The Impact of Human Factors on Digitization: An Eye-tracking Study of OCR Proofreading Strategies

Flavia De Simone, Barbara Balbi, Vincenzo Broscritto, Simona Collina, Roberto Montanari
Università Suor Orsola Benincasa
Centro Scienza Nuova
Naples, Italy
e-mail: flavia.desimone@centroscienzanuova.it
barbara.balbi@centroscienzanuova.it
vincenzo.broscritto@studenti.unisob.na.it
simona.collina@unisob.na.it
roberto.montanari@unisob.na.it

Federico Boschetti, Anas Fahad Kahn
Consiglio Nazionale delle Ricerche
Istituto di Linguistica Computazionale "A. Zampolli"
Pisa, Italy

e-mail: federico.boschetti@ilc.cnr.it
fahad.kahn@ilc.cnr.it

***Abstract*— An Optical Character Recognition (OCR) System is a piece of software that can scan a printed text and translate it into a digital format that can be subsequently edited with a computer. Often the output from OCR software does not correspond closely enough to the original text and a manual correction phase is needed in order to improve accuracy. The aim of the study presented in this paper is to test human strategies in proof correction by means of an eye-tracker. The experiment, which we designed to investigate these strategies consisted in a proofreading task. Participants were divided into two groups: a target group that was trained in how to carry out the task and a control group that had not been so trained. The performances of each group were evaluated in terms of accuracy and time of execution. Results highlighted an effect of learning, an optimization strategy of the target group resulting in higher accuracy and lower time of execution of the task. The practical implications of these results will be discussed.**

Keywords - Optical Character Recognition; Eye-tracking; Reading; Cognitive processe.

I. INTRODUCTION

Of all the processes that impact on the production of resources in the Digital Humanities, the digitization of texts by means of an Optical Character Recognition (OCR) system, and in particular the correction phase, is one of the most complex. This is due both to how resource and time consuming it can be, and also because of the role that human factor issues play in ensuring the accuracy of the final output [1]. On the other hand, the accuracy of digitized corpora is a fundamental requirement for any further phases of analysis and treatment of texts, such as for instance linguistic annotation. To improve the effectiveness of OCR systems, we believe that it is important to study the role of human factors in proofreading activities and to use this information to develop strategies in order to make systems more adaptive to users' needs [2]. The adaptivity of a system, or a machine, is its ability to adapt to its human operator and to thereby reduce his or her cognitive workload [3]. The background to the present study is in human factor psychology, a branch of psychology dedicated to the study of human-machine interaction with a strong connection with cognitive theories [4].

In the following, we present related work needed to place our research and experimental efforts in Section II. We delve into the details of our experimental procedure in Section III.. Finally, we conclude and introduce future work in Section IV.

II. THEORIES AND MODELS

Reading is the complex outcome of a learning process which permits the conversion of a visual representation into a phonological form. The success of the process implies that there is access to a background store of memories containing not only morphological and phonological information but also semantic and syntactic knowledge. Nevertheless, the process is rapid, taking only a matter of milliseconds, it is error free most of the time, and partially unconscious as proved by the Stroop effect [5]. The Stroop paradigm is one of the most commonly used technologies for studying lexical production. The task is very simple: participants are asked to name the color of the ink used to write a word without reading the word. The dependent variable is the time to response. Participants are found to perform the task significantly faster when the color of the ink corresponds to the meaning of the word, despite the fact that they are instructed to ignore the content of the words themselves.

One of the most well-known models for explaining the reading process is the Dual Route Cascaded Model [6], which hypothesizes that two different mechanisms are involved in reading aloud. One mechanism, the lexical route, assumes that an expert reader has a mental representation for every learned word and the visual recognition of the written word directly activates the internal representation thus speeding up the reading process. The other mechanism, the non-lexical route, is applicable to non-words or new words for which a mental representation is not available. In this case, the reader decomposes the words into constituents (graphemes) and then applies graphemes/phonemes conversion rules.

The first step of the reading process in both routes is a visual recognition phase. This is well explained by the model

 ISBN: 978-1-61208-609-5

of word perception developed by McClelland and Rummelhart [7] in which the first level of processing corresponds to visual features that differentiate letters (e.g., the letter E is formed by one vertical and three horizontal tracts). Starting from this theory we decided to work on this visual level to introduce errors into the OCR output, as explained below in Section III-B.

