



Ministry of
Communication



IDRC  CRDI



PAN
Localization
www.PANL10n.net

Pashto Keyboard Layout

Habiburahman Najjullah

Hameedullah Sherani

E-mail:

habib.ghaznawe@acsa.org.af
ghaznawe@gmail.com

hameed.sherania@acsa.org.af
haider sherani@yahoo.com

TABLE OF CONTENTS

Abstract.....	1
Introduction	1
History	2
Definitions	5
First phase of the research	6
Pashto keyboard layouts.....	7
Conclusions.....	15
Problems found in Everson keyboard layout.....	16
Comparison between Microsoft / Everson keyboard layouts	16
Comparison between Liwal / Everson keyboard layouts.....	18
Conclusions.....	18
First part of the frequency analyzer.....	21
Second part of the frequency analyzer	22
Statistical results	24
Frequency analysis in graphical format.....	25
The errors.....	26
Errors table.....	27
Second phase of Pashto keyboard research and development.....	29
Old keyboard layouts	29
Famous keyboard layout design analysis	36
The new keyboard layout development	38
Cold test for the new keyboard layout.....	39
The new keyboard layout made with MSKLC	40
Acknowledgments	42

Abstract:

The lack of patron ship, organizational attention and unified approach towards the development of a standard keyboard layout has brought in chaos and issues.

The 30 years of war has torn the institutions of Afghanistan and thus any developments taking place during this era are due to individual endeavors. These individual endeavors are worth praise and gratitude but lacks a unified and organized vision towards the development of standards in developing local technologies, and thus failed to bring Pashto into the line of developing languages in technology.

The lack of consistency in the work already done in regard of keyboard layout is one of the main initiating factors of this research report.

The inconsistent keyboards developed by individuals and/or organizations sponsored individuals are always based on the old non scientific, non statistical keyboard layouts, and modified copies of Arabic, Farsi keyboard layouts.

This paper highlights the works done by individuals and identifies the demerits of the designed keyboard layouts.

The paper discusses the development and implementation of a scientific keyboard based on the frequency analysis of the Pashto character set. It also compares and evaluates the typing efficiency of this scientific statistics based keyboard.

Introduction

During the 1980's and back, a great need was felt for the digitized production and publications in Pashto. Both the government in Kabul and the afghan jihadist movements in Peshawar were acquiring the ability to produce high quality digitized publications in Pashto.

Newspapers and magazines were typed by the regular typing machines, which lacked the beauty, variety of fonts, attractive design, and easiness of composing work.

Computer systems at that time did not have graphical interface, it was a difficult task to make a new system for Pashto language. People at that time worked hard to design and make a solution for this problem. Therefore it is important to mention the pioneers in this field especially those who contributed their hard work and efforts to the afghan community for free.

Recently work has been done in keyboard layout designing; based on the older keyboard layouts and in order to meet the emergency needs- one of the layouts was approved by ministry of communications of Afghanistan.

As need and requirements of every language and nation vary, the need for more research was felt in order to design a keyboard layout based on scientific research, just Imitating keyboards of other languages doesn't comprehensively address the issues of Pashto language and Pashto typing.

History:

The history of Pashto keyboard layout makers is recorded here in respect to the date of the work done.

An abstract of the received emails about the keyboard layout history is presented here:

Note: Most emails are the exact wordings of the contacted persons with slight modifications.

Mr. Wadan [kader.wadan@titaniaisoft.com]

This is a summary of Mr. Wadan's efforts in Pashto keyboard layout creation.

In 1989 he participated in Berlin- a meeting about the availability of typewriters for the Pashto language. The result of the pre discussions that had taken place in Afghanistan was to order IBM typewriter machines. As he heard about this he suggested not to order typewriter machines but to buy a computer and install proper software for the Pashto language. He mentioned the known advantages of a computer based system over a typewriter and the audience was enthusiastic about the abilities such an alternative brings with it. And so he was asked to provide such a system. At that time he didn't knew that there was no proper program for the Pashto language that fulfill the given requirements. The only possibility existed, was the software developed by the company Gamma Universe. Their program called "Scholar" was text oriented. One had to write encrypted text with English letters from left to right and a second program converted these to Arabic and Pashto letters after it was sent to the printer. Meanwhile –after 1990- they have changed this, so one can read the written text already on screen.

In 1989 he had to provide software based on WYSIWYG (What You See Is What You Get). He decided to develop new software from scratch. He got an invitation from the Kabul University to teach programming and modeling in the C language. His aim was to demonstrate programming strategies in C and develop a word processing program for Pashto. He planned to implement this program in three modules simultaneously during lectures to demonstrate the theoretical techniques.

He developed a prototype for each module to see that wanted goals can be achieved.

The three modules consisted of:

1. **font drawing** :a visual program consisted of a matrix to draw the individual pixels of a font
2. **word processing** :a WYSIWYG based visual program
3. **A printer driver** : The driver was written for 24-Pin printers for the Pashto fonts that he defined by the font drawing and definition module

All three modules were functional, but still they were, prototypes. The keyboard assignment was phonetic it was hardwired in the source code.

He was disappointed that C language was not known, there was no programming experience in Kabul University. So he had to teach the very fundamental aspects of the C-Language. There were no resources for the project he had in mind and so the development of the word processing program for the Pashto language remained at the prototype stage. But it was completely functional software with the necessary features.

There exists an interview with Mr. Wadan on November / December 1989 in AFG-TV, where you can see this program in action.

Mr. Noorulhoda Atel

He started working on the Pashto fonts and keyboards in 1991 using MLS V.3 (Multilingual Scholar), which was a DOS-based basic multilingual text editor, with many useful functions and rather simple to use interfaces for keyboard and font creation and editing. His work was widely used by NGOs working in Peshawar up to 1995. To mention some: Swedish Committee for Afghanistan (SCA), International Rescue Committee (IRC), the Dutch Committee for Afghanistan (DCA), and other composing organizations i.e. Right Type composing center.

In early 1994 he came across a multilingual add on for WordPerfect 5.0 (which also worked with WP 5.1 and he used it with the later version) and started developing Pashto fonts and keyboards for it. The part time development work took about 6 months and was completed in September 1994. DCA was the first organization which started using the adapted WP 5.1 for serious work as soon as it was made available and continued using it up to 1998. The Livestock Program of Afghanistan was the second organization who adopted the software and used to publish multiple books and other publications. Many other NGOs and UN organizations also used the software. Experienced users could use this version of WP 5.1 with a practicality equivalent to very advanced text processing software and using LaserJet printers, it was possible to produce quality publications with this software. Organizations who had expert WP 5.1 users continued using this software up to 2002 even when multilingual operating systems and text processor were available.

Mr. Liwal [info@liwal.net]

Mr. Liwal the owner of the Asia-soft and the current giant of the market didn't provide any information based on the business strategy.

Mr. Sherzad Kamawall [kamawall@hotmail.com]

Kamawall started working on keyboard layout in 2002. At the same time Said Marjan Zazay (Afghanan.net) was also working on his version of a keyboard. Actually it was Zazay who suggested him to program a keyboard.

After a while he found out that another person from NWFP (www.khpalapashto.com) also had made his own keyboard layout.

His biggest problem in creating Pashto layout was the lack of knowledge about character frequencies and detailed knowledge of Pashto alphabets. He mapped the characters at that time to lewal's layout or khpalapashto's as he can't remember exactly.

He also had trouble with the lack of a standard Pashto font. There was "Pashto Kror Asiatype" Font by Liwal. Marjan had also created some beautiful fonts. "Pashto Kror Asiatype" had some bugs (it still has).

The Other figure in Pashto keyboard history that helped Mr. Sherzad was Mr. Abdulhaleem Yousufzay (mashriqsoft jalalabad) in creating a font which he named it Pokhto.ttf. It's now widely used in the internet web pages.

Mr. Said Marjan Zazay [www.afghanan.net]

Said Marjan Zazai is the managing director of AestheTech Software based in Peshawar, Pakistan. AestheTech Software is a software development company which emphasizes on localization. Previously Zazai has been involved in Unicode Font development for Pashto language. While in Kabul he worked voluntarily for ministry of communications of Afghanistan as a Pashto localization expert together with Everson Typography (www.evertype.com) to standardize the Pashto/Dari keyboard layouts and worked on collation and locale for Pashto language. He also due to the business strategy didn't provide any information.

Mr. Mobtakir:

On Monday, May 30, 2005 we got email from Mr. Mubtakir about his request to Microsoft Company to help him legalize his proposed Pashto keyboard to be used both in Afghanistan and Worldwide. But regrettably he did not share with us his proposed Pashto keyboard layout due to his business strategy.

About Mobtakir:

Mr. Mobtakir was an English teacher in Aeronautical Schools in Kandahar/Kabul before joining UNESCO in 1960 as Assistant to Chief of UNESCO Mission. He founded the First Management, Secretarial, and Typing Private School in Kabul, Afghanistan in 1964. He was also importing and distributing typewriters in Afghanistan for the last 26 years

Meeting:

Recently a meeting was arranged with Mr. Mobtakir about the keyboard and the need to make a keyboard with scientific bases. He complained about the current imitated keyboard layouts based on Arabic and Persian keyboards. He encouraged us to start the work for making scientific based Pashto keyboard layout.

Definitions:

Keyboard is a hardware input device consisting of an array of keys that the user presses in order to enter text into the computer.

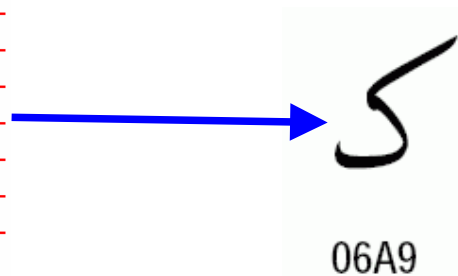
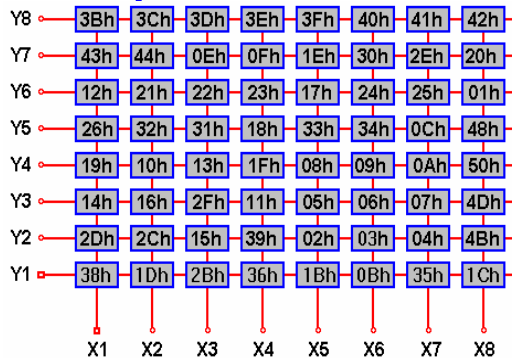
Dead key A key press or modifier-plus-key press combination that produces no immediate effect, but instead modifies the character or characters produced by the next key (called the completer key) that is pressed. In Pashto writing it is was used in Liwal keyboard layout.

