Multilingual Ordering — the European Ordering Rules

Marc Wilhelm KÜSTER

English terminology usually distinguishes between *sorting* and *ordering*. *Sorting* is primarily a service for users to facilitate their access to information by presenting it in a structured and predictable way, e. g. by subdividing the information by subject matter (by having several registers to a book, for instance), having multiple indices in a library etc. *Ordering* — the arrangement of information in alphabetical sequence — is in most cases an integral part of this procedure. It is only ordering that I am going to deal with in this article.

The ABC is one of the things we normally get to know around six or seven and take for granted ever after. Could there be anything more unspectacular than this perhaps not very logical but certainly well-established sequence of letters between *a* and *z*? Why is it worth talking about something that any schoolchild knows by heart?

Depending on what schoolchild, of course. А и б сидели на трубе, the Russian child sings as well, but it would have to continue quite differently. The German pupil learns early on that *ä*, while just as valid a sound as the other two, is somewhat different from *a* and *b* in that it does not file as a letter of its own in a dictionary or telephone book — an experience he or she shares with his or her French counterpart when it comes to accented letters. The Danish student, however, will quite happily search for *ä* after *z*, where no German would expect it, but where a young Tartar or Uzbek would also look first, were he to use a Latin-based dictionary.

In short: Ordering of letters is highly dependent on the cultural expectations. The IT industry has become aware of this and strives to "localize"[1] their software products in this respect. They have investigated

[1] The term *localization*, the adaptation of software to a given set of cultural habits, encompasses many tasks, ranging from fairly straightforward undertakings such as the correct presentation of dates (e. g. 1.12.1999 vs. 12.1.1999, both meaning the very same day), numerals up to complex tasks involving intelligent pattern matching, spell checking, and, last but not least, ordering. Its counterpart, *internationalization*, is in many ways

into usual techniques of ordering as were traditionally used in national standards, libraries, publishing houses etc. In doing so they have also formalized the whole process by inventing metalanguages to deal with this problem.

If we relax the definition somewhat and for a moment take ordering as the arrangement of words in a predefined canonical sequence, then it is as old as writing itself as cuneiform texts from around 3000 BC show:

> [...] der Fund einer Listentafel unter den ältesten Texten [erwies] sich als äusserst bedeutsam, da dies bedeutete, dass die Gattung als solche bereits in der allerersten Phase der Schrift vorhanden war [...][2]

This find is not exceptional. About 650 such samples from the archaic period of Uruk have so far been excavated[3], 525 of these belonging to 14 identifiable *lexical lists*. These lists are (cum grano salis) arranged by topic, the best attested one being the well-known "profession list" *Lú* A.[4] Other lists feature birds, plants, cheese, and the like.

This is hardly surprising, given that writing is an offshot of accounting. The great majority of surviving archaic texts are records of various merchandises and foodstuff and other administrative documents.[5] Future accountants had to learn the correct "orthography" and hierarchy of the words they would have to deal with in their future work. It is therefore no overstatement that ordering was originally a didactic device. It was at the same time more than that:

> Das Listenwerk, für das es in anderen Kulturen keine Parallelen gibt, ist das einzige Ausdrucksmittel der sumerischen Wissenschaft geblieben, die man formal als eine Listenwissenschaft und inhaltlich als Ordnungswissenschaft charakterisieren kann.

---

synonymous, though it focuses more on strategies of how to avoid any undesirable hard-coded cultural expectations in the software itself.

[2] NISSEN (1981), 101. Cf. also CAVIGNEAUX (1983), 612: "L'histoire des listes lexicales commence avec Uruk IV, c'est-à-dire avec la plus vieille écriture connue."

[3] Cf. ENGLUND (1993), 12f.

[4] This list is attested with ca. 130 copies in the archaic period of Uruk alone (cf. NISSEN (1991), 148).

[5] Cf. NISSEN (1991), 57: "Wenn man die Entstehung der Schrift als einen weiteren, allerdings den bedeutsamsten Schritt in der Reihe der Versuche, Kontrollhilfen für die Wirtschaft zu finden, ansieht, dann wird dies dadurch unterstrichen, daß weitaus der größte Teil der inzwischen fast 5.000 Nummern umfassenden Schriftdokumente der Schriftstufen IV und III Verwaltungsvorgänge festhält."