So far, we have described models which have been developed to explain chronometric data and error corpora, but research on reading processes enjoyed a significant boost in the 70s with the introduction of eye-tracking technology into the sciences. Eye-tracking technology has made it possible to study eye movements and to infer underlying cognitive processes, in particular selective attention, that is, the ability to elaborate a stimulus by ignoring all competing stimuli [8]. Although attention can be oriented regardless of eye movement it is more often eye-driven. To better understand this process it is necessary to introduce some additional concepts [9]:

- foveal vision: the fovea is a small region in the retina with a diameter of 1,55 millimeters where visual acuity is at its highest;
- smooth pursuit movements: slow eye movements that follow an object moving in the visual field, keeping it into the fovea;
- saccades: rapid movements of the eyes that change the fixation point;
- fixations: the maintaining of the eyes on a portion of the visual field for a time longer than 250 milliseconds; fixations have two main attributes: location and duration.

Previous studies have used eye-tracking technology in OCR domain. In particular, Rello and Baeza-Yates [10] use the eye-tracker to evaluate the readability of digital texts that contain OCR errors (among other types of errors). Ishiguro et al. [11] apply eye-tracking to study the achievement of multiple tasks at the same time, such as face recognition, object detection and text reading. In order to monitor these activities, each region of interest is processed by the suitable recognizer, which is OCR in the case of text. Buscher et al. [12] use the eye-tracker on OCR documents, in order to annotate which areas are read and which areas are skimmed.

Starting from this scientific background, we developed an eye-tracking study with the specific aim to investigate reading strategies in proofreading task.

III. EXPERIMENT

In the study presented below, we tested for two different aspects:

1- the first related to a question of methodology: the possibility of studying the visual strategies adopted by OCR proofreaders by means of eye-tracking technology;

2- a learning effect: the influence of learning on proofreading strategies.

Towards this aim we compared performances in proofreading tasks of two groups of participants, a group trained for the task and an untrained group, using eye-tracking instruments. We expected that the group trained to the task to be more rapid and accurate in the execution of the task.

A. Participants

Thirty subjects volunteered in the experiment. The age range was between 20 and 45 years old. They all were Italian mother-tongue speakers. They had normal or corrected to normal vision. None of the participants received any money or course credits for participation. In addition, none of them had any previous experience in proofreading. They were equally distributed into two groups: a target group (TG) undergoing a learning phase for the proofreading task before the proper experiment and a control group (CG) that received no training for the task.

B. Materials

The text used for the experiment was an OCR scan extracted from the book “Gomorra” written by Roberto Saviano. We opted for a contemporary Italian text for reasons of ease and familiarity of semantic and syntax.

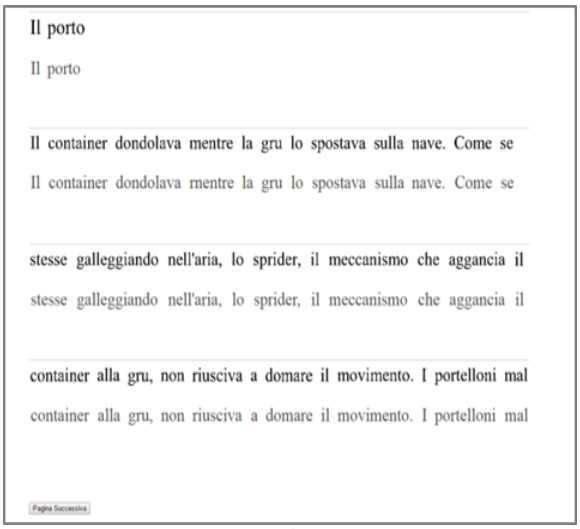

Figure 1: An example of experimental materials

The text was integrated into a web platform and divided in thirty screens: each screen was composed of four sessions and each session was made up of two parts, an image obtained by a high-resolution scan (600 DPI) and a line containing the output of the OCR software, an editable text, where the errors to be corrected by participants could be presented (Figure 1). Orthographical errors were manually inserted in the OCR output by the experimenter. Errors were letters substitutions of two types: “rn” instead of “m” and vice versa, “O” instead of “o” and vice versa[13;14].

