Sans Forgetica: Font Readability Experimental Design

Robert Dykes

Faculty of Human Studies, Jin-ai University

Sans Forgetica: 実験的デザインフォントの可読性

ロバート・ダイクス

仁愛大学人間学部

　本研究ノートは、指標追跡と定量分析による Sans Forgetica の可読性の研究プロジェクトデザインと、その研究を行うことの正当性を示すことを目的とする。Sans Forgetica は「望ましい困難」を生み出すために開発されたフォントであるが、先行研究ではそのような効果が得られるという結果が出ていない。本研究プロジェクトでは、期待される結果が得られない理由のより深い解明を試みる。先行研究では Sans Forgetica の非流暢性を認めている。それではなぜ「望ましい困難」が生み出されないのか。そこで提唱される一つの仮説は、先行研究では非流暢性を決定する定着や読みにくさの評価の正当化の方法が暗示的であることから、実際には Sans Forgetica は非流暢性がないのではないかということである。Sans Forgetica の明示的調査により先行研究とは異なる結果が明らかになるかもしれない。視標追跡と Sans Forgetica の可読性の定量分析を用いることにより必要な明示的データが得られることも期待できる。

キーワード：desirable difficulty, disfluency, Sans Forgetica, readability, eye tracking

Introduction

Sans Forgetica is a relativity new font, released to the public for free download in 2018. The font was "specifically designed to enhance memory retention" (RMIT University, n.p., 2018a). Sans Forgetica was created on the notion that it can improve memory performance when used in learning applications through a concept called desired difficulty. This is when a processing difficulty is introduced at either the encoding or the retrieval process of a learning activity that improves long-term retention (Rosner, Davis, & Milliken, 2015).

Sans Forgetica was created by scientists at RMIT University in Australia (Sans Forgetica, 2018). Dr. Jo Peryman, a member of the design team behind the font, claims that "this if the first time ever that specific principles from psychological theory have been combined with the specific principles from design theory in order to create a font" (RMIT University, n.p., 2018b). However, studies are now coming out that seem to be contradicting the claims of the creators of Sans Forgetica. Taylor, Sanson, Burnell, Wade, and Garry (2020) conclude the font does create a state of disfluency (processing difficulty), but this effect does not lead to a desirable difficulty outcome. Their study included 882 people across four different

experiments using the Arial and Sans Forgetica fonts. Experiment 1 concluded that the Sans Forgetica font was the less fluent of the two fonts. In all three of the remaining experiments, Taylor et al., found no positive memory outcomes, concluding that Sans Forgetica "does not create desirable difficulty" (p.6, 2020). Dykes and Hauca (2020) also reached a similar conclusion when testing for improvements in reading comprehension with Japanese learners of English using Sans Forgetica. This study examined the effects of Sans Forgetica on reading comprehension with Japanese learners of English. Half the subjects were given a short passage in the Century Schoolbook font, the other half were given the same passage in Sans Forgetica. Both groups were later given a short quiz. This was repeated over six administrations. Dykes and Hauca (2020) recorded no significant difference in quiz scores indicating that a desirable difficulty effect did not occur, but a disfluent effect was observed. The difference in reading times between the two fonts returned a significant difference with the Sans Forgetica passages taking longer to finish. Geller, Davis, and Peterson (2020) conducted a 3-part experiment using cue-target pairs with missing letters (the generation effect), cued recall for select words in reading passages, and single-word noun recognition tests. All three experiments failed to show improved learning outcomes for Sans Forgetica concluding "we did not find any evidence for a mnemonic benefit of Sans Forgetica typeface" (Geller, Davis, & Peterson, p.20, 2020).