> Außerhalb der sie ergänzenden Mythendichtungen kannte sie keine feststellenden Sätze und versuchte auch nicht, Erkenntnisse logisch miteinander zu verknüpfen. Damit unterscheidet sie sich von der griechischen und abendländischen Wissenschaft grundsätzlich und kann auch nicht als eine ihrer Vorstufen gelten.[6]

What makes the lists quite genuine precursors to modern alphabetical ordering practices is their longevity and conventional but stable sequence. They were transmitted over centuries and even millenia in all areas where cuneiform script was used, even though language, social structure and habits were often entirely different from those in Sumer around 3000 BC. A prime sample of their transfer to a very different environment is their use in Ebla where a Semitic language was spoken, but pupils still had to copy the Sumerian lexical lists, especially the profession list.[7] They were part of the curriculum for the advanced students and in part signed. Learning and reproducing the Sumerian lexical lists was seen as an essential part of language acquisition.

However, the Eblaites did not leave it at that. Lexical lists of the kind mentioned above are valuable learning aids only for native speakers or advanced learners. Beginners need dictionaries which are arranged by different criteria. Thus, the Eblaite scribes developed[8] a dictionary of considerable length[9] which orders entries by graphemes quite as today. Where necessary, they even used the second grapheme to establish a unique sequence of entries which begin with the same character. Here a sample of entries which start with the character NÌ, the first character of the Eblaite ordering sequence. NÌ was used to form abstracts.[10,11]

---

[6] SODEN (1961), 565.

[7] For the definitive edition of these tablets cf. PETTINATO (1981). For a popular, though slightly dated introduction into the history of Ebla and its discovery cf. PETTINATO (1979).

[8] Rather, the earliest samples of this kind have been excavated in Ebla. It is, of course, perfectly possible that the technology was developed at a different, still undiscovered location.

[9] The reconstructed lexical list has 1457 entries. The definite edition can be found under PETTINATO (1982), 118–130.

[10] Cf. BORGER (1978), 205.

[11] The dictionary also gave Eblaite equivalents to somewhat unusual Sumerian words − not unlike the practice of monolingual dictionaries as the *Rechtschreibduden* which also explains “unusual” words and gives only the orthography of the rest.

nì-sig
nì-sig-sig
nì-⌜gur$_5$⌝
nì-kar
nì-kar-⌜kar⌝
nì-za$_x$
nì-sal
nì-sal-sal [...][12]

The author of this paper agrees with PETTINATO (1979) that

> Il vocabolario è ordinato secondo il principio dei dizionari monolingui e ne presenta gli stessi caratteri di organicità e sistematicità che potremmo definire moderni.[13]

Nonchalantly neglecting all trains of cultural transmission, leaving aside the evolution of alphabetic systems, let us now jump by two and a half millenia to another invention that marks today's ordering practices: the advent of diacritics. Accented letters[14] entered the Greek language around 200 BC and have been a regular feature of Greek orthography since the ninth century AD, though first uses in inscriptions can be dated to 80 AD.[15] Originally intended as reading aids in poetry, they do not constitute new letters but rather specify the details of pronunciation and other distinctive features of existing ones. For this reason, they are often treated not as letters of their own but along with their the corresponding basic letters as can be seen in the encyclopedia of *Suda*.[16]

This tradition was handed over to France and is used in such early examples as Estienne's *Dictionaire Francoiſlatin* (1534) and then Nicot's *Thresor de la langue françoyse* (1606).[17] While examples are not too numerous, both Estienne and Nicot ordered accented letters as if they were normal letters, but used them for distinction of homographs. In this case, the words with accented letters follow those without.[18] This was later

[12] PETTINATO (1982), 119.

[13] PETTINATO (1979), 256.

[14] For the purpose of this article, I shall treat the terms *diacritics* and *accents* as synonymous though it should be understood that this is not permissible for many scripts.

[15] Vergl. JENSEN (1970), 470 and the time-honoured LISKOVIUS (1825), 238f. Diacritics proper were also used in the Brāhmī script, amongst others.

[16] For a good example cf. Suda (1967–1972), 1159f.

