Universität Zürich
Deutsches Seminar
MA-Seminar: Vom analogen zum digitalen Schreiben (FS19)
Prof. Dr. Christa Dürscheid

Schreiben mit Unicode –
Eine kritische Auseinandersetzung.

Abgabedatum: 15. Oktober 2019

Nora Tiziana Parolari
Eliya Margarita Livas

Inhaltsverzeichnis

1 Einleitung 3

2 Aufbau und Organisation eines multilingualen Codierungssystems (Parolari) 4

2.1 Eine Herkulesaufgabe 4

2.2 Das Unicode-Konsortium – ein verantwortungsvoller Gatekeeper? 7

2.3 Das Paradox des „unbekannten Weltstandards“ 14

3 Wie das System Schreibprozesse beeinflusst (Livas) 15

3.1 Kritik am Unicode 15

3.2 Präsuppositionen des Schreibens 18

3.3 Der Unicode als Schreibpraxis 21

4 Fazit 26

5 Bibliographie 27

1 Einleitung

Die vorliegende Arbeit setzt sich mit dem Codierungsstandard Unicode auseinander. Im Zentrum steht dabei die kritische Auseinandersetzung mit diesem global enorm einflussreichen System, das bisher erstaunlicherweise weitgehend unter dem Radar der meisten Nutzerinnen und Nutzer des Unicodes sowie der akademischen Fachwelt geblieben ist. Technische Hintergründe zur Funktionsweise des Codierungsstandards werden, sofern für das grundlegende Verständnis dieser Arbeit notwendig, geliefert – für eine vertiefte Auseinandersetzung mit der Materie ist ein Blick auf die technische Website[1] von Unicode und in die Literatur (siehe Bibliographie) allerdings unerlässlich.

Die Arbeit gliedert sich in zwei Teile. Im ersten Teil wird zunächst auf den Aufbau und die Organisation von Unicode, die Funktionsweise und die Zielsetzung eingegangen, wobei kritisch nach dem Umgang mit Verantwortung gefragt wird. Zum einen geht es um die Frage, wie verantwortungsvoll das Konsortium aufgebaut ist und wie verantwortungsvoll das Konsortium als Gatekeeper bei der Codierung von Zeichen vorgeht. Zum anderen wird nachgeforscht, inwiefern dieser Gatekeeper seiner Verantwortung für eine demokratische Auslegung, Partizipationsmöglichkeiten und Transparenz nachkommt. Der zweite Teil der Arbeit widmet sich dem Unicode als Schreibinstrument. Dabei wird auf Schwierigkeiten eingegangen, die durch kontroverse Codierungsentscheidungen entstanden sind, aber auch auf Schwierigkeiten, die gewissermassen systemimmanent sind, also vom Unicode selbst generiert werden. Es wird thematisiert, wie der Unicode Einfluss auf die Schreibprozesse seiner Nutzerinnen und Nutzer nimmt, um im Anschluss daran mögliche Zukunftsszenarien des Codes zu erläutern.

Mark Davis, der Präsident und Mitgründer von Unicode, der am Google-Forschungsstandort in Zürich Enge tätig ist, wurde für diese Arbeit für ein Interview angefragt. Eine Reaktion darauf ist bis zum Abgabetermin dieser Arbeit leider ausgeblieben.

[1] Vgl. http://unicode.org/main.html <14.10.2019>.

2 Aufbau und Organisation eines multilingualen Codierungssystems (Parolari)

2.1 Eine Herkulesaufgabe

„Everyone in the world should be able to use their own language on phones and computers“[2] – was einfach klingt, hat sich das Unicode-Konsortium zum Ziel gesetzt und sich damit eine Herkulesaufgabe gestellt. Unicode will digitale Kommunikation in allen Sprachen und mittels aller Schriftsysteme der Welt ermöglichen. Dadurch werden schriftliche Texte, seien dies E-Mails, SMS, WhatsApp-Nachrichten oder auch Texte auf Internetseiten wie beispielsweise Wikipedia auf allen Endgeräten, egal in welcher Sprache sie verfasst sind, lesbar – sofern eine geeignete Schriftart (Font) verfügbar ist. Unicode ist also ein Codierungsstandard, den man mittlerweile mit Fug und Recht als Weltstandard bezeichnen kann. Der Erfolg von Unicode besteht darin, dass er Kompatibilitätsprobleme löst, die vorherige Codierungsstandards wie ASCII oder ISO Latin-1 hatten; unlesbare Texte wie *Sch%ne Gr&§e!* gehören dadurch der Vergangenheit an (vgl. Bergerhausen/Poarangan 2011, 018). Die früheren Kompatibilitätsprobleme werden umgangen, indem im Unicode für jedes Schriftzeichen jeden Schriftsystems der Welt ein eindeutiger und einzigartiger Code bestimmt wird. Diese *code points* werden einmalig und für alle Nutzerinnen und Nutzer verbindlich festgelegt; „Insellösungen“ durch unterschiedliche Codierungen ein und derselben Code-Position sind dadurch nicht mehr möglich (vgl. Bergerhausen/Poarangan 2011, 018).

Der Grossbuchstabe *A* im Block *Basic Latin* verfügt so beispielsweise über den Code U+0041, der Kleinbuchstabe *a* im selben Block über den Code U+0061 und so weiter (vgl. Bergerhausen/Poarangan 2011, 077). Codiert werden aber nur Zeichen im Sinne eines abstrakten Konzepts, keine Glyphen, also mögliche typographische Formen des Zeichens (vgl. Bergerhausen/Poarangan 2011, S. 025). Bei einer E-Mail werden demnach nur die Konzepte der Zeichen, die *code points*, um die Welt geschickt und nicht die konkreten Glyphen eines Zeichens, also nicht jedes einzelne Zeichen des Textes in seiner typographischen Form wie beispielsweise in Arial (vgl. Bergerhausen/Poarangan 2011, 025). Das macht das Versenden von Zeichen um einiges einfacher und verbraucht viel weniger Speicherplatz (vgl. Bergerhausen/Poarangan 2011, 025). Die gesendeten Zeichen können grundsätzlich auf jedem Endgerät korrekt angezeigt werden, sofern auf

[2] Vgl. https://home.unicode.org/ <03.09.2019>.

dem Gerät ein Font installiert ist, der das Zeichen wiedergeben kann. Ein grosses *A* wird also immer korrekt als grosses *A* empfangen, auch wenn der Absender beispielsweise mit dem Font Verdana schreibt und ich auf meinem Empfängergerät den Font Comic Sans eingestellt habe. Die korrekte Codierung jedes einzelnen Schriftzeichens der Welt (inklusive aller Emojis) mit einem unverwechselbaren *code point* ist die Herkulesaufgabe von Unicode bei der Etablierung des multilingualen Weltstandards.

Die erste Version des Unicode Standards, Version 1.0.0, wurde im Oktober 1991 veröffentlicht.[3] Die aktuellste Version des Unicode Standards, die am 7. Mai 2019 veröffentlicht wurde, ist die Version 12.1.0.[4] Sie verfügt bereits über insgesamt 137'929 codierte Zeichen. Es ist wichtig zu beachten, dass Unicode nicht Sprachen codiert, sondern Schriftsysteme. Einige Schriftsysteme werden nur von einer einzigen Sprache verwendet, so zum Beispiel das Schriftsystem Hangul, das nur für die koreanische Sprache verwendet wird.[5] Anders sieht dies beispielsweise bei der uns sehr vertrauten lateinischen Schrift aus; diese wird für diverse Sprachen wie Deutsch, Englisch, Französisch und etliche weitere verwendet. Die bis jetzt in Unicode codierten 137'929 Zeichen verteilen sich somit auf rund 150 Schriftsysteme.[6] Durch diese 150 Schriftsysteme können aber eben weit mehr als nur 150 Sprachen wiedergegeben werden. Insgesamt besteht der „Code-Raum“, also die Anzahl möglicher codierbarer Stellen, aus unglaublichen 1'114'112 Zeichen (vgl. Bergerhausen/Poarangan 2011, 029). Es gibt also durchaus noch viel Platz für neue Zeichen – wobei berücksichtigt werden muss, dass die beliebten und mittlerweile geradezu allgegenwärtigen Emojis ebenfalls in Unicode enthalten sind und immer mehr hinzukommen.

Eine wichtige Besonderheit von Unicode besteht darin, dass nicht nur die („lebenden“) Schriftsysteme, die noch heute in Gebrauch sind, codiert werden, sondern auch Schriftsysteme „toter Sprachen“. Als einige wenige Beispiele mögen hier die ägyptischen Hieroglyphen (ca. 3000 v. Chr.-1. Jh.), die phönizischen Schriftzeichen (12.-2. Jh. v. Chr.) oder die Runenschrift (1.-19. Jh.) dienen (vgl. Bergerhausen/Poarangan 2011, 613; 631; 633). Einen Spezialfall bildet hierbei unter anderem die Linearschrift A (ca. 18.-15. Jh. v. Chr.), die in Kreta verwendet wurde und bis heute nicht eindeutig entziffert werden konnte (vgl. Bergerhausen/Poarangan 2011, 623). Die Schriftzeichen dieses Schriftsys-

[3] Vgl. https://www.unicode.org/versions/components-1.0.0.html <03.09.2019>.
[4] Vgl. https://www.unicode.org/versions/Unicode12.1.0/ <03.09.2019>.
[5] Vgl. https://www.unicode.org/standard/supported.html <03.09.2019>.
[6] Vgl. https://www.unicode.org/standard/supported.html <03.09.2019>.

tems wurden also in den Unicode aufgenommen, obwohl die Bedeutung der Schriftzeichen noch nicht gänzlich geklärt ist.

