

Долгова Т. А., доцент

### ЕМКОСТЬ КОМПЬЮТЕРНЫХ ШРИФТОВ ПРИ НАБОРЕ НА БЕЛОРУССКОМ ЯЗЫКЕ

Factors influencing the mathematical expectation of the symbol width in the Byelorussian alphabet are analysed. The number of characters in the line for various fonts is calculated. Comparison with similar parameters of texts in Russian is carried out

Компьютерные технологии допечатной подготовки предоставляют широкий выбор гарнитур для оформления печатной продукции в соответствии с эстетическими требованиями и национальными особенностями различных изданий. В то же время при выборе шрифта следует руководствоваться и экономическими характеристиками.

Величина емкости конкретной гарнитуры позволяет с достаточной точностью определить объем будущего издания. Его можно уменьшить или довести до удобного для печати значения, выбрав шрифт с наиболее подходящими метрическими параметрами.

Число символов строки или полосы набора определенного формата обратно пропорционально математическому ожиданию ширины символа — среднеуточненной ширине  $E$ . Кроме ширины всех знаков заданной гарнитуры и кегля, эта величина непосредственно связана с конкретным алфавитом:

$$E = \sum_{i=1}^N l_i p_i, \quad (1)$$

где  $E$  — среднеуточненная ширина знака, мм;  $l_i$  — ширина знаков языка, мм;  $p_i$  — удельная частота встречаемости знаков алфавита;  $N$  — общее количество знаков, включая прописные и строчные буквы, цифры, знаки препинания и др. Суммарная вероятность появления всех знаков равна единице.

Определение среднеуточненной ширины рассмотрено, например, в [1, 2] применительно к русскому языку. Рассмотрим, как изменятся показатели емкости при наборе на белорусском языке.

Белорусский и русский алфавиты отличаются уже по составу букв: в первом из них есть знаки «і», «ў», «!» и нет «и», «ш», «ъ». Различна и удельная частота встречаемости букв [3]. В табл. 1 приведены значения  $p_i$  и их разность  $\Delta p_i$  (из величины для русского алфавита вычитается соответствующее значение белорусского). При этом буквы «і» и «и» рассматриваются как разные, поскольку ширина этих символов различна, одинаковые по ширине «и» и «й» объе-

динены: частота их встречаемости в русском языке просуммирована, для белорусского языка это величина только для «й». Аналогичным образом символ «ў» объединен с «у» и для белорусского языка рассмотрена суммарная частота их встречаемости. Символы «ъ» и «!» в таблице отнесены к «знакам» ввиду редкой встречаемости.

Прописные буквы и цифры встречаются с одинаковой частотой; частота встречаемости строчных букв в белорусском языке больше, чем в русском, — 0,9252 и 0,9077, а остальных знаков меньше — 0,0502 и 0,0676. Величину  $E$  чаще всего определяют учитывая только строчные буквы, тогда формула (1) имеет вид

$$E = \sum_{i=1}^n l_i p_i / \sum_{i=1}^n p_i, \quad (2)$$

где  $n$  — количество учитываемых символов алфавита.

Для сравнения языков будем использовать набор символов, суммарные удельные частоты встречаемости знаков которого, т. е. знаменатель (2), одинаковы. В данном случае — это строчные буквы и остальные знаки (без прописных букв и цифр). Их частота встречаемости для обоих языков равна 0,975.

Тогда  $\Delta E$  — разность среднеуточненной ширины знака русского и белорусского алфавитов — пропорциональна величине  $\Delta S$ , равной разности числителей формулы (2):

$$\Delta E = E_{\text{рус}} - E_{\text{бел}} = \Delta S / 0,975,$$

$$\Delta S = \left( \sum_{i=1}^n l_i p_i \right)_{\text{рус}} - \left( \sum_{i=1}^n l_i p_i \right)_{\text{бел}}$$

Среди анализируемых символов сильнее всего различаются вероятности появления для букв «а» и «о», в белорусском тексте они составляют около 15% и 3%, а в русском 7,5% и 10%. Буква «і» отличается от своего русского аналога не только встречаемостью (5% и 6,7%), она почти в два раза уже. Все буквы имеют разную частоту появления  $p_i$  и даже при одинаковой ширине  $l_i$  вносят различный вклад в

значение  $E$ . Анализировать, как изменилась доля букв разной ширины, удобнее не по отдельным символам, а по размерным группам.

