Hittite Journal of Science and Engineering, 2016, 3 (1) 23-28
ISSN NUMBER: 2149-2123
DOI: 10.17350/HJSE19030000028

Optimized Rearrangements of Turkish Q and F Keyboards by Means of Language-Statistics and Simple Heuristics

Asım Egemen Yılmaz and Emrah Çiçek
Ankara University, Department of Electrical and Electronics Engineering, Ankara, TURKEY

ABSTRACT

In this study, based on the language (particularly n-gram) statistics of Turkish extracted from a large corpus of meaningful written text of various categories, we try to propose some improvements, namely optimized rearrangements for the Turkish Q and F keyboards via some simple rules and heuristics. Our proposals result in more desirable and efficient keyboard layouts that increase the comfort and the speed of professional typists. The methods and procedures followed throughout this study can be extended and applied for any keyboard of other alphabets and languages.

Article History:
Received: 2016/02/09
Accepted: 2016/06/08
Online: 2016/06/30

***Correspondence to**: Asım Egemen Yılmaz, Ankara University, Department of Electrical-Electronics Engineering, Gölbaşı, Ankara, TURKEY*
Tel: +90 (312) 203 33 00/17 75
Fax: +90 (312) 212 54 80
E-Mail: aeyilmaz@eng.ankara.edu.tr

Key Words:
Language statistics; N-gram Statistics; Keyboard Design; Keyboard Layout; Optimization; Heuristics.

INTRODUCTION

The well-known Q (also known as QWERTY) keyboard has been the de-facto standard of typing machines (typewriters, computers) in majority of the world for more than a century. In the countries applying the Latin alphabet, this keyboard layout is used with minor modifications regarding the local character sets. Even though there exists no published evidence, there is a strong belief that the abnormal layout of the Q keyboard was intentionally preferred in order to decrease the speed of the typists and hence to avoid mechanical jams on the early typewriters. As a matter of fact, the key arrangement is so awkward that very infrequent letters "f" and "j" occupy the default positions of index fingers, namely the most valuable locations on the keyboard. Very frequent letters "a" and "e" seem to be hidden at locations accessible by the little and the middle fingers of the left hand, respectively [1]. Since the era of mechanical typewriters closed long ago, and there are currently no mechanical hurdles, more efficient keyboard designs are now realizable.

With this motivation, in this study, we investigate the possibilities of achievement of more efficient keyboard designs. For this purpose, in Section 2 we first try to formulate the design requirements for a keyboard, which are derived from ergonomic constraints. Then in Section 3, we try to come up with suggestions for improvements in the Turkish F keyboard, which is the unarguable choice of professional typists in Turkey. In Section 4, we apply similar methodology for improvement to the more widely used Turkish Q keyboard. Discussions about the limitations and possible extensions of the current study together with the concluding remarks constitute the content of Section 5.

General requirements for keyboard design

Elimination of the weakness of the Q keyboard has been the concern of many researchers up to now. One of the most significant efforts was by Dvorak [2], who designed the so-called "American Simplified Keyboard" in 1932. The layout was patented in 1936 to the names of Dvorak and his brother-in-law Dealey [3]. Dvorak layout was aimed for professional typing in which 10 fingers are effectively used. For such usage, the areas of responsibilities for the hands and all fingers are as seen in Figure 1. As seen from the figure, the role and responsibility of each hand and each finger is strictly defined throughout the so-called "10-finger typing" technique. At this point, it is worth to state that despite the very common "10-finger typing" terminology, only 8 fingers are used effectively throughout this technique since both

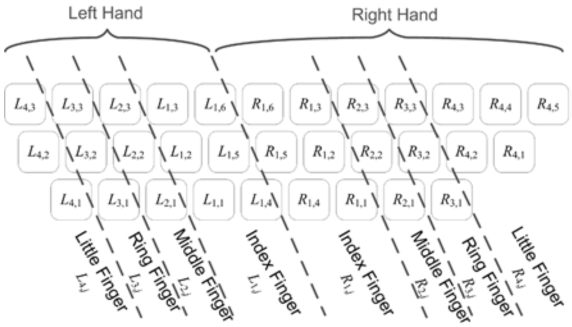

Figure 1. Figure 1. Finger placement and effort distribution for professional "10-finger typing" typing (which utilizes effective usage of only 8 fingers, since the thumbs are only used for pressing the space bar).

thumbs are used only for pressing the space bar.

