

## The Bulgarian Alphabet and Keyboard in the Context of EU Communications

The document addresses the following questions:

1. Origin and development of the Bulgarian alphabet (parallels with other Slavonic and Non-Slavonic alphabets)
2. Bulgarian Keyboard Layouts: history, recent developments, assessments.

### 1. The Cyrillic Alphabet: Historical Notes.

The Cyrillic alphabet as it is used today is a modernized version (based on the 18<sup>th</sup> C. ‘civil script’ introduced by Peter the Great) of the 9<sup>th</sup> century system of letters developed in the Bulgarian capital and literary centre Preslav, on the basis of St. Cyril’s Glagolitic alphabet.

During the 7<sup>th</sup> century, the Thracian-Latin territory of present-day Bulgaria, which had recently been invaded by Slavonic tribes of the “Southern” group, was conquered by the Ancient Bulgarians. Two hundred years later, their state had considerably enlarged its territory and had established itself as one of the three most powerful European monarchies. Until the 9<sup>th</sup> century, the history of this Bulgarian state was exclusively that of the Bulgarian element on the Balkans. The language of this population was the language of the state administration. They had their own system of writing, attested in some 200 inscriptions on stone, but were also familiar with the Greek alphabet. In 852, an ambitious ruler, Boris I, ascended to the Bulgarian throne. At the time of ascending to the throne, the territory of the state was considerable, recently enlarged with conquests in central Europe. While in the lands of Moesia and Thrace the Bulgarians or Thracians were probably not less numerous than the Slavs, these latter were the dominant population of the newly conquered lands. A religious freedom was tolerated, but the power and privileges remained in the hands of the Bulgarian Tangrist aristocracy. The Slavonic population did not identify with this state and was not motivated to preserve its integrity. Boris I had the ambition to unify the different peoples into a strong Christian state, which was to become the cultural centre of the Slavonic world. It is in view of this intention that, immediately after ascending to the throne, the monarch announced his intention to convert himself and the population of the state to Christianity.

Three years later the brothers Constantine and Methodius, originating from the (at the time) predominantly Slavonic area of Thessaloniki, left their positions as Byzantine functionaries and isolated themselves for seven years

in the monastery Polychrone. There, they created the Glagolitic alphabet and translated the sacred books into the South Slavonic dialect which was their mother tongue.

The first mission of the Thessaloniki brothers was, however, to Moravia, not to Bulgaria. It is there that, on the demand of the Moravian prince Rostislav, the first Slavonic literary centre was created. A year later, the population of the Bulgarian state was converted to Christianity. The stifling of a great revolt of the Ancient Bulgarian aristocracy, followed by the massacre of the leaders of the rebels and of large parts of their families marked the end, according to historians, to the hegemony of the Ancient Bulgarians in the government of the state, as well as to the use of Ancient Bulgarian as a language of the state administration.<sup>1</sup>

In 879 the Pope Adrian II proclaimed the Glagolitic writings canonical. The Slavonic dialect of the Bulgarian territory became the fourth language of the Christian service. However, the Moravian mission of the two brothers eventually failed. In 885 the German clergy in Moravia forbid the Slavonic alphabet and liturgy. The Slavonic writings were destroyed; the disciples of Constantine-Cyril and Methodius were banned from the state. Many of them were imprisoned or sold into slavery. Certain managed to escape to Bulgaria, and they found refuge there. Among them were Clement, Naum and Anghelaria. With the help of the Bulgarian rulers, they laid the foundations of the Bulgarian Medieval Literary School with its two centres – the capital Pliska (later moved to Veliki Preslav) and Ohrida. Ironically, it is thus that the non-Slavonic rulers of the First Bulgarian State disseminated developed the Slavonic literary tradition. Christian churches where the Slavonic language was used in the service were built over the whole territory of the state. In conformity with the 8<sup>th</sup> Ecclesiastical Council, schools were opened in every parish.

The Glagolitic alphabet of the first Slavonic texts follows the Ionian (Alexandrian) rules for alphabet creation. It reflects very accurately the phonological system of the South Slavonic dialect; it has a very specific graphical appearance and specific numeric values for the separate letters, differing from those of the Greek alphabet. While this alphabet was readily adopted by the Moravian scholars (and used for centuries later in Dalmatia, Croatia and Bosnia), in Bulgaria it did not take root. The century-long administrative tradition of using Bulgarian and Greek letters, with Greek numerical values, triggered the creation of a ‘compromise’ alphabet, which

---

<sup>1</sup> Ancient Bulgarian continued to be used, though, long after this year, witness of which are inscriptions in this language using the Cyrillic alphabet.

was called 'Cyrillic' in honour of St. Cyril (and, probably, with the purpose of making the shift more palatable). The new alphabet was developed by the literary circle of Preslav under the direction of the son of Boris, Simeon. It bears a close resemblance to the Coptic alphabet. The Cyrillic alphabet follows the Glagolitic one w.r.t. to the letter-phoneme correspondences. However, it introduces a simplified graphical image, closer to the familiar Greek, Coptic, Phoenician one, and restores the familiar values for those letters which bear a similarity to the Greek ones.

