielsweise des Kirchenslawischen oder des Altrussischen stets bedienen wird. Dazu gehören u.a. Texte, Formenparadigmen der Grammatik, das Wissen um die wichtigsten Quellen und deren Editionen, die slavischen Schriften etc. Welchen Gebrauch man von diesen Komponenten macht, liegt also nach wie vor in der Verantwortung des Lehrenden bzw. Lernenden. Kodeks will also kein Online-Lehrbuch sein und die didaktische Nutzung des zur Verfügung gestellten Materials bewußt dem Lehrenden überlassen.

Aus diesen prinzipiellen Erwägungen heraus ergibt sich die Entscheidung, das Material im wesentlichen in deutscher Sprache bereitzustellen, obwohl sich das Internet natürlich hauptsächlich des Englischen bedient. Der Kodeks-Server soll sich aber bewußt an den Bedürfnissen und Erfordernissen des akademischen Unterrichts im deutschsprachigen Raum orientieren, mit dem die Autoren naturgemäß am besten vertraut sind. Um allerdings den anglophonen Web-Surfern wenigstens einen Anhaltspunkt zu bieten, mit welchen Inhalten sie es hier zu tun haben, sind wichtige Überschriften, Stichpunkte, Inhaltsangaben etc. auch in Englisch verfügbar.

## 5.2. Kodeks-Sektionen

Kernstück des Servers sind die mediävistischen Sektionen zu den aufgeführten älteren Sprachstufen (s. Abb.). Jede dieser Sektionen verzweigt sich wieder in mehrere Unterpunkte, vor allem in:

- Schriften und Alphabete
- Grammatik
- Quellen, Texte und Editionen

Einen Schwerpunkt hierbei bietet eine genaue Beschreibung der Quellen, mit denen man es im jeweiligen Sprachbereich zu tun hat, der “Denkmäler” also. Denn wiederum stellt man beispielsweise bei einem Vergleich der verschiedenen Kirchenslawisch-Grammatiken fest, wie unterschiedlich und wie unpräzise die jeweils präsentierten Informationen sein können. Diesem Mangel sollen die in “Kodeks” gegebenen Ausführungen abhelfen, ohne sich jedoch an den Spezialisten wenden zu wollen,

der dieser Informationen sowieso nicht mehr bedarf. Angaben zu den wichtigen Editionen der Texte dürfen hierbei natürlich nicht fehlen.

**Abb. 7** und **Abb. 8** zeigen zwei Seiten aus der Abteilung zu den slawischen Schriften. Die Abbildungen der Schriften selbst sind Graphiken, da die Benutzung unterschiedlicher Fonts im Internet derzeit erst in einer Entwicklungsphase ist und das Funktionieren dieser an sich wünschenswerten Neuerung noch nicht allgemein vorausgesetzt werden kann.<sup>14</sup> Die in **Abb. 8** ganz am Anfang der Seite stehenden stichwortartigen Einträge haben den einzigen Zweck, eine Indizierung dieser Seite durch die Internet-Suchmaschinen zu ermöglichen: Text, der als Graphik vorhanden ist, ist ja nichts weiter als eine Ansammlung von Punkten; er kann weder gesucht noch gefunden werden.<sup>15</sup> Eine Suche nach den aufgeführten Stichwörtern ist aber möglich und müßte zu einem Verweis auf diese Seite führen. Aus dem gleichen Grunde findet sich auch auf der abgebildeten Grammatik-Seite (**Abb. 10**) stichwortartiger Text vor den Paradigmen, die wiederum eine Graphik sind.

---

14 Hierbei konkurrieren zwei verschiedene Technologien miteinander. Die eine Variante spezifiziert die Angabe des zu verwendenden Fonts durch einen entsprechenden HTML-Tag; dies ist übertragungsökonomisch, kann aber natürlich nur bei weitverbreiteten Fonts funktionieren. In der Praxis kann man außer Times (Windows: Times New Roman), Helvetica (Windows: Arial) und Courier im wesentlichen nur die auch als "LaserWriter Plus-Set" bekannten Fonts voraussetzen, die außer den genannten noch Palatino, ITC Bookman, New Century Schoolbook, Helvetica Narrow und ITC Avant Garde ausmachen. Bei allen anderen Fonts riskiert man bei diesem Ansatz, daß auf dem Zielgerät der benötigte Font fehlt.