Keyboard layout is the specification of the physical arrangement of keys on a keyboard and the characters produced when those keys are pressed. It is the software that connects the hardware [Keyboard] to the system. Keyboards of each country may have different locations for the keys arrangement

There are large number of keyboard layouts used for different languages written for different scripts, i.e. Roman script, Arabic script, Han-gul script, Latin script, each have many languages that uses them.

How keyboard works:

- Explanation:
 - User presses a key
 - Key have scan code
 - Keyboard layout software maps the scan code and sends it to the OS then to application.



First phase of the research:**The Idea:**

The idea of developing a Pashto keyboard layout was visualized after a research study of the previous keyboards. After a thorough study of previous keyboards we came to know that these layouts were made without any reasonable bases; but were mere imitation of the Farsi and Arabic keyboards, ignoring and by passing the language contents, the usage frequency of alphabets and the efficiency of Pashto typing.

Beginning of the search

There was a more need for developing a more efficient keyboard based on scientific reasons and statistical analysis. We started collecting information on the structure of certain keyboards like QWERTY, DVORAK, XPeRT, ABKEY keyboards, and the new design of a Hindi keyboard. Many Pashto keyboard layouts were researched as well which are presented later in this report.

As there was an objective of developing a keyboard based on the frequency analysis of Pashto alphabets. The procedures used in the development of the new keyboards were carefully analyzed.

Pashto keyboard layouts:

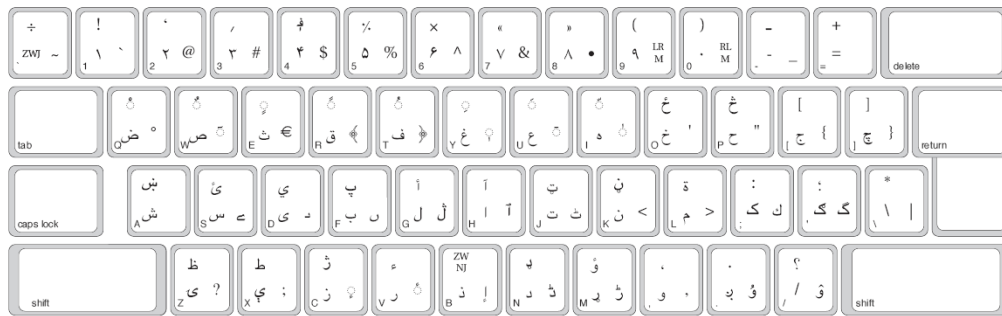


Fig 1. Michael Everson and Roozbeh Pournader layout. (MoC approved).✂

The above layout is implemented by Habiburahman using Microsoft keyboard layout creator (MskLc). This layout was not available for MS windows, and the current available in MS windows is not compatible with Everson keyboard layout.



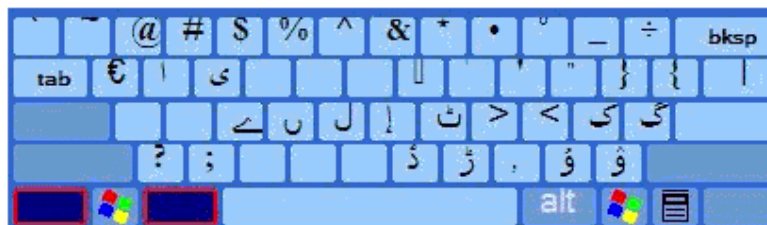
Normal keyboard layout

Fig 2.1: Normal Keyboard Layout



Shift keyboard layout

Fig 2.2: Shift Keyboard Layout



Ctrl+Alt Keyboard layout

Fig 2.3: Ctrl+Alt Keyboard layout

Tolafghan keyboard layout:



د ټول افغان کیبورډ

✂Fig 3: Tolafghan Keyboard Layout (Normal and with Shift pressed)

Tolafghan new keyboard layout: (claims MoC approval)



د افغانستان دمخبرانو وزارت له خوا منل شوی د ټول افغان کیبورډ

✂Fig 4: New Tolafghan Keyboard layout (Normal and shift pressed)

Fig 5: Khpala Pashto Keyboard: ❀1



Fig 5.1 normal state keyboard layout



Fig 5.2 shift state keyboard layout



Fig 5.3 ctrl+alt state keyboard layout

Fig 6: Pashto Keyboard layout developed by Liwal Software: ❀2



Fig 6.1 normal state keyboard layout



Fig 6.2 shift state keyboard layout



Fig 6.3 ctrl+alt state keyboard layout

❀1 source :(<http://www.khpalapashtu.com/sitee/pashtusw/paskeyb.htm>)

❀2 source :(<http://www.liwal.com/windows/pashto/keyboard.htm>)