When, in the beginning of the 11<sup>th</sup> century, the Bulgarian state was subjugated for a period of 170 years by Byzantium, the Slavonic writings were transported to 'Kiev Russia', which became the next great centre of Slavonic letters. It is above all to Medieval Russia that we owe the preservation of a large number of Old Bulgarian manuscripts, of which hundreds of copies were made. As to the original texts which were created in the Russian centres, they demonstrate a considerable departure from the language of the Bulgarian ones. In fact, as far back as the 7<sup>th</sup> century, the Eastern, Southern and Western Slavonic dialects were well differentiated. Even within the South Slavonic group, 'Bulgarian', 'Serbian' and 'Slovenian' already formed separate dialectal groups. As far back as the Old Bulgarian period, as a result of the dominant non-Slavonic factors in the central literary circle of Preslav, the Proto-Slavonic *ь* [y] became *u* [i] – Cf. Russian *рыба* [ryba – 'fish'] / Bulgarian *риба* [riba]. The Proto-Slavonic 'ultrashort' vowels in so-called 'strong' positions changed differently in Bulgarian and in the rest of the Slavonic world. The texts produced in the Pliska-Preslav centre demonstrated a tendency to preserve the middle character of *Ѣ* and not to differentiate the two vowels orthographically. Elsewhere, including the 'Western' centre of Ohrida, they were 'clarified' to *О* (for *Ѣ*) and *Е* (for *Ѧ*).

The modern diversity in the Slavonic alphabets reflects, on the one hand, the differentiation in their phonological systems, which was already apparent in the early years of Slavonic literary tradition, and, on the other hand, the Cyrillic tradition of maintaining a healthy correspondence between phonological and graphical system for each separate language. Even in their diversity, though, Slavonic languages are much closer and mutually understandable to their speakers than languages of the Romance and Germanic groups.

## 2. Modern Slavonic Alphabets and Keyboards

Today the Cyrillic alphabet is used by over 240 million people, mainly in Europe and Asia, speaking, respectively: six Slavonic languages, namely: Bulgarian, Macedonian, Serbian, Russian, Belarusian, Ukrainian; one Persian language (Tadjik), three Turkic languages (Kazakh, Uzbek and Kyrgyz), one Altaic language (Mongolian). Some South Slavonic and all West Slavonic languages use the Latin alphabet, but with special symbols for the specific Slavonic phonemes. Romanian, even though it demonstrates a number of the structural features of Romance languages, has a typically Slavonic phonological system, nearly identical to Bulgarian, and presents them with four types of diacritics.

Romanian, Slovak, Slovenian, Croatian and Polish make use of a Latin QWERTZ type of keyboard, enriched with 2-3 letter keys. Czech alone makes use of a QWERTY keyboard layout, but with a whole additional line of letters numbers are positioned on the QWERTY 'numbers' row.

Serbian and Macedonian use a compromise Cyrillic QWERTZ keyboard with the Q and W substituted with the two ligatures, X used for 'dz', and 2<sup>nd</sup> and 3<sup>rd</sup> rows extended to the right to make room for specific letters.

The standard Russian keyboard bears no resemblance whatever to QWERTY or QWERTZ, although a 'phonetic' QWERTY-type version also exists. The standard Russian keyboard layout is used, more or less faithfully, by the ex-USSR republics, including those with predominantly non-Slavonic population.

The standard Bulgarian keyboard layout bears no resemblance to either the QWERTY or QWERTZ keyboards OR to the Russian one. It was created and adopted a hundred years ago, after carrying out special corpus analysis.

## 3. History of the Bulgarian Keyboard Layout

The first Bulgarian typewriters arrived in Bulgaria at the end of the year 1902, ordered and delivered on the initiative of private and public firms. The typewriters were immediately put to use in the Chamber of Commerce, in ministries, post office administration, even in the Holy Synod. However, all these typewriters had different layouts. Getting to use them efficiently was thus an individual occupation, with little possibilities for professional help. The first users to realize that the effective use of typing ought to be

preceded by professional qualification courses were the stenographers in the Bulgarian National Assembly. They also realized that mass education meant unification of the keyboard layouts and formulated the task of creating a keyboard layout allowing maximally fast typing in the Bulgarian language.

It is thus that in October 1907 Teodor Galabov, at the time head of the Stenographic Office at the National Assembly, carried out with the team of stenographers the first corpus-based study of the Bulgarian language (an investigation based on the analysis of ‘10 000 words, taken from different areas of human life’) in view of the creation of a unified Bulgarian keyboard layout. Two discussion sessions were held in the autumn and winter of the same year. On December 30<sup>th</sup> 1907, the ‘All-Bulgarian Character Net for Typewriters’ was adopted and immediately published (early in 1908). The booklet, carrying the same name, is the first documental source officially regularizing typing in Bulgarian, and also the first guide to 10-finger professional typing. For some reason, this achievement has never received the publicity that it deserves, in Bulgaria or abroad; it was never placed in the context of international typing experience. It should, however, be considered that these Bulgarian stenographers took up the task of creating an ‘ergonomic’ keyboard 30 years before Frank and Lillian Gilbert ever mentioned ‘time and movement’ or developed the notion of effective typing; decades before August Dvorak arrived at the first linguistically supported keyboard for the English language. The major principles, on which the creation of keyboards today rests and to which few other keyboard layouts really answer, were successfully adopted and applied by those Bulgarian pioneers in the field.