Der zweite Ansatz geht deshalb anders vor: Er will alle notwendigen Schriftinformationen mit einer geeigneten Technologie in die Web-Seiten einbetten; dies ist datenintensiver, da die Fontinformationen ja dann ebenfalls übertragen werden müssen, hat aber den Vorteil, daß man von der Softwareausstattung des Zielgerätes unabhängig ist. Allerdings wirft dieser Ansatz auch urheberrechtliche Probleme auf: die Einbettung kommerzieller Fonts in Web-Seiten muß durch die Lizenzbestimmungen des Herstellers ausdrücklich erlaubt sein.

15 Wer als Betreiber eines Servers nicht darauf warten will, daß dieser von den Internet-Indizierungsprogrammen irgendwann über die auf ihn weisenden Links "entdeckt" wird, muß sich aktiv um eine Bearbeitung bemühen; die Suchmaschinen setzen dann sog. *Spider* in Bewegung, die systematisch jedem Link, auf den sie treffen, nachgehen und den gesamten Text des betreffenden Dokumentes indizieren. Da sich das Internet sowohl von der Organisation wie vom Inhalt her täglich ändert, ist die Indizierung und Katalogisierung eine nie endende Aufgabe, die dem aktuellen Stand immer hinterherhinkt.

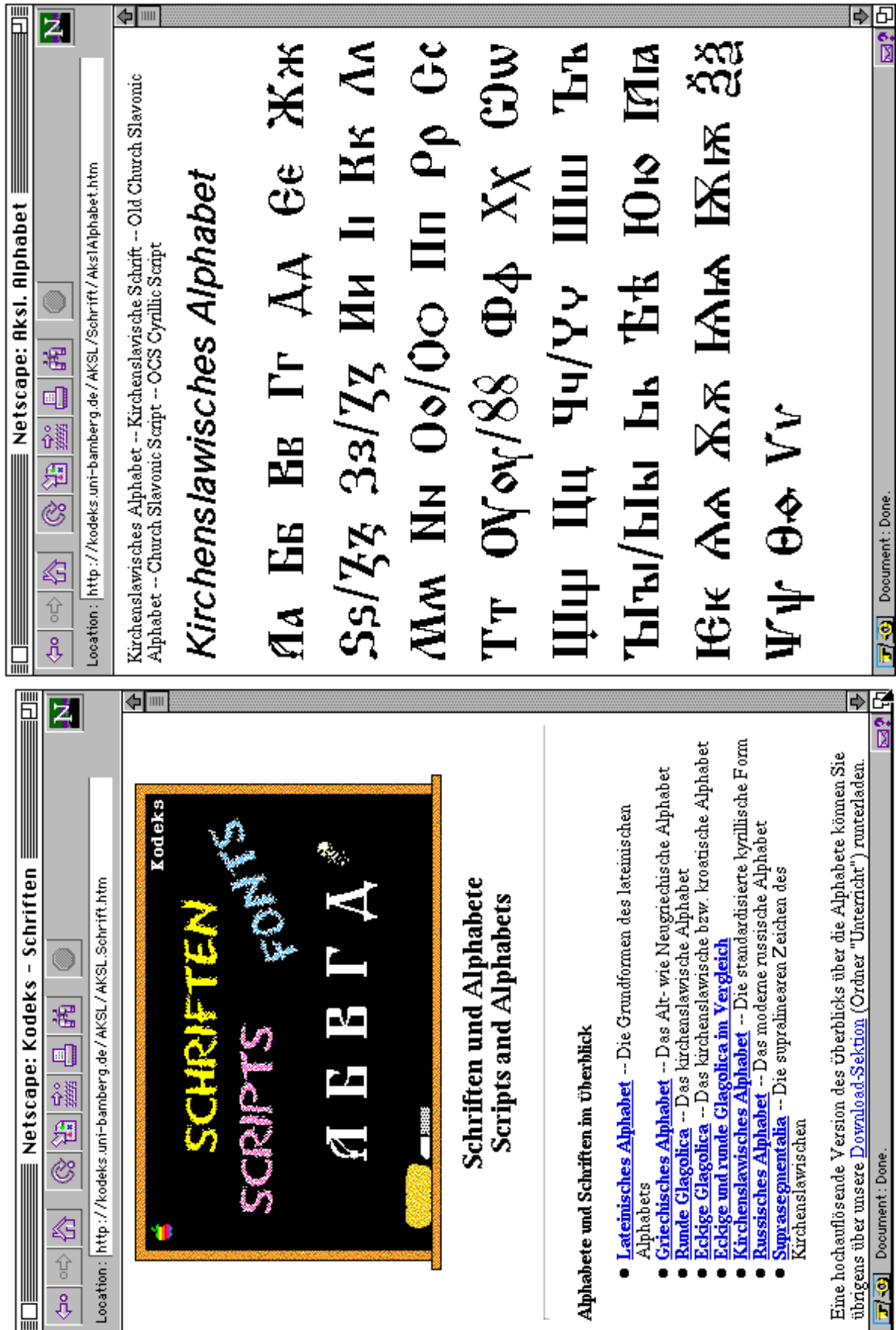


Abb. 7 und 8: Schriften-Übersicht und -Beispiel

The image shows two screenshots from a Netscape browser window. The top screenshot displays a search form titled 'AppleSearch ACGI 1.5.3'. The search input field contains the word 'Glagolica'. Below the search field, there are options for 'Minimum Rank' (set to '\*') and 