

### Die "European Ordering Rules" (EOR; ENV 13710): einheitliche Regeln für das Sortieren multilingualer Daten

Küster, Marc Wilhelm

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Küster, M. W. (2000). Die "European Ordering Rules" (EOR; ENV 13710): einheitliche Regeln für das Sortieren multilingualer Daten. *Historical Social Research*, 25(3/4), 248-257. <https://doi.org/10.12759/hsr.25.2000.3/4.248-257>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

### Die "European Ordering Rules" (EOR; ENV 13710): Einheitliche Regeln für das Sortieren multilingualer Daten

*Marc Wilhelm Küster (Tübingen)\**

#### Einleitung

Standards sind ein Teil unseres täglichen Lebens. Jeder deutsche Schüler weiß, was er sich unter einem DIN A4-Blatt vorzustellen hat. Egal bei welchem Händler man eine Glühbirne kauft, sie passt trotzdem in jede normale Fassung. Ein Meter ist überall auf der Welt genau gleich lang und ein Kilo Äpfel hat überall genau dieselbe Masse. Ein Zug kann ohne Schwierigkeiten quer durch den größten Teil Kontinentaleuropas fahren, weil die Schienen in den meisten Ländern gleich breit sind. Container passen ohne weitere Adaption weltweit in Schiffe, auf Züge und Lastwagen und haben dadurch den Warentransport revolutioniert.

Es gibt alle Arten von Normen, angefangen von inoffiziellen Absprachen bis hin zu nationalen und internationalen Standards, die vom Staat sogar in Gesetzen herangezogen werden können.

#### Institutioneller Hintergrund

Offizielle Standards werden von nationalen Normungsorganisationen wie dem DIN, dem Deutschen Institut für Normung, europäischen und internationalen Organisationen nach genau definierten Regeln verabschiedet. Genau wie sich die ISO für internationale Standardisierung zuständig zeigt, ist es das Comité Européen de Normalisation (CEN) für europäische Standardisierung. Seine Mitglieder sind alle nationalen Normungsorganisationen der Europäischen Union und der EFTA, der European Free Trade Association, sowie die Tschechische Republik. Zu den sogenannten Affiliate members – Staaten die eine volle Mitgliedschaft anstreben – gehören Albanien, Bulgarien, Estland, Kroa-

---

\* Aus dem Protokoll des 78. Kolloquiums über die Anwendung der Elektronischen Datenverarbeitung in den Geisteswissenschaften an der Universität Tübingen vom 5. Februar 2000.

tien, Letland, Litauen, Malta, Polen, Rumänien, die Slowakei, Slowenien, die Türkei und Zypern.

Eine der technischen Komitees (Technical Committees, TCs) von CENs IT-Abteilung, das Information Society Standardization System (ISSS), hat die weitreichende Aufgabe, die europäischen Interessen auf dem Gebiet der Informationstechnologie zu verteidigen und gegebenenfalls Standards zu initiieren. Dieses Komitee läuft unter dem Namen CEN/TC304 Informations- und Kommunikationstechnologie: Lokale europäische Anforderungen und wird von Wolf Arvidson (Schweden) und Þorgeir Sigurðsson (Island) geleitet.

In den Bereich der europäischen Lokalisationsanforderungen fielen von Anfang an auch Fragen paneuropäischen Sortierens, d.h. der Sortierung multilingualer Wortlisten, die von übernationalen europäischen Organisationen wie der Europäischen Kommission, die Hauptkunde und -sponsor dieses Standards ist, erstellt werden. Folglich war die paneuropäische Sortierung integraler Bestandteil des Arbeitsprogramms von CEN/TC304 seit seiner Gründung 1992.

Zwischenzeitlich hatte ISO TC37 "Terminology – principles and coordination" ebenfalls die Absicht geäußert, einen Sortierstandard für bibliographische Zwecke zu erarbeiten, der sich ausschließlich auf die lateinische Schrift beschränken sollte. Dieser Entwurf ist als FDIS 12199 bekannt. Sein Editor, d.h. sein Hauptautor, ist Håvard Hjulstad (Norwegen). Aus europäischer Sicht war die Koordination der verschiedenen Vorhaben von zentraler Bedeutung. 1997 wurde daher unter Förderung der EU und EFTA ein Projektteam ausgeschrieben und schließlich gebildet, dem Håvard Hjulstad als Projektmanager, Marc Wilhelm Küster (Deutschland) als Editor und John Clews (Großbritannien) als Rezensent angehören. Es war vor allem die erwiesene Expertise der TUSTEP-Gruppe als ganzer, unsere große Erfahrung in der Gestaltung von Wörterbüchern, Lexika, Enzyklopädien, Indizes und Registern, die mir den Zuschlag gebracht hat.

