

Transkriptionssysteme und UNICODE | Transcription Systems and UNICODE

Besondere Zeichen in einer standardisierten Welt | Special Characters in a Standardized World

Kurzfassungen | Abstracts

Ulrike Kramer
AT | Wien

Teuthonista I: Teuthonista: Geschichte, WBÖ und DBÖ |
Teuthonista I: Teuthonista: History, WBÖ and DBÖ

The transcription system Teuthonista, which is widely used for High German dialects, can look back to a tradition of more than 100 years in German dialectology. Its name is derived from the dialect journal "Teuthonista" where it was presented by Teuchert in 1924/25. The original script, however, had already been developed by Lenz in 1900, and in the same year it had also been introduced in the journal "Zeitschrift für hochdeutsche Mundarten". Both Lenz and Teuchert came up with a relatively small, basic set of characters, the emphasis being placed on the differentiation of vowel qualities which is regarded to be vitally important in High German dialectology.

A number of doctoral theses, dialect dictionaries and linguistic atlases have taken up Teuthonista and adapted it for their own system of transcription, among them projects as renowned as the Bavarian linguistic atlases, the Atlas of historical German dialects in the Czech Republic (ADT), the Bavarian Dictionary (BWB) and also the Dictionary of Bavarian Dialects in Austria (WBÖ) and its database (DBÖ).

Even though the name itself is never officially mentioned in the WBÖ, the script that both the dictionary and the database are based on can be easily identified as Teuthonista.

Throughout the years the Teuthonista script has been further developed by different kinds of projects, as a consequence more special characters have come into existence. Like many other projects the WBÖ and the DBÖ thus use a few special characters which have not been encoded in Unicode, yet. These characters are needed to ensure that the most accurate transcriptions of Austrian Bavarian dialects are rendered and documentation will show that they actually do turn up in several sources and records.

**Rosemarie
Spannbauer-Pollmann**
DE | Passau

Teuthonista II: Teuthonista und Sprachatlanten |
Teuthonista II: Teuthonista and Linguistic Atlases

The transcription system Teuthonista has been widely used for the documentation of dialects in the German-speaking countries over the last 45 years. Thus, a huge data base has been created in the different types of archives (paper, CDs and other disks) and part of it has also found its way into printed publications like language atlases. But the biggest amount of the precious data is still waiting to be evaluated and analysed and to be used by researchers and laypersons. For this purpose, Teuthonista characters have to be implemented in the font-system Unicode in order to be available for every user and on every platform (like Unix, Macintosh, Apple...). It will be possible to access, analyse and ultimately publish the data in different forms, if not - they will be lost because they will have lost their accessibility.

Alois Dicklberger
DE | Regensburg

Teuthonista III: Teuthonista und ADT |
Teuthonista III: Teuthonista and ADT

The tradition of the phonetic transcription system Teuthonista goes back to the beginning of the dialectology of the Romania. Through the Swiss project of the „Sprach- und Sachatlas Italiens und der Südschweiz“ this system of speech notation was brought to the „Schweizerdeutschen Sprachatlas“ and the “Atlas Linguistique de l’Alsace”. The present projects of dialect atlases at a small scale in Austria, Italy, Czech Republic and Germany use mostly a more or less modified version of Teuthonista. Until now these atlases were published in a printed

form and the individual projects could find their own solution to represent the sonic signs, it was the job of the printing shop.

With the development of the PC systems, the importance of the individual processing of the phonetic transcription signs come to the fore. Meanwhile there exist TEX based Teuthonista word processing systems and Teuthonista fonts which work on MS Windows platform. These different and individual systems hinder a cross-system and a platform independent exchange of linguistic information. A future presentation of the results of research in dialectology have to use the opportunities which is provided by the internet but this is impossible as long as the Teuthonista is not standardized in Unicode.

Karl Pentzlin
DE | Schongau

Teuthonista IV: Teuthonista und UNICODE |
Teuthonista IV: Towards a Unicode Encoding of Teuthonista

Teuthonista is a German dialect writing system employing numerous combinations of letters and diacritical marks stacked vertically or even side-/ downwards to its symbols. Although this is not uncommon for some South Asian scripts already encoded in Unicode, this is uncommon for the Latin script in its extent it is used for Teuthonista.

Thus, a proposal to encode Teuthonista must convince the Unicode Consortium that Teuthonista is in fact a set of characters (i.e. a valid addition to the Latin script), rather than being a notation system like musical notes.