'Maximum Returns' (set to '10'). A section titled 'Select Info Sources to search...' includes several checkboxes: 'AltRuss' (checked), 'AbsSloven' (checked), 'Computer' (checked), 'Text\_Corpora' (checked), 'Electronic\_Separata' (checked), and 'AKSL' (checked). At the bottom of the form, there is a checkbox for 'Show matching terms in bold type (slower)' and a 'Perform Search' button.

The bottom screenshot shows the results page for the search of 'Glagolica'. It features a table with columns for grammatical cases and their corresponding forms. The table is organized into sections for 'N Sg.', 'N Pl', and 'Altern.'. The 'N Sg.' section lists forms for cases K, G, D, A, I, L, and V. The 'N Pl' section lists forms for cases G, D, A, I, and L. The 'Altern.' section lists alternative forms for the genitive case (Г - ж/с, К - ч/ц, Х - ш/с). The browser's address bar shows the URL 'http://kodeks.uni-bamberg.de/AKSL/Grammatik/AkslSubstMask.htm'.

	K	g/kX	K'	J	C
N Sg.	РАСЪ	ЕОГЪ	КОНЬ	ИРАИ	ААААА
G	РАСА	ЕОГА	КОНИ	ИРАИ	ААААА
D	РАСОУ	ЕОГОУ	КОНО	ИРАО	ААААА
A	РАСЪ	ЕОГЪ	КОНЬ	ИРАИ	ААААА
I	РАСОАА	ЕОГОАА	КОНЕАА	ИРАЕАА	ААААА
L	РАСЪ	ЕОГЪ	КОНИ	ИРАИ	ААААА
V	РАСЕ	ЕОРЕ	КОНО	ИРАО	ААААА
N Pl					
G	РАСИ	ЕОСИ	КОНИ	ИРАИ	ААААА
D	РАСВАА	ЕОГВАА	КОНЕАА	ИРАЕАА	ААААА
A	РАСЫ	ЕОГЫ	КОНЬ	ИРАИ	ААААА
I	РАСЫ	ЕОГЫ	КОНИ	ИРАИ	ААААА
L	РАСЪА	ЕОГЪА	КОНЪА	ИРААА	ААААА
N/A/V Du	РАСА	ЕОГА	КОНА	ИРАИ	ААААА
G/L	РАСОУ	ЕОГОУ	КОНО	ИРАО	ААААА
D/I	РАСВАА	ЕОГВАА	КОНЕАА	ИРАЕАА	ААААА
Altern.		Г - ж/с К - ч/ц Х - ш/с			