The following principles were adopted in the design of the keyboard layout: 1/ Minimum movement from the 2<sup>nd</sup> (basic) row. 2/ If movement is necessary, then preferably move to the third, not to the first row. 3/ Lowest frequency letters assigned to the 3<sup>rd</sup> row; 4/ Rhythmic writing was achieved by means of hand alternations (necessitating a corpus study of digraphs). This requirement was largely met by placing vowels on the left-hand side of the keyboard. 5/ Next: most people have a more active right hand. While ensuring relative balance for the two hands, a more active right one was still assumed. 6/ Not all fingers are equally agile. More active forefingers and middle fingers were assumed, with lesser toil for the fourth finger and the little finger. 7/ Finally, using the same finger to write strings of letters is very ineffective. Thus, even in the absence of hand alternation, finger alternation at least was ensured.

It can be seen from the above list how complex, expertise- and time-consuming a task the creation of effective keyboard actually is. In the

context of the mass use of QWERTY/QWERTZ-type keyboards, even in the Cyrillic world, one could ask oneself why such work was carried out a hundred years ago. Because, in 1907, QWERTY already had over 30 years of history. Clearly, a century ago Bulgarian stenographers were convinced that, however well adapted, the QWERTY layout could not lead to effective writing of Bulgarian texts. They were also probably well aware that this layout is not even adequate for the needs of English (the story of the creation of QWERTY is well known). Indeed, when typing in an English text with the QWERTY layout, only 32 per cent of the strokes fall on the 2<sup>nd</sup> (basic) row. Most strokes fall on the third one, and more than 16 % - on the undesirable 1<sup>st</sup> one! Not more than 100 English words can be written without leaving the 2<sup>nd</sup> row. The reason is that about all high-frequency words in English are placed on the 'undesirable' rows. The basic row includes two of the least frequent letters (J and K) and does not include any of the most frequent ones (E,T,O). Although over 40 % of the letters in an English text are vowels, only one of the five English vowels appears on the basic row (but: within the scope of the 'little' finger of the left hand!). Actually, QWERTY assumes a left hand, considerably more active than the right one and little fingers, as active as forefingers. Another shortcoming of QWERTY is the absence of satisfactory hand alternation. Over 3000 high-frequency words are typed in with the left hand alone, about 300 – with the right hand alone.

About any alternative keyboard would allow better performance than that! Thus, to make an obvious comparison, the Dvorak layout places 8 of the most frequent English letters (all vowels plus T, H, N) on the basic row. The least frequent consonants are on the first row (V, J, K, Q, X). It is thus that 70 per cent of the strokes fall on the basic row, 22 per cent – on the third one and only 8 per cent – on the first one. With the Dvorak layout, thousands of English words can be written without moving the fingers from their position on the basic row. 'Jumps' from the 1<sup>st</sup> to the 3<sup>rd</sup> row and back, quite frequent with QWERTY, practically do not occur with Dvorak.

In its 100-year long history, the Bulgarian keyboard layout has often been characterized as a work of genius. Its effectiveness has been proved in one of the most convincing tests for keyboards: competitions between keyboard operators. To this day, only partial changes have been made to the layout, mainly in the 4<sup>th</sup> (numerical) row. In 1978, this layout became a Bulgarian State Standard. During the 80es, typing was a compulsory subject not only for those secondary school students specializing in administration, but also for a large number of normal and foreign language schools.

#### 4. E-communication and Problems with the Use of Cyrillic Characters

The early years of e-communication imposed a number of restrictions on the use of Cyrillic characters. Above all, e-mails and chat rooms had to make use of Latin characters. This situation was used by proponents (often acting under external pressure) of the introduction of Latin as a ‘parallel’ alphabet, i.e. of a gradual transition to a Latin-type of alphabet. Not any type of Latin, though (a Rumanian or Czech alphabet with their numerous diacritics could very well serve the purpose for Bulgarian), but, simply, the English alphabet. This drive initially took the form of waves of ‘globalization’ campaigns, cresting around the time of the celebrations of the Cyrillic alphabet and the brothers Cyril and Methodius (May 15 – 24). It later developed into ‘literacy’ campaigns, as Bulgarian users soon found out that the English alphabet could not perform the function of a Bulgarian one and tampered with it in different ways. Thus, to make up for the lack of specific letters, the new ‘alphabet’ incorporated the number keys and other available characters: 4 (Bulg. ‘chetiri’) became Ч ‘ch’, 6 (Bulg. ‘shest’) became Ш ‘sh’; ‘@’ became the middle vowel ‘Ъ’. The “useless” letters Q, W and X also acquired different functions Я, Ю, Ы. The process was chaotic and every chat room had its own ‘transliteration’, which soon acquired the status of jargon. A large part of the young population of the country thus became used to the QWERTY keyboard. Since, in the period following the democratic changes in the country, keyboard writing was not considered, by any government, an element of basic literacy, the nearly 100-year-old tradition of typing education was discontinued. Bulgarian children were not taught at school to use the Bulgarian standard keyboard. When they wanted to type in Cyrillic letters, the easiest option was to use the so-called ‘phonetic layout’, based on the English QWERTY.