В самом общем случае все символы условно можно разбить на три группы: узкие, нормальные (средние), широкие.

Для большинства текстовых гарнитур к широким можно отнести буквы: ж, м, ф, ы, ю, ш, щ; к узким относятся в основном знаки препинания и буквы: г, з, і. В табл. 2. представлена частота встречаемости символов для таких групп.

Как видно, в белорусском языке чаще встречаются широкие и узкие символы, а в русском больше средних. Разность  $\Delta p$  для узкой группы больше, но их ширина в несколько раз меньше.

Таким образом, суммарный вклад в величину  $\Delta E$  символов, для которых  $\Delta p < 0$  (узкие и широкие), сопоставим со вкладом символов нормальной группы, для которых  $\Delta p > 0$ . Пусть  $x$  — средняя ширина букв нормальной группы, аналогичную величину для узкой группы можно принять равной  $x - a$ , для широкой —  $x + a$ . Параметр  $a$  характеризует интервал варьирования ширины в пределах группы. Тогда  $\Delta S$  можно оценить с помощью выражения:

$$\Delta S = 0,069x - 0,045(x - a) - 0,024(x + a) \approx 0,02a.$$

Для многих гарнитур средняя ширина близка к величине полукруглой шпации, узкие буквы соотносят с 1/4 кегля, широкие — с 3/4 кегля.

Тогда, если  $a = 1/4$  кегля, то от средней ширины 1/2 кегля разность  $\Delta S$  составляет 1%. Это значит, что разница между среднеуточненной шириной символа русского и белорусского языков  $\Delta E = \Delta S / 0,975$  сравнима с величиной 0,01E, при десятом кегле это значение порядка 0,02 мм.

Можно сделать первоначальный вывод, что среднеуточненная ширина символа белорусского алфавита не сильно отличается от русского.

Как известно, даже для текстовых шрифтов при одинаковом кегле для разных гарнитур размеры одних и тех же букв могут заметно отличаться, не являются постоянными и группы, куда входят буквы с одинаковой шириной. Например, в гарнитуре Garamond фирмы Monotype буквы «а» и «э» относятся к узким, а «и» довольно широкая: при 10-м кегле их размер 1,4 мм и 1,9 мм при среднеуточненной ширине 1,7 мм. А для гарнитуры без засечек Arial (Monotype), для которой  $E = 1,9$  мм, «а» и «и» имеют одинаковую ширину 2 мм, что заметно больше, чем 1,8 мм для «э».

Учесть все особенности шрифта можно только используя точные расчеты  $E$  по формуле (2), где каждая строчная буква рассматривается отдельно.

Таблица 1  
Частота встречаемости букв

Символ	$p_i$		$\Delta p_i$
	рус. яз	бел. яз	
а	0,0748	0,1552	-0,0804
б	0,0147	0,0178	-0,0031
в	0,0415	0,0290	0,0125
г	0,0160	0,0191	-0,0031
д	0,0275	0,0272	0,0003
е,ё	0,0775	0,0382	0,0393
ж	0,0089	0,0058	0,0031
з	0,0155	0,0243	-0,0088
і	0,0000	0,0509	-0,0509
и, й	0,0782	0,0120	0,0662
к	0,0338	0,0335	0,0003
л	0,0433	0,0335	0,0098
м	0,0277	0,0264	0,0013
н	0,0589	0,0631	-0,0042
о	0,1020	0,0341	0,0679
п	0,0244	0,0249	-0,0005
р	0,0436	0,0463	-0,0027
с	0,0490	0,0382	0,0108
т	0,0549	0,0345	0,0204
у, ў	0,0256	0,0443	-0,0187
ф	0,0017	0,0018	-0,0001
х	0,0090	0,0112	-0,0022
ц	0,0043	0,0231	-0,0188
ч	0,0127	0,0142	-0,0015
ш	0,0077	0,0102	-0,0025
щ	0,0035	0,0000	0,0035
ы	0,0181	0,0456	-0,0275
ь	0,0174	0,0118	0,0056
э	0,0021	0,0105	-0,0084
ю	0,0055	0,0069	-0,0014
я	0,0079	0,0316	-0,0237
сумма	0,9077	0,9252	-0,0175
Прописн. б.	0,0236	0,0235	0,0001
Цифры	0,0011	0,0011	0,0000
Др. знаки	0,0676	0,0502	0,0174
Сумма всех	1,0000	1,0000	0,0000