[17] Cf. ESTIENNE (1534) and NICOT (1606). A good study into the beginnings of French lexicography is to be found under WOOLDRIDGE (1977).

[18] Samples include in ESTIENNE (1534) words like Coſte vs. Coſté (139), in NICOT (1606) Felle vs. Fellé (282) and again Coſte vs. Coſté (152).

systematized in the *multilevel ordering approach*, i. e. by the introduction of several ordering levels, which is the corner stone of modern ordering.

The general idea is to ignore diacritics, non-letters and casing for comparison at level 1 and to mirror letters with diacritics to the corresponding basic letters. E. g. *é*, *ë* and *e* are equivalent on level 1 as are *a* and *A*. Similarly, "special letters" such as *ð* or *ß* are treated as what is seen as their corresponding base form.

It must, however, be stressed that the very concept of "basic letters" is language-dependent. In Danish, *å* is clearly as "basic" a letter as *a;* and *a̋* decomposes to ´ and *å*[19] rather than ´° and *a*.[20] *þ* (Thorn) is a perfectly legitimate basic letter in Icelandic where it is the last letter of the alphabet. In German catalogues, however, it is treated as equivalent to *th* on level 1. Even if the repertoire of basic letters is identical, it can still be arranged differently.

Only if words are identical on level 1, they are distinguished on level 2. E. g., *(il) aime* and *(il a) aimé* are perfectly identical on level 1, but distinct on level 2. Another example from German is the often quoted word pair *Masse* vs. *Maße*. *ß* is treated as equivalent to *ss* on level 1, though ordered after sharp s on level 2. Countless other examples could be added.

The relative order of diacritics is again language-dependent. There are no universal rules whether the acute precedes the gravis or vice versa.[21]

Level 3 distinguishes words that are identical on level 2 by taking (amongst others) casing into account. In Germany, lowercase precedes uppercase so that *(das) Maß* follows *(er) maß*. Reverse conventions are used, e. g., in Denmark and in some English dictionaries.

Level 4 finally looks at non-letter characters that may be present in the word (e. g. *its* vs. *it's*). There are few established sequences here. Some dictionaries even prefer to deal with this kind of characters on level 1.

---

[19] Historically, the "ring" is not a diacritic at all, but an *o* jumped upwards. In 1948 it replaced the old spelling *aa*. Politiken (1999) unambiguously calls *å* "det 29. bogstav i alfabet" (1591).

[20] Therefore, the term *basic letter* is often avoided and *first level letter* used instead. It is more exact and often preferable, though less intuitive.

[21] An unpublished study by Bunz and Küster (document No. CEN/TC304 N773) has looked in detail how different Greek dictionaries deal with diacritics. It can be made available upon request.

It is this behaviour that is mirrored in the “tailorable”, i. e. adaptable, ordering mechanisms which allow the exact specification of the ordering sequence by the user through a metalanguage.

To define a common denominator, Canada proposed an international meta-ordering standard which is currently under development in ISO, the international organization for standization.[22] Its metalanguage is POSIX-inspired, though any other syntax that can be unequivocally mapped to its metalanguage is equally permissible. I have written a filter that does this mapping to the TUSTEP ordering language which is one of the earliest implementations of the multilevel ordering mechanism for the academic environment.

The suite of `#sort`-modules in TUSTEP were developed back in the early 70s — and are incidently still conformant to the forthcoming standard. I shall use them to explain how the concept of tailorable multilevel ordering can be transposed to the computer age. At the same time, the code also demonstrates TUSTEP’s modular approach: instead of invoking a monolithic ordering program, TUSTEP features three different modules which can be combined with other commands for more complicated tasks, e. g. the handling of complex databases or varied registers.

For demonstration purposes, the mechanism is tailored to match a slightly simplified version of the *norme française sur le classement alphabétique des dénominations* (NF Z 44–001: Novembre 1995)[23]:

```
#create,target
#presort,source,-std-,mode=-,erase=+,parameter=*
ssl       50 50 50
xs1       ~%>@~~
xs1       ~%>@>@~~

xs2       ~%/~1z~
xs2       ~%\~2z~
xs2       ~%>>~3z~
xs2       ~%:~4z~
xs2       ~%>@~>=01z~
```

[22] The draft standard 14651 is currently in the FCD-stage (FCD: final comittee draft) and is handled in ISO IEC/JTC1/SC22/WG20.