Abb. 9 und 10: Suchformular und Formenparadigma

**Abb. 9** zeigt das Suchformular des Kodeks-Servers, mit dem innerhalb des Servers nach beliebigen Stichwörtern aus dem Text gesucht werden kann. Das Beispiel zeigt, daß als Suchbegriff *Glagolica* eingegeben wurde, nach dem im Bereich AKSL (Altkirchenslawisch) gesucht werden soll. Die Eingabe von Wortbestandteilen ist zulässig, also z. B. *Glagoli*. Als Ergebnis der Stichwortsuche erhält man eine Liste von – mit Links versehenen – Dokumenten, in denen das gesuchte Wort enthalten ist. Der Inhalt des Servers wird von dem Programm *AppleSearch Server* in regelmäßigen Abständen automatisch indiziert; der Index liegt also bereits fertig vor und braucht bei einer Stichwortsuche nur abgefragt zu werden.

### 5.3. Ergänzende Funktionen

Zu den Kernsektionen des Kodeks-Servers kommen bereits jetzt einige ergänzende Bereiche hinzu, die sich auf lange Sicht als nicht weniger wichtig erweisen könnten.

- Der Bereich *Text-Corpora* enthält derzeit bereits etliche russische belletristische Texte, z.T. als Spiegel anderer Server.<sup>16</sup> Diese Textdateien (im MS-Word-Format) können frei heruntergeladen und für wissenschaftliche Zwecke frei benutzt werden. Wer selbst – unter Beachtung aller Urheberrechte – Texte zur Verfügung stellen kann, möge den Autor kontaktieren.

- Der Bereich *Digitale Sonderdrucke* enthält elektronische Separata der an “Kodeks” beteiligten Autoren. Dies sind die vollständig formatierten Druckvorlagen von Artikeln, die mit einem speziellen Viewer (für Macintosh und Windows), der ebenfalls auf dem Server vorhanden ist, gelesen werden können. Diese Dokumente sind unabhängig von den auf dem Zielgerät installierten Schriften, da die entsprechenden Informationen mit in die Dokumente eingebettet sind. Erzeugt wurden diese Dokumente mit einer Software namens *Common Ground*.<sup>17</sup> Auch hier gilt: “Kodeks” ist offen für die Aufnahme solcher elektronischer Sonderdrucke auch von Kollegen, kann quasi als Repositorium dienen.

---

16 Einen Teil der Texte hat Elke Mann (Universität Potsdam) nach expliziter Freigabe durch die Urheber zur Verfügung gestellt, wofür ihr an dieser Stelle noch einmal herzlich gedankt sei.

17 Als Alternative zu *Common Ground* bietet sich noch das von der Firma Adobe kreierte *pdf-Format* an (*pdf* = *portable document format*), das jedoch mindestens in den ersten Versionen nicht gut für Fremdsprachen geeignet war. Die *Viewer* für dieses Format sind gleichfalls frei aus dem Internet ladbar, die Erzeugung der Dateien geschieht jedoch mit Software, die kostenpflichtig ist.

• Die Abteilung *Computerwissen für Slavisten* enthält Materialien, die speziell Slavisten das Verständnis für die besonderen Probleme der Datenverarbeitung mit slavistischem Inhalt erleichtern sollen (vgl. **Abb. 11**). Gegenwärtig sind hier vor allem etliche Codierungstabellen zu den wichtigsten Code-Pages auf Windows bzw. Mac und Internet einsehbar. Sie entsprechen von ihrer Art her der oben abgebildeten KOI-8-Codierungstabelle.