Fig 7: Universal word 2000 Keyboard layout: ✨

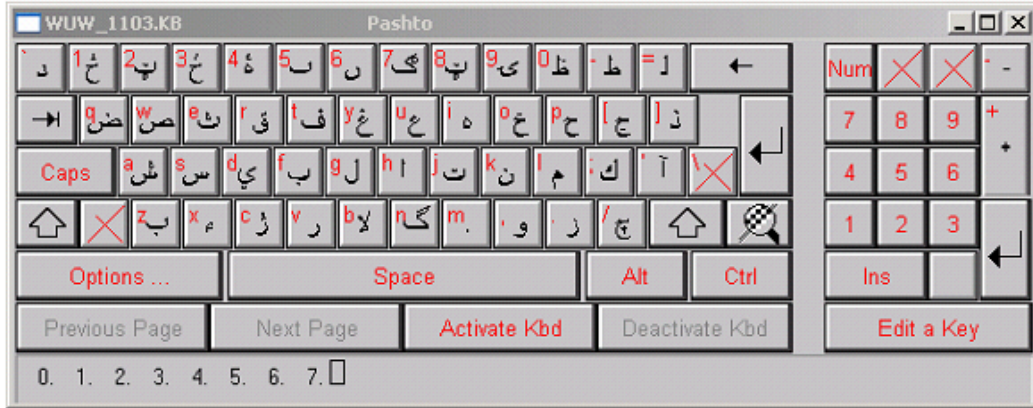


Fig 7.1 normal layout

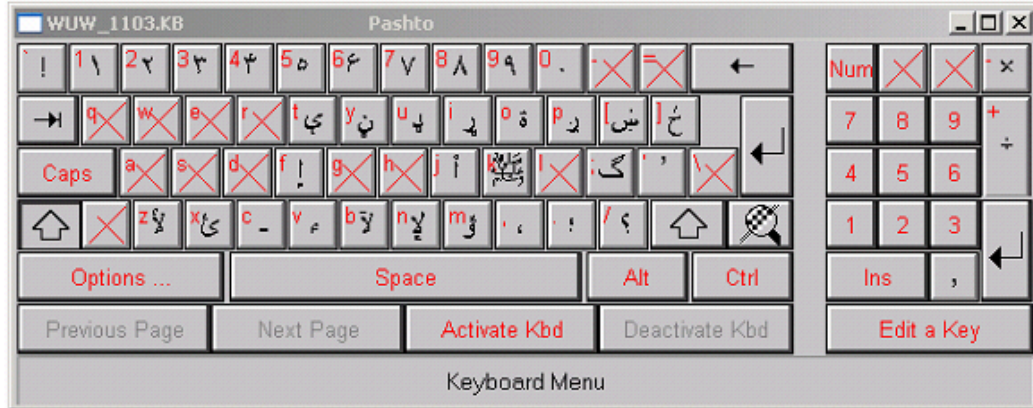


Fig 7.2 shift layout



Fig 7.3 ctrl+alt layout



Fig 7.4 ctrl layout

Fig 8: Gamma universe keyboard layout: ✨



Fig 8.1 normal layout



Fig 8.2 shift layout



Fig 8.3 ctrl layout

Fig 9: Unitype Global Writer: ✨1



Fig 9.1 normal layout



Fig 9.2 shift layout

OLPC XWindow based Walter Bender Layout: ✨2

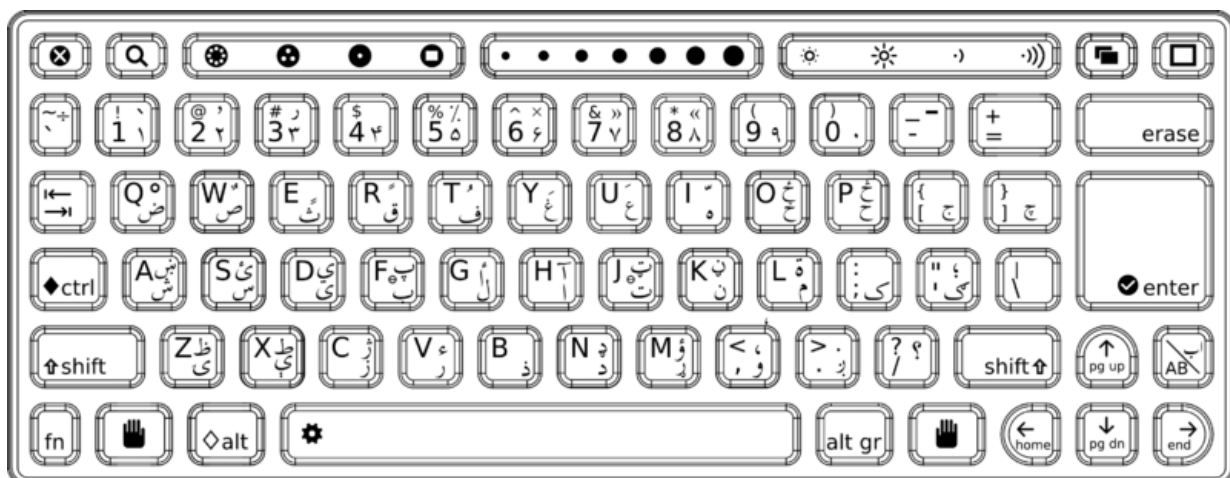


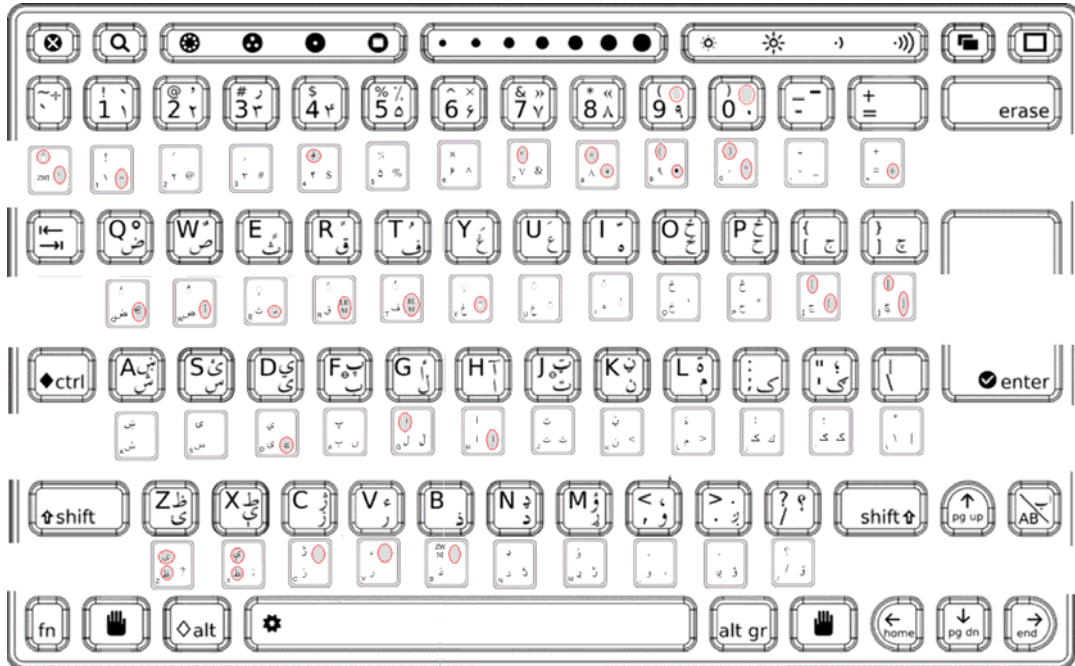
Fig: OLPC Walter Bender Keyboard

✨1 source: (<http://www.unitype.com/globalwriter.htm>)

✨2 (http://wiki.laptop.org/go/Pashto_Keyboard)

When comparing the official Pashto keyboard layout for Afghanistan (made by Everson, Roozbeh, Marjan Zazai, & Tamim Noori) and the Pashto keyboard layout made by Walter Bender. There were many differences.

For details please check the following table and image:



(yellow color shows the difference with walter keyboard layout.)

	Unshifted	Shifted	AltGr
TLDE	200D	0654	0060
E01	06F1	0021	007E
E02	06F2	066C	0040
E03	06F3	066B	0023
E04	06F4	AFGHANI	0024
E05	06F5	066A	0025
E06	06F6	00D7	005E
E07	06F7	00BB	0026
E08	06F8	00AB	066D
E09	06F9	0029	2022
E10	06F0	0028	00B0
E11	002D	0640	005F
E12	003D	002B	00F7
D01	0636	0652	20AC
D02	0635	064C	0671
D03	062B	064D	0649
D04	0642	064B	200E
D05	0641	064F	200F
D06	063A	0650	0653
D07	0639	064E	ZWARAKAY
D08	0647	0651	0670
D09	062E	0681	0027
D10	062D	0685	0022
D11	062C	005D	007D
D12	0686	005B	007B
C01	0634	069A	<FREE>
C02	0633	06CD	<FREE>

C03	06CC	064A	06D2
C04	0628	067E	06BA
C05	0644	0623	06B7
C06	0627	0622	0625
C07	062A	067C	0679
C08	0646	06BC	003E
C09	0645	0629	003C
C10	06A9	003A	0643
C11	06AB	061B	06AF
B01	0638	0626	003F
B02	0637	06D0	003B
B03	0632	0698	<FREE>
B04	0631	0621	<FREE>
B05	0630	200C	<FREE>
B06	062F	0689	0688
B07	0693	0624	0691
B08	0648	060C	002C
B09	0696	002E	06C7
B10	002F	061F	06C9
BKSL	005C	002A	007C
SPCE	0020	200C	00A0

Note:

- Key AD01 shifted status key is mentioned by Walter to be Arabic Sukun which is 0652 Unicode but is mentioned FE7E by Walter table.
- Key AE04 shifted status is the Afghani currency sign which is 060b Unicode, but is mentioned e60b by Walter table.

From the market

The market was checked for old type writers and found that most look like the following enhanced image:

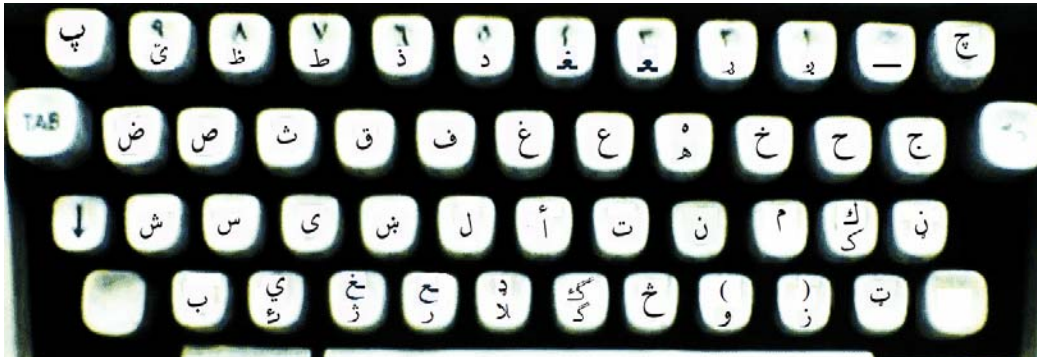


Fig 10

The original image:



Fig 11

Conclusions:

This survey conclude that

- There is no standard keyboard layout.
- The layouts are not based on scientific reasons.

These inconsistent keyboard layouts create problems on the user side and also adversely affects in development of a unified non variant language content and corpus.

Problems found in Everson keyboard layout:

- Based on Arabic keyboard layout.
- He made two more layouts for Dari and Uzbek that could be handled in Pashto keyboard layouts. For more explanation check the following image:

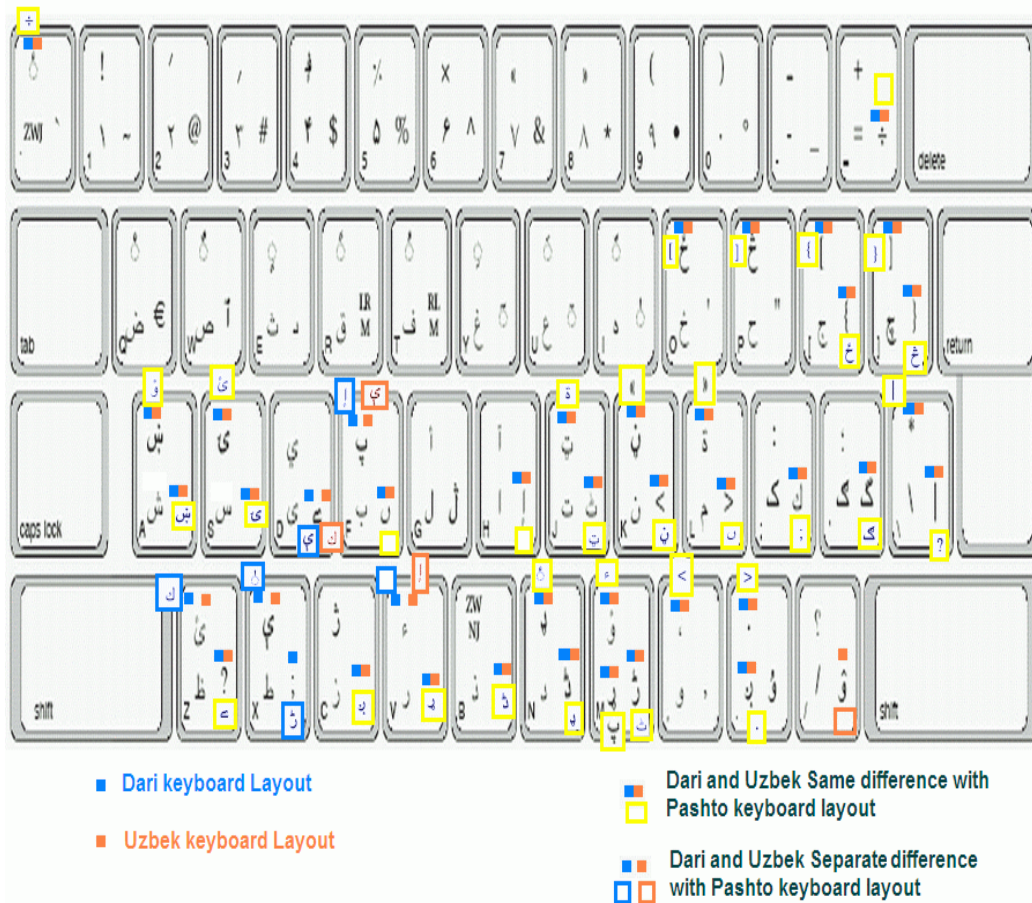


Fig 12: Everson Keyboard layout

The colored squares indicate the scattered characters position that shows how hard it is for the typist to remember the different layouts for the different languages. And that it all can be under Pashto keyboard layout.

Comparison between Microsoft / Everson Keyboard layouts:

- This document shows the difference in layout between Microsoft keyboard layout and Everson which is also known by “TolAfghan keyboard layout”
- The keyboard layout out by Michael Everson and his associates (Roozbeh Pournader, Zazai) is the layout that was recognized and attested by Ministry of Communications in Afghanistan.
- Microsoft keyboard layout is therefore not compatible with the Everson and associates keyboard layout.