Around the year 2000 a group of mathematicians, directed by the head of the NATO office in Sofia, Lyubomir Ivanov, made several attempts to impose ‘standards of literacy’ in using the English alphabet for writing Bulgarian texts. They organized presentations at the Bulgarian Academy of Sciences and in a number of universities but did not find direct support in academic circles. In 2003, two ministers from one of the governing parties in the country (the ‘National Movement *Simeon of Saxe-Coburg-Gotha*’) used their power to fund two parallel projects. The Minister of Education and Sciences Daniel Valchev funded a project entitled ‘Unified Solution for the Problems Arising from the Use of the Cyrillic Alphabet’. The project was aimed at the creation of a unified standard of transliteration based on the

English alphabet, to be mirrored in the creation of an English/Bulgarian QWERTY-based keyboard layout. As could be expected, the project did not succeed to fulfill these ambitions. A revised 'phonetic keyboard' was proposed as a second national standard, and a slightly revised transliteration procedure was put forward, with French-style transliteration substituted with an English-style one. The tasks of the project (creating a unified Bulgarian standard of transliteration for documents and keyboards based on the English alphabet) being by definition unattainable, the authors parted in the process, the linguists working on the transliteration refusing to accept the suggestions of the mathematicians. Thus, the project failed to offer a unified standard.

Even though the outcome of the transliteration project was not adopted as a national standard and a different transliteration standard is, even today, still active, the project was followed by another one, 'Understandable Bulgaria', financed this time by 'Saxe Cobourg Gotha' Minister of State Administration, Nikolai Vasilev (who had directly participated, as chairperson, in the discussion of the transliteration proposals). As part of this new ambitious, large-scale project, all Bulgarian name-plates with names of towns, villages, streets etc. are being substituted with new ones, featuring the (non-standardized) English-style transliteration.

For the duration of the 'Unified solution...' project, the authors organized a discussion forum with the purpose of popularizing the idea of a transition to the English alphabet and introducing 'standards of literacy' in writing Bulgarian texts with this alphabet. However, reactions were very strongly unfavorable. The organizers of the campaign were ridiculed, even verbally abused by their young 'targets', and soon had to give up this line of action.

The authors of the revised keyboards and the experts assessing the results of their work at the Bulgarian Institute of Standardization were two nearly identical sets of people, numbering 4. Shortly following the 'expert approval', but before official publication, the proposals for a new 'phonetic keyboard layout' and a revised version of the established (1978) keyboard layout were adopted by Microsoft in their 'Vista'. This gave rise to a wave of protests from users of both keyboard layouts. The large number of letters of protest forced the Bulgarian Institute of Standardization to postpone the publication of the keyboard standards and to organize an open public discussion. The discussion was held a couple of weeks ago, on June 19<sup>th</sup> 2008, in one of the conference halls of Sofia University, in the presence of experts in the fields of standardization, transliteration, keyboard layouts,

administration. The objections to the adoption of the proposed standards were as follows:

- the proposed standard does not come in answer to user demand;
- the standard was not performed by specialists in the area and was not professionally elaborated; as a result, the keyboard layouts do not answer to conditions of ergonomicity and ease of use.
- was not subjected to expert discussion in the TC-80 of BIS prior to preliminary adoption by the committee;
- was not accepted by a sufficiently representative committee of specialists; to sum up, the whole procedure was carried out ‘in the dark’;
- the project has a clear drive to the popularization of the English alphabet as a parallel system of writing Bulgarian texts – which was defined as conflicting with Bulgarian national interests and European language policy.

The adoption of the proposals was put to the vote and they were unanimously rejected. The decisions of the conference were: 1/ to adopt as keyboard layout standard for Bulgarian the 1978 keyboard layout for typewriters with, possibly, and after open discussion, necessary minor changes in the non-letter and non-punctuation keys; 2/ to wait for the outcome of the MEEK WS and possible follow-ups before considering an additional ‘phonetic’ standard, not necessarily based on the English QWERTY.

The conference agreed that the use of keyboard standards is largely an educational problem and that the adoption of standard keyboard layouts ought to be backed by educational measures.

## 5. 100 years later: the ANABELA Study

In view of the public discussion of the proposals for new keyboard layout standards and on the occasion of the 100<sup>th</sup> anniversary of the standard Bulgarian keyboard layout, the Association for the National Bulgarian Electronic Archive (ANABELA) and the Association of Stenographers and Keyboard Operators carried out a corpus-based investigation of Bulgarian texts in order to assess, 100 years after its creation, the merits of the standard keyboard (BSK) against the proposed QWERTY-based new ‘phonetic’ one (NPK), claimed by its authors to equal the BSK in effectiveness.

For the purposes of the study, the ANABELA provided data from two corpora: the 50 million running words general corpus of the Bulgarian

language (author: senior researcher Dr. Kiril Simov) and a 1 million running words corpus of texts from the register of state administration (author: Associate professor Dr. Maria Stambolieva). ANABELA also performed specific analysis of the corpora on the basis of criteria, defined by the National Association of Stenographers and Computer Operators. These included establishing letter frequencies, digraph frequencies, graphic word frequencies.