• Die *Download Area* (vgl. **Abb. 14**) umfaßt alle oben genannten Bereiche, dazu weitere, vornehmlich diverse Mac-Software, die wissenschaftlichen Zwecken dient, u.a. Fonts, Programme zur Textmanipulation, Internet-Software usw.

• Mit den *Links* wird auf ähnlich gelagerte oder ergänzende Projekte im Internet hingewiesen (vgl. **Abb. 13**).

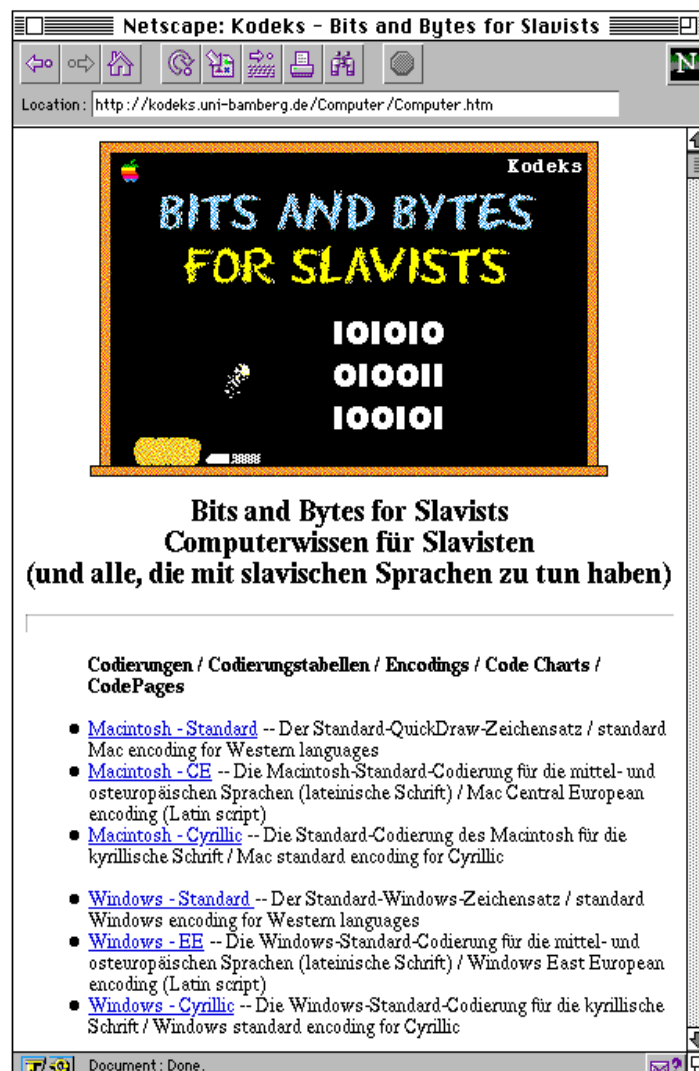


Abb. 11: Sektion Computerwissen für Slawisten



## 6. Verwandte Projekte

Im Internet gibt es natürlich eine Reihe von Projekten, die das “Kodeks”-Projekt in der einen oder anderen Richtung ergänzen und erweitern. Auf sie sei an dieser Stelle wenigstens kurz eingegangen.

### *Corpus Cyrillo-Methodianum Helsingiense*

■ <http://www.ling.helsinki.fi/dept/slav/ccmh>

Das *CCMH* will – unter Leitung von Jouko Lindstedt – die wichtigsten kirchenslawischen Texte in elektronischer Form bereitstellen. Unter anderem stehen bereits jetzt der Codex Marianus und der Codex Suprasliensis zur Verfügung. Alle Texte sind im 7-Bit ASCII, können also zwischen allen Computerplattformen ausgetauscht werden. Dies bedeutet aber zugleich, daß die verwendete Codierung keinerlei Codierungsstandard für das Kyrillische entspricht. Dadurch, daß z. B. das Vorliegen eines Großbuchstabens einfach durch ein spezielles Zusatzzeichen signalisiert wird, können die Großbuchstaben benutzt werden, um Zeichenvarianten und zusätzliche Zeichen zu kodieren. Es ist klar, daß eine solche Codierung nicht die Präzision ersetzen kann (und auch nicht ersetzen soll), die eine gedruckte Edition bietet. Die Texte sind denn auch eher dazu gedacht, sie ergänzend zur Buchausgabe zur Verfügung zu haben.