Fig 13
Microsoft normal
keyboard layout



Fig 14
Everson normal
keyboard layout



Fig 15
Zazai normal
keyboard layout



Fig 13.1
Microsoft
normal
keyboard
layout



Fig 14.1
Everson
normal
keyboard
layout

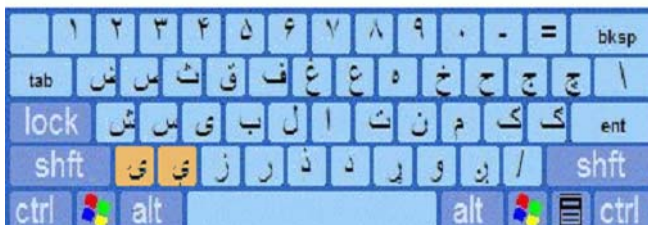


Fig 15.1
Zazai normal
keyboard
layout

in microsoft the characters:
 ي ي
 are in place of the characters
 ظ ط
 in everson keyboard layout



Fig 13.2
Microsoft Shift
keyboard layout



Fig 14.2
Everson Shift
keyboard layout



Fig 15.2
Zazai Shift
keyboard layout



Fig 13.3
Microsoft
shift
keyboard
layout



Fig 14.3
Everson
shift
keyboard
layout



Fig 15.3
Zazai
shift
keyboard
layout

microsoft KeyB layout	everson KeyB layout
÷	ا
\$	پ
»	خ
»	خ
ئ	ی
ظ	ئ
ط	ی



Fig 13.4
Microsoft Alt
keyboard layout



Fig 14.4
Everson Alt
keyboard layout



Fig 13.5
Microsoft
ALT
keyboard
layout



Fig 14.5
Everson
ALT
keyboard
layout

microsoft KeyB layout	everson KeyB layout
~	~
esc	esc
F1	F1
F2	F2
F3	F3
F4	F4
F5	F5
F6	F6
F7	F7
F8	F8
F9	F9
F10	F10
F11	F11
F12	F12
@ # \$ % ^ & +	@ # \$ % ^ & +
- =	- =
bksp	bksp
tab	€
0	ا
1	ی
2	ء
3	ں
4	ل
5	ا
6	ث
7	>
8	<
9	گ
0	گ
1	?
2	;
3	:
4	!
5	د
6	ڈ
7	ڑ
8	,
9	ؤ
0	ؤ
Windows	Windows
alt	alt
Windows	Windows
menu	menu

Comparison between Everson / Liwal Keyboard layouts

- The keyboard layout by Michael Everson and his associates (Roozbeh Pournader) is the layout that was recognized and attested by Ministry of Communications in Afghanistan.
- Liwal keyboard layout is therefore not compatible with the Everson and associates keyboard layout. And also it is not based on scientific bases.



Liwal and (Michael Everson, Roozbeh Pournader, Said Marjan Zazai, & Mohammad Tamim Noori) Keyboard Layout comparison

Fig 16

The yellow highlighted characters show the difference of Liwal keyboard layout with Everson keyboard layout.

Conclusion:

- The keyboard layout by Michael Everson and his counterparts (Roozbeh Pournader), Liwal, Zazai, Sherzad, and others – are based on Arabic and Persian keyboard, keyboards are developed without any research or study of the Pashto characters frequencies which is a fundamental step towards the development of modern keyboard layout.

The first part of the frequency analyzer:

The first part of the program is used to analyze the Pashto characters frequency. This program reads Pashto text files and decomposes the Text into characters. It then checks each character repetition and counts them. The output is then shown in a text box which generates an html file for the users and a text file as an input for another program to be used in the second phase.

Language and Tools Used

The program was developed using VB.Net 2005 Beta version, but due to some problems with the beta version VB.Net 2003 was used. No code has been written for error handling. Users need to be careful while using the program or might end up with errors. This Program only deals with analyzing the text files.

Using the analyzer

Pashto text files are needed as an input for this program. Once the input is provided it starts analyzing it, it might take from 5 seconds up to 3 minutes or more depending on the system speed and the text files size.

Once the program finishes analyzing, two files are generated as outputs (one html file for the user to read and one text file for another program for further processing).

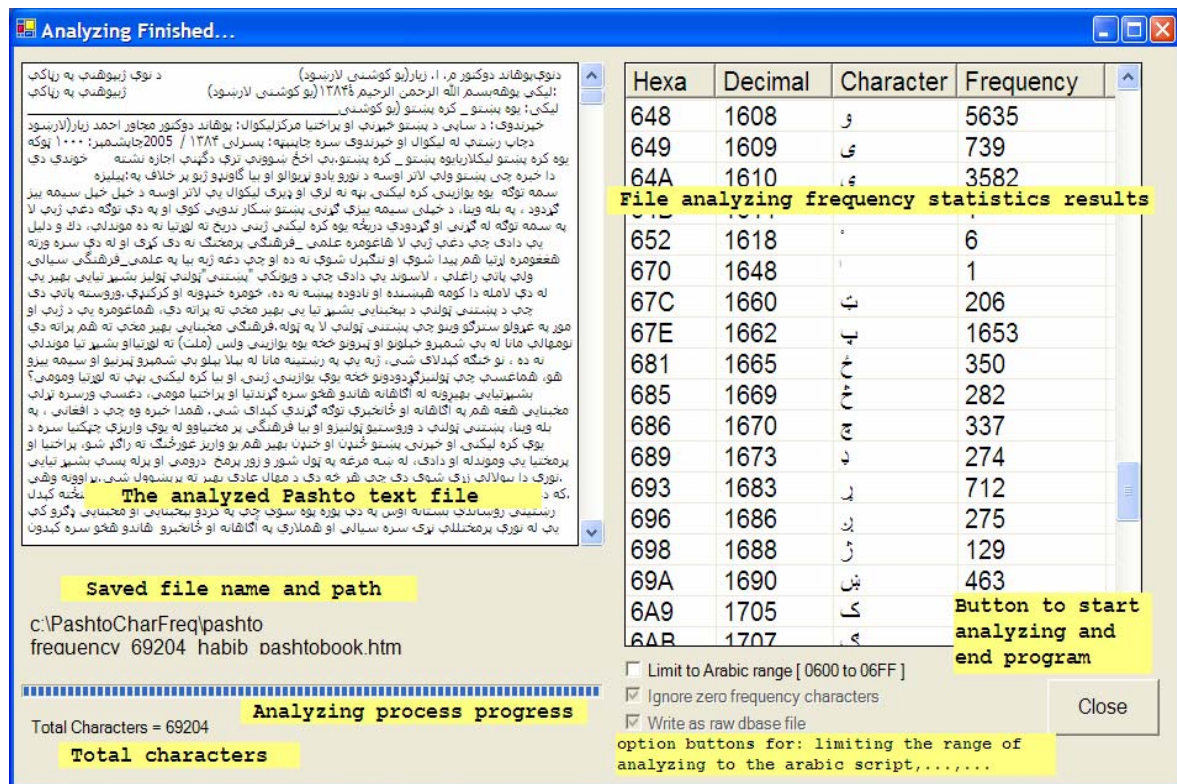


Fig 17: Interface of the analyzer.

The Researcher Name...

Please Enter your name:

The first inputbox for entering the researcher name

OK

Cancel

Fig 18: Input box for researcher name

The Source of the Search

Please Enter the Pashto source name:

The second inputbox for entering the source of the search (book or article name)

OK

Cancel

Fig 19: Input box for research source

The output file: (if you are reading this document in .doc format you can double click to open this attachment)

1 -



pashto frequency_69204_habib_pashtobook.htm
 Type: HTML Document ,Date Modified: 11/17
 Size: 8.75 KB

pashto frequency_69204_habib_pashtobook.htm

2 -



pashto frequency_69204_habib_pashtobook.txt
 Type: Text Document
 Size: 1.10 KB

pashto frequency_69204_habib_pashtobook.txt

Testing and Verification

A basic test was conducted in order to verify the results of the program with different text and different users. Copies of the software were provided to linguists for the reason. It was found that the results generated were correct.

The input Corpus

We need to have a large amount of digitized text in Pashto that's to be analyzed. Large amount of the text will help identify the correct frequencies of the characters. The aim was to analyze up to 50 MB of Pashto text files, but unfortunately there were not enough reachable formats available.

The e-corpus used for analyzing

Digitized text had to be searched and then to be used as an input. Computer science magazine and other internet resources were researched. Some websites offered Pashto books and articles in html format, which were then converted to text files.

The total amount of text Mr. Habiburahman was able to get was about 12 MB. Many organizations were visited for getting some digitized text only one organization provided 28 MB of Pashto text, but was not suitable for the research because it was in a conversational format, and had a lot of repeated text, which would give wrong statistical results.

As Pashto has some variance in writings, the same as the variance in speaking, due to different dialects of the different tribes and regions, it would have been better to analyze more than 12MB of the text to reflect more accurate results.

Second part of the Frequency analyzer:

The second part of the program is used to collect the analyzed results of the first program and show the overall statistics of all analyzed files in one file.

Languages and Tools Used

This part of the program was also developed using VB.Net 2003, contains no error handling. This program only deals with the collection of the analyzed data.

Using this part of the Analyzer

After analyzing the Pashto text files with the first part of the analyzer, this second part of the analyzer reads and collects the statistics of these files.

This program sums up the character statistics of the files being analyzed. The output of this program is a single html file, which shows the files processed and the total statistics. As many files can be added for analysis using the “ACCUMULATE” button and then “MAKE HTML” button to show the statistical results of the frequencies.

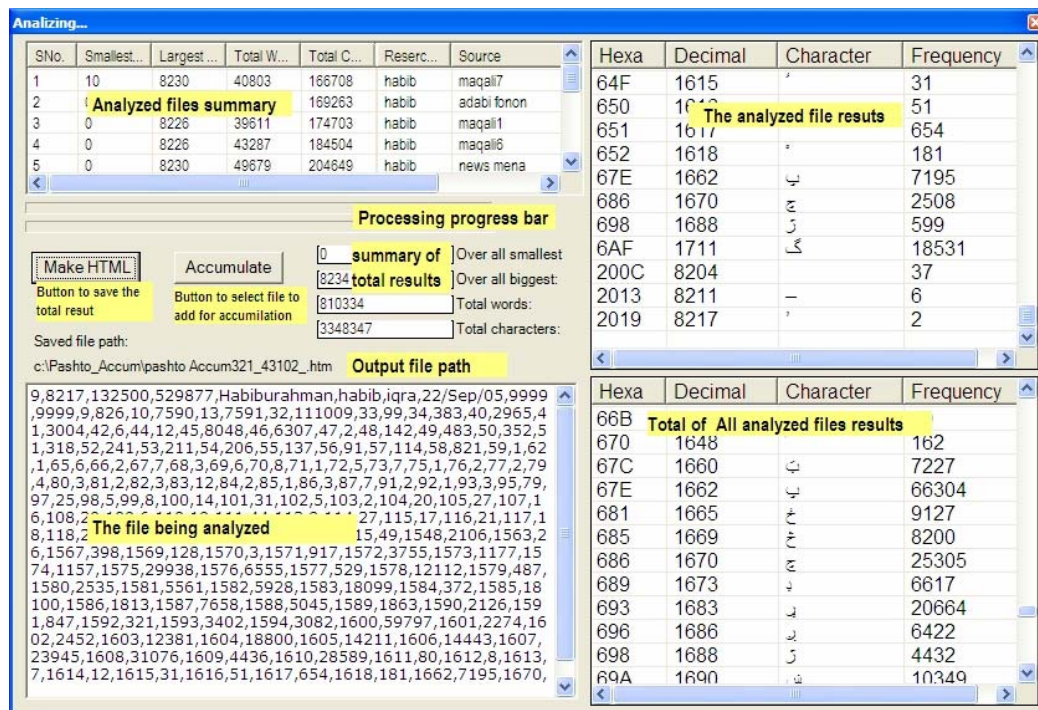
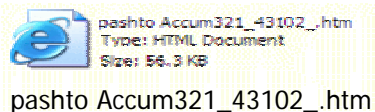


Fig 20: The second part of the analyzing software.