Dvorak layout causes less finger motion, increases the typing speed and reduces the typing errors compared to the standard Q keyboard. Studying the letter frequencies in English and the human physiology (particularly the hands), Dvorak came up with the following principles for an ideal keyboard design [3]:

- There should be a balanced distribution between two hands (making the typing activity more rhythmic, increasing speed, reducing errors and fatigue).
- For maximum speed and efficiency, the most common letters and bigrams should be the easiest to type. This means that they should be on the home (i.e. the middle) row, which is where the fingers rest, and under the strongest fingers (As a matter of fact, about 70% of keyboard strokes on the Dvorak Simplified Keyboard are performed on the middle row).
- The least common letters should be on the bottom row, which is the hardest row to reach.
- The right hand should perform (slightly) more of the typing, because most people are right-handed.
- Bigrams should not be typed with adjacent fingers.
- Stroking should generally move from the edges of the board to the middle. An observation of this principle is that, for many people, when tapping fingers on a table, it is easier going from little finger to index than vice versa. This motion on a keyboard is called inboard stroke flow.

Meanwhile, Dvorak pointed out the major problems of the Q keyboard as follows (where the numerical figures of merit were not identified by Dvorak himself but some other researchers much later) [2]:

- Many common letter combinations require awkward finger motions.
- Many common letter combinations are typed with the same finger.
- Many common letter combinations require a finger to jump over the home row.
- Many common letter combinations are typed with one hand while the other stays in idle mode.
- Many common letter combinations are typed by adjacent fingers, which is slower than using other fingers.
- For an English text, about 56% of typing is performed with the left hand, which is the weaker hand for most people.
- For an English text, about 30% of typing is performed on the lower row, which is the slowest and most difficult row to reach.
- For an English text, about 52% of keyboard strokes are performed in the top row, requiring the fingers to travel away from the home row most of the time.

Table 1 summarizes the improvement of Dvorak layout compared to the standard Q keyboard (for English). It should also be noted that the Q keyboard causes an effort load of 56% for the left hand; whereas, Dvorak layout assigns 56% effort load to the right hand, which is stronger for majority of the people.

Table 1. Key stroke distributions for the Q and Dvorak simplified keyboards (per statistics of English).

Row	*Original Q Keyboard*	*Dvorak Simplified Keyboard*
Top	*52%*	*22%*
Middle	*32%*	*70%*
Bottom	*16%*	*8%*

As a matter of fact, trying to achieve more ergonomic key stroke distributions (considering the keyboard rows and attaining the highest load to the middle row; and also considering the effort loads of both hands) have so far constituted the main aim of some previous studies. In [4], Wagner et al. applied Ant Colony Optimization for re-arrangement of the standardized keyboards based on English, German and French keyboards, separately. Via simple rules and heuristics, Abbasov et al. applied a similar approach for an ideal keyboard arrangement for Azerbaijani Turkish in [1]; Dasgupta et al. conducted a similar study for Bangla language in [5]. Deshwal and Deb applied Genetic Algorithm for the solution of the same problem for Hindi language in [6]; meanwhile, Malas et al. applied it for Arabic in [7]. With the same motivation, in this study we try to achieve more-optimized rearrangements of Turkish F and Q keyboards. For this purpose, we use the statistics of Turkish and a simple heuristic to be described in the upcoming sections.

In the literature, there are also other sorts of studies regarding keyboard design:

- One of the major topics is achievement of the optimized industrial design of the keyboards with the concerns regarding the human posture and the positioning of wrists, fingers, tendons, etc.
- Another area of interest for many researchers is to achieve optimal arrangements for the so-called "Single Finger Keyboards", which have evolved throughout the development of technology:
- In the last two decade, Single Finger Keyboards were in the form so that multiple characters were assigned to limited number of solid keys. The most common examples are the ones on the traditional phones and old-generation cellular phones (e.g. where the key "2" is attained to the letters "A", "B" and "C"). The main concern was to find the ideal layouts for such keyboards to decrease the single finger (usually assumed to be the thumb of the right hand) in such studies.
- Especially after the development of the smart phones and tablet PCs, these keyboards evolved in a way that there are sufficient number of keys (compared to the number of letters in the alphabet) which are soft and reconfigurable. Again, the main concern in current studies is to find the ideal layouts for such keyboards to decrease the single finger (usually still assumed to be the thumb of the right hand).