Håvard Hjulstad hatte die Aufgabe, die paneuropäischen Sortierregeln mit FDIS 12199 zu koordinieren, während ich die europäischen Interessen in der ISO/IEC JTC1/SC22/WG20 vertrat und sicherstellte, dass sie eine echte Anwendung der 14651 (s.u.) werden würde – zusätzlich zu der Aufgabe, den europäischen Standard überhaupt erst einmal zu schreiben.

Wie von der Projektdefinition verlangt arbeiteten wir auf der Basis eines vorgegebenen Zeichensatzes, der lateinische, griechische und kyrillische Buchstaben umfasste, die in Büro- und Verwaltungsanwendungen benötigt werden.

Im Frühsommer 1998 wurde CEN/TC304 ein erster Entwurf vorgelegt, der im Juli 1998 modifiziert in die erste Abstimmungsrunde ging. Deren erfolgreiches Ergebnis führte direkt zur Endabstimmung, die ebenfalls erfolgreich war.

Im Dezember 1999 wurden die paneuropäischen Sortierregeln offiziell als sog. Europäische Vornorm (ENV) 13710 verabschiedet. Wenn sich die Vornorm, die bereits in sich ein gültiger Standard ist, in der Anwendung bewährt,

wird sie in längstens drei Jahren eine in Europa allgemein verbindliche Norm sein.

Da die Norm die Hürden eines jeden Standards durchlaufen musste und somit von den europäischen Staaten mitgetragen werden muss, galt es, sie für alle Beteiligten akzeptabel zu gestalten. Dazu gehört gerade auf der europäischen Bühne natürlich auch die Politik – jede Norm, gerade im kulturellen Bereich, ist immer auch ein Politikum.

### Warum?

Warum aber gerade einen Standard zur Anordnung des ABCs? Das Alphabet gehört zu den Selbstverständlichkeiten dieses Lebens über die wir selten nachdenken. Wer würde so etwas Banales wie das sprichwörtliche "Von A bis Z" hinterfragen wollen? Wir lernen das ABC in der ersten Klasse der Grundschule und danach ist es genauso Teil unseres geistigen Repertoires wie das kleine Einmaleins. Eine Forschungsaufgabe auf Grundschulniveau?

Dennoch: Die meisten von uns werden sich an das leichte Gefühl der Überraschung erinnern, als wir im Religionsunterricht zum ersten Mal den biblischen Ausdruck "Ich bin das A und das O, der Erste und der Letzte, der Anfang und das Ende" (Off. 22, 13) gehört haben, gefolgt von der obligatorischen Referenz auf die alten Griechen und ihre Angewohnheit, das Alphabet mit  $\Omega$  anstelle eines Z zu beenden.

Die Griechen hielten dagegen ihre Reihenfolge für die Natürlichste der Welt. Einige sahen in ihr sogar etwas, was die ewigen Harmonien des Himmels widerspiegelte – die Anzahl der Buchstaben zwischen  $\alpha$  und  $\omega$  war in ihren Augen "ebensogroß wie [die Anzahl der Töne] auf der Flöte vom tiefsten zum höchsten Ton, und diese Zahl ist der des Himmels ganzen gleich." (Aristoteles: Metaphysik, Buch 14,6, zitiert nach Horst Seidel: Aristoteles' Metaphysik. Zweiter Halbband: Bücher VII(Z)-XIV(N). In der Übersetzung von Hermann Bonitz. Felix Meiner Verlag : Hamburg, 1980. Fairerweise muss gesagt werden, dass Aristoteles sich von dieser Position distanziert.)

### Eine Reise durch Europa

Leider ist mit diesem offensichtlichen Unterschied erst der Anfang der Schwierigkeiten umrissen. Nicht nur hat jede Schrift eine eigene Sortierreihenfolge, die meisten Schriften haben deren viele widersprüchliche.