Note: The output file is attached with this report.

The output file: (if you are reading this document in .doc format you can double click to open this attachment)

1 –



The statistical results:**Analyzed output of about 6,000,000 Pashto characters corpus.**

Although the aim was the analysis of characters up to 25,000,000 but due to the lack of digitized Pashto text, we stopped at 6 Million characters; other issues arose as well during this research which are mentioned later in this document.

The analyzing of 12MB of Pashto text resulted in

Total words: 1892096

Total characters: 7769909

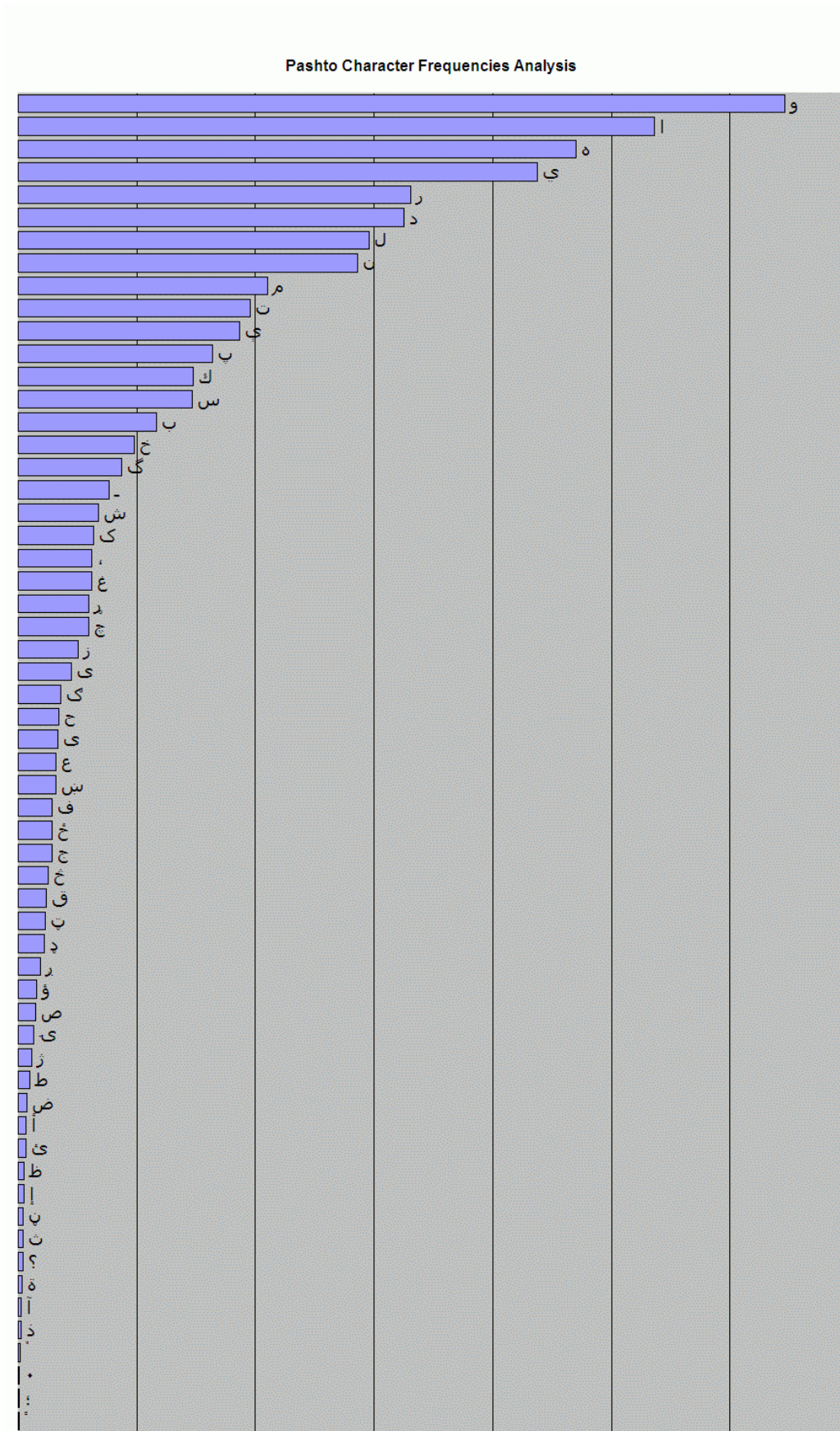
In unrefined details:

Hex	Dec	Char	Freq				
60C	1548	،	61906	651	1617	ٸ	1754
61B	1563	؛	852	652	1618	ٸ	555
61F	1567	؟	3683	654	1620	ٸ	1
621	1569	ء	594	660	1632	ٸ	1004
622	1570	آ	2729	661	1633	ٸ	757
623	1571	أ	6718	662	1634	ٸ	448
624	1572	ؤ	15820	663	1635	ٸ	416
625	1573	!	4562	664	1636	ٸ	297
626	1574	ئ	6391	665	1637	ٸ	248
627	1575	ا	536317	666	1638	ٸ	152
628	1576	ب	116296	667	1639	ٸ	183
629	1577	ة	3065	668	1640	ٸ	259
62A	1578	ث	195259	669	1641	ٸ	314
62B	1579	ث	4067	66A	1642	ٸ	37
62C	1580	ج	28553	66B	1643	ٸ	19
62D	1581	ح	33879	66C	1644	ٸ	1
62E	1582	خ	97524	670	1648	ٸ	166
62F	1583	د	325401	674	1652	ٸ	2
630	1584	ذ	2505	67C	1660	ٸ	23070
631	1585	ر	330636	67E	1662	ٸ	164079
632	1586	ز	50192	681	1665	ٸ	28558
633	1587	س	146314	685	1669	ٸ	25602
634	1588	ش	67426	686	1670	ٸ	59287
635	1589	ص	14265	689	1673	ٸ	21714
636	1590	ض	7487	693	1683	ٸ	59653
637	1591	ط	9664	696	1686	ٸ	18957
638	1592	ظ	4589	698	1688	ٸ	11543
639	1593	ع	32109	69A	1690	ٸ	31596
63A	1594	غ	61868	6A9	1705	ٸ	63158
640	1600	-	76620	6AB	1707	ٸ	36242
641	1601	ف	28905	6AF	1711	ٸ	87108
642	1602	ق	23336	6BC	1724	ٸ	4198
643	1603	ك	147492	6C0	1728	ٸ	174
644	1604	ل	295740	6CC	1740	ٸ	33087
645	1605	م	210099	6CD	1741	ٸ	13206
646	1606	ن	286241	6D0	1744	ٸ	186809
647	1607	ه	470605	6D2	1746	ٸ	7
648	1608	و	646235	6F0	1776	ٸ	502
649	1609	ى	45217	6F1	1777	ٸ	573
64A	1610	ي	437435	6F2	1778	ٸ	355
64B	1611	ء	769	6F3	1779	ٸ	302
64C	1612	ء	54	6F4	1780	ٸ	167
64D	1613	ء	22	6F5	1781	ٸ	294
64E	1614	ء	209	6F6	1782	ٸ	139
64F	1615	ء	474	6F7	1783	ٸ	151
650	1616	ء	152	6F8	1784	ٸ	170
				6F9	1785	ٸ	265

✂ Table. 1: results of the analysis before further processing.

Frequency analysis of the characters in graphical format:

Fig 21: Pashto characters analysis in graphical format.



Note: The excel file of this document is attached with this report.



Fig 21: Pashto characters analysis in graphical format (Continued)

Note: The excel file of this document is attached with this report.

The errors:

After analyzing the results, the following errors were found:

- The Yeh 7 shapes and wrong usages.
- The Kaf 2 shapes and wrong usages.
- The Gaf 2 shapes and wrong usages.
- The numbers are used in 2 formats.

For more details see the chart:

- Blue** color to show the YEH's errors
- Orange** color to show the KAFs errors
- Yellow** color to show the GAFs errors
- Green** color to show the Number errors

✦ **Table 2: Errors found**

Hex	Decimal	Character	Frequency	Hex	Decimal	Character	Frequency
648	1608	و	646,235	6CC	1740	ی	33,087
627	1575	ا	536,317	639	1593	ع	32,109
647	1607	ه	470,605	69A	1690	ین	31,596
64A	1610	ی	437,435	641	1601	ف	28,905
631	1585	ر	330,636	681	1665	خ	28,558
62F	1583	د	325,401	62C	1580	ج	28,553
644	1604	ل	295,740	685	1669	ش	25,602
646	1606	ن	286,241	642	1602	ق	23,336
645	1605	م	210,099	67C	1660	ت	23,070
62A	1578	ت	195,259	689	1673	د	21,714
6D0	1744	ی	186,809	696	1686	ر	18,957
67E	1662	پ	164,079	624	1572	و	15,820
643	1603	ك	147,492	635	1589	ص	14,265
633	1587	س	146,314	6CD	1741	ی	13,206
628	1576	ب	116,296	698	1688	ژ	11,543
62E	1582	خ	97,524	637	1591	ط	9,664
6AF	1711	گ	87,108	636	1590	ض	7,487
640	1600	-	76,620	623	1571	أ	6,718
634	1588	ش	67,426	626	1574	ئ	6,391
6A9	1705	ک	63,158	638	1592	ظ	4,589
60C	1548	،	61,906	625	1573	!	4,562
63A	1594	غ	61,868	6BC	1724	ن	4,198
693	1683	ر	59,653	62B	1579	ث	4,067
686	1670	چ	59,287	61F	1567	؟	3,683
632	1586	ز	50,192	629	1577	ة	3,065
649	1609	ی	45,217	622	1570	آ	2,729
6AB	1707	ک	36,242	630	1584	ذ	2,505
62D	1581	ح	33,879	651	1617	،	1,754
				660	1632	،	1,004