Our study does not have any concerns regarding to human posture; it also does not deal with Single Finger Keyboards. Hence it should not be confused with studies of these sorts.

Improvement in Turkish F keyboard

Motivated from the fact that the Q keyboard is not appropriate for the statistics and characteristics of Turkish, and inspired from the methodology of Dvorak, Yener et al. conducted research for achievement of an optimum keyboard layout for Turkish between 1955-66 [8]. This layout, which is referred to as the "F keyboard", has been the cult but unarguable choice of professional typists (or touch-typists, or 10-finger typists) such as secretaries, clerks, etc. The layout has been declared to be a national standard in 1974 [9]. It should also be noted that Marsan has initiated a similar study in France in order to achieve an optimum keyboard layout for French in 1970s, and the outcome of this study (so called "Marsan" keyboard) has been declared as a national standard in 1987 in France.

For the construction of the layout, Yener et al. performed a dictionary-based frequency analysis on 30,000 words; and identified the occurrence frequencies of each letter. Based on these statistics and using the guidelines of Dvorak (mentioned in Section II), they constructed the layout. An additional design decision was to distribute all

Figure 2. Layout of the original F keyboard (letter frequencies obtained from dictionary based analysis - adapted from [9]).

vowels inside the left hand's area of responsibility. Since each syllable in Turkish has to include exactly one vowel, this yields a balanced distribution between the left and the right hands. The F keyboard layout is seen in Figure 2 together with the letter frequencies per dictionary-based analysis.

The F keyboard unarguably fits much better to Turkish compared to the Q keyboard, since:

- it increases the key stroke distributions in the home (middle) row,
- it decreases the responsibilities of weaker fingers (such as the little and ring fingers), and
- it assigns slightly much more responsibility to the right hand.

As a matter of fact, in organizations such as typing contests, F keyboard users have so far outperformed to the Q keyboard users numerous times. However, we identify a major deficiency in the design procedure of the F keyboard: dictionary based frequency analysis. In fact, in order to achieve an optimum keyboard layout, the letter statistics shall be extracted from a sufficiently large corpus of meaningful Turkish texts of various categories rather than a dictionary. The major reasons for this can be listed as follows:

- In daily life, a typist would deal with typing meaningful texts, not the words listed dictionary.
- In daily usage, the frequencies of the words are not identical; and hence considering them as equally-frequent (as in the dictionary based analysis) is misleading.
- Turkish is an agglutinative language, and the majority of the words in meaningful texts are inflected, not in their stem forms as in dictionaries. From this point of view, dictionary based analysis is again misleading.

In [10-11], we have performed such a text based analysis

Table 2. Letter frequencies in Turkish (results obtained via dictionary based analysis [9] and text based analysis [10-11]).

Letter	*Dictionary Based Analysis [9]*		*Text Based Analysis [10-11]*		*Relative Frequency Change (%)*
	Frequency	*Order*	*Frequency*	*Order*	
a	*14.34%*	*1*	*11.46%*	*1*	*20.07%*
b	*2.02%*	*16*	*2.67%*	*15*	*-32.37%*
c	*1.25%*	*22*	*0.92%*	*25*	*26.58%*
ç	*1.15%*	*25*	*1.05%*	*23*	*8.73%*
d	*2.47%*	*13*	*4.60%*	*8*	*-86.39%*
e	*8.88%*	*2*	*9.07%*	*3*	*-2.11%*
f	*1.23%*	*23*	*0.49%*	*28*	*60.03%*
g	*1.12%*	*26*	*1.15%*	*20*	*-2.68%*
ğ	*0.66%*	*27*	*1.05%*	*22*	*-58.40%*
h	*1.95%*	*18*	*1.11%*	*21*	*43.02%*
ı	*4.13%*	*11*	*4.56%*	*9*	*-10.47%*
i	*7.29%*	*4*	*9.32%*	*2*	*-27.85%*
j	*0.07%*	*29*	*0.05%*	*29*	*26.47%*
k	*7.38%*	*3*	*4.65%*	*7*	*36.96%*
l	*5.72%*	*6*	*6.40%*	*6*	*-11.95%*
m	*6.14%*	*5*	*3.51%*	*11*	*42.79%*
n	*4.47%*	*9*	*7.42%*	*4*	*-66.00%*
o	*1.98%*	*17*	*2.58%*	*16*	*-30.30%*
ö	*0.50%*	*28*	*0.77%*	*27*	*-54.31%*
p	*1.30%*	*21*	*0.87%*	*26*	*32.87%*
r	*4.74%*	*8*	*7.04%*	*5*	*-48.59%*
s	*4.32%*	*10*	*3.15%*	*13*	*27.07%*
ş	*1.55%*	*20*	*1.53%*	*18*	*1.48%*
t	*5.27%*	*7*	*3.60%*	*10*	*31.64%*
u	*2.70%*	*12*	*3.14%*	*14*	*-16.43%*
ü	*2.47%*	*14*	*1.92%*	*17*	*22.11%*
v	*1.22%*	*24*	*1.01%*	*24*	*17.08%*
y	*2.06%*	*15*	*3.32%*	*12*	*-61.56%*
z	*1.68%*	*19*	*1.50%*	*19*	*10.66%*