Beschränken wir uns zunächst einmal auf den Teil Europas, der die lateinische Schrift verwendet. Wie allgemein bekannt, benutzen die meisten europäischen Sprachen zusätzliche Buchstaben, die es im englischen Grundalphabet nicht gibt; einigen Sprachen wie etwa dem Litauischen fehlen im Gegenzug Buchstaben, die im Englischen völlig normal sind.

Wir wollen uns jetzt nicht akribisch durch akademische Tabellen durchquälen. Gehen wir stattdessen auf eine alphabetische Stippvisite quer durch Europa. Beginnend mit Aarfit in Grönland fahren wir nach Ansbach in Deutschland, Århus in Dänemark, Azincourt in Frankreich, Lladrós in Spanien, Løberg in Norwegen, Lobersberg in Österreich, Zürich in der Schweiz und Żyrardów in Polen. Damit Ende der Unternehmung? Nein, sondern ein kleines Gedankenspiel: der gleiche Vortrag in Dänemark.

Dort würde die selbe Reise von Ansbach über Azincourt, Lladrós, Lobersberg, Løberg, Żyrardów und Aarfit nach Århus führen. Warum das?

Werfen wir einen kurzen Blick auf das dänische Alphabet: Zunächst einmal kennt es drei zusätzliche Buchstaben nach z, nämlich æ, ø und å. Weiterhin versteht es die Buchstabenkombination aa als äquivalent zu å und sieht y und ü als Varianten des selben Buchstabens an. Nach diesen Regeln tauchen Zürich, Aarfit und Århus alle nach Żyrardów auf unserer Reiseroute auf.

Noch weiter nördlich, in Island, verlief unsere Reise noch anders: Aarfit, Århus, Ansbach, Azincourt, Lladrós, Lobersberg, Løberg, Zürich und Żyrardów. Nach der aktuellen Praxis haben fast alle Buchstaben ihren eigenen Platz im Alphabet, einige wie å und ä fallen jedoch aus dem Rahmen und werden im konkreten Fall wie a behandelt. Wiederum folgen drei Buchstaben auf z, nur sind es hier þ, æ und ö / ø. In einigen wichtigen isländischen Wörterbüchern wie dem Íslensk Orðabók fyrir skóla og skrifstofur (Reykjavík 1994) fehlen einige englische Buchstaben wie das w, d und ð. (Eth, ein stimmhaftes /th/) werden als gleichwertig betrachtet, genau wie v und w in den wenigen Fremdwörtern, in denen w verwendet wird. Das führt zu Reihenfolgen wie vb, WC, vé.

Ähnliche Beispiele kann man auch im Süden Europas finden. Spanien behandelt traditionell sowohl ch und ll als auch ñ als eigene Buchstaben, die nach c, l bzw. n einsortiert werden, so dass plötzlich Løberg und Lobersberg vor Lladrós zu stehen kommen.

In der jüngsten Zeit wurde diese Praxis revidiert und die Sonderrolle von ch and ll aufgegeben. Der aktuellste Druck (1999) des maßgeblichen spanischen Wörterbuchs, des Diccionario de la lengua española der königlichen spanischen Akademie, folgt allerdings noch den alten Vorgaben. Nach einer gängigen These sind solche Änderungen der letzten Jahre ein direktes Resultat unzureichender Sortierwerkzeuge, die nicht in der Lage waren, mit Digraphen zurechtzukommen und dadurch für die Erhaltung der kulturellen Vielfalt in Europas abträglich waren.

Sprachen können verschiedene Auffassungen davon haben, was zu Sortierzwecken für einen Buchstaben kennzeichnend ist – einerseits können Digraphen (Zweibuchstabengruppen) wie das spanische ll, das tschechische ch oder das sorbische Dž als ein Buchstabe gewertet werden, andererseits können einzelne Buchstaben für zwei stehen. Außerdem werden sprachspezifische Buchstaben nach divergierenden Strategien behandelt: In einigen Sprachen

tauchen sie mit eigenem Eintrag im Wörterbuch auf (oft, aber keineswegs immer, nach z), in anderen werden sie bei mehr oder minder ähnlichen Buchstaben einsortiert. Obendrein gehen die meisten Sprachen unterschiedlich mit Fast-Homographen wie role und rôle oder aime und aimé um, wenn die Buchstaben mit Diakritika keinen eigenen Wörterbucheintrag haben.