The two keyboard layouts were compared for:

1. Ease of highest-ranking letters entry. The top 10 letters were consecutively checked, for each of the keyboard layouts, for : keyboard register (1, 2, 3, 4), activated hand (R, L) and activated finger (1, 2, 3, 4, 5).
2. Ease of highest ranking letter combinations (digraphs) entry. 30 top ranking digraphs were assigned finger-hand-register indexes; these were compared for the two keyboard layouts.
3. Ease of entry for top rank graphic words (for frequency indexes over '10' in the corpus of state administration documents).

Letters.

The first 20 letters show very close parallels in the two corpora. The ranks of the 1<sup>st</sup>, 5<sup>th</sup>, 6<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, 9<sup>th</sup>, 10<sup>th</sup>, 14<sup>th</sup>, 15<sup>th</sup>, 16<sup>th</sup>, 17<sup>th</sup>, 20<sup>th</sup> letters are stable. Slight variations can be noticed in the relative ranks of the vowels И, Е, О; the consonants К, Л, П; Б and Г, as shown in Table 1 below.

Table 1. Letters: occurrences in corpus

General corpus: (50 million running words)		к	12249525
		л	11928785
а	45133223	п	10662252
и	33040769	м	9035809
е	32341653	з	8075900
о	32290450	я	6592038
		ь	6183145
т	27918353		
н	26267961	б	5868185
р	18780674	г	5280405
с	18583841		
в	16326925	у	4903144
д	12792758		

State administration corpus (1 million running words)	Л	182502
	П	175009
а	К	168690
е	М	136508
о	З	123860
и	Я	95188
	Ь	84001
т		
н	Г	82114
р	Б	71054
с		
в	У	63921
д		

Table 2 below defines each letter in terms of its position on the keyboard rows and the hand and finger to which it is assigned (assuming 10-finger typing).

Table 2. Letter entry: Finger (1–5)&Hand (Left, Right), Register (1–3)

	BSK	NPK
Rank 1.		
А	[3L, 2]	[5L, 2]
Rank 2-4.		
И	[2L, 3]	[3R, 3]
Е	[3L, 3]	[3L, 3]
О	[2L, 2]	[4R, 3]
Rank 5-10.		
Т	[2R, 2]	[2L, 3]
Н	[3R, 2]	[2R, 1]
Р	[3R, 1]	[2L, 3]
С	[3R, 3]	[4L, 2]
В	[4R, 2]	[4L, 2]
Д	[4R, 3]	[3L, 2]
Rank 11-13.		
К	[2R, 3]	[3R, 2]
Л	[4R, 1]	[4R, 2]
П	[2R, 1]	[4R, 3]
Rank 14-17.		

М	[4R, 2]	[2R, 1]
З	[4R, 3]	[5L, 1]
Я	[4L, 2]	[5R, 3]
Ъ	[2L, 1]	[2R, 3]

Rank 18-19.

Б	[5R, 1]	[2R, 1]
Г	[2R, 2]	[2L, 2]

Rank 20.

У	[4L, 3]	[2R, 3]
---	---------	---------

Comparing the 10 highest rank letters:

1.1. The *row assignment* is comparable, slightly superior in the case of the BSK: 5 letters in the 2<sup>nd</sup>, base position, row, 4 in the 3<sup>rd</sup> and only 1 in the 1<sup>st</sup> (Compare 4 letters in the 2<sup>nd</sup> row, 5 letters in the 3<sup>rd</sup>, 1 letter in the 1<sup>st</sup> row for the NPK).

1.2. The *finger assignment* is clearly better in the case of the BSK: 3 letters assigned to a forefinger, 5 – to a middle finger, 2 – to a ring finger. None of the 10 highest rank letters is assigned to a fifth finger. In the case of the NPK layout, the highest rank letter (А) is assigned to a fifth finger, while the other letters are evenly distributed among 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> fingers.

1.3. The 10 highest-rank letters of the Bulgarian alphabet involve a more *active right hand* using the BSK :4 letters are assigned to the left hand vs. 6 – to the right hand. Compare with 7L: 3R in the case of the NPK.

The 5<sup>th</sup> finger of the left hand for the undoubtedly most frequent Bulgarian letter, assigned by both the standard Phonetic keyboard and the NPK, is definitely not acceptable.

Finally, while the BSK keeps all letter keys within easy reach of the two hands, both existing versions of the “phonetic” keyboard involve displacements from the standard letter key positions (Cf. the position of Ю, Ъ, accented И, Ш, Я in the NPK layout).

## 5.2. Letter Combinations.

Bulgarian does not have diphthongs. It does not have digraphs in the phonetic sense (2 letters: 1 sound). Iotation of [a], [o] and [u] is relatively restricted. The majority of syllables are open, of the CV type, followed by CCV and CVC. 3-consonant clusters are not very frequent. It is in view of

these facts that the BSK was designed – with consonants and consonant groups assigned to the more active right hand, single vowels assigned to the left hand, consonant-vowel alternations reflected in right hand-left hand alternations. Table 3 below presents the most frequent Bulgarian 2-letter combinations (digraphs), indexed for finger, hand and keyboard row.

Table 3. Two-letter combinations.