for evaluation of the *n*-gram statistics for Turkish; where the contents and details of the corresponding corpus used throughout the analysis can be found. As seen in Table 2, the letter frequencies are quite different compared to the results obtained via dictionary based analysis. Particularly,

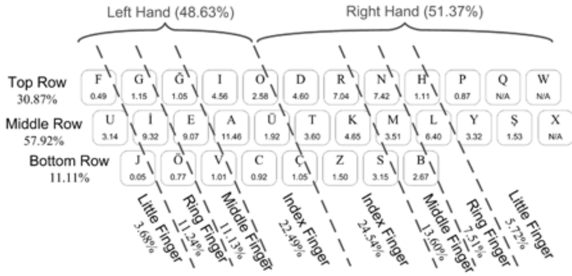

Figure 3. Layout of the original F keyboard (letter frequencies obtained from our text based analysis).

dictionary based analysis show that the most frequent 8 letters are "a, e, k, i, m, l, t, r"; whereas they are "a, i, e, n, r, l, k, d" for the text based analysis. The "relative frequency change" quantity seen in Table 2 is defined as follows:

Relative Frequency Change = 100 × [(Dictionary Based Analysis Frequency – Text Based Analysis Frequency) / Dictionary Based Analysis Frequency]

where negative values express reduction in frequency, and positive values express increasing frequency.

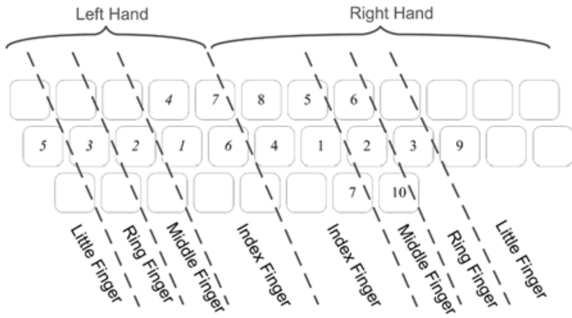

Figure 4. Procedure followed during the design of the original F keyboard: The numbers on the keyboard identify the "value" of the relevant position.

Considering the letter frequencies per text based analysis, the performance of the F keyboard regarding the key stroke distributions seems to be degraded as seen in Figure 3. The usage rate of the top row is in fact 30.87% (not 24.03%); and the usage rate of the home (middle) row is

Table 3. Improvements in hand/finger and row usage rates for the F keyboards (original versus our modified version).

	Original F Keyboard			*Modified F Keyboard*		
Finger	*Left Hand*	*Right Hand*	*Total*	*Left Hand*	*Right Hand*	*Total*
Index	*22.49%*	*24.54%*	*47.03%*	*22.49%*	*25.00%*	*47.49%*
Middle	*11.13%*	*13.60%*	*24.73%*	*11.38%*	*11.73%*	*23.11%*
Ring	*11.24%*	*7.51%*	*18.75%*	*10.99%*	*8.28%*	*19.27%*
Little	*3.68%*	*5.72%*	*9.40%*	*3.68%*	*5.55%*	*9.23%*

Row	*Original F Keyboard*	*Modified F Keyboard*
Top	*30.87%*	*22.67%*
Middle	*57.92%*	*65.76%*
Bottom	*11.12%*	*11.48%*

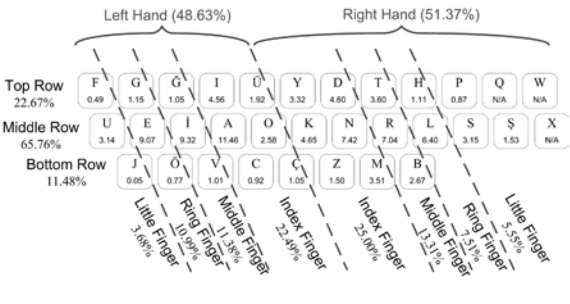

Figure 5. Layout of the modified F keyboard (letter frequencies obtained from text based analysis).