Das deutsche Scharf-S (ß) wird im Gegenzug für Sortierzwecke immer einem Doppel-S (ss) gleichgesetzt.

Sind die Unterschiede in den europäischen Ländern, die die lateinische Schrift nutzen, schon bedeutend, werden sie immens auf dem Gebiet der früheren Sowjetunion. Man vergisst zu leicht, in welchem Maß dieses Land ein Vielvölker- und Vielsprachenstaat war (und es Russland nun ist).

Die meisten der vielen dort lebenden Sprachen nutzen oder nutzten Varianten der kyrillischen Schrift mit einer beachtlichen Anzahl von Sonderbuchstaben für spezielle Phoneme. Nach den dramatischen Veränderungen der frühen Neunziger sind aktuelle Informationen über die Lage oftmals schwer zu bekommen.

Die Sortierregeln variieren drastisch von Sprache zu Sprache. Ähnlich der Situation in Teilen Westeuropas fügen einige Sprachen, wie das Tartarische, Sonderbuchstaben nach dem letztem Buchstaben des russischen Alphabets, dem я ein, z.B. ... я, Ә, Ө, Ү, Ж, Һ. Andere, wie das Kasachische, sortieren sie neben den Buchstaben ein, denen sie ähnlich sehen (z.B. Һ und Һ) oder die für ein verwandtes Phonem stehen (z.B. а und Ә). Selbst innerhalb eines Ansatzes sind die konkreten Realisierungen im Regelfall inkompatibel. Außerdem wurden eng verwandte Sprachen mit sehr unterschiedlichen Alphabeten und Sortierreihenfolgen beglückt – einige sehen dies als Auswirkung von Stalins Strategie des 'Divide et impera' die ihr Bestes tat, ein einheitliches Nationalgefühl der vielen unruhigen Stämme im Süden Russlands zu verhindern.

Das sollten genug Beispiele sein, um das generelle Problem zu illustrieren: Alphabetische Anordnungen gehören zu den grundlegenden Koordinaten einer Kultur und werden infolgedessen, wie die meisten Kulturschätze, eifersüchtig verteidigt. Sie differieren drastisch zwischen den verschiedenen Kulturen und können sogar religiöse (so z.B. im Alten Testament und noch verstärkt in der Kabbala) und philosophische Konnotationen haben.

Besonderheiten in Schrift, Zeichensatz und der Sortierreihenfolge werden oft als zentrale Bestandteile der nationalen und kulturellen Identität angesehen, ähnlich wie Fahnen und andere Symbole. Aus diesem Grund ist der Wunsch, sie zwischen den Staaten zu harmonisieren, im Regelfall gering. Wenn man daher Daten für eine bestimmte Benutzergruppe sortiert, muss man deren Wünsche respektieren und die Datenverarbeitung an ihre Gegebenheiten anpassen.

## Software

Nicht nur im Spanischen hat unzureichende Software in der Vergangenheit viel dazu beigetragen, das kulturelle Erbe Europas zu beschädigen. Weitverbreitete Krücken wie der MS-DOS sort-Befehl haben im Wesentlichen nach der Zeichenkodierung sortiert – eine Strategie, die nicht einmal für das Englische gut funktioniert, da sie a nach Z einordnet, von allen anderen Sprachen einmal abgesehen. Es ist sogar vorgekommen, dass solche Werkzeuge für die Herstellung von Wörterbüchern und Registern herangezogen wurden.

Glücklicherweise haben sich einzelne Gruppen immer gegen diese Form von Kultur-Barbarentum gewehrt. Bereits in den frühen Siebziger Jahren haben Wilhelm Ott und Kuno Schälkle einen vollständigen Satz frei konfigurierbarer Sortierwerkzeuge bereitgestellt, die nun als svorbereite (Sortier-Vorbereite), rvorbereite (Register-Vorbereite), saufbereite (Sortier-Aufbereite), raufbereite (Register-Aufbereite) und sort (die eigentliche Sortierung) Teil von TUSTEP sind. Diese Werkzeuge werden in vielen Projekten genutzt, u.a. auf dem Gebiet der Lexikographie, und implementieren ein mächtiges und schnelles Mehr-Ebenen-Sortierverfahren. Im Zusammenspiel dieses Moduls mit weiteren bietet TUSTEP alles, was man für anspruchsvolle Herstellung von Wörterbüchern, wissenschaftlichen Editionen und Database publishing benötigt. Weitere Informationen finden Sie auf der TUSTEP-Homepage.