To accomplish this, there an exhaustive list of the building blocks of Teuthonista must be supplied, together with the rules how these building blocks are used to combine to the complex Teuthonista symbols.

Also, this must be done in a way that these building blocks and rules are compatible to the mechanisms defined in the Unicode standards, as well as to the unwritten practices of encoding.

Some special Teuthonista symbols are presented with analyses how such symbols can be decomposed into building blocks which can be encoded.

A preliminary character set based on the "Einführung in den Sprachatlas der Deutschen Schweiz" (R. Hotzenköcherle, Bern 1962) is presented informally, which can be used as base for further work on the Unicode encoding of Teuthonista.

Peter Weiss
SL | Ljubljana

ZRCola: Entwicklung und Zukunft des unicodebasierten
Eingabesystems ZRCola | [The Transcription System ZRCola](#)

Das Eingabesystem ZRCola wurde von mir am Wissenschaftlichen Forschungszentrum der Slowenischen Akademie der Wissenschaften und Künste (ZRC SAZU) in Ljubljana in Laufe der letzten sieben Jahren entwickelt. Es ermöglicht die Eingabe äußerst komplexer Texte, ihren problemlosen Austausch zwischen verschiedenen Gruppen über Sprachgrenzen hinweg, den Druck sowie das Erstellen von Datenbanken. ZRCola hat sich beim internationalen linguistischen Großprojekt *Slawischer Sprachatlas* sehr bewährt, was sich auch darin spiegelt, daß der zweite Teil seines Namens das russische Projektkürzel enthält (*Obščeslavjanskij lingvističeskij atlas*).

Das System besteht aus vier Komponenten:

- (1) die Schrift 00 ZRCola.ttf, die seit kurzem auch in Italic vorliegt (die beiden Boldvarianten sind erst teilweise ausgearbeitet);
- (2) ein Programm für das Zusammenstellen bzw. Zerlegen der Sonderzeichen;
- (3) ein Programm für die Anpassung (Erweiterung) der Tastatur;
- (4) eine umfassende Dokumentation des Systems.

Wenn Sonderzeichen schon komplex sind, sollte man sie zerlegen können, um sie leichter in den Griff zu bekommen. Das Zeichen \tilde{o} beispielsweise kann auf diese Weise in ein Grundzeichen *o* und die Diakritika $\tilde{\cdot}$ zerlegt werden. Diese Bausteine sind überschaubar und leicht zu beherrschen. Das Programm berücksichtigt darüber hinaus auch in den einzelnen Sprachen gängige zusammengesetzte Zeichen, z. B. \ddot{o} (+ $\tilde{\cdot}$). Das System kann laufend erweitert werden. Unsere Sammlung neuer Zeichen wird damit immer reicher. Vollkommenheit wird angestrebt, obschon diese letztlich nie erreicht wird.

Vorteile für die Benutzer: Das System funktioniert im Programm Word, das in Europa sehr verbreitet ist. Neue Zeichen können laufend vorgeschlagen werden. Schrift und System sollten zentral verwaltet werden – dies bedeutet unter Umständen einen gewissen Zwang, was aber andererseits seine Vorteile hat.

Was die Zukunft betrifft: Im privaten Unicode-Bereich gibt es nur noch wenig Platz (Ende Oktober 2008: 3315 von 6400 Positionen vergeben), worunter die Überschaubarkeit leidet. Heutige gängige Programme sind noch

nicht auf diese neue Situation vorbereitet: weder MS Office, das natürlich Kompatibilität mit älteren Programmen ermöglichen muß, weder DTP-Programme und Programme zum Erstellen von Fonts (Font Creator z. B. unterstützt nur die erste Ebene der Unicode mit etwa 65.000 Zeichen, FontLab nur ein paar Tausend Zeichen, die Schrift 00 ZRCola umfaßt zur Zeit jedoch bereits 6370 Zeichen). Alle diese Programme müssen aufeinander abgestimmt sein, um effektiv zu sein. Bis dahin aber sollen alle Möglichkeiten der PUA auf der Basisebene ausgenutzt werden, in der Sprachwissenschaft weit verbreitete respektive bei einigen linguistischen Projekten (z. B. WBÖ) verwendete Zeichen im Unicode ihren standardisierten Platz finden und die Fonts auf neue Zeiten vorbereitet werden. Bis dahin ordiniere ich weiter im PUA-Wartezimmer.