61B	1563	٤	852	66A	1642	٪	37
64B	1611	٥	769	64D	1613	٨	22
661	1633	١	757	66B	1643	٩	19
621	1569	٦	594	6D2	1746	٤	7
6F1	1777	١	573	674	1652	٧	2
652	1618	٠	555	654	1620	٧	1
6F0	1776	٠	502	66C	1644	٤	1
64F	1615	٧	474				
662	1634	٢	448				
663	1635	٣	416				
6F2	1778	٢	355				
669	1641	٩	314				
6F3	1779	٣	302				
664	1636	٤	297				
6F5	1781	٥	294				
6F9	1785	٩	265				
668	1640	٨	259				
665	1637	٥	248				
64E	1614	٧	209				
667	1639	٧	183				
6C0	1728	٥	174				
6F8	1784	٨	170				
6F4	1780	٢	167				
670	1648	١	166				
650	1616	٧	152				
666	1638	٦	152				
6F7	1783	٧	151				
6F6	1782	٩	139				
64C	1612	٧	54				

Second phase of Pashto keyboard research and development

It is known that the typist speed and accuracy is affected by the distribution of Characters over a keyboard. If the most used characters are scattered away the typist speed will be lower, and there will be more stress on the typist nerves and fingers.

In this phase we are going to make a new Pashto keyboard layout design. This design will be based on the first phase research output by knowing Pashto characters frequencies. Because every finger has specific strength and accessibility, for example the index finger is more prominent in accessibility and strength so it should have the highest frequency letters. The keyboard rows as well vary in accessibility. The middle row is the most accessible. The higher frequency letters will be placed in the base row/middle row/home row. This will result in less stress and less time consumption.

This new keyboard layout is cold tested for efficiency using the frequency of the characters and the movement of the fingers. The statistical calculation of the time consumed is attached here with this report. For the reason of practical testing, the keyboard is published and is tested through new keyboard learners in order to know the efficiency and learning curve (time and speed) of the new layout.

Target Keyboard Layout provision:

- To be easier to learn for beginners.
- Produce a higher typing speed for a typical computer user.
- Usage of Agile Inner Fingers (forefinger and middle finger).
- Reduction of typing errors.
- Reduction in Finger Reach and stress, minimize the typing effort.
- Less fingers travel.

Previous known keyboard layouts:

Dvorak keyboard ✂

The Dvorak Simplified Keyboard is a keyboard layout patented in 1936 by Dr. August Dvorak, [an educational psychologist and professor of education at the University of Washington in Seattle, Washington], and William Dealey as an alternative to the more common QWERTY layout. This keyboard was a result of significant ergonomic research and is known to outperform the standard QWERTY keyboard.

- August Dvorak 1936. Increased accuracy in typing by almost 50 % and speed by 15-20 %
- Fingers stay on the home row 70% of the time. The world record speed on Dvorak is 225 wpm.
- Dvorak estimated that the fingers of an average typist in his day travelled between 12 and 20 miles on a QWERTY keyboard; the same text on a Dvorak keyboard would require only about one mile of travel.
- Dvorak believed that 'hurdling' and awkward keystroke combinations were responsible for most common typing errors.



Dvorak Keyboard Layout

Fig 22

The Dvorak layout was designed to address the problems of inefficiency and fatigue which characterized the QWERTY keyboard layout. The QWERTY layout was introduced in the 1860s, being used on the first commercially-successful typewriter, the machine invented by Christopher Sholes. The QWERTY layout was designed so that successive keystrokes would alternate between sides of the keyboard so as to avoid jams. Some sources also claim that the QWERTY layout was designed to slow down typing speed to further reduce jamming.



Mrs. Barbara Blackburn of Salem, Oregon can maintain 150 wpm for 50 min (37,500 key strokes) and attains a speed of 170 wpm using the Dvorak Simplified Keyboard (DSK) system. Her top speed was recorded at 212 wpm. Source: Norris McWhirter, ed. (1985), *THE GUINNESS BOOK OF WORLD RECORDS*, 23rd US edition, New York: Sterling Publishing Co., Inc. ❀

Fig 23: Mrs. Barbara Blackburn, the World's Fastest Typist

QWERTY keyboard



QWERTY (pronounced / kwerti/) is the most common modern-day keyboard layout on English-language computer and typewriter keyboards. It takes its name from the first six letters seen in the keyboard's top first row of letters. The QWERTY design was patented by Christopher Sholes in 1868 and sold to Remington in 1873, when it first appeared in typewriters. It was designed to "slow down" typing, to prevent the types from jamming. The QWERTY keyboard is also a commonly used nickname to name the English language keyboard.



Fig 24: Sholes-Glidden typewriter



Fig 25: QWERTY keyboard layout

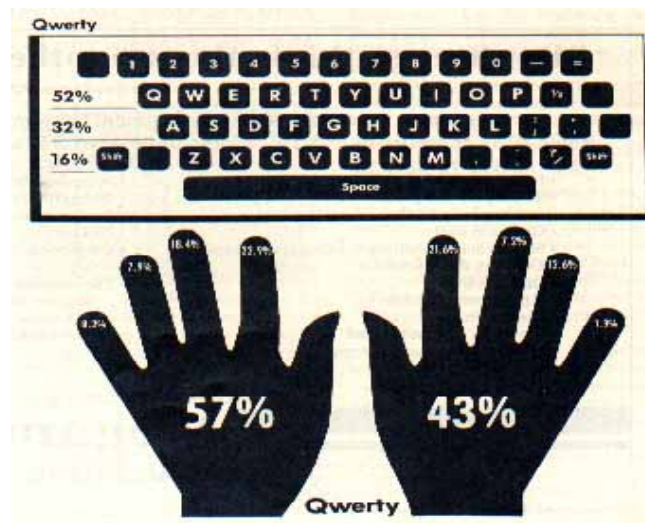


Fig 26: QWERTY keyboard layout efficiency

Dvorak				Qwerty			
Distance:	10.90m	Number row:	0.0%	Distance:	17.80m	Number row:	0.0%
Same hand:	19.83%	Top row:	26.88%	Same hand:	41.47%	Top row:	51.80%
Same finger:	2.131%	Home row:	68.19%	Same finger:	7.540%	Home row:	31.63%
Total keys:	708	Bottom row:	4.918%	Total keys:	708	Bottom row:	16.55%
Finger:	8% 9% 11% 16% --	14% 14% 13% 10%		Finger:	8% 6% 18% 21% --	16% 10% 12% 4%	

Fig 27: Comparison between Dvorak and Qwerty layouts.

English alphabets frequency:

Letter	Frequency
A	%8.167
B	%1.492
C	%2.782
D	%4.253
E	%12.702
F	%2.228
G	%2.015
H	%6.094
I	%6.966
J	%0.153
K	%0.772
L	%4.025
M	%2.406
N	%6.749
O	%7.507
P	%1.929
Q	%0.095
R	%5.987
S	%6.327
T	%9.056
U	%2.758
V	%0.978
W	%2.360
X	%0.150
Y	%1.974
Z	%0.074

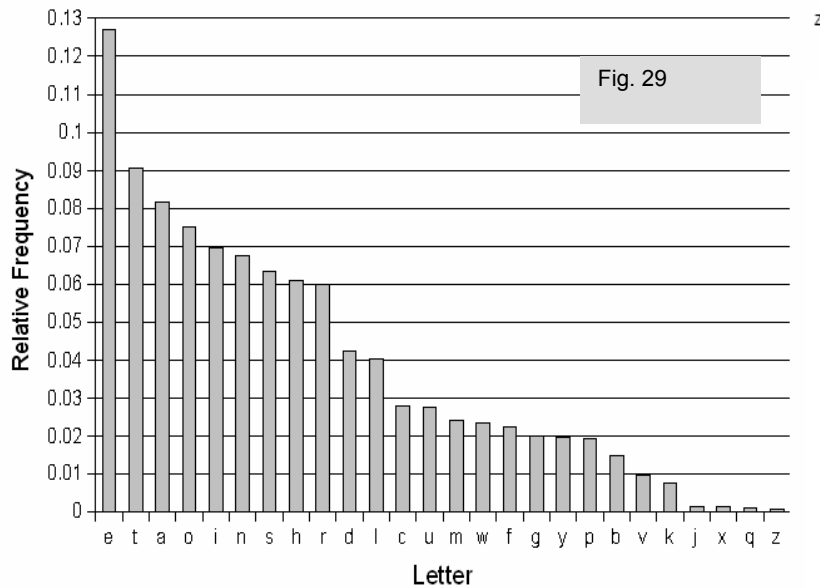
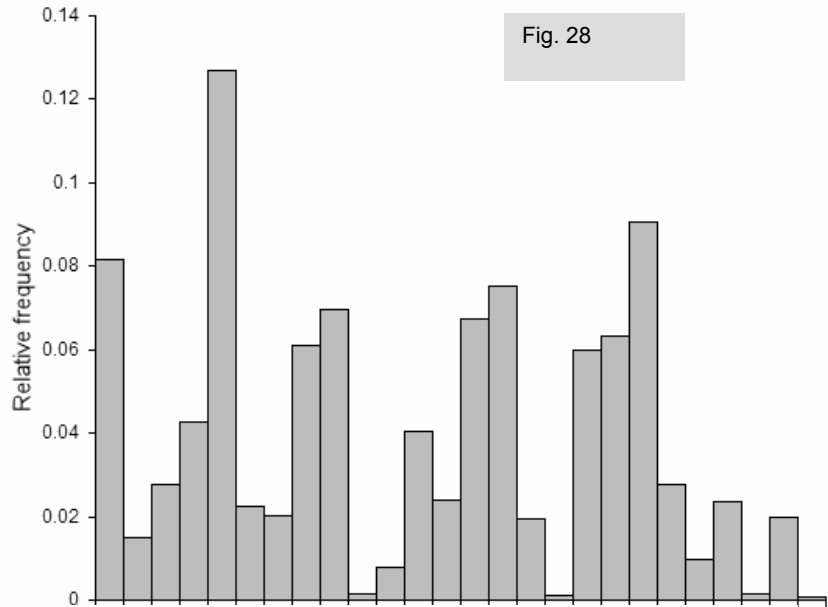