Таблица 2  
Суммарная частота встречаемости букв в группах

Группа	$\Sigma p_i$		$\Delta = (\Sigma p_i)_{рус} - (\Sigma p_i)_{бел}$
	рус. яз.	бел. яз.	
узкие	0,0991	0,1445	-0,0454
нормальные	0,8031	0,7342	0,0689
широкие	0,0731	0,0967	-0,0232

## Сравнение показателей емкости шрифтов белорусского и русского алфавитов

Гарнитура (фирма)	ширина $E$ (10 пт), мм		емкость $e$ (6 кв.), знаков		$\Delta e$ , знаков	$\Delta E$ , %
	Рус. яз.	Бел. яз.	Рус. яз.	Бел. яз.		
TimesNewRoman (Monotype), V. 2.0	1,7654	1,7132	61	63	2	2,9%
BookAntiqua (Monotype)	1,9664	1,9174	55	56	1	2,5%
Garamond (Monotype)	1,703	1,6323	63	66	3	4,1%
LazurskiCTT (ParaGraph)	1,7764	1,7307	61	62	1	2,6%
Peterburg CTT (ParaGraph)	1,8725	1,8286	57	59	2	2,3%
Literaturnaya CTT (Poligrafmash, Intermicro)	1,9562	1,9553	55	55	0	0%
Arial (Monotype)	1,9423	1,8906	55	57	2	2,7%
PragmaticaCTT (ParaGraph)	2,001	1,9440	54	55	1	2,8%
TextBookC (ParaGraph)	1,8182	1,7674	59	61	2	2,8%

В таблице 3 приведены такие результаты при 10-м кегле для разных гарнитур (в скобках указана фирма-производитель или держатель лицензии). Для наглядности для двух алфавитов посчитаны емкости строки  $e$  длиной 6 квадратов по формуле  $e = (18,05 \cdot 6) / E$  и их разность  $\Delta e$ . Последний столбец таблицы содержит относительное отклонение среднеуточненной ширины символа белорусского языка по сравнению с русским.

Полученные для двух языков точные значения мало отличаются друг от друга, емкость шрифта для белорусского алфавита чуть выше, чем у русского.

Расчеты, основанные на величине  $E$ , определяемой рассмотренным способом, не могут учесть всех особенностей верстки (расстановка переносов, компенсационная верстка и др.); считается, что реальный объем может отличаться от расчетного примерно на 5% (в сторону увеличения или уменьшения) [4]. Поэтому различие между среднеуточненной шириной символа двух языков, которое не превышает

5%, практически не будет оказывать влияние на реальный объем издания.

Таким образом, для текстов на белорусском языке можно применять те же значения емкости либо, для особенно узких шрифтов и малых кеглей, увеличивать их на несколько процентов.

## Литература

1. Волкова Л. А., Ревякова О. Н. Исследование параметров компьютерных шрифтов // Полиграфия. — 2001. — № 1. — С. 42–43.
2. Долгова Т. А. Учет вертикальной емкости при определении экономичности компьютерных шрифтов // Труды БГТУ. Сер. IX. Издательское дело и полиграфия. — 2003. — Вып. XI. — С. 43–48.
3. Справочник технолога-полиграфиста: В 3-х ч. Ч. 1. Наборные процессы / М. В. Шульмейстер, Г. А. Таль. — М.: Книга, 1981. — 255 с.
4. Ревякова О. Н. Моделирование технологии подготовки изданий на донaborной стадии: Автореф. дис. ... канд. наук: 05.02.13 / МГУП. — М., 2002. — 16 с.