**Abb. 12** zeigt im Vergleich (a) die Version, wie sie im Internet auf dem Bildschirm erscheint, (b) die gleiche Passage mit (teilweiser) Rückübertragung in eine kirchenslawische Schrift, schließlich (c) die Jagićsche Edition.

Ob es ein Fortschritt ist, derart in sieben Bit kodierte Texte online zur Verfügung zu haben, welchen Nutzen man aus derartigen Texten ziehen kann – derartige Fragen können an dieser Stelle nicht ausführlicher diskutiert werden.

Eine Reihe weiterer Texte sind von den Herausgebern der Buchausgaben elektronisch aufbereitet worden, so die Kiever Blätter, die Freisinger Denkmäler u. a. m.; die Codierung folgt dabei dem von Helsinki vorgegebenen Muster. Links findet man ebendort.

### *Obshtezhitie*

■ <http://www.ceu.hu/medstud/obsht.htm>

Diese Webseite “zur Erforschung kyrillischer und glagolitischer Manuskripte und früher Drucke” wurde als Folge der Konferenz in Blagoevgrad 1995 eingerichtet, deren Proceedings von BIRNBAUM et al. (1995) herausgegeben wurden; sie wird von Ralph Cleminson betreut. Die Webseite versammelt wichtige Informationen über mediävistische Konferenzen und sonstige Aktivitäten, über einschlägige Publikationen, bietet Links zu wichtigen Projekten und ähnliches mehr.

1052310 -to na tE  
1052400 ostavi tu dar& tvoi pr@d& ol&tarem& . J Sed& pr@Zde s&miri sE .  
1052410 s& bratrom& svoim& . i togda priSed& prinesi dar& tvoi ::  
1052500 \*bOdi uv@qajE sE s& sOp\$rem\$ svoim\$ skoro . don\$deZe esi na pOti  
1052510 s& nim\$ . da ne pr@dast& tebe sOdii . J sOdii tE pr@dast& sluD@ .  
1052520 J v& tem\$nicO v&vr&Zet& tE .  
1052600 amin\$ !gljO teb@ ne izideSi ot& tOd@ . don\$deZe v&zdasI  
1052610 posl@d\$niI kodrant& ::  
1052700 \*slySaste @ko reCeno by!s drev&nim& . ne pr@ljuby s&tvoriSi .  
1052800 az& Ze !gljO vam& . @ko v\$S@k& iZe v\$z\$rit& na ZenO s& poxotijO .  
1052810 juZe ljuby s&tvori s& nejO v& !srdci svoem\$ .

1052310 -то на тѧ  
1052400 остави тоу даръ твои прѣдъ олѣтаремъ . и шедъ прѣжде съмири сѧ .  
1052410 съ вратромъ своимъ . и тогда пришедъ принеси даръ твои ::  
1052500 Бжди оувѣщаѧ сѧ съ сѧпъремъ своимъ скоро . донѣдеже еси на пѣти  
1052510 съ нимъ . да не прѣдастъ тебе сѧдни . и сѧдни тѧ прѣдастъ слоуѣѣ .  
1052520 и въ темьницѧ вѣврѣжетъ тѧ .  
1052600 аминь ! глѣж тебѣ не изидеши отъ тѣдѣ . донѣдеже въздаси  
1052610 послѣдънии кодрантъ ::  
1052700 Слѣшасте ѣко речено въ!с дrevънимъ . не прѣлюбы сътвориши .  
1052800 азъ же ! глѣж вамъ . ѣко въсѣкъ иже възъритъ на женѧ съ похотнѧ .  
1052810 юже любви сътвори съ неѧ въ !срдци своемъ .