If Sans Forgetica is failing at what it was designed to do, why continue studying it further? The answer is, to better understand why Sans Forgetica is not achieving the outcomes it was designed for. The next step in examining Sans Forgetica is to take a closer look at some of the processing and encoding that Sans Forgetica is supposed to have a desirable difficulty effect upon. This research design plan aims to better understand San Forgetica better by investigating what

is occurring with the readability of Sans Forgetica by conducting a quantitative analysis on the Sans Forgetica's effect on reading comprehension. In order to ascertain some new feedback and information about Sans Forgetica's effects the proposed study will employ eye tracking hardware and software. The data to be collected and examined will include fixation time, saccade length, words per minute (WPM), re-fixations, and words between fixations. The proposed project will examine both native speakers of English and Japanese non-native speakers of English. This study will also serve as a replication study of Screws (2016), a study that examined a common sans serif font, and a serif font (see figure 1).

Figure 1: Sans serif vs serif font

Readability and legibility of text has been researched for well over 100 years. Tarasov, Sergeev, and Filimonov (2015) traced textbook legibility studies as far back as 1881. In modern studies the two main font types that are most often compared are serif and sans serif fonts (other fonts include script, display, and specialty fonts like dingbats). One area of contention is that studies are not always in agreement if there is a significant difference in serif and sans serif fonts when other variables are equalized. It should be noted that all the following studies discussed below examined fonts displayed on a digital screen as opposed to print material. Screws (2016) found no significance difference when measuring the readability of a serif and sans serif font based on the following tests, fixation time, saccade length, WPM, re-fixations, and words between fixations. Hojjati and Muniandy (2014) conducted a study comparing a serif and sans serif font and found a significant difference determining

that Verdana (sans serif) was the better font for long text. The measurements came from reading times and short quiz results. Morrison and Noyes (2003) found a significant difference in readability between a serif and sans serif font. Bernard, Lida, Riley, Hackler, and Janzen (2002) compared eight fonts, four serif and four sans serif fonts, each at three different font sizes (10, 12, and 14pt.) and found no significant difference in reading efficiency between any of the fonts at any of the sizes. Josephson (2008) compared two serif and two sans serif fonts and found no significant difference between the fonts, but noted that this could be because of the of low sample size used in the research, however; they concluded, similar to Hojjati and Muniandy (2014), that the Veranda font (sans serif) appeared to be the more favorable font. Veranda had the shortest reading times and the fewest regressions. Contradictory to two other tests, Times New Roman (serif) had the smallest number of fixations while Veranda had the highest indicating in this measure Times New Roman was the easiest to read, while Veranda was the hardest. Finally, Dogusoy, Cicek, and Cagiltay (2016) report that in their research they found that participants were able to read faster and more accurately with a sans serif font, but fixations between the serif and sans serif fonts were the same. It should be noted that Dogusoy, Cicek, and Cagiltay (2016) did not provide data on statistical significance.

The number of contradictory studies comparing common "sans" adds serif and serif fonts is quite long. This is where Sans Forgetica is possibly unique. It is anything but a common font. Dykes and Hauca (2020) found that it impairs reading speed. Geller, Davis, and Peterson (2020) determined it was indeed a disfluent font. Taylor et al. (2020) also concluded the same. But the reason why font studies are returning contradictory results and why Sans Forgetica is not producing a desirable difficulty effect may be linked. Some font studies are using implicit measurements, while some are using explicit measurements. The same is

happening with Sans Forgetica and this may explain some of the mystery of font (non)disfluency and Sans Forgetica missing desirable difficulty effect. Geller, Davis, and Peterson (2020) hypothesize that desired difficulty may not be occurring with Sans Forgetica because while these studies are indeed reporting it as a disfluent font, it may in fact not be disfluent after all. Geller, Davis, and Peterson (2020) emphasize that for a desirable difficulty to occur you need a *perceptually disfluent stimulus,* and this is not what these studies are measuring or observing when examining Sans Forgetica. There could be a reason for this:

> Drawing valid conclusions about disfluency, then, requires the use of objective disfluency measures. In many studies, perceptual disfluency is tested subjectively (via JOLs [judgments of learning] or difficulty ratings), but never explicitly tested. Thus, it could be that the failure to observe an effect in the current set of studies is because Sans Forgetica typeface is simply not perceptually disfluent. (Geller, Davis, & Peterson, 2020, p. 21)

It is the aim of this proposed project to have another look at Sans Forgetica and its disfluent status, this time not testing through JOLs or difficulty ratings, but through various eye tracking measurements.