Table 3: English alphabets frequencies

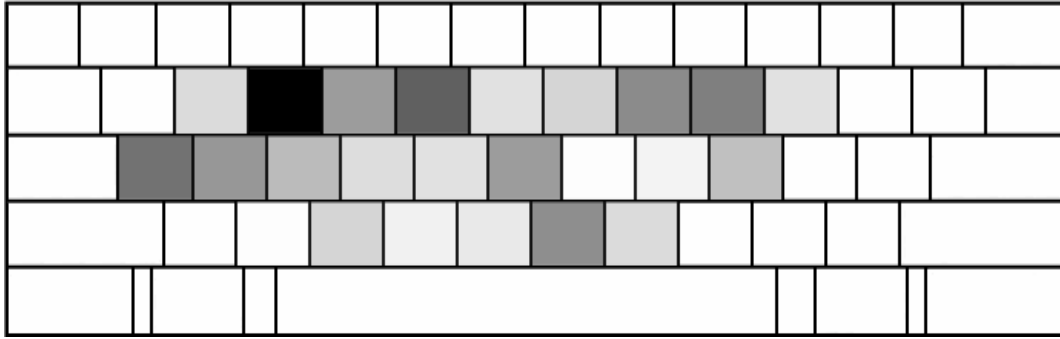
QWERTY Keyboard Layout (Characters Frequency)

Fig 30: The darker colors reflect higher frequencies. ✂1

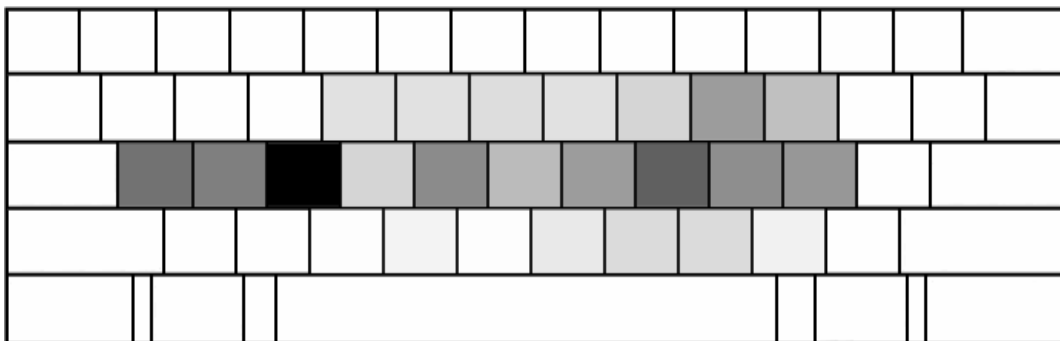
DVORAK Keyboard Layout (Characters Frequency)

Fig 31: the darker colors reflect higher frequencies. ✂1

The above images show where the most frequently used keys lie on different keyboard layouts.

What's wrong with the QWERTY layout? ✂2

- It places very rare letters in the best positions, so your fingers have to move a lot more .
- It suffers from a high same finger ratio that slows down typing and increases strain .
- It allows for very long sequences of letters with the same hand (e.g. "sweaterdresses").
- It was designed to prevent the keys from sticking, without any consideration to ergonomic or efficiency aspects .
- It was designed so the word "typewriter" could be typed on the top row to ease demonstrations .
- It suffers from an extremely high ratio of home-row-jumping sequences e.g. ("minimum")
- QWERTY is very boring to learn because very few meaningful words can be formed with the keys on home row. Thus typing tutors typically have students typing nonsense for the first several lessons.

✂1 Source (<http://forum.colemak.com/index.php> - <http://web.syr.edu/~rcranger/dvorak/narativ4.html>)

✂2 Source (<http://forum.colemak.com/index.php>)

What's wrong with the Dvorak layout? ✂

- The main problem with Dvorak is that it's too difficult and frustrating to learn for existing QWERTY typists because it's so different from QWERTY.
- It is based on English as spoken in 1934.
- A new typist may take several months to really become fluent on Dvorak layout.

Pashto Typing:

Typing in Pashto in contrast to English doesn't have a long history. Still foundations are being laid down and is the right time to take into consideration scientific and logical measures for any steps to be taken in the development of techniques and technologies in Pashto Language.

So we are in need of developing a keyboard based on characters' frequencies. Following is the calculation of Pashto characters frequency and frequency based analysis of previous keyboards.

The different colors used in the frequency tables and in the Keyboard layouts are clarified in the file attached with the report.

Pashto characters frequency:

Frequency	Character	color	Decimal	Hexa
646,235	و		1608	648
536,317	ا		1575	627
470,605	ه		1607	647
437,435	ي		1610	64A
330,636	ر		1585	631
325,401	د		1583	62F
295,740	ل		1604	644
286,241	ن		1606	646
210,650	ک		1603	643
210,099	م		1605	645
195,259	ت		1578	62A
186,809	ی		1744	6D0
164,079	پ		1662	67E
146,314	ځ		1587	633
123,350	ګ		1711	6AF
116,296	ب		1576	628
97,524	ښ		1582	62E
78,304	ی		1609	649
76,620	ـ		1600	640
67,426	ښ		1588	634
61,906	،		1548	60C
61,868	ښخ		1594	63A
59,653	ر		1683	693
59,287	چ		1670	686
50,192	ز		1586	632
33,879	ح		1581	62D

32,109	ع		1593	639
31,596	ښ		1690	69A
28,905	ف		1601	641
28,558	خ		1665	681
28,553	ج		1580	62C
25,602	څ		1669	685
23,336	ق		1602	642
23,070	ښ		1660	67C
21,714	د		1673	689
18,957	ږ		1686	696
15,820	و		1572	624
14,265	ص		1589	635
13,206	ی		1741	6CD
11,543	ژ		1688	698
9,664	ط		1591	637
7,487	ض		1590	636
6,718	ا		1571	623
6,391	ی		1574	626
4,589	ظ		1592	638
4,562	!		1573	625
4,198	ن		1724	6BC
4,067	ث		1579	62B
3,683	؟		1567	61F
3,065	ه		1577	629
2,729	آ		1570	622
2,505	ذ		1584	630

Table: 4 Pashto characters frequencies

Pashto characters frequency

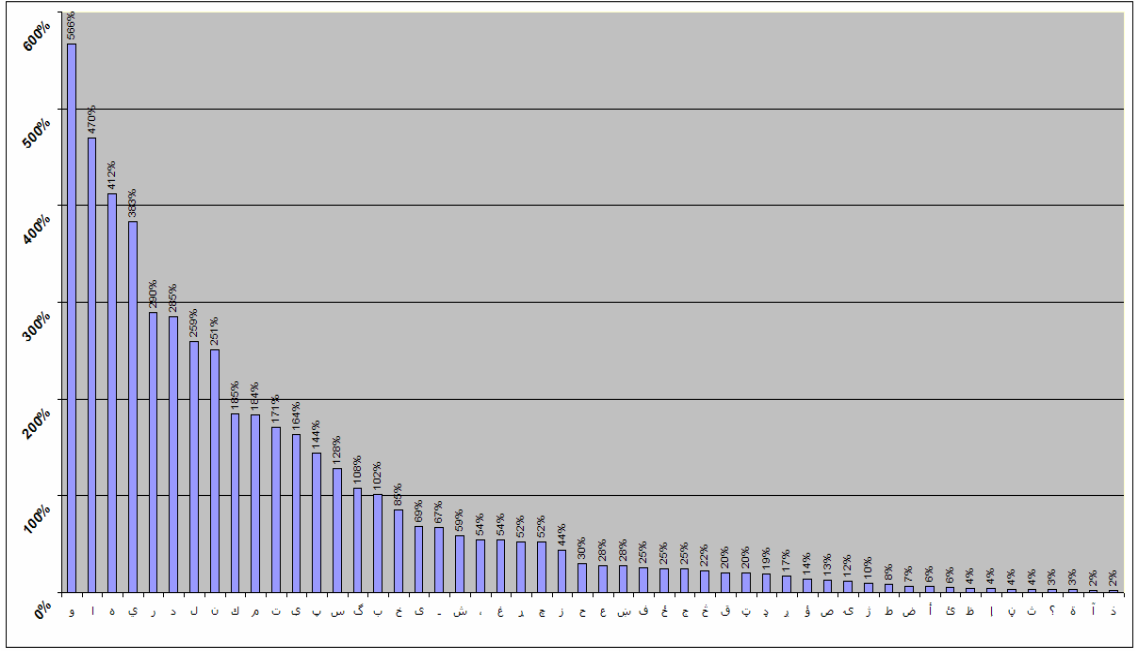


Fig 32: The analysis reflection in graphical format.

Analyzed corpus summary:

Over all Total words: 1892096

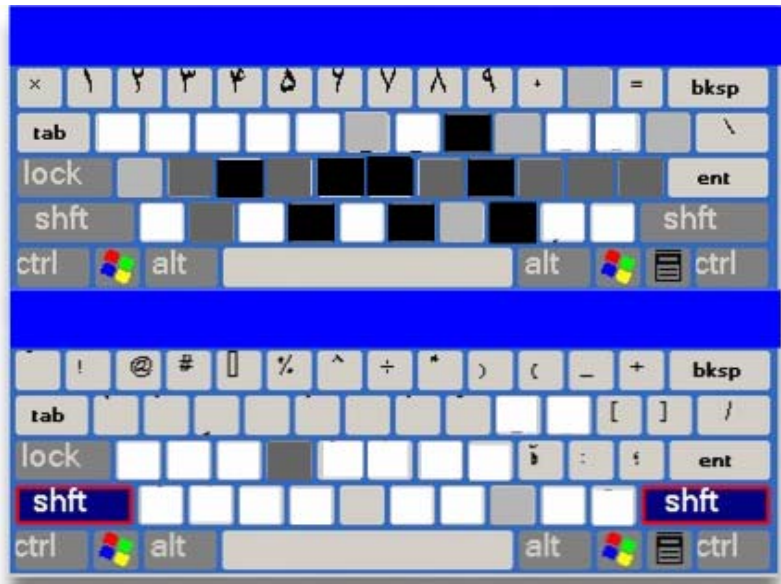
Over all Total characters: 7769909

Famous keyboard layouts design analyzing:

(Everson, Liwal, Tolafghan) keyboard layouts have the same position for the top 8 Pashto characters:

(و ا ه ي ر د ل ن)

For the typing machine the top used characters are more scattered to keep typing slow and prevent typing heads from jamming



د افغانستان دمخابراتو وزارت له خوا منل شوی د ټول افغان کيښورچ

Fig 33: Most frequently used keys on TolAfghan keyboard layout:

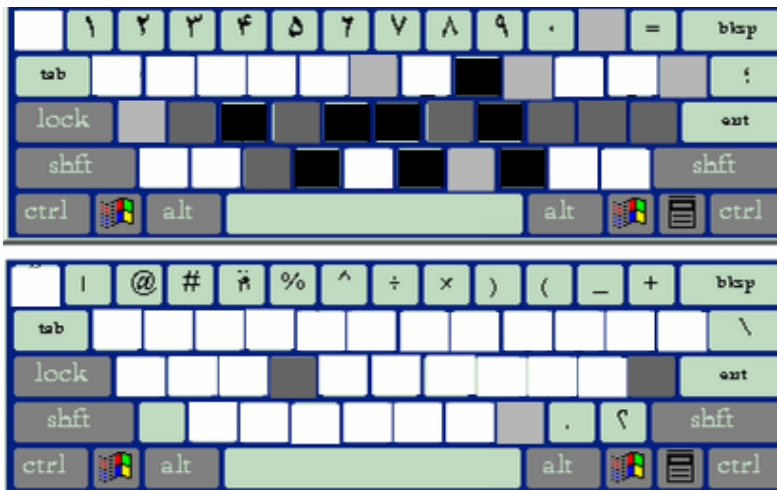


Fig 34: Most frequently used keys on Liwal keyboard layout

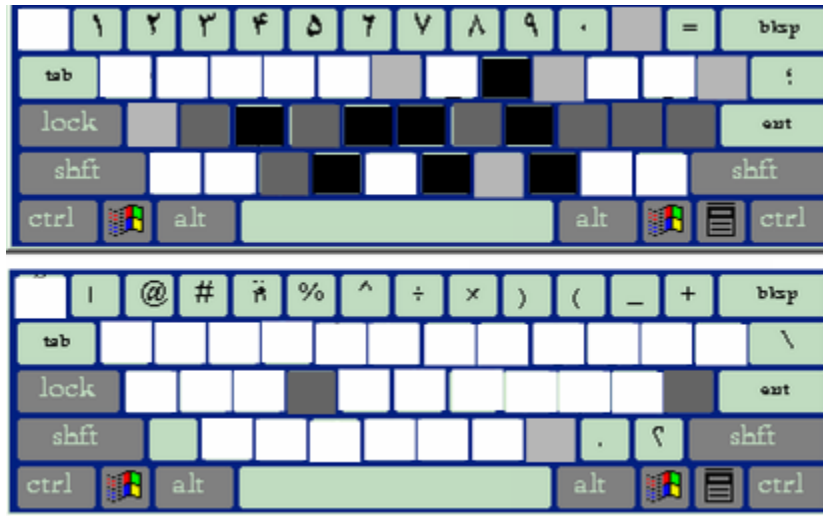


Fig 35: Most frequently used keys on Everson keyboard layout

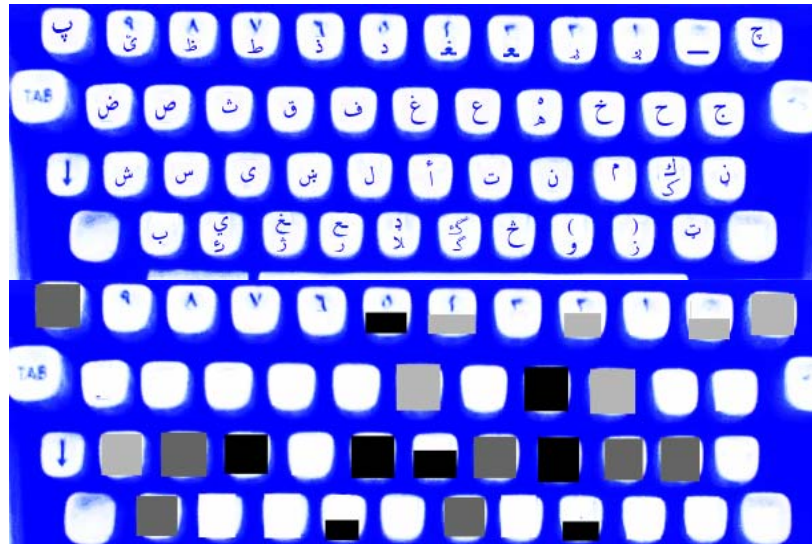


Fig 36: Most frequently used keys on Old Pashto Typewriter keyboard layout

The new keyboard layout development

The following two keyboard layouts were proposed based on the frequencies of the characters, the colors specify different frequencies of the characters.

Note: See the attachment for frequency color specifications.

Proposed Pashto keyboard layout without considering the old layout



Fig 37

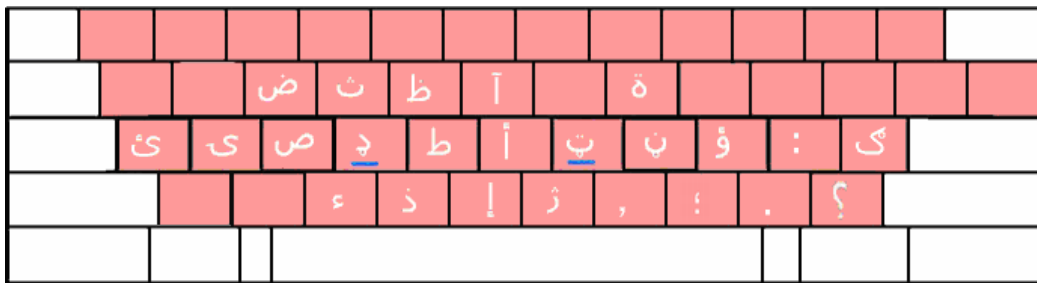


Fig 38

Proposed Pashto keyboard layout considering the old layout



Fig 39



Fig. 40

The cold test conducted for the new keyboard layout bear the following results:

Sample text analysis for new keyboard layout efficiency

Pashto text source: "adabi fonon" ادبي فنون

Author: ع . بينوا

Total words= 42848

Total characters= 169263

Source: http://library.tolafghan.com/adabi_funoon/larlik.shtml

Frequency Analyzer software made by Habiburrahman.

Movement's efficiency:

[new layout: 45,655]

New keyboard layout time estimation:

In Minutes 354.82

In Hours 5.91

Movement's efficiency:

[old layout: 102,348]

Old keyboard layout time estimation: based on 40 words per minute

In Minutes 795.43

In Hours 13.26

Based on the background research, knowledge, the already proposed layout and the scientific inferences a new keyboard layout was designed using MSKLC (Microsoft Keyboard Layout Creator) version 1.3.4.022.



Fig. 41.1 the new keyboard layout in normal status



Fig. 41.2 the new keyboard layout in shift status

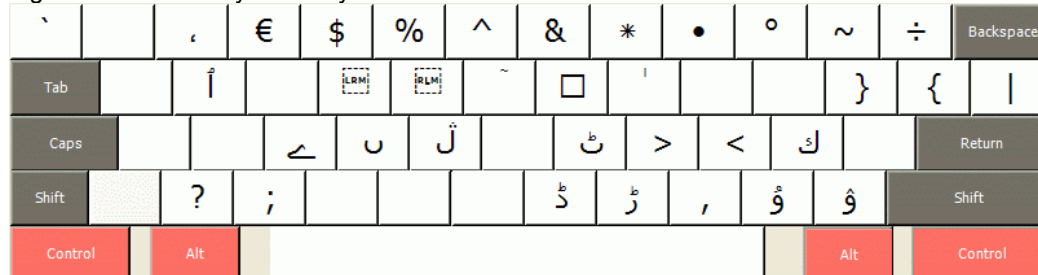


Fig. 41.3 the new keyboard layout alt-ctrl status

Key assignment for Pashto, using ISO/IEC 9995 notations.

	Unshifted	Shifted	AltGr
TLDE	0027	0654	0060
E01	06F1	0021	
E02	06F2	0040	066c
E03	06F3	0023	20ac
E04	06F4	060b	0024
E05	06F5	066a	0025
E06	06F6	00D7	005E
E07	06F7	00BB	0026
E08	06F8	00AB	066D
E09	06F9	0029	2022
E10	06F0	0028	00B0
E11	0640	005F	007E
E12	003D	002B	00F7
D01	0634	0652	<FREE>
D02	0649	064C	0671
D03	06D0	0636	<FREE>
D04	067E	062B	200E
D05	002D	0638	200F
D06	063A	0622	0653
D07	062A		0659
D08	0633	0629	0670
D09	0628		<FREE>
D10	062E	0022	<FREE>
D11	0686	005D	007D
D12	0693	005B	007B
C01	0645	0626	<FREE>
C02	062F	06CD	<FREE>
C03	064A	0635	06D2
C04	0647	0689	06BA
C05	0644	0637	06B7
C06	0646	0623	
C07	0648	067C	0679
C08	0627	06BC	003E
C09	0631	0624	003C
C10	06A9	003A	0643
C11	06AF	06AB	
B01	0696	200D	003F
B02	0642	200C	003B
B03	0641	0621	<FREE>
B04	0639	0630	<FREE>
B05	069A	0625	<FREE>

B06	0632	0698	0688
B07	062D	066B	0691
B08	0681	061B	002C
B09	062C	002E	06C7
B10	0685	061F	06C9
BKSL	060C	002A	007C
SPCE	0020	00A0	<FREE>

Acknowledgments:

The work of this research report has been extensively carried by Habiburahman Najiullah. It's due to all his endeavors that this work has been possible. Other colleagues and peers contributed to the very foundation of this research. Omar Mansoor Ansari the then leader of the project was highly effective at initiating the work successfully. In addition we would like to extend our gratitude and thankfulness to the following for their cooperation and support.

- Majeedullah Qarar
- Pan localization
- Ustaz Mohammad Asif Samim
- Sharifullah Mahboob
- Da ulomo academy (academy of sciences Afghanistan)
- Haron Wardak for his help in searching and finding some keyboard layouts.

Hameedullah Sherani
Country Project Leader
PANL
MoCIT, Kabul, Afghanistan.