Zudem werden in den Unicode grundsätzlich alle Zeichen aufgenommen, die in anderen, bereits existierenden Zeichencodes enthalten sind (vgl. Bergerhausen/Poarangan 2011, 047). Dadurch ist Unicode rückwärtskompatibel zu allen bereits bestehenden Zeichencodes und kann diese mittelfristig ersetzen (vgl. Bergerhausen/Poarangan 2011, 047). Das dies in nicht allzu ferner Zukunft Realität werden könnte, ist durchaus wahrscheinlich: Unicode gilt jetzt schon als Standardprogramm der multilingualen Textverarbeitung und wird auf der ganzen Welt, auf allen möglichen technischen Geräten und für diverse Formen der schriftlichen Kommunikation verwendet. Konkurrenz in derselben Grössenordnung hat Unicode zudem kaum zu fürchten; höchstens bei einzelnen Schriftsystemen gibt es konkurrierende Codierungssysteme, die sich aber nur auf ebenjenes Schriftsystem beziehen und nicht multilingual funktionieren.[7]

Durch die Codierung aller Zeichen aller Schriftsysteme der Welt hat Unicode nicht nur ungeheuren Einfluss auf die globale Kommunikation, Unicode kann möglicherweise auch dazu beitragen, dass Minderheitensprachen gefördert werden: Indem ausnahmslos alle Zeichen jeden Schriftsystems codiert werden sollen, werden auch Minderheitensprachen berücksichtigt. Dadurch sind diese Minderheitensprachen, sofern geeignete Fonts entwickelt sind, auf allen möglichen technischen Endgeräten verfügbar. Sprecherinnen und Sprecher von Minderheitensprachen haben so die Möglichkeit, mit ihren technischen Geräten in ihrer eigenen Sprache schriftlich zu kommunizieren, ohne dabei beispielsweise auf die Schriftzeichen ihrer ehemaligen Kolonialmächte zurückgreifen zu müssen (vgl. Bergerhausen/Poarangan 2011, 042). Unicode kann also die Emanzipation und Stärkung der kulturellen Identität bei Sprecherinnen und Sprechern von Minderheitensprachen beeinflussen (vgl. Bergerhausen/Poarangan 2011, 042). Zudem kann die Codierung von Schriftsystemen in Unicode möglicherweise dazu beitragen, dass diese nicht verloren gehen: Die codierten Zeichen bleiben erhalten, auch wenn ein Schriftsystem nicht mehr aktiv benutzt wird. Allgemein hat Unicode einen grossen Einfluss auf Forschungsbeiträge in Bereichen wie Linguistik, Geschichte, Archäologie und ähnlichen mehr. Mit Unicode lassen sich viel einfacher Beiträge verfassen, die verschiedene Schriftsysteme thematisieren oder auf Zeichen sehr alter Schriftsysteme ein-

[7] Vgl. http://chinesemac.org/pages/character_sets.html <09.10.2019>.

gehen. Zudem können solche Zeichen oder Forschungsbeiträge problemlos versandt, online publiziert und vor allem von Suchmaschinen gefunden werden.

2.2 Das Unicode-Konsortium – ein verantwortungsvoller Gatekeeper?

Dieser weltweite Erfolg von Unicode und der damit einhergehende Einfluss bringen eine enorme Verantwortung bei der Codierung von Zeichen mit sich, denn: Was einmal codiert ist, wird nie mehr aus Unicode gelöscht.

Jede noch so seltene oder exotische Sprache will codiert sein – und zwar korrekt. Dies erfordert genaue Recherchen und einen präzisen, verantwortungsvollen Umgang mit den zu codierenden Zeichen durch das Unicode-Konsortium. Das Unicode-Konsortium ist ein Zusammenschluss von Computer-Unternehmen, Software-Entwicklern, Datenbanken, Regierungsorganisationen, Forschungseinrichtungen, internationalen Agenturen, verschiedenen Nutzergruppen und interessierten Einzelpersonen.[8] Das Konsortium besteht seit 1991, allerdings mit wechselnden Mitgliedern, und ist ein Non-Profit-Unternehmen, das sich durch Mitgliederbeiträge und Spenden finanziert.[9] Die Höhe der Mitgliederbeiträge pro Jahr orientiert sich dabei am jeweiligen Status der Mitgliedschaft; möglich sind die Optionen *full* (21'000$), *institutional* ($14,000 / $10,000 for Qualified Educational Institutions), *supporting* ($8,750 / $5,000 for Non-Profit), *associate* ($2,900 / $1,500 for Non-Profit), *individual* (75$), *student* (35$) und *liaison* (by invitation).[10] Für langfristige Mitgliedschaften über mehrere Jahre hinweg gibt es Preisreduktionen.[11] Wie in untenstehender Tabelle ersichtlich wird, haben diejenigen Mitglieder, die am meisten zahlen, auch die meisten Rechte und somit den grössten Einfluss auf die Arbeit des Konsortiums.

Comparison of Membership Levels[12]

	Full	**Inst**	**Supp**	**Assoc**	**Liais**	**Indiv/Student**
Technical Committee Meetings						
Participate	✓	✓	✓	✓	✓	
Vote Weight*	1	1	½	0	0	n/a

[8] Vgl. https://www.unicode.org/consortium/consort.html <10.09.2019>.
[9] Vgl. https://www.unicode.org/consortium/consort.html; http://www.unicode.org/consortium/levels.html <10.09.2019>.
[10] Vgl. http://www.unicode.org/consortium/levels.html <10.09.2019>.
[11] Vgl. http://www.unicode.org/consortium/levels.html#compare <10.09.2019>.
[12] Vgl. https://unicode.org/consortium/levels.html <10.09.2019>.

Full Member Consortium Meetings						
Participate & Vote	✓					
Representation						
Primary Representative or Member	1	1	1	1	1	1
Alternate Representatives	4	4	2	2	1	0
Technical Committee Email List Participation						
Primary Representative or Member	✓	✓	✓	✓	✓	✓
List Participation for Member Employees	✓	✓	✓	✓	✓	
Member Acknowledgement						
Marquee Listing on Home Page	✓	✓	✓	✓		
Link to Member's Website	✓	✓	✓	✓	✓	
Discounts						
Conference Discounts for Member	✓	✓	✓	✓		✓
Conference Discounts for Member Employees	✓	✓	✓	✓		
Logo Use Fee Waiver	✓	✓	✓	✓		

* Entities that are part of a Control Group, as defined by the Unicode bylaws, will be limited to collectively no more than one vote in Unicode technical committees.

Full member-Status und somit Stimmrecht besitzen aktuell Adobe, Apple, Facebook, Google, Huawei, IBM, Microsoft, das Ministerium für religiöse Stiftungen und Religionsangelegenheiten des Sultanats von Oman, Netflix, Oracle und SAP.[13] Die *institutional* und *supporting members*, die ebenfalls über Stimmrechte verfügen, setzen sich aktuell aus den Regierungen von Bangladesch, Indien und Tamil Nadu (Südindien), der Universität von Kalifornien (Berkeley), Monotype Imaging und Emojipedia zusammen.[14]

Das Konsortium und vor allem natürlich die stimmberechtigten Mitglieder entscheiden im *unicode technical committee* (UTC) über die Gestaltung, Wartung und Qualität des Unicode-Standards.[15] Es sind also vor allem grosse Firmen, Institutionen oder Regierungsorganisationen, die sich die relativ teuren Beiträge für stimmberechtigte Mitglieder leisten und durch ihre Arbeit im UTC massiven Einfluss auf Schriftsysteme, Sprachen und Kulturen nehmen können (vgl. Gillam 2003, 58). Sie sind es, die in den vier-

[13] Vgl. http://unicode.org/consortium/members.html <12.09.2019>.
[14] Vgl. http://unicode.org/consortium/members.html <12.09.2019>.
[15] Vgl. https://unicode.org/consortium/consort.html <16.09.2019>.

bis fünfmal jährlich stattfindenden Treffen über Änderungen und Ergänzungen des Standards diskutieren, Vorschläge einbringen, eingereichte *proposals* zur Aufnahme neuer Zeichen prüfen und neue Versionen des Standards herausgeben (vgl. Gillam 2003, 58). Übliche Änderungen des Unicode-Standards benötigen einen Mehrheitsbeschluss im UTC, grössere, richtungsweisende Veränderungen eine Zweidrittelmehrheit (vgl. Gillam 2003, 58).

Die Monopolstellung, die Unicode bei der Codierung von Zeichen, insbesondere durch ihren Fokus auf die multilinguale Textverarbeitung, inne hat, ist einerseits ein klarer Vorteil. Alle Zeichen werden nach den gleichen Kriterien codiert und es gibt nur eine einzige globale Anlaufstelle in dieser Grössenordnung für *proposals*. Dadurch ist es auch unwahrscheinlich, dass es zu Wettbewerben und Konkurrenzsituationen kommt, zumindest was ganzheitliche Codierungssysteme anbelangt. Andererseits beinhalten Monopole von Natur aus ein gewisses Risiko. Bei Unicode entscheidet eine vergleichsweise ziemlich kleine Gruppe von Menschen über ein System, das globale Auswirkungen auf die Kommunikation der Menschheit hat. Ein bedeutender Umstand dürfte in diesem Zusammenhang aber sein, dass Unicode nicht kommerziell tätig ist.