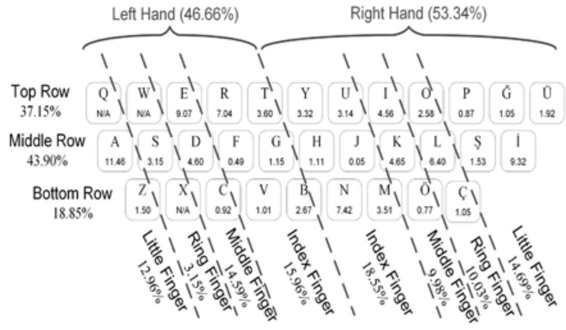

Figure 6. Layout of the original Turkish Q keyboard (letter frequencies obtained from text based analysis).

57.93% (not 63.77%). Hence, the layout shall be reconsidered by using the text based analysis results.

Even though not explicitly published, it can be deduced that Yener et al. followed the procedure pictorially depicted in Figure 4. The numbers on the keyboard in this figure identify the "value" of the relevant position. Namely, the number 1 denotes the most valuable position since it is the anatomically most accessible one. Hence, the more frequent a letter occurs, the more valuable of a position it gets. Inside the left hand's area of responsibility, they have assigned the vowels (starting from the most frequent) to the positions seen in the figure. Similarly, they have assigned the consonants (starting from the most frequent) to the positions seen in the figure inside the right hand's area of responsibility.

We implement a simple heuristic applying the same procedure, but this time relying on the text based statistics. This yields the layout given in Figure 5. With this layout, the home (middle) row usage rate is increased to 65.76%, and the top row usage rate is decreased to 22.67% keeping the bottom row usage rate at almost the same level. This layout also preserves the responsibilities of the left and the right hands and the fingers at reasonable levels. Compared results are summarized in Table 3.

Improvement in Turkish Q keyboard

Even though the Q keyboard is well known to be inefficient, it has dominated the market and economically outperformed all alternatives. Also in Turkey, a slightly modified version of it (by addition of the Turkish special characters) seen in Figure 6 is widely used.

In Germany, in order to increase the efficiency of the Q keyboard, the positions of the letters "z" and "y" was interchanged. In order to distinguish these two layouts, the original Q keyboard is referred to as the QWERTY layout, while the German-modified version of it is referred to as the QWERTZ layout. Inspired from this, we now try to improve the row usage rates and hand/finger responsibilities by interchanging the positions of some particular letters.

As seen in Figure 6, the original Q keyboard layout yields quite high little finger usages (which is the weakest finger), quite low index finger usages (which is the strongest finger); almost same middle and top row usages, and quite high bottom row usage. The following observations can be made:

- "f" and "j" occupy the most valuable positions (middle row and index finger), even though they have quite low frequencies.
- "a" and "i" are at hardly accessible positions even though they have the highest frequencies.
- "e", "r" and "n" have considerable frequencies, but they are not positioned at the middle row.
- "l" has a considerable frequency, but assigned to a relatively weak finger, the ring finger.

Hence for a new keyboard layout, the following letters are positionally interchanged: "a↔f", "i↔j", "e↔d", "n↔h", "r↔g", "l↔k"; which yields the layout seen in Figure 7. By these modifications, the middle row usage rate is raised to 60.57% (from 43.90%), the top and bottom row usages rates are reduced to 26.79% (from 37.15%) and 12.54% (18.85%), respectively. Index finger usage rates are dramatically

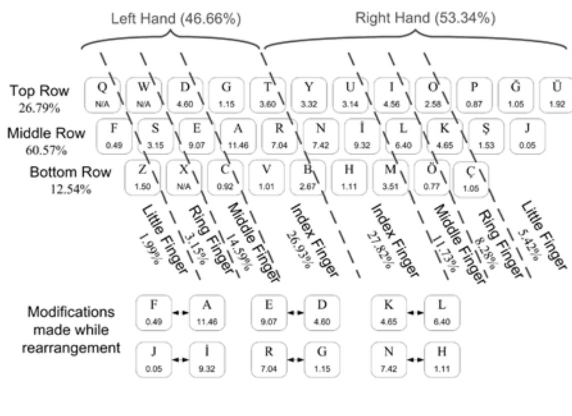

Figure 7. Layout of the original Turkish Q keyboard (letter frequencies obtained from text based analysis).