[23] The simplification consists mainly in having curtailed the full list of diacritics and ligatures to increase clarity.

```
rlf       0 1 0                                                    12

xs3       ~>*~>=011~                                               14
xs3       ~<*~>=012~
                                                                   16
a1      :  abcdefghijklmnopqrstuvwxyz
*eof                                                               18

#sort,-std-,-std-,1+150,+,1+150                                    20

#copy,-std-,target,erase=+,mode=+                                  22
```

This looks more complicated than it is. It is unnecessary to discuss details here, so I would just like to draw attention to a few of the principal points. `%>@` (4) signifies “any diacritical mark” and `%/` resp. `%\` (7–8) are mnemonic expressions for acute and gravis respectively, `%>>` for the circumflex and `%:` for the trema.

The procedure is simple enough: In the `xs1`-lines (4f) the `presort`-program is instructed to ignore diacritical marks on the first ordering level. For the second level (`xs2`, 7–11) accents are sorted in the given order and evaluated in backward sequence (12).[24] Further diacritical marks could be added as needed. On a third level we distinguish between small and capital letters, ordering the latter after the former.[25] `a1` (17) lists the collating sequence which consists in this case of the ordinary Latin alphabet which it would also have defaulted to.[26] Letters and further symbols could be ordered in any desired sequence, catering e. g. for the placement of *ä* after *z* in Swedish or *ll* between *l* and *m* in Spanish.

TUSTEP’s object-oriented approach even allows to embed this kind of ordering sequences in the document itself.

---

[24] This speciality of French ordering reflects French phonetic structure: the last diacritic (e. g. *é* in *côté*) carries greatest weight in pronunciation. To use the standard example: a source file containing the words *cote, coté, côté, côte, Cote, Coté, Côté, Côte* is transformed correctly to a target file with *cote, Cote, côte, Côte, coté, Coté, côté, Côté.*

[25] For French both small before capital and capital before small are permissible.

[26] Different alphabets could be given for any of the ordering keys.

For monolingual texts[27] it is quite easy to achieve the expected ordering sequence by "tailoring" an artificial default assumption to the task at hand. What, however, should we do in the case of *multilingual* corpora with potentially conflicting cultural expectations?

Let us, for simplicity's sake, focus on a single example: building registers to a book the contents of which covers a number of languages. Let us furthermore only concentrate on a subject index.

An obvious choice would be to do one register per language. Each register could then be "culturally correct". This approach has, however, several drawbacks:

- To extract words automatically into the correct register, all entries would have to be tagged in a higher-level protocol, for example some SGML-derivative (e. g. using the `lang`-attribute in TEI). ISO IEC 10646 / Unicode only knows implicit "script-tagging", not language tagging. To add it requires considerable extra work.
- The potential user would be expected to know all relevant cultural expectations to locate a term in the respective registers. This is unlikely to be the case.
- A user would have to glance through all registers to be sure he has covered all occurrencies of a concept.

These drawbacks make it in many cases desirable to collate all register entries into a single one, employing an artificial but universally understandable and acceptable default sequence.

To find such a compromise sequence for pan-European purposes the European Commission decided to fund a project team under the auspices of the Technical Committee 304 "European Localization Requirements" of the European standards organization CEN, the "editor" (i. e. principal author of the resulting European standard) of which is the author of this paper.[28] We were tasked with dealing with a given UCS character repertoire that covers apart from Latin the Greek and the Cyrillic scripts. At a later stage Armenian and Georgian are to be added to cover the main

---

[27] Databases with words and expressions from more than one language are here taken as a special case of monolingual texts if at any given time only words from one language are presented to the end user.

[28] For more information see also `http://www.stri.is/TC304/EOR`.

European scripts still in use today. The standard is intended to facilitate record keeping in commercial and administrative environments − quite the same motivation that stands behind the original invention of ordering. Let's see if it likewise allows use in more varied environments.

Instead of many details here the main design principles:
- − access to information must be facilitated as much as possible;
- − the user is not assumed to be versed in the technicalities of ordering;
- − the rules should be derived from standardized rules and common practice in a large number of European languages without giving preference to the rules of any language or languages in particular;
- − the result should be as consistent as possible.