Ungefähr einmal pro Jahr wird vom Konsortium eine neue Version des Unicode-Standards veröffentlicht, die neue Zeichen und/oder Emojis enthält.[16] Das Konsortium hat demnach eine gewisse „Gatekeeper-Funktion“ inne, wie Christa Dürscheid das in ihrem Aufsatz „Bild, Schrift, Unicode“ (vgl. Dürscheid 2018, 279-284) treffend beschreibt. Wie verantwortungsvoll ist nun aber das Unicode-Konsortium als Gatekeeper aufgebaut? Und wie verantwortungsvoll geht der Gatekeeper seiner Aufgabe nach?

Das Konsortium wird, wie zuvor erwähnt, hauptsächlich von grossen IT-Firmen getragen, die sich eine *full*, *institutional* oder *supporting membership* leisten können. Unicode wirbt auf seiner Homepage mit ausführlich dargelegten Argumenten um solche Firmen. Die Kernpunkte der Argumentation sind folgende: Durch eine Mitgliedschaft im Unicode-Konsortium können 1. von Unternehmen getätigte Investitionen in die Softwareentwicklung geschützt werden, 2. kann der Unicode-Standard so erweitert werden, dass er die eigenen Anforderungen erfüllt, 3. können moderne und globale Applikationen vorangetrieben werden und 4. kann eine Mitgliedschaft die Führungsrolle des Unternehmens demonstrieren.[17] Dass der Unicode-Standard tatsächlich so erweitert werden kann, dass er den eigenen Interessen nützt, zeigt der Fall von Google: Google ist

[16] Vgl. https://unicode.org/history/publicationdates.html <16.09.2019>.
[17] Vgl. https://www.unicode.org/consortium/why_join.html <17.09.2019>.

seit 2005 Mitglied des Konsortiums, zunächst als *associate member*, seit 2006 allerdings als *full member*.[18] Google hatte grosses Interesse daran, seinen E-Mail-Service Gmail auch in Japan anbieten zu können; um damit aber in Japan erfolgreich zu sein, war es notwendig, kompatible Emojis anbieten zu können, die dort bereits weit verbreitet waren (vgl. Bracher 2017, 7). Als eines der wichtigsten Mitglieder brachte Google den Vorschlag, die japanischen Bildzeichen in den globalen Unicode-Standard aufzunehmen, im Konsortium ein – aus dem eigennützigen Motiv heraus, sich damit den japanischen Markt erschliessen zu können (vgl. Bracher 2017, 7). Dem Vorschlag wurde stattgegeben und die Emojis sind seit der Unicode Version 6.0 aus dem Jahr 2010 verfügbar.[19] Die Aufnahme von Emojis in den Unicode war dabei keineswegs unumstritten und wirft auch heute noch kontroverse Fragen auf, wie beispielsweise folgende: Sind Emojis für die Kommunikation notwendig? Sind Schriftzeichen und Emojis gleichberechtigt? Kann die Codierung von Emojis im Unicode jemals als beendet erklärt werden? Und so weiter...

Unter den stimmberechtigten Mitgliedern des Konsortiums befindet sich aktuell nur eine einzige Institution, die linguistische Abteilung der Universität von Kalifornien in Berkeley, deren Vertreter einen dezidiert linguistischen Hintergrund besitzen.[20] Angesichts der Tragweite von Unicode und seines erheblichen Einflusses auf den Sprachgebrauch ist dies ein erschreckender Umstand. Allerdings kann er etwas relativiert werden: Unter den *executive officers* (Geschäftsführern) von Unicode befindet sich Peter Constable, ein studierter Linguist, der früher in der Lehre und Forschung des SIL International[21] tätig war.[22] Zu den technischen Leitern des Konsortiums gehören zudem Deborah Anderson vom linguistischen Institut der Universität von Kalifornien (Berkeley), die die SEI (*Script Encoding Initiative*) leitet, Ken Whistler, Ph. D. der Universität von Kalifornien (Berkeley) in Linguistik und Ken Lunde, Ph. D. der Universität von Wisconsin-Madison, ebenfalls in Linguistik.[23] Nichtsdestotrotz ist die Anzahl der Linguis-

---

[18] Vgl. https://www.unicode.org/history/contributors.html <17.09.2019>.
[19] Vgl. https://emojipedia.org/unicode-6.0/ <17.09.2019>.
[20] Vgl. https://www.unicode.org/consortium/members.html; https://www.unicode.org/history/contributors.html <20.09.2019>.
[21] Das SIL international (*Summer Institute of Linguistics*) ist eine christliche, wissenschaftlich arbeitende Nichtregierungsorganisation, die sich für die Erforschung, Förderung und den Erhalt ethnolinguistischer Minderheiten einsetzt. Das SIL ist zudem *associate member* von Unicode. Vgl. https://www.sil.org/about/discover <09.10.2019>.
[22] Vgl. https://unicode.org/consortium/directors.html; https://unicode.org/consortium/members.html <20.09.2019>.
[23] Vgl. https://unicode.org/consortium/directors.html <20.09.2019>.

tinnen und Linguisten im Konsortium, gemessen am erheblichen Einfluss des Unicode-Standards auf die weltweite Kommunikation und Forschung, als zu gering zu bewerten.

Die Codierung von Zeichen kann, wie Unicode selbst schreibt, Jahre dauern.[24] Dies liegt unter anderem daran, dass die Arbeit des Unicode-Konsortiums mit der Arbeit der ISO (*International Organization for Standardization*) synchronisiert werden muss (vgl. Gillam 2003, 58f). Grössere Probleme bzw. Kontroversen bei der Codierung gab es bisher vor allem bei der sogenannten Han-Vereinheitlichung der CJK-Zeichen, also der Vereinheitlichung von chinesischen Hanzi-, japanischen Kanji- und koreanischen Han-ja-Zeichen (vgl. Korpela 2006, 201).[25] Es wird unter anderem kritisiert, dass die Vereinheitlichung und Codierung der Zeichen nicht immer nachvollziehbar sei und die Zeichen teilweise zu stark und mit wenig Rücksicht auf deren konkrete Verwendung in den jeweiligen Sprachen vereinheitlicht würden. Im „Unicode Cookbook for Linguists" werden zudem verschiedene sogenannte *pitfalls* veranschaulicht (vgl. Moran/Cysouw 2018, 19-35). *Pitfalls* sind Elemente von Unicode, die den Nutzerinnen und Nutzern womöglich unerwartet und nicht intuitiv erscheinen – die Autoren bewerten diese aber ausdrücklich nicht eigentlich als Fehler von Unicode, sondern als Entscheidungen, die wohl aus berechtigten Gründen so getroffen wurden (vgl. Moran/Cysouw 2018, 19). Als Beispiele werden unter anderem die Umstände genannt, dass im Unicode nur einzelne Zeichen und keine Grapheme codiert werden oder dass die Namen einiger codierter Zeichen missgedeutet werden können (vgl. Moran/Cysouw 2018, 21, 25f).
Das Buch „Unicode explained" (vgl. Korpela 2006) nimmt die Arbeit des Unicode-Konsortiums ebenfalls kritisch unter die Lupe. Dabei werden Aspekte diskutiert wie die allgemein hohe Komplexität des Codes und die grosse Anzahl codierter Zeichen, die mögliche Ineffizienz aufgrund der zwei *bytes* pro Zeichen, der Vorwurf der Begünstigung westeuropäischer Sprachen und Ähnliches mehr (vgl. Korpela 2006, 198-206). Im Grossen und Ganzen wird die Kritik aber, wie im „Cookbook for Linguists", relativiert – das Unicode-Konsortium scheint bei seinem zweifelsohne sehr komplexen Projekt sehr vieles richtig gemacht zu haben. Hinzu kommt, dass es Probleme gibt, die fälschlicherweise Unicode angelastet werden, obwohl Unicode gar keine Schuld trägt: Zum

[24] Vgl. https://unicode.org/pending/proposals.html <21.09.2019>.
[25] Vgl. https://de.lexipedia.org/wiki/Han-Vereinheitlichung. Eine Auseinandersetzung mit der Problematik bietet https://web.archive.org/web/20100328042929/http://tclab.kaist.ac.kr/~otfried/Mule/unihan.html <21.09.2019>.

Beispiel unterstützen äusserst beliebte und oft verwendete Schriftarten wie Arial, Times New Roman und Verdana kein Rumänisch – obwohl Rumänisch eine Sprache mit einem vergleichsweise relativ einfachen Schriftsystem darstellt. Dieser Nachteil, dass nicht alle *fonts* mit allen Zeichen kompatibel sind, und die daraus entstehenden Unannehmlichkeiten für die Benutzerinnen und Benutzer sind aber kein Fehler von Unicode, sondern Sache der Typographen (vgl. Kappenberg 2007, 40).

Ein weiterer wichtiger Punkt bei der Betrachtung des Unicode-Konsortiums als Gatekeeper ist folgender: Übernimmt der Gatekeeper die Verantwortung für eine demokratische Auslegung, Partizipationsmöglichkeiten und Transparenz? Ein zentraler Bestandteil der Arbeit des Unicode-Konsortiums bildet die Begutachtung von *proposals*. *Proposals* sind Vorschläge zur Aufnahme neuer Zeichen[26] und Emojis[27] in den Unicode und können grundsätzlich von jeder und jedem eingereicht werden. Die Ansprüche sind allerdings ziemlich hoch:

> **Requirements of Proposal Form**[28]
> The proposal summary form requires the following information (paraphrased):
> - the repertoire, including proposed character names;
> - the name and contact information for a company or individual who would agree to provide a computerized font (True Type or PostScript) for publication of the standard;
> - references to dictionaries and descriptive texts establishing authoritative information;
> - names and addresses of appropriate contacts within national body or user organizations;
> - the context within which the proposed characters are used (for example, current, historical, and so on);
> - especially for sporadic additions, what similarities or relationships the proposed characters bear to existing characters already encoded in the standard.