### Der Final Draft International Standard (FDIS) 14651

In der Mitte der Achtziger wunderte sich der Franko-Kanadier Alain LaBonté über die scheinbar willkürliche Reihenfolge von Homographen in französischen Wörterbüchern, die aber trotzdem intuitiv von allen Französisch-Sprechenden gefunden werden konnten. Er fand heraus, dass diese Methode vollständig deterministisch war, war aber gleichzeitig schockiert, als er herausfand, wie inkonsequent die meisten kommerziellen Sortierprogramme damit umgingen. Er schloss daraus auf das dringende Desiderat eines internationalen Standards, der es erlauben würde, eine allgemein gültige Methodik zur Sortierung mehr-schriftlicher Texte zu definieren.

1992 wurde das Projekt selbst der Arbeitsgruppe "Internationalisierung" mit dem wohlklingenden Namen "ISO/IEC JTC1/SC22/WG20" zugeordnet, die Teil des Komitees ist, das sich in der ISO mit Programmiersprachen beschäftigt. Das Projekt, und damit der zukünftige Standard, heißt 14651.

Ein solcher Standard wird sich notwendigerweise auf die methodischen Anforderungen an eine Mehr-Ebenen-Sortierung konzentrieren, nicht etwa eine allgemeingültige "Welt-Sortierung" schaffen wollen. Der Fokus des Projekts liegt damit auf Anpassbarkeit, also darauf, mit Hilfe dieser Methodik nationale, kulturelle oder sogar persönliche Präferenzen eindeutig und maschinenlesbar

spezifizieren zu können. Dazu entwickelte die Arbeitsgruppe eine Syntax, die aufwärtskompatibel zum POSIX-Standard ISO/IEC 9945 Informationstechnik – Betriebssystem-Schnittstelle für die Austauschbarkeit von Systemen (POSIX). Teil 1: 1996, Teil 2: 1995, ist, obwohl es sich dabei nur um eine Beispielsyntax handelt. Bewusst wurde darauf geachtet, dass andere Syntaxformen wie die von svorbereite oder die des Unicode Collation Algorithm ebenfalls gültige Ausdrucksformen der selben Methodik sind.

Als Service für den Nutzer gehört zum Standard eine Art von Default, eine Common Template Table, um die Standard-interne Terminologie zu verwenden, die eine halbwegs vernünftige Reihenfolge für Unicode-Zeichen enthält, die nicht von einer bestimmten Anpassung betroffen sind. Der französische Sortierstandard NF Z 44-001 (November 1995) z.B. spezifiziert im Detail eine Zeichenreihenfolge für die lateinische Schrift. Er schweigt hingegen über die Sortierung von chinesischen Ideogrammen oder von georgischen Buchstaben, wo es vernünftig sein könnte, die Standardannahmen unverändert zu lassen.

Leider hat dieser Standard bis zu seiner Fertigstellung viel Zeit in Anspruch genommen. Inzwischen steht die Abstimmung zum DIS (draft international standard) an, und es ist zu erwarten, dass er vor Jahresende ein offizieller ISO-Standard sein wird.

### Paneuropäische Sortierregeln

Der zu erwartende ISO-Standard zur Sortiermethodik ist ein großer Schritt vorwärts, löst aber nicht alle Probleme in Europa. Die Situation ist nun immerhin für nationale Gegebenheiten geklärt, die in dem 14651-Mechanismus spezifiziert werden können.

Es gibt hingegen viele Fälle, in denen das nicht ausreicht. Ein klassisches Beispiel sind multinationale Organisationen wie die Europäische Union, die nicht einfach der Praxis eines Mitgliedsstaates zum Schaden aller anderen folgen kann. Sie müssen außerdem Daten nicht nur in lateinischer und griechischer Schrift, sondern auch in Kyrillisch handhaben können. Weitere Beispiele, für die nationale Standards nicht ausreichen, stellen Proceedings paneuropäischer Konferenzen oder Berichte über multilaterale Verhandlungen dar.

Weniger offensichtlich, aber nicht weniger wichtig ist die Verwendung der paneuropäischen Regeln für Zeichen, die nicht von direkter Relevanz für eine spezifische Sprache sind. Z.B. ist es aus griechischer Sicht unproblematisch, lateinische und kyrillische Buchstaben nach den EOR zu sortieren.