Some of the principles proved to be not 100% compatible, but we tried our best to find a reasonable compromise.

For the Latin script only the letters *a* to *z* and *þ* are considered to be basic letters.[29] Letters with diacritical marks are unified with their basic letters on level 1, letters which could for our purpose be regarded as modified letters are unified with their basic form, e. g. the letters *đ* and *ð* with *d*. An order of diacritics that satisfies most needs is now established.

For Greek, the situation is in so far simpler as there is only one dominant language that uses that script and therefore less need for compromises. There was still a lot of discussion on some points such as the best sequence of diacritics for Greek in conjunction with that for Latin or the treatment of the final sigma. Further minor peculiarities are the presence of the numeric characters and of the digamma in the list of basic letters, but they were considered essential by Greece.[30] The resulting sequence is α β γ δ ε ϝ ς ζ η θ ι κ λ μ ν ξ ο π ϙ ρ σ τ υ φ χ ψ ω ϡ.

The most problematic case proved to be that of the Cyrillic letters. Cyrillic is used to write a great number of languages of very different origin. The by far biggest of these is, of course, Russian, the sequence of which should be respected in any case. Generally speaking, the order of

---

[29] They are called "first level letters" in the text of the draft standard.

[30] These characters are, of course, archaic letters that have ceased to be used as such as early as in the pre-Classical period. Nowadays, the numerals ς, ϙ and ϡ appear mainly in legal documents. The digamma ϝ is not used in any living language.

the Slavonic languages is, with few minor exceptions, compatible, but that of the diverse Turkic, Mongolic, and Uralic languages is not. In addition, some of the Central Asian follow-up states such as Uzbekistan are currently in process of switching to Latin-based alphabets. Reliable and reasonably up-to-date information is not always easy to get hold of. It took some major effort and, after some time, cooperation with the Russian standards organization ГОСТ[31] to establish a sensible sequence.

The aim was to minimize the need for adaptation for the biggest languages which use Cyrillic letters and continue to do so.

Unfortunately, there is little current population data available for the area of the former Soviet Union, the most recent census accessible being the one of 1979.[32] It lists some 137 million Russians, 42 million Ukrainians, 12.5 million Uzbeks, 9.5 million Belorussians, 6.5 million Kazakhs, 6.3 million Tartars, 5.5 million Azerbaidzhanis, 2.8 million Armenians, 2.8 million Georgians and 2.2 million Moldavians, to give only the ten largest ethnic groups. Of these, Armenians and Georgians use, of course, their own scripts, and Uzbeks, Azerbaidzhanis and Moldavians will write in variants of the Latin script in the future.[33]

However, even this seemingly trivial task is full of pitfalls as we can see when we have a closer look at the alphabets in question:[34]

Russian: а б в г д е (ё) ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ы ь э ю я[35]

Ukrainian: а б в г ґ д е є ж з и і ї й к л м н о п р с т у ф х ц ч ш щ ь ю я[36]

Belorussian: а б в г д (дж) (дз) е ё ж з і й к л м н о п р с т у ў ф х ц ч ш ы ь э ю я[37]

---

[31] My specific thanks go to Mr. Yakupov and Mr. Bolotov from ParaType. Mr. Yakupov is co-chair of the Russian character coding committee.

[32] Cf. KOZLOV (1988), 94ff.

[33] Moldavia already used the Latin script in 1991 (cf. Akademia de Ştiinţe 1991), whereas the 1978 dictionary still employed Cyrillic letters (cf. Akademija de Štiince 1978). Uzbekistan and Azerbaidzhan prescribe the Latin script from 2000 onwards.

[34] I would like to give my heartfelt thanks to Prof. Mangold who gave me access to his unique library of dictionaries and grammars from which many of these samples are culled.

[35] OZHEGOV (1995), 12. The pre-Revolutionary Russian alphabet contained, of course, more letters.

[36] Novyj orfohrafičnyj slovnyk (1997), 2.

[37] LOBAN (1990), 6. For variants cf. MIXNEVICH (1992), 204 and GRABCHYKAU (1991), 8.