Die *proposals* sind ein Mittel, neue Zeichen oder auch ganze *scripts* zur Codierung vorzuschlagen, und ermöglichen dadurch die Partizipation am Unicode, ohne zwingend zahlendes Mitglied sein zu müssen. Unterstützung bietet dabei die bereits zuvor erwähnte SEI (*Script Encoding Initiative*) des Departements für Linguistik der UC Berkeley, deren erklärtes Ziel es ist, die Vorbereitung von *script proposals* so zu finanzieren und zu fördern, dass die *proposals* ohne grössere Überarbeitung vom Unicode-Konsortium übernommen werden können.[29] Dies kommt insbesondere Minderheiten-

[26] Vgl. https://unicode.org/pending/proposals.html <24.09.2019>.
[27] Vgl. https://www.unicode.org/emoji/proposals.html <24.09.2019>.
[28] Vgl. https://unicode.org/pending/proposals.html <24.09.2019>.
[29] Vgl. https://linguistics.berkeley.edu/sei/ <21.09.2019>.

sprachen zu Gute, die vom linguistischen und technischen Know-how der SEI profitieren können (vgl. Bergerhausen/Poarangan 2011, 042). Von Zeit zu Zeit ersucht das Unicode-Konsortium zudem die breitere Öffentlichkeit um Rückmeldung zu bestimmten *proposals*, um mehr Informationen über deren praktische Auswirkungen zu erhalten und um die Überprüfung technischer Details auf eine grössere Personengruppe ausweiten zu können.[30]

Am 17. Juli 2019 lancierte Unicode eine neue Website[31], die grundsätzlich moderner und übersichtlicher gestaltet ist als die alte. Gerade für Laien ist die neue Website (auch als *public site*[32] bezeichnet) in puncto Transparenz sicherlich gut geeignet für einen ersten unkomplizierten Einblick in die Thematik. Inhaltlich stösst die neue Website aber sehr schnell an ihre Grenzen; will man bei der Recherche etwas in die Tiefe gehen, wird man über Links rasch wieder auf die alte Website[33] geleitet, die nun als *technical site*[34] bezeichnet wird. Dort lassen sich allerdings alle möglichen Informationen finden zum Unicode, dem Konsortium, zu verwirklichten, abgelehnten oder sich in der Pipeline befindenden *proposals*.[35] Unicode ist demnach in vielerlei Hinsicht sehr transparent. Übersichtlich ist die alte Website aber nur bedingt und sie wurde dementsprechend berechtigterweise schon als „hidden temple“[36] bezeichnet. Es sind zwar unheimlich viele Informationen verfügbar, diese müssen auf der komplexen Website aber erst einmal (wieder-)gefunden werden – für Aussenstehende und Laien ein nicht ganz einfaches Unterfangen. Die Komplexität der Website widerspiegelt aber nicht nur die Komplexität des Unicodes selbst, sondern auch die Vielfalt der Nutzerinnen und Nutzer: Laien ebenso wie Expertinnen und Experten sollen die Website nutzen können und zwar sowohl als Informationskanal als auch als Austauschplattform und Arbeitsinstrument. So gesehen ist der Schritt zur Aufteilung in eine *public* und eine *technical website* sicherlich sinnvoll und im Hinblick auf die Nutzerinnen und Nutzer mit unterschiedlichen Ansprüchen realisiert worden. Die stark vereinfachte *public website* hat also ganz konkret

[30] Vgl. https://www.unicode.org/review/ <21.09.2019>.
[31] Vgl. https://home.unicode.org/ <18.09.2019>.
[32] Vgl. https://home.unicode.org/basic-info/how-to-use-this-site/ <24.09.2019>.
[33] Vgl. https://unicode.org/main.html <18.09.2019>.
[34] Vgl. https://home.unicode.org/basic-info/how-to-use-this-site/ <24.09.2019>.
[35] Vgl. https://www.unicode.org/alloc/Pipeline.html <18.09.2019>.
[36] Vgl. https://medium.com/@maggieshafer/unicode-a-story-of-corruption-connection-and-smiling-poo-598295e4af9d <18.09.2019>.

zum Ziel, eine breitere Öffentlichkeit zu erreichen und deren Partizipation zu ermöglichen.[37]

2.3 Das Paradox des „unbekannten Weltstandards“[38]

Was bei Unicode erstaunt, ist der Umstand, dass er bisher weitgehend unbekannt geblieben ist. Unicode ist bereits jetzt ein Weltstandard, der von einem enormen Teil der Menschheit tagtäglich genutzt wird und der vor allem durch die Emojis gewissermassen in aller Munde ist: Die Emoji-Updates werden mittlerweile sehnlich erwartet und in den Medien kommentiert (vgl. Dürscheid 2018, 281). Trotz dieser Umstände wissen die wenigsten, wie Unicode aufgebaut ist, funktioniert und was die eigentliche Zielsetzung des Konsortiums ist. Dass vielen Laien den Begriff Unicode nur im Zusammenhang mit der jährlichen Veröffentlichung der Emoji-Updates kennen, dürfte für das Unicode-Konsortium Fluch und Segen zugleich sein: Fluch deshalb, weil das eigentliche „Kerngeschäft“, die Codierung von Zeichen, trotz seiner grossen Wichtigkeit wenig Beachtung findet – Segen deshalb, weil von der Aufmerksamkeit, die zunächst den Emojis zufällt, möglicherweise auch die Codierung von Zeichen profitieren kann.

Gerade für Minderheitensprachen ist die geringe Aufmerksamkeit und das fehlende wirtschaftliche Interesse grosser Firmen ein Problem: Die Codierung ist abhängig von Mitgliederbeiträgen und Spenden – ansonsten können Schriftsysteme nicht oder nur in langwierigen Prozessen codiert werden. Um dem entgegenwirken zu können, wurde die „adopt a character“-Aktion ins Leben gerufen (vgl. Bracher 2017, 6). Durch die so generierten finanziellen Mittel werden Unicode-Projekte finanziert – aktuell beispielsweise die Erforschung und Codierung von Maya-Hieroglypheninschriften.[39] Auch hier zeigt sich allerdings der Emoji-Trend: Obwohl diverse Zeichen wie Buchstaben, Zahlen, Interpunktionszeichen, Umlautzeichen oder Symbole adoptiert werden könnten[40], entscheiden sich die meisten für die Adoption eines Emojis.[41] Firmen unterstützen die „adopt a character“-Aktion zudem durchaus nicht nur aus Selbstlosigkeit: Die Sponsoren werden neben dem adoptierten Emoji, das meist in enger Beziehung zur Firma oder

---

[37] Vgl. https://home.unicode.org/the-unicode-consortium-launches-new-website-in-celebration-of-world-emoji-day-2/ <24.09.19>.
[38] Bezeichnung übernommen von Bergerhausen/Poarangan, 2011, 009.
[39] Vgl. https://home.unicode.org/adopt-a-character-grant-to-support-maya-inscriptional-hieroglyphs/ <07.10.2019>.
[40] Vgl. https://www.unicode.org/consortium/choosing.html <07.10.2019>.
[41] Vgl. https://www.unicode.org/consortium/adopted-characters.html <07.10.2019>.

Marke steht, namentlich genannt und die Firmenhomepage direkt verlinkt[42] – eine willkommene Werbeplattform.

Es ist sicherlich wünschenswert, dass der „unbekannte Weltstandard“ mehr ins Bewusstsein der Öffentlichkeit dringt. Durch die zuvor erwähnte neue *public website* ist in dieser Hinsicht ein erster Schritt in die richtige Richtung unternommen worden. Nichtsdestotrotz bleibt definitiv noch viel Luft nach oben – vor allem auch von Seiten der Akademikerinnen und Akademiker verschiedener Fachbereiche. Es erstaunt, dass Unicode bisher nicht grössere Aufmerksamkeit in der Fachwelt erhalten hat und nicht breiter diskutiert worden ist. Um bei Unicode tatsächlich als stimmberechtigtes Mitglied Einfluss nehmen zu können, sind selbstverständlich die notwendigen finanziellen Mittel aufzubringen, was wahrscheinlich für viele ein Mitgrund für die mangelnde Partizipation darstellt. Zudem sind mittlerweile die wichtigsten und meistverbreiteten Schriftsysteme ohnehin codiert – für die noch ausstehenden Schriftsysteme für ausgestorbene oder Minderheitensprachen fehlt womöglich vielen das notwendige und langfristige Interesse und/oder Know-how. Nichtsdestotrotz wäre eine Auseinandersetzung mit der Thematik von Unicode durch die Fachwelt sehr erwünscht und wird über kurz oder lang auch nicht mehr umgangen werden können. Einer solchen Auseinandersetzung stünde auch nichts im Wege – eine Mitgliedschaft ist nicht zwingend notwendig, um sich mit der Thematik auseinanderzusetzen. Will man aber, über *proposals* hinaus, aktiv Einfluss nehmen auf die Entwicklung des Unicodes, ist eine Mitgliedschaft nach wie vor unumgänglich.