Table 4. Improvements in hand/finger and row usage rates for the Q keyboards (original versus our modified version)

	Original Turkish Q Keyboard			Modified Turkish Q Keyboard		
Finger	Left Hand	Right Hand	Total	Left Hand	Right Hand	Total
Index	15.96%	18.55%	34.51%	26.93%	27.82%	54.75%
Middle	14.59%	9.98%	24.57%	14.59%	9.98%	24.57%
Ring	3.15%	10.03%	13.18%	3.15%	10.03%	13.18%
Little	12.96%	14.69%	27.65%	1.99%	5.42%	7.41%

Row	Original Turkish Q Keyboard	Modified Turkish Q Keyboard
Top	37.15%	26.79%
Middle	43.91%	60.57%
Bottom	18.85%	12.54%

increased, meanwhile the ring and the little finger usages are dramatically decreased with these modifications. Effects of the modifications are summarized in Table 4.

CONCLUSION

In this study, we demonstrated that it is possible to increase the effectiveness of the keyboard layout with a similar approach applied in [1, 4-7]. By using the language statistics, we tried to make improvement suggestions for both the Q and F keyboards. Our proposals seem to have more fair key stroke distributions (in terms of row usages and hand/finger responsibilities).

In fact, it is not desired to have the elements of popular bigrams positioned at the area of responsibility of the same finger. Such a requirement converts the keyboard design to a multiobjective optimization problem. In our ongoing research studies, we try to formulate the problem in this manner and try to find the Pareto fronts via metaheuristics such as Genetic Algorithms and Ant Colony Optimization algorithm by also considering the bigram statistics obtained for Turkish in [10-11].

ACKNOWLEDGEMENTS

The authors would like to express our gratitude to the anonymous reviewers for their precious time and effort spent throughout their careful examination and evaluation of the manuscript as well as their guidance and helpful comments for improvement.

REFERENCES

1. Abbasov A, Hajiyev A, Afandiyev G. Statistical approach for an optimal placement of the letters of different alphabets on a computer keyboard. Applied and Computational Mathematics 8 (2009) 36–41.
2. Cassingham RC. The Dvorak Keyboard, Freelance Communications, California (USA), 1986.
3. Dvorak A, Dealey W. American simplified keyboard. U.S. Patent Nr: 2,040,248, 1936.
4. Wagner MO, Yannou B, Kehl K, Feillet D, Eggers J. Ergonomic modelling and optimization of the keyboard arrangement with an ant colony algorithm. Journal of Engineering Design 14 (2003) 187–208.
5. Dasgupta T, Basu A, Das A, Mandal P. Design and evaluation of Bangla keyboard layouts. Paper presented at the IEEE Students' Technology Symposium, Kharagpur, India, 3–4 April. IEEE Publication Department, New York (NY), USA, pp. 248–254, 2010.
6. Deshwal PS, Deb K. Ergonomic design of an optimal Hindi keyboard for convenient use. Paper presented at the IEEE Congress on Evolutionary Computation, Vancouver BC, Canada, 16–21 July. IEEE Publication Department, New York (NY), USA, pp. 2187–2194, 2006.
7. Malas TM, Taifour SS, Abandah GA. Toward optimal Arabic keyboard layout using genetic algorithm. Paper presented at the 9th International Middle Eastern Multiconference on Simulation and Modeling (MESM 2008), 26–28 August. IEEE Publication Department, New York (NY), USA, pp. 1–4, 2008.
8. Anonymous. The struggle of the F keyboard (in Turkish). Electrical Engineering 431 (2007) 36.
9. Turkish Standards Institute. Turkish standards - basic keyboard layout (in Turkish). 1991.
10. Çiçek E, Yılmaz AE. A study on the n–gram and syllable based statistical properties of Turkish (in Turkish). Paper presented at the 3rd Engineering and Technology Symposium (MTS3), Ankara, Turkey, 24–25 April. Çankaya University Publications, Ankara, Turkey, pp. 68–77, 2010.
11. Çiçek E, Yılmaz AE. A new Morse code scheme optimized according to the statistical properties of Turkish. Turkish Journal of Electrical Engineering and Computer Sciences 21 (2013) 804–811.