Digraph and frequencies		Fingers and keyboard rows	
		BSKeyboard	NPK keyboard
на	8745512	[3R, 2], [3L, 2]	[2R, 1], [5L, 2]
та	5315014	[2R, 2], [3L, 2]	[2L, 3], [5L, 2]
то	5309415	[2R, 2], [2L, 2]	[2L, 3], [4R, 3]
ни	5231782	[3R, 2], [2L, 2]	[2R, 1], [3R, 3]
ат	4960094	[3L, 2], [2R, 2]	[5L, 2], [2L, 3]
ст	4754920	[3R, 3], [2R, 2]	[4L, 2], [2L, 3]
ра	4670790	[3R, 1], [3L, 2]	[2L, 3], [5L, 2]
те	4540845	[2R, 2], [3L, 3]	[2L, 3], [3L, 3]
ва	4184135	[4R, 2], [3L, 2]	[2L, 1], [5L, 2]
ен	4011182	[3L, 3], [3R, 2]	[3L, 3], [2R, 1]
ит	3826383	[2L, 3], [2R, 2]	[3R, 3], [2L, 3]
от	3738225	[2L, 2], [2R, 2]	[3R, 3], [2L, 3]
пр	3706451	[2R, 1], [3R, 1]	[4R, 3], [2L, 3]
ре	3520766	[3R, 1], [3L, 3]	[2L, 3], [3L, 3]
но	3462865	[3R, 2], [2R, 2]	[2R, 1], [3R, 3]
ан	3398216	[3L, 2], [3R, 2]	[5L, 2], [2R, 1]
за	3328827	[4R, 3], [3L, 2]	[5L, 1], [5L, 2]
по	3285929	[2R, 1], [2L, 2]	[4R, 3], [3R, 3]
да	3284242	[3R, 3], [3L, 2]	[3L, 2], [5L, 2]
не	3119811	[3R, 2], [3L, 3]	[2R, 1], [3L, 3]
ли	2622831	[4R, 1], [2L, 3]	[4R, 2], [3R, 3]
се	2555190	[3R, 3], [3L, 3]	[4L, 2], [3L, 3]
ти	2300146	[2R, 2], [2L, 3]	[2L, 3], [3R, 3]
че	1970506	[5R, 2], [3L, 3]	[5L, 3], [3L, 3]
ви	1863665	[4R, 2], [2L, 3]	[2L, 1], [3R, 3]
ме	1791825	[4R, 2], [3L, 3]	[2R, 1], [3L, 3]
из	1640316	[2L, 3], [4R, 3]	[3R, 3], [5L, 1]
си	1580741	[3R, 3], [2L, 3]	[4L, 2], [3R, 3]
го	1507553	[2R, 2], [2L, 2]	[2L, 2], [4R, 3]
аз	1373131	[3L, 2], [4R, 3]	[5L, 2], [5L, 1]

The data presented in the table above can be summarized as follows:

- 2.1. *Hand alternation.* With two exceptions, all digraphs are of the CV or VC type – hence the 28 cases of hand alternation for the BSK, against only two cases of right-hand activation only. In the case of the NPK keyboard, however, the two hands alternate in only 13 out of 30 cases.
- 2.2. *Finger alternations.* In the two cases where there is no hand alternation using the BSK, there is unproblematic finger alternation (3<sup>rd</sup>, 2<sup>nd</sup> or 2<sup>nd</sup>, 3<sup>rd</sup>). With the NPK keyboard, in two important cases (the preposition ‘za’ (for) and the 1<sup>st</sup> p. sg. personal pronoun ‘az’) there is no finger alternation, the finger being the 5<sup>th</sup> finger of the left hand, with displacement from the 2<sup>nd</sup> to the 1<sup>st</sup> row or backwards.
- 2.3. *Rows.* With the BSK, the ten highest-rank digraphs involve use of the base row in 16 cases, of the 3<sup>rd</sup> row in 3 cases and of the undesirable 1<sup>st</sup> row – in only 1 case. With the NPK keyboard, the figures are: 6 letters in the base row, 10 in the 3<sup>rd</sup> row and 1 in the 1<sup>st</sup> row. In general, using the BSK for the digraphs will involve over 50 % of base row strokes, 33% strokes in the 3<sup>rd</sup> row and just over 15% for the 1<sup>st</sup> row. With the NPK, the majority of strokes fall in the 3<sup>rd</sup> row, the 1<sup>st</sup> and 2<sup>nd</sup> rows being of comparable use: Cf. Table 4 below.

Table 4. Rows.

	BSK	NPK
1-10 digraph		
row 1	1	4
row 2	16	6
row 3	3	10
11-20 digraph		
row 1	4	4
row 2	11	4
row 3	5	12
21-30 digraph		
row 1	1	4
row 2	7	5
row 3	12	11

2.4. *Finger activity.* Using the BSK, the digraphs in Table 3 involve the use of the active forefingers and middle fingers in 52 out of the 60 cases. With the NPK keyboard, these fingers are activated in only 39 out of 60 cases. The little finger (of the right hand) will be activated only once using the BSK. The little finger (of the left hand) is very active with the NPK, taking care of nearly 25% of the strokes – Cf. Table 5 below.