3 Wie das System Schreibprozesse beeinflusst (Livas)

3.1 Kritik am Unicode

Seit der Unicode und das Unicode-Konsortium ins Leben gerufen wurden, haben sich viele kritische Stimmen über die Vor- und Nachteile von Unicode geäussert. Viele Kritiken setzen berechtigterweise an der Verantwortung an, die im Umgang mit Unicode gegeben ist. Der Unicode selbst hat, wie an einigen Beispielen ersichtlich werden wird, politische Schlagkraft. Dass der Unicode in jedem Fall Arsenal in technischer Form darstellt, war wohl nicht allen immer bewusst. Sinnfällig wird dies indirekt etwa an den

[42] Vgl. https://www.unicode.org/consortium/adopted-characters.html <07.10.2019>.

Debatten um die Einführung von Swastiken: Es handelt sich dabei um Zeichen, welche in vielen asiatischen Ländern Glück und Gesundheit symbolisieren, im 20. Jahrhundert jedoch als Symbol des Nationalsozialismus in Deutschland von sich reden machten. Nachdem Microsoft Office 2003 die Swastiken als Teil des Unicodes in die Schriftart „Bookshelf Symbol 7“ aufgenommen hatte, wurden Beschwerden dagegen laut. Danach entfernte Microsoft Office die Swastiken wieder, mitsamt einem Davidstern (vgl. Maher 2014).

Das Unicode-Konsortium selbst hat sich seinem Leitspruch entsprechend – „*universal, uniform, unique*“[43] – zum Ziel genommen, alle Schriftzeichen der Welt in den Unicode aufzunehmen[44]; Platz genug ist in dem 1996 erweiterten Code-Raum dafür (vgl. Bergerhausen/Poarangan 2014, 029). Dieselbe Zielsetzung zeichnet sich als Trend auch für die ab 1995 aufgenommenen, in der Version 6.0 von 2010 offiziell als „Emoji“ bekanntgewordenen Bildzeichen[45] ab und wird in Zukunft wohl auch die stark kritisierte Han-Vereinheitlichung (vgl. Cheong 1999; Topping 2001; Kubota 2002) betreffen: Viele Schriftzeichen des chinesischen, japanischen und koreanischen Sprachraumes (CJK) gründen auf denselben Wurzeln der Han-Dynastie des 3. Jahrhunderts. In der Han-Vereinheitlichung wurden die CJK-Zeichen mit gemeinsamer Wurzel jeweils dem gleichen unverwechselbaren *code point* zugeordnet. CJK-Nutzer müssen nun in den betreffenden Regionen mit einer spezifisch ‚japanischen‘, ‚chinesischen‘ oder ‚koreanischen‘ Schriftart arbeiten, um das für die Region gebräuchliche, passende Schriftzeichen verwenden zu können. Nicht allein der Umstand, dass der CJK-Nutzer sich in hohem Masse mit dem Schriftstil seiner Region identifiziert, ist prekär. Selbst wenn ein japanischer Schreiber einwilligen würde, eine Hongkonger Schriftart zu nutzen, würde er fallweise die gesuchten Schriftzeichen nicht erkennen: Obwohl von den gleichen Han-Schriftzeichen ausgegangen wird, können sich die Glyphen in den jeweiligen Schriften wesentlich voneinander unterscheiden, sodass sie Angehörige verschiedener Schriftkulturen nicht ohne weiteres – oder gar nicht – als dasselbe Schriftzeichen identifizieren würden. Die erhoffte gegenseitige Verständigung durch die Han-Vereinheitlichung, abgeleitet aus der gemeinsamen Vergangenheit der Han-Dynastie, ist in diesem Fall nicht gegeben. Zudem existieren auch innerhalb der gleichen Schriftkultur unterschiedliche Glyphen desselben Schriftzeichens, deren Einsatz von Kontext, Anlass und An-

[43] Vgl. http://www.unicode.org/history/summary.html <10.10.2019>.
[44] Vgl. http://www.unicode.org/standard/WhatIsUnicode.html <10.10.2019>.
[45] Vgl. https://www.unicode.org/emoji/charts/emoji-versions.html <10.10.2019>.

gemessenheit abhängig ist. Auch dieser Sachlage wurde in der Han-Vereinheitlichung nicht Rechnung getragen.

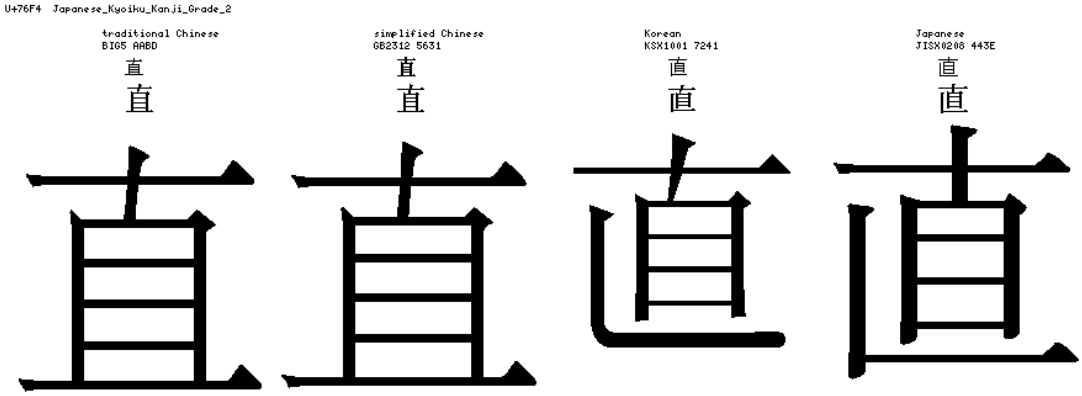

Abbildung 1: Das Han-Schriftzeichen mit der Code-Bezeichnung U+76F4 steht für die Bedeutung <gerade>. Die beiden chinesischen Schriftzeichen links sind dem koreanischen und japanischen Leser allerdings in den meisten Fällen nicht geläufig (vgl. Kubota 2002).[46]

Die Inkorporierung sämtlicher Schrift- und Bildzeichen dieses Planeten[47] scheint den Unicode zu rehabilitieren und auch den Kritikern, welche vor den negativen Folgen von Internationalisierung und Globalisierung warnen, Einhalt zu gebieten; denn schliesslich trägt Unicode mit der damit einhergehenden Förderung von Minderheitensprachen zu einem grösseren sprachlichen Pluralismus bei. Gegen die Verallgemeinerung von Auswirkungen der Globalisierung argumentiert auch Geert Lovink in seinem essayistischen Beitrag „Maschinen sprechen eigentlich Deutsch“, der 2007 im Sammelband „McDeutsch“ (hrsg. von Krystian Woznicki) erschienen ist: „Die Annahme, dass mit dem Internet alle auf einmal anfangen Englisch zu schreiben, stimmt einfach nicht. Reden, vielleicht. Lesen, ja, wer weiss. Aber bestimmt nicht Schreiben.“ (Lovink 2007, 60) Lovink interessiert sich dabei für „die gängige These, das Englische reisse die Weltherrschaft an sich und das Internet sei sein Vehikel – das Ganze soll aus der Netzperspektive angegangen werden.“ (ebd.)

Seine Beobachtung zum Schreiben scheint zunächst zuzutreffen (vgl. auch Bergerhausen/Poarangan 2014, 042), fragt aber im essayistischen Format nicht nach den Hintergründen. Zum einen ist Unicode im Jahr 2007 noch weniger internetweit etabliert als

[46] Vgl. http://www.unicode.org/cgi-bin/GetUnihanData.pl?codepoint=76F4 <10.10.2019>.
[47] Vgl. http://linguistics.berkeley.edu/sei/index.html <10.10.2019>. Das Projekt der „Script Encoding Initiative (SEI)“ unter der Leitung von Dr. Deborah Anderson setzt sich insbesondere für die Kodierung von Minderheitensprachen ein (vgl. Dürscheid 2018).

heute. Zum anderen ist der Einfluss des Englischen im Internet nicht ausschliesslich auf lexikalischer Ebene zu suchen, sondern, da die Kommunikation über digitale Geräte läuft, in den Strukturen, über das sich das Internet und die Internetsprache konstituiert. Auch „das Englische“ und „das Internet“ sind diffuse Begriffe: Zu fragen wäre, welche Faktoren mit Bezug auf die Sprache genau Einfluss auf welche Bereiche des Internets nehmen. Interessant ist die Sonderstellung, die Lovink dem Schreiben zuspricht. Sollten „das Englische“ oder andere präskriptive Faktoren einen Einfluss auf den Sprachgebrauch über das Internet ausüben, so bedeutet das, dass sich dieser Einfluss entweder nur schwach auf das Schreiben auswirkt, sich im Schreiben an Stellen manifestiert, welche sich dem Betrachter nicht auf den ersten Blick erschliessen oder dass der Prozess stark verlangsamt stattfindet.

3.2 Präsuppositionen des Schreibens

Der Unicode scheint – heute und mindestens in der Theorie – die lückenlose digitale Kommunikation verschiedener Schriftgemeinschaften überall auf der Welt, international und interkulturell, zu ermöglichen. Letzte Hürden präsentieren sich nur noch in den scheinbar rein technischen Mängeln, welche durch die richtige Anwendung des Unicodes zu beseitigen wären.