C. Equipment

To control visual behavior of participants and acquire eye movement data an eye-tracker was used. We opted for a remote, non-contact system, FaceLab (Figure 2), as it was suitable for use in a controlled environment, a laboratory, with a task presented via computer desktop. The system

consists of an infrared pod and two cameras posed on the desk at the base of the computer monitor. Before data acquisition, the eye-tracker is calibrated and cameras position is adjusted for each participant to increase data accuracy.

One great advantage of this system, compared to a wearable system, is the stability and accuracy of the resulting data due to the upright and stable position of participants. This favors calibration and avoids the kind of data loss that might result from wireless connection issues.

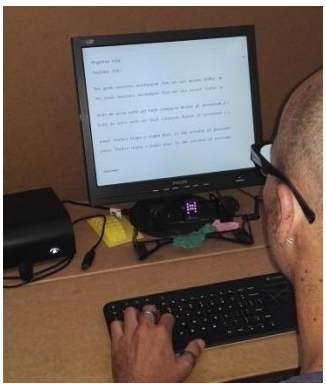

Figure 2. FaceLab 5.0

D. *Procedure*

The experiment was carried out in a quiet ad-hoc prepared room. Participants were seated in front of a 17" computer monitor with a maximum resolution of 1280x1024 at 60 Hz. The two groups of participants received different instructions. Participants in the TG were familiarized with the task in a learning phase in which they could correct thirty screens of text before starting with the experiment proper. They were informed that in the experiment they would find the same errors as in the learning phase. The entire procedure lasted about one hour. Participants in the CT were asked to pass directly to the experimental phase without any specific training. This experiment lasted about half an hour.

E. *Results*

To evaluate the accuracy of performances the number of detected errors has been acquired and analyzed. A statistical test on frequencies (chi-squared) revealed a significant effect of learning (Figure 3): participants inserted in the TG corrected a significant major number of errors respect to CG ($p<.004$).

To compare the strategies adopted by the two groups we focused on three metrics: the time to complete the task, the mean number of fixations and the mean fixation duration [15]. To extract metrics about fixations we designed two Areas of Interest (AOIs) for each of the sessions into which the text is divided, the image and the OCR editable text (see Section III-B). Statistical tests on the means (t-test) highlighted that the two groups adopted two different visual strategies: TG tended to be more rapid in executing the task compared to CG ($p<.06$) because the participants mostly focused their attention on the OCR output as can be inferred by an higher number of fixations ($p<.02$; Figure 4) and a longer fixation duration ($p<.0003$; Figure 5).

IV. CONCLUSION

Taken together the data confirm the two hypotheses: the suitability of the adopted methodology to study human-factor issues in digital era and the learning effect, an advantage in terms of accuracy and time of execution, on proofreading strategies. Our next step will consist in the application of the same methodology to study the strategies adopted by expert proofreaders; in addition, we are also looking into the possibility of verifying the strategies adopted by proofreaders to correct different types of errors (syntactic, semantic). The main aim of this direction of study is to use the knowledge acquired to design and develop OCR systems that are ever more adaptive to human users' needs.

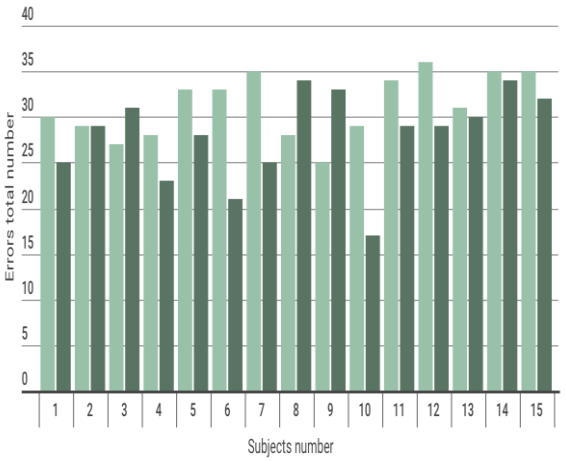

Figure 3. Number of detected errors for each subject.