Research Project

1. The first stage of this proposed research project is to run a replication study of Screws (2016) (with native speakers of English, same as the Screws study), but replace the common sans serif font used in that study with Sans Forgetica (see figure 2).

Screws (2016) found no significant differences theorizing that there just is not enough difference between common sans serif and serif fonts. The hypothesis this proposal holds is the opposite, that Sans Forgetica will be different and unfamiliar enough

that it will produce a significance difference in the eye tracking tests.

2. The second stage of the proposed research project will be to repeat stage 1, but with Japanese learners of English (JLE).

The hypothesis for stage 2 is that for non-native English speakers both fonts, even a common serif font, is in fact not common enough in daily life and daily usage for most JLE. This will result in no significant difference between the chosen serif font and Sans Forgetica.

3. Stage 3 will consist of a comparative analysis between both sets of data from stage 1 and 2. This will provide some insights into explicit readability information for a common serif font and Sans Forgetica between native and non-native speakers of English.

For stage 3 the hypothesis is that the JLE will have lower words per minute, shorter saccade length, fewer words between fixations and higher fixation time, re-fixation count (regression) for both the serif and Sans Forgetica font.

ABCDEF vs ∧3C↺Ǝ∃Ƒ

Figure 2: Serif vs Sans Forgetica

Eye tracking measurements

The three main measures of most applied linguistics eye tracking researchers are saccades, fixations, and regressions (Conklin & Pellicer-Sánchez, 2016). Saccades and fixations are automatic, physiological responses. Saccades are the movements of the eyes, in this context, the movements from one word to another. Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka, and Van de Weijer (2011) explain that "[s]accades are very fast – the fastest movement of the body can

produce – typically taking 30-80 ms to complete" (p23). This project will measure saccade length, distance not time, measured in computer screen pixels as the unit of measure. The next major measure is fixations, "the interval between the eyes' movements, when the eyes stop, are called fixations" (Conklin, Pellicer-Sánchez, & Carrol, 2018, p.1). A Fixation is "the time of the fovea alignment of an object exceeds 100ms, and in the meantime, the watched object will be fully processed, with a clear image formed at the fovea" (Zhan, Shen, & Wang, 2014, p.273). Fixation time (or fixation duration) will be measured in time in milliseconds (ms) with expected values to fall between 100ms and 500ms. Fixation distance will also be measured using words between fixations as a unit of measure. Regression (or re-fixations) is the third main measurement. A regression is a type of saccade when the reader moves their eyes back to a previously encountered section of text (Conklin, Pellicer-Sánchez, & Carrol, 2018). Not a main eye tracking measurement, but still a valuable data point is words per minutes (WPM) and it is an indicator of reading speed. See figure 3 for an example heat map showing scan path, fixations, and regressions.
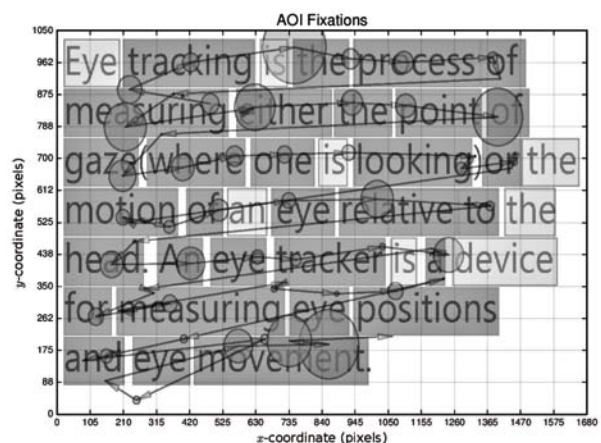

Figure 3: heat map showing scan path, fixations, and regressions
Adapted from Screws (2016)