Abbildung 2: Viele Fehler entstehen durch die Inkompatibilität unterschiedlicher Zeichenstandards (vgl. Bergerhausen/Poarangan 2014, 018). Mark Davis, Präsident des Unicode-Konsortiums, kommentiert dies hier auf seinem Twitter-Account mit einem Seufzer.[48]

Die Diskussion um Unicode als Gatekeeper (vgl. Dürscheid 2018; „Unicode and Totality“, Lennon 2010, 168) scheint sich zunehmend von den Schriftzeichen auf die Emojis zu verlagern. Gewisse Problematiken, die in der Schreibproduktion auftreten, sind jedoch, so die These, nicht auf einen fehlerhaften Umgang mit dem Unicode zurückzu-

[48] Vgl. https://twitter.com/mark_e_davis <10.10.2019>.

führen: Sie werden durch den Unicode selbst generiert. Im Folgenden stelle ich einige Überlegungen zu dieser These an.

Der Unicode ordnet den jeweiligen Schriftzeichen eine bestimmte Stelle in einer Matrix (*code point*) zu (vgl. Margulies 2009, 115). Die digitalen Schriftzeichen entsprechen dabei den zu Ende geschriebenen Schriftzeichen des analogen handschriftlichen Schreibens: Sie liegen in ihrer endlichen Form vor, wie dies in einem von Hand verfassten Manuskript der Fall ist, aber auch in einem gedruckten Buch. Die Schriftzeichen in ihrer endlichen Form an einem Ort abzubilden, wo es um deren Darstellung geht, ist gerechtfertigt. Doch muss klar sein, dass die digitalen Schriftzeichen aus dem Unicode nicht dazu benutzt werden, diese im bereits geschriebenen Zustand zu präsentieren, sondern zum Schreiben: Ihr Einsatz erfolgt *während* des Schreibprozesses, nicht danach.

Von Zeichen auszugehen, die während des Schreibens bereits ihre endliche Form aufweisen, gestaltet sich für Schriftsysteme des lateinischen Alphabets (auch römisches Alphabet genannt) vergleichsweise unproblematisch: „Text-processing technology arose in the English-speaking world and assumed as a norm the use of the Roman alphabet with few or no diacritics.“ (Hyman/Scharf 2011, 114) Das Schreiben in Sprachen, die sich des lateinischen Alphabets bedienen, lässt sich auch in einer Matrix bzw. in einem Raster darstellen: Auf syntagmatischer Ebene erfolgt ein imaginärer Platzhalter für einen Buchstaben (oder für eine Leerstelle oder ein Satzzeichen). Auf paradigmatischer Ebene würde das passende Zeichen eingefügt. So kann nach einem <S_> beispielsweise entschieden werden, ob nun mit <p> oder <t> fortgefahren wird, ob aus dem <S_> ein <Spiel> wird oder ein <Stiel>. Strukturell gesehen funktioniert der Unicode auf dieselbe Weise: Einer unverwechselbaren Stelle im Unicode wird ein ganz bestimmtes Zeichen zugeordnet. Für Leerstellen für kommende Schriftzeichen ist ebenfalls vorgesorgt.

Diese Vorgehensweise geht solange gut, wie in einem Schriftsystem operiert wird, in dem auf jede Leerstelle exakt ein Zeichen kommt und syntagmatisch Zeichen um Zeichen fortgeschritten wird. Doch wie verhält es sich mit Sprachen, deren Schriftsysteme anders aufgebaut sind? Deren Schriftzeichen ambig sind? In denen auf eine Leerstelle im Syntagma mehr als ein Zeichen kommt?

Ein solches Beispiel repräsentieren die Umlaute der deutschen Sprache. Das deutsche Schriftsystem bedient sich zwar ebenfalls des lateinischen Alphabets, enthält jedoch auch die Umlaute <Ä, ä, Ö, ö, Ü, ü> sowie das grosse und das kleine Eszett <ß, ß>. Auf dieser Ebene finden sich die meisten Fehler, die ihren Ursprung nicht immer in der Inkompatibilität verschiedener Zeichenstandards haben, sondern in ein und demselben

Standard auftreten können, in diesem Fall im Unicode-Standard. Sie gründen im Verständnis der Umlaute selbst: Was beispielsweise ist ein <ä>? Ist es ein <a> mit zwei Punkten? Oder stellt das <ä> ein selbstständiges Zeichen dar, da es auf einem verwandten, aber doch eigenen Laut beruht, losgelöst vom <a>? Die meisten Muttersprachler würden wohl für beide Perspektiven Argumente finden.

Auch die technischen Lösungen für dieses Problem sind entsprechend vielseitig. Unicode offeriert für diese Zwecke zwei Möglichkeiten: Eine vorkombinierende und eine dekombinierende Variante. Letztere bricht Schriftzeichen jeweils auf ihre Basiseinheiten herunter: Das <a> sowie die zwei Punkte werden jeweils separat kodiert und erst im Text zum Zeichen <ä> kombiniert.[49] Unicode nennt diesen Block „Combining Diacritical Marks".[50] In der anderen Variante liegt das Zeichen <ä> ohne weitere Untereinheiten vor: Es wird hier als „LATIN SMALL LETTER A WITH DIAERESIS"[51] bezeichnet, obwohl manche mit Fug und Recht behaupten würden, dass es sich hier definitiv nicht um einen ‚lateinischen Kleinbuchstaben' handelt, sondern um ein Schriftzeichen des deutschen Schriftsystems. In den Bezeichnungen des Zeichens <ä> spiegelt sich eine Problematik, die auch auf technischer Ebene wirkt. Welcher *code point* nun beim <ä> abgerufen wird, hängt von der jeweiligen Schriftverarbeitung ab, die zum Einsatz kommt. Je nachdem, ob die Schriftart (*font*) des gesendeten Dokumentes auch auf dem Empfängergerät installiert ist und welche Zeichen die Glyphen jener Schriftart umfassen, gelingt der Transfer, oder auch nicht. Doch geht es mir hierbei nicht um die Qualität der Schriftart, darum, wie gut sie ausgebaut ist, um möglichst viele Zeichen darstellen zu können. Tatsache ist zum einen, dass ausgeklügelte Schriftarten notwendig sind, um den Unicode-Standard unterstützen zu können. Zum anderen handelt es sich beim Zeichen <ä> de facto um rund drei verschiedene *code points*, und zwar <ä>, <a> und <¨>, welche drei völlig voneinander isolierte Einheiten darstellen. Das Zeichen <ä>, welches in vielerlei Hinsicht komplexere und vielschichtigere Züge trägt, als sich zunächst erahnen lässt, wird im Unicode[52] auf drei Aspekte des Buchstabens reduziert, welche zudem starr in den Code eingebunden sind und nur minimale Modifikationen erlauben. Im analogen Schreiben stellt sich das Problem nicht, alle Varianten des oben genannten Beispiels können im Schreibakt mitgedacht werden, während im

[49] Weitere Beispiele werden in folgendem Blog beschrieben: http://www.swiftkurs.de/zeichen-und-unicode/ <12.10.2019>. Diese finden in der Programmiersprache Anwendung.
[50] Vgl. https://unicode.org/charts/PDF/U0300.pdf <10.10.2019>.
[51] Vgl. https://decodeunicode.org/en/u+000E4 <10.10.2019>.
[52] Vgl. ebd.

Moment der Enkodierung des Schriftzeichens für den Schreiber bereits gedacht wurde. Der Unicode in Form einer Matrix zwingt so den Programmierer der Schriftverarbeitung, sich im vorliegenden Beispiel für eine Perspektive entscheiden zu müssen, nämlich für die vor- oder dekombinierte Variante des Zeichens <ä>. Dieses Phänomen tritt auch in anderen Sprachen mit diakritischen Zeichen auf.

Ein weiteres Beispiel repräsentieren die 214 Radikale der chinesischen Schrift, welche in teils abgewandelter Form auch in der japanischen und koreanischen Schrift (CJK-Zeichen) gebräuchlich sind und oft die Basis für komplexere Schriftzeichen formen. Im Unicode werden diese als sogenannte „Kangxi Radicals" in einem eigenen Block geführt.[53] Dieselben Zeichen sind noch einmal separat im Block der „CJK Unified Ideographs" enthalten.[54] In letzterem Block stehen die Zeichen allerdings nicht für ihre Eigenschaft als Radikale, sondern werden als unabhängige Schriftzeichen gelistet, welche eine eigenständige Semantik aufweisen. Auch in diesem Beispiel werden jeweils zwei verschiedene Bedeutungseinheiten eines Schriftzeichens durch den Unicode voneinander isoliert.

In der Monographie „Linguistic Issues in Encoding Sanskrit" gelangen Peter M. Scharf und Malcolm D. Hyman in der Enkodierung des Sanskrits zu einem ähnlichen Schluss:

> „Encoding standards for Sanskrit that are based on Devanāgarī or Romanization inherit the deficiencies inherent in the underlying scripts. They suffer from ambiguity and redundancy by departing from a one-to-one correspondence and by inconsistency in the basis for encoding." (2011, 114)

Scharf und Hyman führen demnach bestimmte Kodierungsprobleme im Sanskrit ebenfalls auf die Eins-zu-eins-Korrelation im lateinischen Alphabet zurück, dessen Prinzip präsuppositiv im Unicode umgesetzt ist.

3.3 Der Unicode als Schreibpraxis

Die Theorie hinter Unicode wirkt sich nicht nur auf den Programmierer aus, sondern gerade auch auf den Nutzer, der heute infolge des sich etablierenden Unicode-Standards seinen Schreiballtag mehr als je zuvor mit der täglichen Anwendung des Unicodes bestreitet. Der digitale Schreibprozess hat den Nutzer vieler Überlegungen enthoben, welche mit dem analogen Schreibakt einhergehen und denen sich nun der Programmierer ausgesetzt sieht. Der Nutzer profitiert dafür von der Akzeleration des Schreibprozesses.