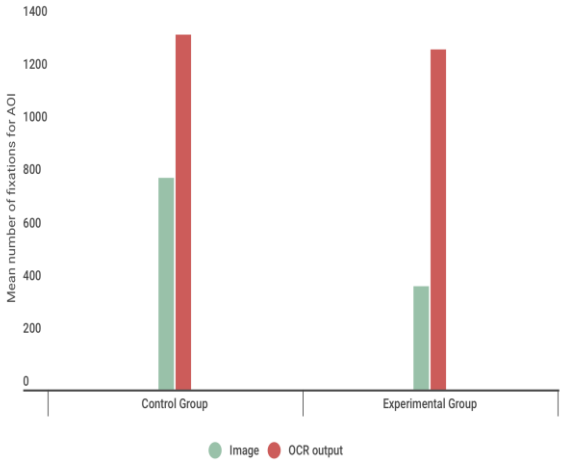

Figure 4. Mean number of fixations per AOIs (image: green bar; OCR output: red bar) for each group

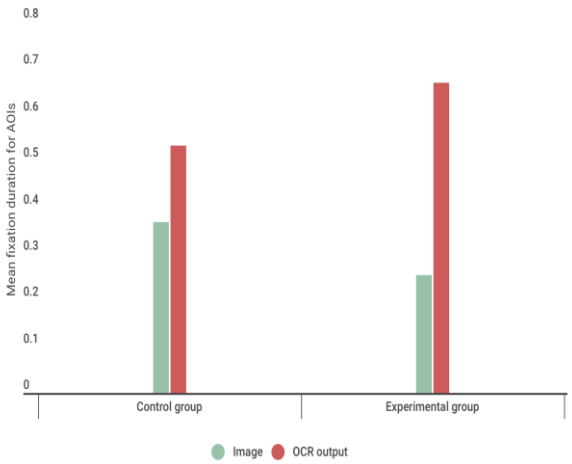

Figure 5. Mean fixation duration per AOIs (image: green bar; OCR output: red bar) for each group

REFERENCES

[1] G. Nagy, "At the frontiers of OCR". Proceedings of IEEE, 80 (7), pp. 1093-1100, 1992.

[2] H. Lieberman, F. Paternò, M. Klann and V. Wulf, "End-user development: An emerging paradigm". In End user development. Springer Netherlands, pp.1-8, 2006.

[3] L. A. Zadeh, "On the definition of adaptivity". Proceedings of the IEEE, 51(3), pp. 469-470, 1963.

[4] C. D. Wickens, J. G. Hollands, S. Banbury and R. Parasuraman, Engineering psychology & human performance. Psychology Press, 2015.

[5] J. Stroop, "Studies of interference in serial verbal reactions". Journal of Experimental Psychology, 18 (6), pp. 643-662, 1935.

[6] M. Coltheart, K. Rastle, C. Perry, R. Langdon and J. Ziegler, "DRC: A dual route cascaded model of visual word recognition and reading aloud". Psychological Review, 108 (1), pp. 204-256, 2001.

[7] J. L. McClelland and D. E. Rumelhart, . "An interactive activation model of context effects in letter perception: Part 1. An account of basic findings". Psychological Review, 88, pp. 375-407, 1981.

[8] L. Pei – Lin, "Using eye tracking to understand learners' reading process through the concept-mapping learning strategy". Computers & Education, 78, pp. 237-249, 2014.

[9] K. Rayner, "Eye movements in reading and information processing: 20 years of research". Psychological Bulletin, 124(3), pp. 372-422, 1998.

[10] L. Rello and R. Baeza-Yates, "Lexical quality as a proxy for web text understandability". Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion). ACM, New York, NY, USA, pp. 591-592, 2012.

[11] Y. Ishiguro, A. Mujibiya, T. Miyaki and J. Rekimoto. "Aided eyes: eye activity sensing for daily life". Proceedings of the 1st Augmented Human International Conference (AH '10). ACM, New York, NY, USA, Article 25, pp. 7, 2010.

[12] G. Buscher, A. Dengel, L. van Elst and F. Mittag, 2008. "Generating and using gaze-based document annotations". In CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08). ACM, New York, NY, USA, pp. 3045-3050, 2008.

[13] D. Fiset et al., "Features for identification of Uppercase and lowercase letters". Psychological Science, 19 (11), pp. 1161-1168, 2008.

[14] F. Peressotti and J. Grangier, "The role of letter identity and letter position in orthographic priming". Perception & Psychophysics, 61 (4), pp. 691-706, 1999.

[15] M. Just and P. Carpenter, "Eye Fixations and Cognitive Processes". Cognitive Psychology, 8, pp. 441-480, 1976.

 ISBN: 978-1-61208-609-5