Table 5. Finger activity.

(Fingers)	BSK				NPK			
	2	3	4	5	2	3	4	5
1-10 digraph	7	12	1	0	10	3	2	5
11-20 digraph	8	11	1	0	7	7	2	4
21-30 digraph	8	6	5	1	4	8	4	4

### 5.3. Graphic words.

Table 6 presents a list of highest-rank graphemes (over 10 occurrences) in the one-million Bulgarian corpus from the register of state administration. A relatively small percentage of the forms pose problems for the efficient 10-finger typing with the Bulgarian Standard Keyboard. With the new phonetic keyboard, however, professional 10-finger entry can be difficult (\*) or very difficult (\*\*).

Table 6. List of highest-rank graphemes in the 1 million Bulgarian corpus from the register of state administration.

АЗ **	БЪДЕ**	ВКЛЮЧИТЕЛНО**
АКО	БЪЛГАРИЯ**	ВНИМАНИЕ
БАНКА*	БЮДЖЕТ**	ВОДА
БЕ	БЯХА*	ВРЕМЕ**
БЕЗ*	В	ВСИЧКИ**
БЕШЕ*	ВАС	ВТОРА**
БИ	ВАШАТА*	ВЪВ
БИХ*	ВАШЕТО*	ВЪЗДЕЙСТВИЕ**
БЛАГОДАРЯ	ВЕЧЕ*	ВЪЗДЪРЖАЛИ*
БЪДАТ**	ВИ	ВЪПРОС

ГЛАСУВА  
ГОВОРИ  
ГОДИНА  
ГОСПОДА  
ГОСПОДИН  
ДА\*  
ДАЛИ\*  
ДЕЙНОСТ\*\*  
ДЕЙСТВИЕ\*\*  
ДЕН\*\*  
ДЕЦА\*\*  
ДНЕС  
ДО  
ДОБРЕ  
ДОКЛАДЧИК\*\*  
ДОКУМЕНТ\*\*  
ДОСТА  
ДУМАТА\*\*  
ДУМИТЕ\*\*  
ДЪРЖАВАТА\*  
ЕДВА\*\*  
ЭКСПЕРТ\*  
ЗА\*\*  
ЗАКОН\*\*  
ЗАМЕСТНИК\*\*  
ЗАТОВА\*\*  
ЗАЩО\*\*  
ЗАЯВИТЕЛЯ\*\*  
ЗДРАВЕ\*  
ЗНАЧИТЕЛНО\*\*  
ИЗПОЛЗВАНЕ\*\*e  
ИЗПЪЛНЕНИЕ\*  
ИКОНОМИКА\*\*  
ИЛИ\*  
ИМ\*\*  
ИМАТЕ\*\*  
ИМЕННО  
ИНФОРМАЦИЯ  
ИСКАМ\*\*  
КАЖА\*  
КОГАТО  
КОЕТО  
КОЛЕГИ\*  
КОМИСИЯ\*\*  
КОНКУРЕНЦИЯ  
КОЯТО\*  
КЪМ\*  
МАРТ\*\*  
МЕЖДУ\*  
МЕСЕЦ\*\*  
МИНИСТЪР\*

ПРЕДСЕДАТЕЛ  
МИСЛЯ  
МОЖЕ  
МОЛЯ\*\*  
МУ\*\*  
МЯСТО  
НАРОДЕН  
НАС  
НАЧИН\*\*  
НЕ  
НЕОБХОДИМ\*\*  
НЕЩО\*  
НИЕ  
НО  
НЯКОЙ\*  
НЯКОЛКО\*\*  
НЯМА\*  
ОБРАЗОВАНИЕ\*\*  
ОБХВАТ\*\*  
ОБЩИНА\*  
ОБЩИНСКИ\*\*  
ОБЩО\*  
ОРГАН\*  
ОТ  
ОТГОВОР\*\*  
ОЩЕ  
ПАЗАР\*\*  
ПОВЕЧЕ  
ПОЛИТИКА\*\*p  
ПОСОКА\*\*r  
ПРЕДВИД\*\*r  
ПРЕДЛОЖЕНИЕ\*\*  
ПРЕДПРИЯТИЕ\*\*  
ПРЕДСЕДАТЕЛ\*\*  
ПРЕДСЕДАТЕЛЮ\*\*  
ПРЕЗ\*  
ПРИЕМА  
ПРИЕТ  
ПРОГРАМА  
ПРОТИВ  
ПЪРВА\*\*  
ПЪРВИ\*\*  
ПЪРВО\*\*  
РАЗБИРА\*\*r  
РАЗДЕЛ\*\*  
РАЗЛИЧЕН\*  
РАЗПОРЕДБИ\*\*  
РАЗРЕШЕНИЕ\*\*  
РЕД\*\*  
РЕДАКЦИЯ\*\*  
РЕЗУЛТАТ\*