[53] Vgl. http://www.unicode.org/charts/PDF/U2F00.pdf <11.10.2019>.
[54] Vgl. http://www.unicode.org/charts/PDF/U4E00.pdf <11.10.2019>.

Letzterer Vorteil verschaffte der Schreibmaschine im 19. Jahrhundert schnell den Status des „unentbehrlichen Werkzeug[s]“ für den „Einsatz in Büros“ (Ludwig 1994, 62). Die Entscheidungsprozesse, welche heute beim Programmieren und Enkodieren anfallen, oblagen im 19. und 20. Jahrhundert den Konstrukteuren der Schreibmaschinen. Um keine riesigen Geräte produzieren zu müssen, wurden die mechanischen Vorgänge weitestgehend vereinfacht: Auch die Konstrukteure hatten sich folglich für eine technische Lösung zur Realisierung von Buchstaben mit diakritischen Zeichen zu entscheiden. Genau wie es heute bei Unicode der Fall ist, wurden auch in der Schreibmaschinenmechanik mehrere Varianten umgesetzt: Der Nutzer konnte Buchstaben mit diakritischem Zeichen mit einer sogenannten Tottaste simultan umsetzen. Aufwendiger gestaltete sich die Möglichkeit, genau ein Schriftzeichen im Dokument zurückzugehen, um einem Buchstaben ein diakritisches Zeichen nachträglich hinzuzufügen. Der Nutzer musste sich also praktisch schon beim Kauf der Maschine für eine Perspektive auf die diakritischen Zeichen entscheiden.

Die Problematik des einseitig orientierten Schreibens schien jedoch im Zeitalter der Schreibmaschine nicht direkt auf, da die unterschiedlichen Lösungen – hier z.B. für die Umsetzung von Umlauten –, die in die Schreibmaschinen implementiert waren, nicht versuchen mussten, miteinander zu kommunizieren, wie es beim heutigen Internet der Fall ist: Das Internet kommuniziert nicht über Dokumente, sondern über Arten des Schreibens.

Der Unicode ist letztlich nichts anderes als eine standardisierte Form des Schreibprozesses. Dem reflektierten Nutzer mussten aber die Defizite des Schreibmaschinensystems gegenüber des handschriftlichen Schreibens mehr oder weniger bewusst gewesen sein: Die Matrix geht an dieser Stelle aus dem Akt der Simplifizierung aufgrund der angestrebten „Handhabbarkeit“ (Ludwig 1994, 62) der Schreibmaschine hervor, dem der Akt der Differenzierung und Vielfalt aus linguistischer Perspektive vorzuziehen wäre. Die Vorteile der Schnelligkeit und damit natürlich auch der Wirtschaftsfähigkeit im Zeitalter der Industrialisierung überwogen jedoch.

Die Geschichte des Schreibmaschinenschreibens beginnt aber eigentlich schon mit der Erfindung des Buchdrucks: „Die Schreibmaschine führte das Prinzip des Druckens, das bisher ausschliesslich bei der Multiplikation von Texten genutzt worden war, in die Praxis des Schreibens generell ein.“ (Ludwig 1994, 62) Der Druck fand zunächst vor allem in Asien und danach in Europa Verbreitung (vgl. Ludwig 1994, 60). Doch war der Druck einerseits regional beschränkt, andererseits ging man für den Textdruck fast im-

mer vom handschriftlichen Manuskript aus. Heute wird eine interlinguale Vernetzung angestrebt; was aber noch viel wichtiger ist: Der Prozess des Druckens, im Sinne einer simultanen Reproduktion eines Schriftzeichens, hat sich in den Prozess des Schreibens integriert. Während sich der Buchdruck zum Schreiben im lateinischen Alphabet (welches ohne diakritische Zeichen auskommt) sehr gut eignet, bildet die Matrix des Buchdrucks jedes andere Schriftsystem, mit dem geschrieben werden soll, nur unzureichend ab. Auf diese Weise erscheint der Aufstieg der angloamerikanischen Weltmacht in anderem Licht: Das im lateinischen Alphabet implizierte Verständnis der Verschriftlichung einer Sprache, kombiniert mit den Vorteilen des Druckverfahrens, ergaben eine durchschlagende Wirkung. Doch wird dies nicht der einzige Faktor sein, weshalb sich Englisch als „Sprache der Technologie“ (Lovink 2007, 60) durchgesetzt hat.

Der Nutzer verfügt heute über eingeschränkte Möglichkeiten im Gebrauch der Tastatur: Meist wird ihm das Schriftzeichen in endlicher Form vorgeschlagen. Interessanterweise bietet das Unternehmen Apple, dessen Interfaces auf die intuitive Bedienung ausgerichtet sind, die Möglichkeit an, z.B. das Zeichen <ä> über die Taste <A> zu wählen, indem diese lange gedrückt wird. Auch hier muss eine zusätzliche technische Stütze zum Unicode bereitgestellt werden, welche die Zweidimensionalität der Matrix im Nachgang zur Mehrdimensionalität umfunktioniert. Dies bedeutet nicht, dass Unicode die Vielschichtigkeit und Mehrdeutigkeit einer Sprache, eines Schriftsystems oder eines Schriftzeichens gänzlich unreflektiert belässt: Unicode versucht im grossen Freiraum, der in den noch unbelegten Blöcken verfügbar ist, möglichst alle Bestandteile eines Zeichens aufzunehmen: In den Blockbezeichnungen schimmert das Reflexionsmoment hindurch. Auch in der Anordnung der Zeichen innerhalb eines Blocks wird, wenn auch inkonsistent, eine sinnvolle Ordnung angestrebt. Trotzdem hindert der Unicode durch seinen eigenen Aufbau daran, diese Reflexionsmomente auch technisch umsetzen zu können. Die Struktur der Matrix holt sie immer wieder ein.

Dies zeigt sich auch an einem von Bernd Kappenberg beschriebenen Beispiel des Schweizer Fernsehens (vgl. Kappenberg 2012, 225), dessen Fall auf der Website „NETSCOUT“ (hrsg. von Andreas Renggli) unter dem Titel „Neue Rechtschreibung für fremdsprachige Familiennamen“ wie folgt zusammengefasst wird:

> „Das Schweizer Fernsehen hat die redaktionseigene Auslegung der deutschen Rechtschreibung überarbeitet. Ab sofort gilt für fremdsprachige Nachnamen die lautnahe Schreibweise. Sie folgt den phonetischen Regeln, verzichtet jedoch auf diakritische Zeichen. [...] Auch der Suchmaschinenoptimierung scheint die

Massnahme zu dienen: Mit ‚Lindsey Won' erscheint der Sender bei Google zuoberst, mit ‚Lindsey Vonn' nur auf Platz zwei."[55]

Abbildung 3: Lindsey Vonn heisst in der „redaktionseigenen Auslegung" (Kappenberg 2012, S. 225) Lindsey Won.[56]

Der Grund, weshalb Stichworte ohne diakritische Zeichen von Suchmaschinen leichter gefunden werden, kann mit den Kodierungsprozessen von Unicode erklärt werden: Der Unicode stellt für die meisten Schriftzeichen ohne diakritisches Zeichen einen einzigen *code point* bereit, für Schriftzeichen, in denen zwei Zeichen desselben Paradigmas auf eine Stelle im Syntagma kommen, jedoch mehrere. Suchmaschinen müssen also in ihrer Suche von komplexeren Masken ausgehen, in denen alle Schreibvarianten, im Sinne der Kombination von Unicode-Zeichen, berücksichtigt werden. Die Entscheidung zur vereinfachten Suche liegt in diesem Fall jedoch bei der Redaktion und ist nicht durch Unicode selbst generiert. Komplexere Masken liegen im Bereich des Möglichen und es läge in der Verantwortung der Redaktion, den andersgearteten Aufbau von verschiedenen Sprach- und Schriftsystemen zu berücksichtigen. Als greifbare Lösung für die Zweidimensionalität der Matrix schaffen also Programme Abhilfe, welche die verloren gegangenen Beziehungen der Schriftzeichen untereinander wiederherstellen und so

[55] Vgl. https://www.netscout.ch/20110311/schweizer-fernsehen-neue-rechtschreibung-fremdsprachige-familiennamen <11.10.2019>.
[56] Vgl. ebd.

Verbindungen schaffen, die das Raster des Unicodes nicht bieten kann. Dies verkompliziert die ganze Sache natürlich enorm und stellt den Mehrwert des Unicodes infrage, da ursprünglich ebenjene „Insellösungen“ (Bergerhausen/Poarangan 2014, 018) der Grund für die Einführung des Unicodes waren.

Die letzte Frage muss denn schliesslich lauten: Ist der Mensch ein tippendes Wesen? Programme, an denen gearbeitet wird, welche die Involvierung der Handschrift beinhalten, die über Tablets in digitale Schriftzeichen umgewandelt wird, mögen die Defizite beim Schreiben mit Unicode beheben können. Der Vorteil der Schnelligkeit in der Produktion wird dabei verloren gehen, derjenige der Schnelligkeit bei der Verarbeitung eines gewissen Prozentsatzes an Information wird dafür weiterbestehen (Big Data). Wahrscheinlicher ist jedoch, dass sich die individuelle Handschrift in der Welt der digitalen Schriften nach wie vor behaupten wird, da durch die Handschrift eine Fülle an zusätzlichen Informationen transportiert wird, was sich am einzigartigen Ausdruck einer jeden Handschrift ablesen lässt. Dieser Umstand spielt für die vernetzende Kommunikation in einem sozialen Gefüge eine massgebliche Rolle; besonders, wenn man bedenkt, dass erst ein Bruchteil dessen bekannt ist, wie Informationsaustausch in der menschlichen Interaktion zustande kommt.

