Are justification and hyphenation good or bad for the reader? First results

Leila Akhmadeeva, Rinat Gizatullin and Boris Veytsman

1 Introduction

Since the early days of typography text justification was considered a necessary feature of well-typeset text. The type block should be rectangular and grey. Even the "New Typography" by Jan Tschichold [5], while rejecting many dogmas of classical typesetting, still retained text justification and, as a consequence, hyphenated text.

While hyphenated text has obvious aesthetic advantages, typography is not just about aesthetics: its main purpose is to help the author to convey her thoughts to the readers. Is hyphenated and justified text really better for the reading and comprehension?

In this work we try to answer this question.

2 Experimental methods

The experimental methods were a variation of the scheme used in our previous papers [1, 2, 7]. A group of $N = 300$ healthy volunteers (Bashkir State Medical University undergraduates) was given two texts, A and B. Each text had 282 words, typeset with LaTeX using ParaType Serif fonts. Half of the participants got text A justified and text B ragged right, while the other half got text B justified and text A ragged right. The participants were asked to read the text. After a minute they marked their current reading position. Immediately after the reading the participants were given a multiple choice test (10 questions with 4 variants of answers to choose from). To test the long-term memory, we repeated the test 60 minutes later. We compared the differences between the justified and ragged right tests.

3 Results

The difference between the results for justified and ragged right texts for the same subjects is shown in Figures 1 and 2. The results suggest that justified texts give slightly higher reading speed and slightly fewer correct answers on the delayed test (note that Figure 2 shows the *difference* between justified and ragged texts). However, the effect is small: we will see below that it is smaller than the difference between the texts A and B themselves and the individual differences between the subjects. To quantify the effect we use a Bayesian technique [3, 4].

In our model we assume that each participant has an individual reading speed and individual probability of correctly answering a question. Besides these individual propensities, there are corrections for the texts (A or B) and typesetting (ragged or justified), common for all participants. These corrections are what we want to determine.

More formally, let us introduce the parameters

$$\delta_a = \begin{cases} 1, & \text{Text } A \\ -1, & \text{Text } B \end{cases} \qquad \delta_j = \begin{cases} 1, & \text{Justified text} \\ -1, & \text{Ragged text} \end{cases} \tag{1}$$

Then we will model the reading speed v as a normal distribution with the mean

$$v = v_{\text{ind}} + \frac{1}{2}(\delta_a v_a + \delta_j v_j) \tag{2}$$

and standard deviation σ. Here v_{ind} is the individual reading speed, v_a is the difference between text A and text B, and v_j is the difference between justified and ragged texts. We need to estimate v_a and v_j.

To estimate the probability of a correct answer on a test (either immediate or delayed) we use the log odds function [6]:

$$L = \ln\left(\frac{p}{1-p}\right) \tag{3}$$

where p is the probability of correctly answering. When p changes between 0 and 1, L changes between $-\infty$ and $+\infty$.

Then we can write down the mean log odds as

$$L = L_{\text{ind}} + \frac{1}{2}\left(\delta_a L_a + \delta_j L_j\right) \tag{4}$$

where the parameters have the same meaning as in equation (2). We use separate estimates for immediate and delayed test.

The *a priori* distribution for the parameters is the following:

1. Normal for v_{ind} and L_{ind} with mean from the data and deviation equal to 100 times the data deviation.
2. Normal for v_a, v_j, L_a and L_j with zero means and deviation equal to 100 times the data deviation.
3. Uniform for all standard deviations from 1/1000 to 1000 times the data deviation.

We found that a number of participants had reading speeds higher than 282 words per minute, so they were able to read the whole text before the time was up. We used the censoring methods for Bayesian analysis to overcome this limitation [4].

We used multiple chain Monte Carlo simulations (16 chains with 10,000 samples each) for each model.