РЕПУБЛИКА\*\*  
СА\*  
САМО\*  
СЕ  
СЕГА\*  
СЛЕД\*\*  
СЛЕДВА\*\*  
СЛУЖБА\*  
СМЕТКА\*  
СПОРЕД\*\*s  
СРЕДСТВА\*\*  
СРЕЩУ  
СТАВА\*  
СТЕ\*  
СТОКИ\*\*  
СТРАНА  
СЪБРАНИЕ\*\*  
СЪДЪРЖАНИЕ\*  
СЪМ\*\*  
СЪЩО\*  
ТАЗИ\*\*  
ТАМ  
ТЕ  
ТЕЗИ\*  
ТОЗИ\*  
ТРЯБВА\*  
ТУК  
ТЪЙ\*  
ТЪРСЕНЕ  
ТЯ  
УВАЖАЕМИ\*\*  
УСТРОЙСТВО\*\*  
УЧАСТИЕ\*\*  
УЧЕБНИК\*\*  
ЧАС\*\*  
ЧАСА\*\*  
ЧАСОВИ\*\*  
ЧАСТ\*\*  
ЧЕ  
ЧРЕЗ\*\*  
ЩЕ\*

Forms marked with one star (\*) present one or all of the following difficulty types: no hand alternation and difficult finger alternation; 4<sup>th</sup>-5<sup>th</sup> finger activity; keyboard “jumps” between the lowest and highest rows. These graphemes (45) amount to 26% of the forms listed.

Forms marked with two stars (\*\*) present one of the following difficulty types: no finger alternation; reaching for the leftmost corner of the 4<sup>th</sup> row (Ю) or for the rightmost corner (Щ, Я), necessitating eye control. These graphemes (79) amount to 45% of the forms listed.

Thus, over 70% of the highest-rank graphemes in our corpus would pose problems for professional entry using the NPK keyboard.

With the BSK, the same criteria marked only 10 graphemes with one star (включително, време, въздържам, въпрос, предложение, предприятие, председател, председателю, разбира, търсене) and 16 graphemes – with two stars (днес, държава, едва, здраве, използване, информация, председател, нас, отговор, предвид, следва, служба, според, средства, срещу, уважаеми). The problematic cases thus amount to only 15% of the listed graphemes.

The analysed data indicate that the NPK keyboard layout is not adequate to the task of the efficient entry of Bulgarian texts. While the creation of a unified keyboard for Bulgarian Cyrillic and one or more Latin-based alphabets is no doubt a task well worth investing in, it is a task which is still to be carried out – on the basis of the definition of well-motivated effectiveness criteria and priority hierarchies and of serious corpus analysis.

## 6. Alphabets, Keyboard Layouts, Writing Tuition and the EU Language Policy.

6.1. The desirability of finding alternatives or parallel solutions to the present colourful keyboard situation stems from: 1/ the substitution of handwriting with typing as the dominant ‘writing’ activity, with the necessity to recommend to users effective, ergonomic keyboard layouts, ensuring maximum speed of information entry with a minimum of strain and negative physical consequences; and 2/ the EU language policy, encouraging the study of many languages, not simply one or two international language (s) – with the ensuing problem of designing maximally economical keyboards, allowing ease of information entry with a minimal mnemonic effort.

6.2. As was pointed out above, there is no such thing as an abstract ‘Latin’ alphabet – alphabets are language-specific; similarly, there is no unified ‘Cyrillic’ alphabet. Transliteration standards therefore have a choice between favouring one particular language as their target (for Bulgarian, this was French in the past, nowadays – English), or else (a more reasonable solution), preferring international standards (such as ISO-9). At the moment, ISO-9 does not offer a

satisfactory option for Bulgarian because of the non-letter equivalents proposed for the letters Ъ and Ь. The Bulgarian Institute for Standardization has, however, announced its intention to apply for respective changes in this standard. If this becomes a fact, a basis will be created for a reversible 1:1 transliteration, making use of diacritics for the specific Slavonic letters. There is nothing undesirable in this: ALL Slavonic languages making use of a Latin-type have diacritics, and there is no reason why Bulgarian should form an exception. Letter combinations do not yield unambiguous reversibility.

6.3. There is no necessary dependence between language group and alphabet. Many of the ex-Soviet republics used alternatively Arabic, Latin and Cyrillic. Rumanian used Cyrillic until the 20<sup>th</sup> century. A well-designed transliteration standard can easily demonstrate that one can switch from one to the other and back without great effort.

6.4. Reversible transliteration can form the basis for partially phonetic-type Latin/Cyrillic/Greek keyboards for EU users. This, however, can clearly not be done on the basis of the English alphabet and QWERTY keyboard: keyboards will inevitably have to be extended, with longer letter rows, function keys for diacritics, or both.

6.5. There is no widely accepted keyboard layout which could serve as an obvious candidate for keyboard unification. A closer look at the so-called QWERTY layout shows that not much of it remains stable from language to language (neither the Q, nor the W, nor the Y, for sure). The situation is even more varied for the languages using the Cyrillic alphabet.

6.6. QWERTY-type keyboards are not designed on the basis of quantitative analyses of language data; their efficiency is not a compliment to 21<sup>st</sup> century humanity ☺.

6.7. The similarities between the Bulgarian Standard Keyboard and the Dvorak layout demonstrate that, broadly, the task of designing an effective keyboard layout for more than one language is not an impossible one. A set of unified principles applied to languages of the same (Indo-European) group, increasingly enriched with international lexis and standardized European jargon, might well yield very similar results.