Im Dezember 1999 wurde die europäische Vornorm ENV 13710, besser bekannt als die europäischen Sortierregeln (European Ordering Rules (EOR)), von den nationalen Standardisierungsorganisationen des CEN verabschiedet. Aus verschiedenen Gründen ist dies ein Meilenstein in der Geschichte multilingualer Datenverarbeitung.



## Designprinzipien der European Ordering Rules

Die europäischen Sortierregeln dienen als Kompromiss für paneuropäisches Sortieren. Das allein impliziert, dass die resultierende Reihenfolge nicht mit einer existierenden Sortierreihenfolge identisch sein darf, sie müssen allerdings universell nachvollziehbar sein. Der Endnutzer braucht kein detailliertes Wissen über verschiedene europäische Sortierpraktiken zu verfügen, um mit dem Standard zurechtzukommen. Technisch muss es eine Anpassung an die 14651 sein.

Die EOR beschreiben ein Mehr-Ebenen-Sortierverfahren. Unterscheidungen auf der ersten Ebene sind dabei natürlich die wichtigsten und behandeln Buchstaben (und Symbole), die ihre eigenen Wörterbucheinträge haben. Unterscheidungen auf Ebene 2 basieren auf Diakritika oder varianten Formen wie dem ð, das auf der ersten Ebene oft als eine Variante des d angesehen wird, oder, in seltenen Fällen, auf beidem. Die dritte Ebene unterscheidet nach Kapitalisierung – Englisch polish (polieren) gegenüber Polish (polnisch) oder Deutsch arm gegenüber Arm. Die vierte Ebene schließlich kümmert sich um Leerzeichen und Symbole und trennt it's von its oder Mac Duffin von MacDuffin. Weitere Ebenen werden manchmal benötigt, um nach Typographie – z.B. Goethes Divan (sein West-östlicher Divan) gegenüber Goethes Divan (sein Sofa) -oder relativ zur Grundlinie – z.B. in Formeln wie  $m^2$  gegenüber  $m_2$  – zu differenzieren.

Für die lateinische Schrift entschieden wir uns, "ungewöhnliche" Buchstaben so weit wie möglich mit Buchstaben des englischen Alphabets zu vereinheitlichen, vor allem auf der Basis ihres Aussehens, was glücklicherweise weitgehend mit einer Vereinheitlichung nach phonetischen Kriterien zusammenfällt. (Ausnahmen sind z.B. das grönländische  $\kappa$  (Kra), das mit q transliteriert wird, aber von jedem, der das nicht weiß, für eine Variante von k gehalten wird, mit dem wir es deshalb auch schweren Herzens vereinheitlicht haben.) Das funktionierte gut für alle Buchstaben außer dem isländischen Thom ( $\mathfrak{þ}$ ) dessen normale Transliteration als /th/ in weiten Kreisen nicht bekannt ist und das auch keinen Buchstaben hat, mit dem man es vernünftigerweise visuell gleichsetzen könnte. Thorn blieb damit als eigener Buchstabe nach z. Insgesamt ergab sich die Reihenfolge der lateinischen Buchstaben als a b c d e f g h i j k l m n o p q r s t u v w x y z  $\mathfrak{þ}$ .

Griechisch machte weniger Schwierigkeiten, da es vor allem von einer Sprachgemeinschaft und in einem Land verwendet wird. Das einzige größere Problem waren die Zahlzeichen  $\zeta$  (Stigma),  $\text{Ϝ}$  (Koppa) und  $\text{Ϟ}$  (Sampi) sowie das Digamma F. Das Digamma, Vorfahre des lateinischen Buchstabens F, wird in keiner modernen Sprache verwendet (es taucht allerdings häufig in sprachwissenschaftlichen Studien auf), während die Zahlzeichen noch heute in Gesetzestexten u.ä. auftauchen können.

Die resultierende Reihenfolge hätte die griechischen Sophisten wohl sehr erstaunt  $\alpha \beta \gamma \delta \epsilon \text{ F } \zeta \eta \theta \iota \kappa \lambda \mu \nu \xi \omicron \pi \text{ Ϝ } \rho \sigma \tau \upsilon \phi \chi \psi \omega \text{ Ϟ}$ : Das griechische

Alphabet endet mit Sampi – wie es gegenwärtiger griechischer Praxis entspricht.