Conclusion

If we run the same eye tracking procedures with Sans Forgetica, it will either show a disfluent font compared to a common serif font or a non-disfluent font. Geller, Davis, and Peterson state that many recent studies hold the conclusion that "perceptual disfluency has little impact on actual memory performance" (2020, p.21). If the eye tracking data analysis returns significant results that Sans Forgetica is a disfluent font, that would support the above theory. If the eye tracking data analysis does not return significant results, then that outcome would support the Geller, Davis, and Peterson theory that "Sans Forgetica typeface is simply not perceptually disfluent" and not "sufficiently difficult in order to trigger increased metacognitive monitoring and control processes" (2020, p.21).

References

Bernard, M., Lida, B., Riley, S., Hackler, T., & Janzen, K. (2002). A comparison of popular online fonts: Which size and type is best. *Usability news, 4*(1), 2002.

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research,* 32(3), 453-467.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge University Press.

Dogusoy, B., Cicek, F., & Cagiltay, K. (2016, July). How serif and sans serif typefaces influence reading on screen: An eye tracking study. In *International Conference of Design, User Experience, and Usability* (pp. 578-586). Springer, Cham.

Dykes, R., and Hauca, M. (2020). Sans Forgetica: Typography's Effect on ESL/EFL Reading Comprehension. In D. Shaffer & J. Kimball (Eds.), KOTESOL Proceedings 2019 (pp. 53-60). Korea TESOL.

Geller, J., Davis, S. D., & Peterson, D. (2020). Sans forgetica is not desirable for learning. Available at https://psyarxiv.com/ku5bz/download?format=pdf.

Hojjati, N., & Muniandy, B. (2014). The Effects of Font Type and Spacing of Text for Online Readability and Performance. *Contemporary Educational Technology,* 5(2), 161-174.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford.

Josephson, S. (2008). Keeping your readers' eyes on the screen: an eye-tracking study comparing sans serif and serif typefaces. *Visual communication quarterly,* 15(1-2), 67-79.

Morrison, S., & Noyes, JM. (2003). A comparison of two computer fonts - serif versus ornate sans serif. *Usability News.* Available at http://psychology.wichita.edu/surl/usabiltynews/52/UK_font.htm.

RMIT University (2018a). https://youtu.be/CEXzSehu5HM Accessed on August 22nd, 2020.

RMIT University (2018b). https://youtu.be/PO2Eo6D5tNQ Accessed on August 22nd, 2020.

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica, 160,* 11-22. https://doi.org/10.1016/j.actpsy.2015.06.006

Sans Forgetica (2018). https://sansforgetica.rmit.edu.au/ Accessed on August 22nd, 2020.

Screws, J. (2016). Quantitative Analysis of Font Type's Effect on Reading Comprehension. Clemson University.

Tarasov, D. A., Sergeev, A. P., & Filimonov, V. V. (2015). Legibility of textbooks: a literature review. *Procedia-Social and Behavioral Sciences, 174,* 1300-1308.

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: the (lack of) effect of Sans Forgetica on memory. *Memory,* 1-8.

Zhan, W., Shen, Z., & Wang, J. (2014). Research on Eye Movement Tracking in ESL Reading. *Sensors & Transducers, 173*(6), 272.

SUMMARY

This research note is the research project design and justification for conducting an eye tracking and quantitative analysis on the readability of Sans Forgetica. Sans Forgetica is a font designed to evoke a desirable difficulty effect; however, early studies have shown that Sans Forgetica is not producing this outcome. The proposed project wants to attempt to better understand why this may be happening. Studies do agree that Sans Forgetica is a disfluent font, so why

is desirable difficulty not occurring? One theory put forward is that Sans Forgetica is in fact not disfluent, as observations up till now have been implicit, applying judgments of learning or difficulty ratings to determine disfluency. Explicit examinations of Sans Forgetica may reveal a different outcome. An eye tracking and quantitative analysis on the readability of Sans Forgetica could provide the explicit data needed.