12

MATĚ. V 23—32.

(Микл. I.) то на тѧ 24 остави тоу даръ твои прѣдъ олѣтаремъ . и шедъ прѣжде съмири сѧ .  
съ вратромъ своимъ . и тогда пришедъ принеси даръ твои :: 25 Бжди оувѣщаѧ сѧ съ сѧпъремъ своимъ скоро . донѣдеже еси на пѣти съ нимъ . да не прѣдастъ тебе сѧдни . и сѧдни тѧ прѣдастъ слоуѣѣ . и въ темьницѧ вѣврѣжетъ тѧ . 26 аминь глѣж тебѣ не изидеши отъ тѣдѣ . донѣдеже въздаси послѣдънии кодрантъ :: 27 Слѣшасте ѣко речено въ!с дrevънимъ . не прѣлюбы сътвориши . 28 азъ же глѣж вамъ . ѣко въсѣкъ иже възъритъ на женѧ съ похотнѧ . юже любви сътвори съ неѧ въ срдци своемъ . 29 лице же око твоѧ

Abb. 12: Matthäus V, 23–28; (a) 7-Bit-Codierung des CCMH, (b) Rückübertragung in kirchenslawische Schrift, (c) Jagićsche Edition

LITERATUR

Apple Computer:

1992 *Guide to Macintosh Software Localization*. Reading, MA: Addison-Wesley.

Birnbaum, D. et al. (eds.):

1995 *Computer Processing of Medieval Slavic Manuscripts. Proceedings. First International Conference, 24–28 July, 1995, Blagoevgrad, Bulgaria*. Sofija.

Cölfen, E., Cölfen, H., Schmitz, U.:

1996 *Linguistik im Internet. Das Buch zum Netz – mit CD-ROM*. Opladen–Wiesbaden: Westdeutscher Verlag.

Feldmann, D., Neumann, F. W., Rommel, Th. (eds.):

1996 *Anglistik im Internet (Anglistische Forschungen, Bd. 240)*. Heidelberg: C. Winter.

Jagić, V. (ed.):

1883 *Quattuor Evangeliorum versionis palaeoslovenicae Codex Marianus glagoliticus (...)*. Berlin (Nachdruck Graz 1960).

Kempgen, S.:

1989 Neue Medien für den Russischunterricht (Computergestütztes Lehren und Lernen mit HyperCard). In: W. Girke (Hg.), *Slavistische Linguistik 1988*, München, 89–112.

1995 Complex Script Systems on Today's Personal Computers. In: *Computer Processing of Medieval Slavic Manuscripts, Proceedings, First International Conference, 24–28 July, 1995, Blagoevgrad, Bulgaria*, Sofija, 68–78.

Microsoft Corporation

1985–94 Benutzerhandbuch Microsoft MS-DOS 6.22. [o.O.]

Peek, R. P., Newby, G. B. (eds.):

1996 *Scholarly Publishing. The Electronic Frontier*. Cambridge/MA–London: MIT Press.

Sheldon, K. M.:

1991 ASCII goes global. *BYTE* 7, 108–116.

The Unicode Consortium

1991–92 *The Unicode Standard. Worldwide Character Encoding. Version 1.0*. Vols. 1–2. Reading, MA: Addison-Wesley.

RESUME

The present article consists of two parts. In the first part, the author outlines the special difficulties that slavists have to deal with today when they exchange texts which employ different character encoding standards. While several work-arounds can be used, the real solution will be the implementation of the Unicode standard on all major platforms (Unix, Windows, Mac). In the second part of the article, the author presents the “Kodeks” server (<<http://kodeks.uni-bamberg.de>>) which aims to help students of Slavic philology master the medieval languages they have to learn (Old Church Slavonic, Old Russian). The article is richly illustrated with screen-shots of web pages as served out by the ‘Kodeks’ server.