Als strittigste Frage erwies sich die Standardsortierung für die kyrillischen Buchstaben. Kyrillisch wird heutzutage von mehr als fünfzig Sprachen verwendet -wie oben gesagt, mit völlig unterschiedlichen Ansätzen. In fast allen kyrillischen Alphabete werden sämtliche Buchstaben, einschließlich derjenigen mit Diakritika, als Grundbuchstaben behandelt. Oft werden allerdings auch Di- und Trigraphen als eigene Buchstaben eingeordnet, vor allem in kaukasischen Sprachen. Eine pan-kyrillische Sortierung hat somit weitreichende, auch politische, Implikationen.

Daher fühlte ich mich unwohl dabei, diese Fragen selbst zu entscheiden, zumal ich kein Spezialist für die verschiedenen türkischen, mongolischen und kaukasischen Sprachen bin, um die es geht. Kontakte mit TOCT, der russischen Standardisierungsorganisation, erwiesen sich als schwierig und für Monate passierte nichts, so dass ich doch selbst eine erste Version vorbereiten musste. Als man mich nach Dutzenden von Anläufen schließlich doch mit den entsprechenden Experten verband, verbesserte sich die Situation entscheidend. Emil Yakupov, Ko-Vorsitzender des russischen Komitees zur Zeichenkodierung und Direktor von ParaType, und Sergei Bolotov, ein Sprachwissenschaftler und einer seiner Angestellten, zeigten sich überaus hilfsbereit, so dass mein Entwurf verbessert werden konnte und nun in einem wesentlich höheren Maß sowohl Beziehungen zwischen Sprachen als auch phonetische Ähnlichkeiten berücksichtigt. Ich möchte ihnen an dieser Stelle für die fruchtbare Kooperation danken.

Die so entstandene Liste der kyrillischen Buchstaben ist ziemlich lang und trotzdem unvollständig, da der Zeichenvorrat, auf dem die EOR aufbauen, die kirchenslawischen Buchstaben nicht enthält. Mit dieser Einschränkung kann sie aber ein gewisses Maß an Endgültigkeit für sich beanspruchen: а ă ä å æ б в г ф г д ђ з е ё ж ж ж з э с з и й і ї ј к к љ к к л л љ м н њ н њ о ö ø ð п р с с т т ћ у ў ү ү ѱ х х ѣ ц ч ч ч ч е е ц ш щ ы ь Ѣ ь э ю я q I

Die Sortierreihenfolge der Schriften untereinander ist: Lateinisch, Griechisch, Kyrillisch.

Aus urheberrechtlichen Gründen werde ich nicht weiter in Details zur Behandlung von Buchstaben wie ð und ß gehen, die auf Grundbuchstaben abgebildet werden. Ich werde auch nicht die Tabellen mit der Reihenfolge der Diakritika reproduzieren. Es sei immerhin so viel gesagt, dass auf der dritten Ebene die European Ordering Rules Klein- vor Großbuchstaben sortieren, wie es europäischer Mehrheitspraxis entspricht.

Der Standard selbst ist über die CEN Mitgliedsorganisationen erhältlich, deren Adressen bei CEN zu finden sind. Mehr Information lässt sich auch bei STRi finden.

## Pläne für die Zukunft

Das Projekt ist bereits jetzt verlängert und wird auf alle europäischen Schriften – es fehlen noch Armenisch und Georgisch – erweitert werden. Auch wird das Zeichenrepertoire vergrößert werden und beispielsweise die altkirchenslawischen Buchstaben enthalten. Pläne, ebenso die hebräische Schrift abzudecken, die eine so große Rolle in der europäischen Geschichte gespielt hat und für europäische Sprachen wie Jiddisch benötigt wird, erhielten aber leider nicht die notwendige Unterstützung – was ich persönlich überaus bedauere.

Das Projektteam wird auch die Akzeptanz der Europäischen Sortierregeln überwachen und nach möglichen Fällen Ausschau halten, die erst die Praxis zeigen kann. Dabei wird es eng mit Wissenschaft, Industrie und anderen Anwendern zusammenarbeiten. Das Ende dieses Projekts ist gleichzeitig der Beginn für neue Aktivitäten – die Arbeit an paneuropäischer Lokalisierung hat gerade erst begonnen.