The results are plotted in Figures 3, 4 and 5. On the figures we plot the probability distributions of the parameters in equations (2) and (4); the x axis

Figure 1: Distribution of reading speed and test results

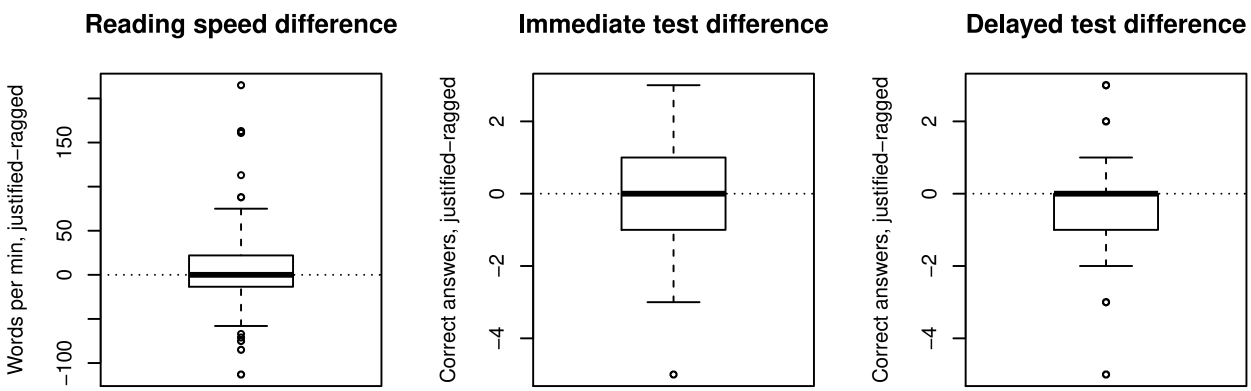

Figure 2: Difference of speed of reading and test results for justified and ragged right texts for the same subject

shows the values of the parameter, while the y axis shows the relative probability of this value according to the simulations. In all figures the first panel represents the individual differences, the second panel the differences between texts A and B, and the last panel the difference between the justified and ragged right texts. On the last panel we plot the zero line (no difference) and 95% interval for the parameters. The significance test of the usual statistics corresponds to the 95% interval being completely to one side of the vertical zero line [4].

The results show that the individual differences in all models dominate the other factors. The difference between the justified and ragged texts is small. However on the 95% level we can say that justified texts are being read faster than the ragged right (by about 7 words per minute), and, even more interesting, the results for delayed tests are better for the ragged right texts. If we convert the log odds to the number of correct answers, we can see that on a 100-question test with 90% correct answers the difference would be about 4 points.

4 Discussion and conclusions

We see that there is a small, but persistent difference between justified and ragged right texts: the former are read slightly faster, but on the delayed tests (when the text is committed to long term memory) give slightly worse results.

Does this mean that one should typeset exam materials in the ragged right fashion? Not necessarily. We do not know whether this effect is specific for our population: Cyrillic readers in Russian, with a significant proportion having Russian as the second language (many students of Bashkir State Medical University have Tatar or Bashkir as their first language). Still, our findings are very intriguing and should be further investigated. One way of interpreting the results might be the interplay between visual image of a word and its commitment to memory: a justified text has hyphenated words with “broken” visual image. If this is the case, the effect should be more pronounced for languages with longer words like Russian and German than for languages with shorter words like English.