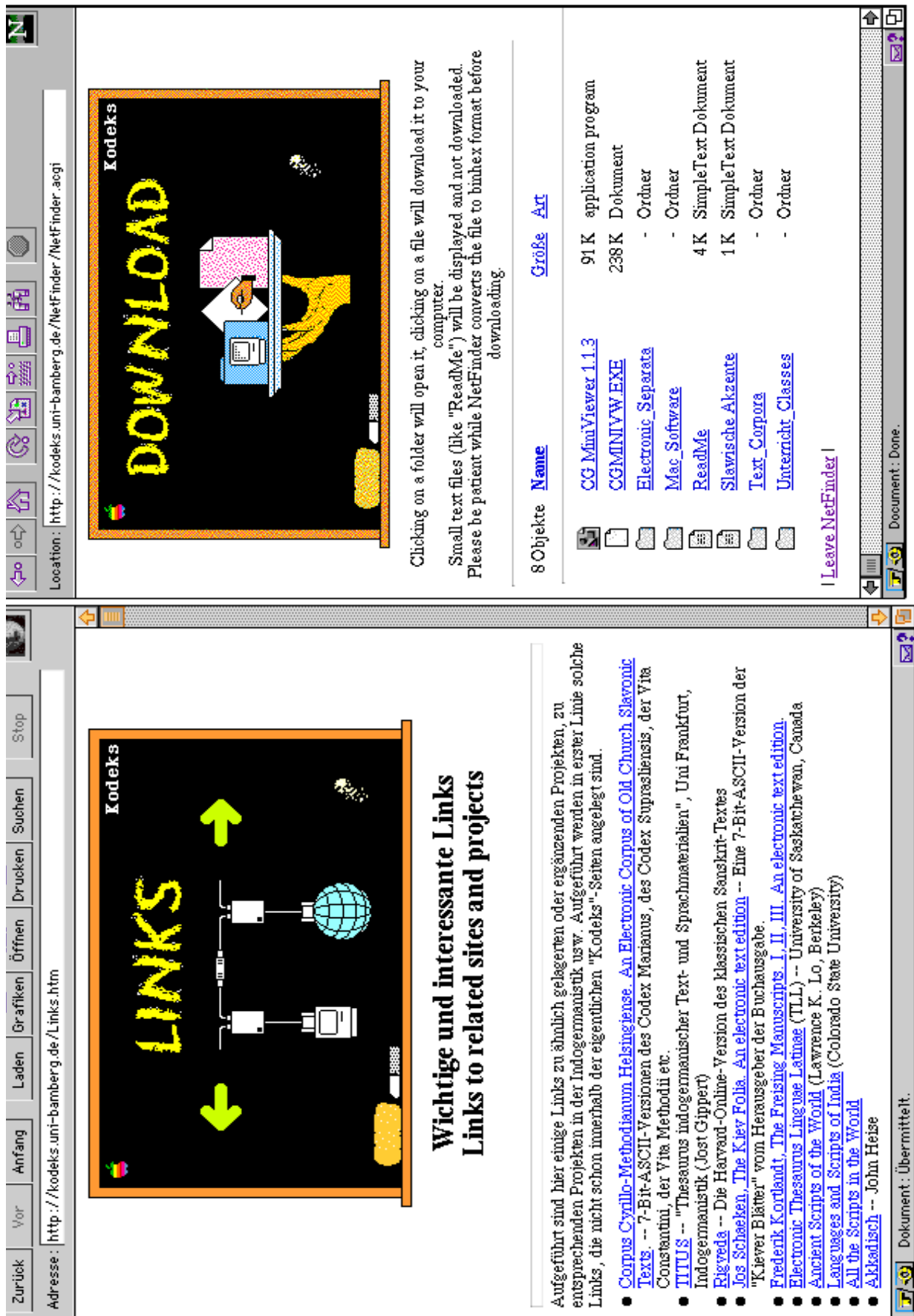


Abb. 13 und 14: Links und Download-Sektion