Gábor Hosszú
HU | Budapest

Standardisierung des Szeklerischen-Ungarischen Rovas |
Standardization of Szekler-Hungarian Rovas Script

The presentation introduced the *Szekler-Hungarian Rovas* (in German Szeklerisch-ungarische Runenschrift) script an interesting part of the Hungarian cultural heritage. It is a pronunciation and character based script with extensive use of the ligatures and ideograms.

The right-to-left direction is the dominant; however left-to-right is also widely used. The boustrophedon and the top-to-bottom directions, as well as the organic motif-building usage in the fine arts are also possible. However, it is not frequently applied. The general rule is that any of these directionalities are acceptable. According to the character facing, the main principle is that the faces of the characters are opposite (mirrored images) in the two kinds of horizontal writing directions.

The origin of the Szekler-Hungarian Rovas has not clarified scientifically, but its existence and usage is already firmly proved from as early as the 9th century. However, there are numerous similar but unproved evidences from earlier ages, too. Since than up to now, its usage is continuous without any break and it never became an extinct script. From the 19th century it became more and more popular. Nowadays the use of the Szekler-Hungarian Rovas is very extensive, e.g. there are Rovas script in the public places on street-plates that means that many people knows about the Rovas writing and a lot of them can read it. Another example of the increasing use of this script is the many books written in Szekler-Hungarian Rovas script. There are more and more villages and towns that write their name on the town-border-plates with Latin and Szekler-Hungarian Rovas scripts, too.

The unification and then the standardization process of the Szekler-Hungarian Rovas script have a long history. It based on scientific research works and agreements of the different Rovas schools. The latest step in standardization of the Szekler-Hungarian Rovas script was the "Alive" Rovas Symposium (4th October, 2008), where 59 participants (experts, teachers, researchers) from different countries discussed the proposal for encoding the Szekler-Hungarian Rovas in the BMP and the SMP of the Universal Character Set (UCS). Based on voting procedure the Symposium accepted it and then it was submitted:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3527.pdf>

The presentation also shows the character repertoire of the standard proposal with the proposed Unicode code points. After this many old relics and contemporary documents are presented including mural inscription from 1497, the first Rovas teaching book from 1598, alphabets from various ages, maps of languages created by the German philologist Gottfried Hensel, Crossword puzzles from 1991, phone card with Rovas inscription released in 2002, many books, posters, journals in Rovas, etc.

Christoph Praxmarer
AT | Innsbruck

Digitalisierung des EDD: Wright's Transkriptionssystem |
Digitizing the EDD: Wright's Phonetic Script

The *English Dialect Dictionary (EDD)* was published between 1898 and 1904. "The dictionary includes, so far as is possible, the complete vocabulary of all English dialect words which are still in use or are known to have been in use at any time during the last two hundred years in England, Ireland, Scotland and Wales" (Wright, v). It was initiated by Walter Skeat, who founded the English Dialect Society (EDS) in 1873 for the purpose of collecting dialect information. Wright's sources for the *EDD* include the publications of the EDS, overall 80 volumes of dialect glossaries, additional glossaries, literary sources and a network of informants. The *EDD* was a huge project at its time and today it is still an important source for historical variety studies.

Our small team in Innsbruck has been working on an online version of the *EDD (EDD-Online)* for over two years and a beta version has been available since August. The aim of the project is to provide a sophisticated search interface for linguists that allows for complex searches. It is possible to search within different parts of the

dictionary entries (ie. citations) and to improve search results with filters such as, for instance, dialect areas (ie. Yorkshire), part of speech (ie. adverb) or usage label (ie. colloquial).

So far, the implementation of the phonetic transcriptions in *EDD-Online* have been postponed because of certain problems involved. The OCR software we used was not able to read the phonetic script. Considering that for an estimated third of all of the over 60.000 headwords of the dictionary a phonetic transcription is provided, a lot of manual correction work is required. Practical considerations and technical restrictions concerning our web interface and database have lead to the decision that we will use a system based on a virtual keyboard for the input of phonetic symbols and images for the output.

References

EDD-Online. Retrieved 5 November 2008. <<http://eddonline.uibk.ac.at>>.

Wright, Joseph. *The English Dialect Dictionary*. 6 vols. Oxford: Henry Frowde, 1898-1905.