Für den Unicode präsentieren sich in puncto Internationalisierung und Globalisierung im Sinne einer weltweiten Vernetzung drei unterschiedliche Zukunftsszenarien: Würden die zur Zeit notwendigen, individuellen Modifikationen an den einzelnen Rechnern durch Zusatzprogramme zur differenzierten Nutzung von Unicode vorgenommen, würde dies nach wie vor die Effektivität des Unicodes infrage stellen – sieht man von der Inventur-Leistung ab (vgl. Weingarten 2002; vgl. Bergerhausen/Poarangan 2014, 011), die Reminiszenzen an das 19. Jahrhundert weckt. Würde der Unicode unmodifiziert weitergeführt, nähme das Schreckgespenst ‚Uniformierung der Vielfalt von sprachlichen Strukturen‘ Gestalt an, da sich weltweit Schriftsysteme an die kodierungssystembedingte Struktur der Matrix assimilieren würden. Als dritte Option kann eine weltumspannende flüssige Kommunikation imaginiert werden, die auch auf strukturell-technischer Ebene jeder Eigenheit einer Sprache gerecht wird und so die Vielfalt der Sprachwelt stärkt. Die Voraussetzung dafür ist eine flexible Technologie, welche es vermag, sich an die fortwährenden Entwicklungen einer Sprache und an deren wechselnden Formenreichtum zu adaptieren und die gleichzeitig das gesamte vorhandene Wissen um ein Schriftsystem und dessen Zeichen vereint – eine Technologie der Zukunft.

4 Fazit

Der Blick auf die Organisationsstrukturen des Unicode-Konsortiums zeigt, dass sich das Konsortium in vielen Aspekten seiner Rolle als Gatekeeper bewusst ist: Durch seine Anlage als Non-Profit-Organisation, die demokratische Auslegung durch Abstimmungsverfahren, Partizipationsmöglichkeiten durch Mitgliedschaft oder durch Vorschläge zur Aufnahme neuer Zeichen (*proposals*) und die grosse Transparenz versucht es, seiner Verantwortung als „unbekannter Weltstandard" (Bergerhausen/Poarangan 2011, 009) nachzukommen. Entscheidungen in der Zeichenkodierung trifft es unter Miteinbezug möglichst vieler Faktoren. Dennoch existiert in einigen Bereichen Nachholbedarf: Die Mitgliedschaften und damit die Stimmberechtigung in wichtigen Beschlüssen werden durch Mitgliederbeiträge geregelt. Dies führt dazu, dass sich das Konsortium mehrheitlich aus Unternehmen zusammensetzt, welche oft eigene Interessen verfolgen, die Zahl der Linguistinnen und Linguisten dagegen deutlich beschränkt ist. Um einen verantwortungsvollen Umgang mit dem Unicode gewährleisten zu können, wäre eine verstärkte Beteiligung von Mitgliedern mit linguistischem Hintergrund wünschenswert. Dies würde auch einen reflektierteren Umgang mit verschiedenen Kodierungsfragen ermöglichen; gerade auch mit jenen, welche durch systemimmanente Vorgänge entstehen, wie z.B. die Kodierung verschiedener Ebenen von Schriftzeichen in ein System, das sich aus isolierten *code points* konstituiert. Aussagen darüber, wie sich Unicode auf den Schreibakt und unser Schreibverhalten auswirkt, setzen eine eingehende Beschäftigung mit den Grundpfeilern von Unicode voraus. An dieser Stelle seien insbesondere Wissenschaftler diverser Fachrichtungen dazu aufgefordert, sich kritisch mit dem Unicode auseinanderzusetzen, aktiv Vorschläge oder Korrekturen einzubringen und sich im Idealfall sogar als *institutional* oder *full member* im Konsortium zu engagieren. Dies wäre ein Schritt in die richtige Richtung, wodurch zukünftig den sich stellenden Herausforderungen mit produktiven Lösungen begegnet und der Verantwortung nachgekommen werden kann, die mit einem Weltstandard einhergehen.

5 Bibliographie

Bergerhausen, Johannes/Poarangan, Siri (2011): decodeunicode – Die Schriftzeichen der Welt. Mainz: Hermann Schmidt.

Bracher, Katharina (2017): Emojis – Nachricht mit Gefühl. In: NZZ, 08. Januar 2017, S. 1-11. Online unter: https://nzzas.nzz.ch/gesellschaft/emojis-nachricht-mit-gefuehl-ld.1336511 <09.10.2019>.

Cheong, Otfried (1999): Han Unification in Unicode. Online unter: https://web.archive.org/web/20100328042929/http://tclab.kaist.ac.kr/~otfried/Mule/unihan.html <15.10.2019>.

Dürscheid, Christa (2018): Bild, Schrift, Unicode. In: Mensching, Guido/Lalande, Jean-Yves/Hermes, Jürgen/Neuefeind, Claes (Hrsg.): Sprache - Mensch - Maschine. Beiträge zu Sprache und Sprachwissenschaft, Computerlinguistik und Informationstechnologie für Jürgen Rolshoven aus Anlass seines sechsundsechzigsten Geburtstages. Köln: Universität zu Köln, S. 269-285.

Gillam, Richard (2003): Unicode Demystified. A Practical Programmer's Guide to the Encoding Standard. Boston etc.: Addison-Wesley.

Hyman, Malcom D./Scharf, Peter M. (2011): Linguistic Issues In Encoding Sanskrit. Delhi etc.: The Sanskrit Library.

Kappenberg, Bernd (2007): Zeichen setzen für Europa. Der Gebrauch europäischer lateinischer Sonderzeichen in der deutschen Öffentlichkeit. Online unter: https://www.mediensprache.net/networx/networx-49.pdf <15.10.2019>. In: Networx 49. ISSN: 1619-1021. Projekt Sprache@web.

Kappenberg, Bernd (2012): Zeichen setzen für Europa. Der Gebrauch europäischer lateinischer Sonderzeichen in der deutschen Öffentlichkeit. Stuttgart: ibidem.

Korpela, Jukka K. (2006): Unicode explained. Beijing etc.: O'Reilly.

Kubota, Tomohiro (2002): Han Unification. Online unter: http://www8.plala.or.jp/tkubota1/unicode-symbols-unihan.html <15.10.2019>.

Lennon, Brian (2010): In Babel's Shadow. Multilingual Literatures, Monolingual States. Minneapolis/London: University of Minnesota.

Lovink, Geert (2007): Maschinen sprechen Deutsch. Kritische Medienpraxis zwischen Amsterdam, Sydney und Berlin. In: Woznicki, Krystian (Hrsg.): McDeutsch. Protokolle zur Globalisierung der deutschen Sprache. Berlin: Kulturverlag Kadmos Berlin, S. 58-60.

Ludwig, Otto (1994): Geschichte des Schreibens. In: Günther, Hartmut/Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler For-

schung. 1. Halbband (Handbücher zur Sprach- und Kommunikationswissenschaft [HSK] 10.1). Berlin/New York: De Gruyter, S. 48-65.

Maher, Jennifer Helene (2014): Decoding What is Good in Code. Toward a Metaphysical Ethics of Unicode. In: Limbu, Marohang/Verhulsdonck, Gustav (Hrsg.): Digital Rhetoric and Global Literacies. Communication Modes and Digital Practices in the Networked World. United States of America: IGI Global, S. 80-96.

Margulies, Simon B. (2009): Digitale Daten als Quelle der Geschichtswissenschaft. Eine Einführung (Kölner Beiträge zu einer geisteswissenschaftlichen Fachinformatik 2). Hamburg: Dr. Kovač.

Moran, Steven/Cysouw, Michael (2018): The Unicode Cookbook for Linguists. Managing writing systems using orthography profiles (Translation and Multilingual Natural Language Processing 10). Berlin: Language Science Press.

Shafer, Maggie (2015): Unicode: A story of corruption, connection and smiling poo. Online unter: https://medium.com/@maggieshafer/unicode-a-story-of-corruption-connection-and-smiling-poo-598295e4af9d <18.09.2019>.

Topping, Suzanne (2001): The Secret Life of Unicode: A Peek at Unicode's Soft Underbelly. Online unter:
http://www.btetrud.com/Lima/The%20Secret%20Life%20of%20Unicode.pdf
<15.10.2019>.
Abstract online unter: http://unicode.org/iuc/iuc20/a306.html <15.10.2019>.

Weingarten, Rüdiger (2002): Der Computer als Schriftenmuseum. Latinisierung von Schriften durch computertechnische Zwänge?. In: Ehlich, Konrad/Greber, Erika/Müller, Jan-Dirk (Hrsg.): Materialität und Medialität von Schrift. Schrift und Bild in Bewegung. Bielefeld: Aisthesis, S. 165–182.

Abbildungsverzeichnis

Abbildung 1: CJK-Schriftzeichen U+76F4. In: Kubota, Tomohiro (2002): Han Unification. Online unter: http://www8.plala.or.jp/tkubota1/U76F4.png <15.10.2019>.

Abbildung 2: Twitter-Account von Mark Davis. Online unter: https://twitter.com/mark_e_davis <15.10.2019>.

Abbildung 3: Screenshot eines Beitrags des Schweizer Fernsehens zu Lindsey Vonn. Online unter:
https://www.netscout.ch/20110311/schweizer-fernsehen-neue-rechtschreibung-fremdsprachige-familiennamen <15.10.2019>.