Kazakh: а ә б в г ғ д е ё ж з и й к қ л м н ң о ө п р с т у ұ ү ф х h ц ч ш щ ъ ы і ь э ю я[38]

Tartar: а б в г д ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ы ь э ю я ә ө ү җ ң h[39]

These are only five of well over fifty languages that must be considered. Most alphabets have their idiosyncrasies, and it is often the small languages such as Abkhazian (some 90.000 speakers) that sport the widest variety of non-Russian letters.[40] Nevertheless, there are two general rules:

- non-Russian letters are normally ordered either directly after similar Russian letters (as in Kazakh) or after я (as in Tartar);
- almost[41] all letters are treated as basic letters.[42] Often, digraphs and trigraphs are also treated as basic letters (e. g., in many Caucasian languages).

Obviously, a pan-European ordering standard can only choose one option or the other. We decided to follow the former Soviet practice to regard all Cyrillic letters save some special cases as basic letters. Following Kazakh tradition, we also incorporated the non-Russian letters next to their Russian counterparts, unless the tradition for characters which appear only in one language demanded a different position.

This proposal was then sent to ГОСТ for comment. A number of improvements and corrections were asked for[43] and implemented, but the guiding principles set out above were in line with Russian demands.

From a logical point of view the pan-Cyrillic ordering sequence leaves a lot to be desired. It is not consistent in its treatment of letters with

---

[38] DUKENBAEV (1989), 1.

[39] Cf. Sovremennyj tatarskij jazyk (1969), 112 and DAWLETSCHIN (1989), 7.

[40] For a most revealing study into the social and political background of these developments, cf. BALDAUF (1993).

[41] One of the few exceptions is the Russian letter ё which is normally subsumed under е but which is also rarely used outside children's books and dictionaries where it functions primarily as a pronunciation aid.

[42] This is an old feature of Slavic languages, stemming from Hus' "Orthographia Bohemica" which introduced a system of diacritics to write the Slavic languages — resulting in true letters, not only reading aids. Cf. also SCHRÖPFER (1968).

[43] Some of the changes resulted in phonetically identical or similar letters being put in closer vicinity.

diacritics nor in its ordering according to either phonetic or graphic criteria. However, standards are compromises, and user expectations rarely accord with purely logical considerations.

Only time will tell if the decisions are sensible for a commercial and administrative environment. Only continuing discourse can clarify to what degree the European ordering rules can be fruitful for academic projects on multilingual corpora.

Bibliography

Akademia de Ştiinţe (1991): Akademia de Ştiinţe a R.S.S. Moldova (ed.): Dicţionar ortografic cu elemente de ortoepie şi morfologie. Chişinău: Redacţia Principală a Enciclopediei Sovietice.

Akademija de Štiince (1978): Академия де Штиинце а рсс молдовенешть (ed.): Дикционар ортографик ал лимбий молдовенешть. Кишинэу: Редакция принчипалэ а енчиклопедией советиче молдовенешть, [2]1978.

BALDAUF (1993): Baldauf, Ingeborg: Schriftreform und Schriftwechsel bei den muslimischen Rußland- und Sowjettürken (1850−1937): Ein Symptom ideengeschichtlicher und kulturpolitischer Entwicklungen. Budapest: Akadémiai Kiadó.

BORGER (1978): Borger, Rykle: Assyrisch-babylonische Zeichenliste. Kevelaer / Neukirchen-Vluyn.

CAVIGNEAUX (1983): Cavigneaux, A.: Lexikalische Listen. In: *Reallexikon der Assyriologie*, Vol. 6, 609–641.

DAWLETSCHIN (1989): Dawletschin, T., Dawletschin, I., Tezcan, S.: Tatarisch-deutsches Wörterbuch. Wiesbaden: Otto Harrassowitz.

DUKENBAEV (1989): Дукенбаев, С. Х.: Казахско-русский школьный словарь. Для школ с русским языком обучения. Алматы: Мектеп.

ENGLUND (1993): Englund, Robert K.; Nissen, Hans J.: Die lexikalischen Listen der archaischen Texte aus Uruk. Berlin: Gebr. Mann Verlag (= Archaische Texte aus Uruk, Vol. 3).