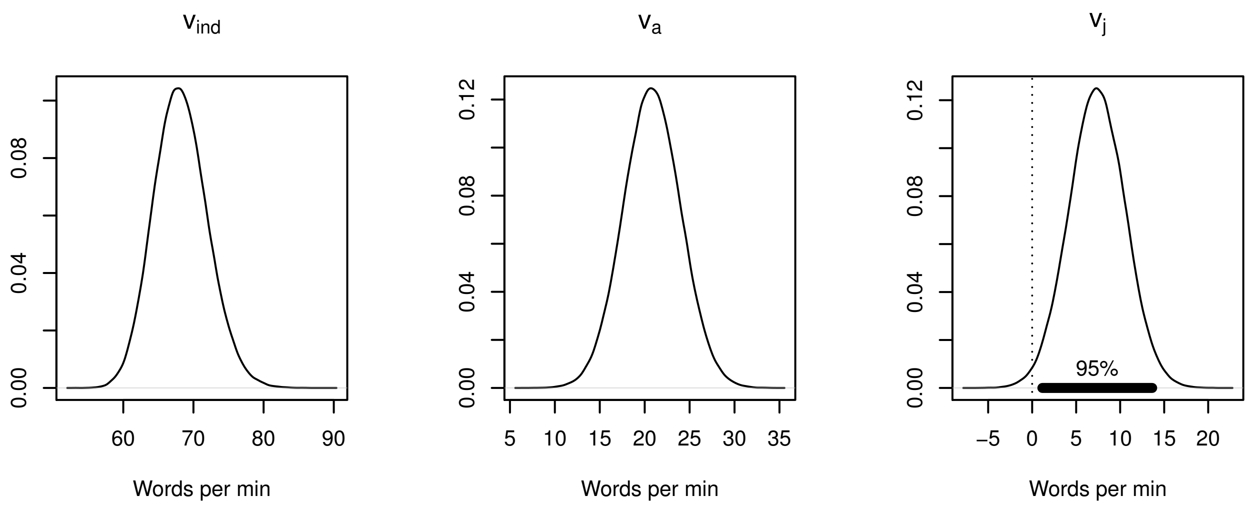

Figure 3: Bayesian estimate for the speed of reading model

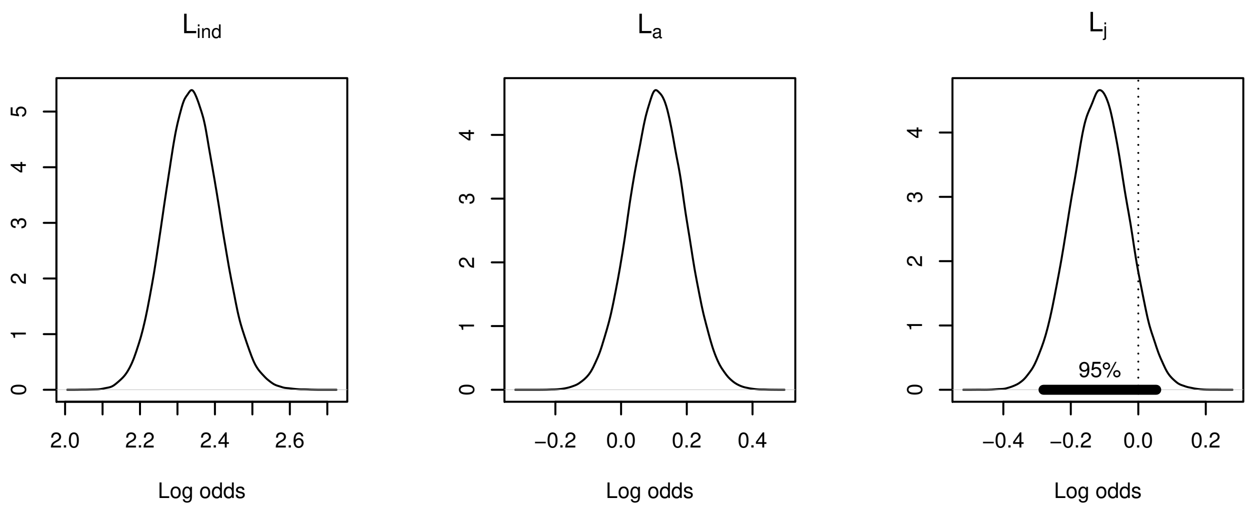

Figure 4: Bayesian estimate for the immediate test model

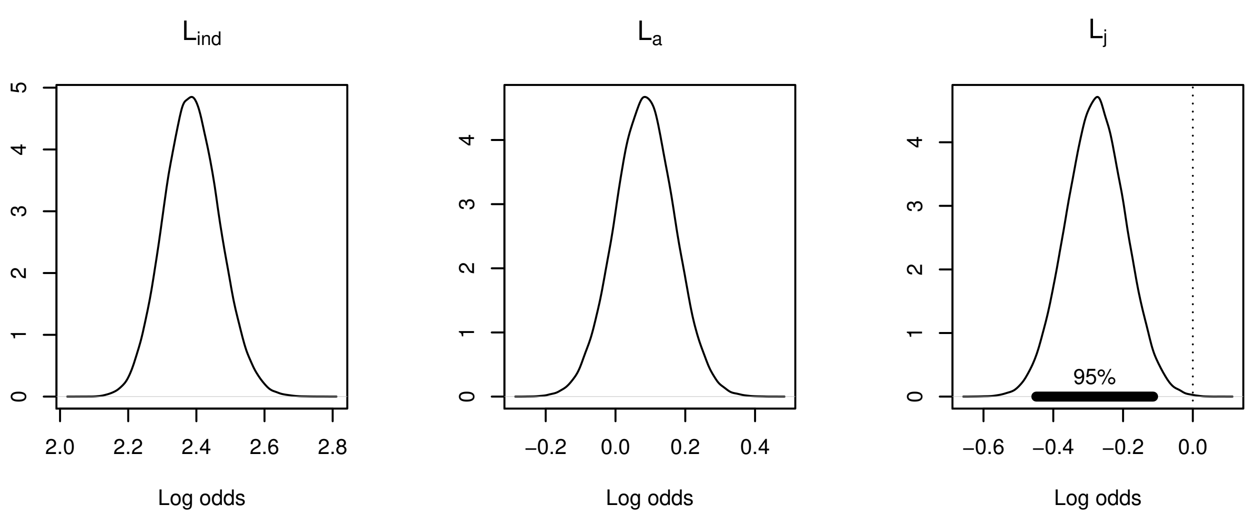

Figure 5: Bayesian estimate for the delayed test model

Acknowledgments

The Bayesian simulations for this work were performed on the George Mason University ARGO computing cluster.

One of the authors (RG) gratefully acknowledges the student travel grant from The European Academy of Neurology (EAN) which allowed him to present these results at the 2nd Congress of EAN (Copenhagen, May 2016).

As usual, the audience at TUG 2016 gave us many interesting ideas, which we will try to use in further research.

References

[1] Leyla Akhmadeeva, Ilnar Tukhvatullin, and Boris Veytsman. Do serifs help in comprehension of printed text? An experiment with Cyrillic readers. *Vision Research*, 65:21–24, 2012.

[2] Leyla Akhmadeeva and Boris Veytsman. Typography and readability: An experiment with post-stroke patients. *TUGboat*, 35(2):195–197, 2014. `http://tug.org/TUGboat/tb35-2/tb110akhmadeeva.pdf`.

[3] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer, Dordrecht; Heidelberg; London; New York, 2009.

[4] John K. Kruschke. *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan*. Academic Press, second edition, 2014.

[5] Jan Tschichold. *The New Typography*. University of California Press, Berkeley and Los Angeles, CA, 1998.

[6] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Statistics and Computing. Springer, New York, fourth edition, 2010.

[7] Boris Veytsman and Leyla Akhmadeeva. Towards evidence-based typography: First results. *TUGboat*, 33(2):156–157, 2012. `http://tug.org/TUGboat/tb33-2/tb104veytsman-typo.pdf`.

◇ Leila Akhmadeeva
Bashkir State Medical University, 3 Lenina Str., Ufa, 450000, Russia
`la (at) ufaneuro (dot) org`
`http://www.ufaneuro.org`

◇ Rinat Gizatullin
Bashkir State Medical University, 3 Lenina Str., Ufa, 450000, Russia

◇ Boris Veytsman
Systems Biology School & Computational Materials Science Center, MS 6A2, George Mason University, Fairfax, VA, 22030, USA
`borisv (at) lk (dot) net`
`http://borisv.lk.net`