ESTIENNE (1534): Estienne, Robert: Dictionaire Francoiſlatin avtrement dict: Les mots Francois, auec les manieres d'vſ diceulx, tourner en Latin. Paris: Imprimerie de Robert Estienne.

GRABCHYKAU (1991): Грабчыкаў, С. М.: Беларуска-рускі слоўнік. Мінск: Народная Асвета, [3]1991.

JENSEN (1970): Jensen, Hans: Sign, Symbol and Script. An account of man's efforts to write. London: Allen and Unwin, [3]1970.

KOZLOV (1988): Kozlov, Viktor: The peoples of the Soviet Union. Introduced by Michael Rywkin. Translated by Pauline M. Tiffen. London: Hutchinson / Indiana University Press.

LISKOVIUS (1825): Liskovius, Karl Fr. Sal.: Über die Aussprache des Griechischen und über die Bedeutung der Griechischen Accente. Leipzig: Johann Ambrosius Barth.

LOBAN (1990): Лобан, М. П.; Суднік, М. Р.: Арфаграфічны слоўнік. Мінск: Народная Асвета.

MIXNEVICH (1992): Міхневіч, А. Я., Навічэнка, Н. М.: Англа-беларускі размоўнік. Мінск: Навука і тэхніка.

NICOT (1606): Nicot, Jean: Thresor de la langue françoyse, tant ancienne que moderne. Paris: David Douceur.

NISSEN (1981): Nissen, Hans J.: Bemerkungen zur Listenliteratur Vorderasiens im 3. Jahrtausend (gesehen von den Archaischen Texten von Uruk). In: Cagni, Luigi: La lingua di Ebla. Napoli, 99–108.

NISSEN (1991): Nissen, Hans J.; Damerow, Peter; Englund, Robert K.: Frühe Schrift und Technik der Wirtschaftsverwaltung im alten Vorderen Orient. Informationsspeicherung und -verarbeitung vor 5000 Jahren. Bad Salzdetfurth: franzbecker.

Novyj orfohrafičnyj slovnik (1997): Новий орфографічний словник української мови. 50000 слів; граматика в таблицях. Київ: Видавництво Аконіт.

OZHEGOV (1995): Ожегов, С. И.; Шведова, Н. Ю: Толковый словарь русского языка. Москва: АЗЪ, ^{3}1995.

PETTINATO (1979): Pettinato, Giovanni: Ebla. Un impero inciso nell'argilla. Milano.

PETTINATO (1981): Pettinato, Giovanni: Testi lessicali monolingui della biblioteca L. 2769. Napoli (= Materiali epigrafici di Ebla, Vol. 3).

PETTINATO (1982): Pettinato, Giovanni (ed.): Testi lessicali bilingui della biblioteca L. 2769. Napoli (= Materiali epigrafici di Ebla, Vol. 4).

Politiken (1999): Politikens Nudansk Ordbog med etymologi. Aalborg: Politikens Forlag.

SCHRÖPFER (1968): Schröpfer, Johannes: Hussens Traktat 'Orthographica Bohemica'. Die Herkunft des diakritischen Systems in der Schreibung slavischer Sprachen und die älteste zusammenhängende Beschreibung slavischer Laute. Wiesbaden: Otto Harrassowitz.

SIDNEY ALLEN (1974): Sidney Allen, W.: Vox Graeca. A guide to the pronunciation of classical Greek. Cambridge: Cambridge University Press, ^{2}1974.

SODEN (1961): Soden, Wolfram von: Sumer, Babylon und Hethiter bis zur Mitte des zweiten Jahrtausends v. Chr. In: *Propyläen Weltgeschichte*, Vol. 1. Berlin: Ullstein Verlag.

Sovremennyj tatarskij jazyk (1969): Современный татарский литературный язык. Лексикология, фонетика, морфология. Москва: Издательство наука.

Suda (1967–1972): Adler, Ada (ed.): Suidae Lexicon. A-Ω. Index. Stuttgart: Teubner (Editio stereotypa editionis primae, 1928–1938) (= Lexicographi Graeci, Vol. I).

WOOLDRIDGE (1977): Wooldridge, Terence Russon: Les débuts de la lexicographie française. Estienne, Nicot et le Thresor de la langue françoyse 1606. Toronto et Buffalo: University of Toronto Press.