



HAL
open science

States of the text: fixation, legitimation and multimodal publication

Antoine Fauchié

► To cite this version:

Antoine Fauchié. States of the text: fixation, legitimation and multimodal publication. Digital Humanities 2022, Alliance of Digital Humanities Organizations (ADHO), Jul 2022, Tokyo, Japan. pp.458-459. hal-04641926

HAL Id: hal-04641926

<https://hal.science/hal-04641926v1>

Submitted on 28 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

containing a 6+ letter word. The literary section of Brown corpus that includes fiction written by non-dyslexic writers was used as a reference corpus. Brown corpus is freely searchable through Sketch Engine and is accessible for downloading. We also compiled another contemporary reference for contemporary non-dyslexic American writers as a second reference corpus. De Luca et al. (2008) suggested that word length selectively influenced word recognition in impaired versus skilled readers, regardless of the moderating action of sublexical, lexical, and semantic factors. In this study, the linguistic analysis of the concorded lines retrieved from Butler's novels, aims to test whether Butler's reading deficits affected her selection of lengthy words. The categorization of 6-letter and 9+letter words demonstrated that most lengthy words were either nominalized or jargon terms. The use of lengthy verbs was, however, inconspicuous. The phonotactic probability of the most frequently used words was complex in nouns and modifiers but not in verbs. The study suggests that Butler's choice of lengthy words is a coping mechanism she used to regulate her image as a dyslexic writer. Although De Luca et al. (2008) conducted a reading-task study, the reading and comprehension are neuro-linguistically dependent on the morphological complexity of the written text. Crisp et al (2011) concluded that the severely impaired phonology and the representation of semantic-phonological impairment demonstrate that semantic representations are central to reading in the face of phonological impairment. Implications inferred from this experiment are ushered towards effective computation of word imageability, frequency of use, selection of word senses, word familiarity, word length, semantic diversity and phonological neighborhood density.

Bibliography

- Crisp, J., Howard, D., & Lambon Ralph, M. A.** (2011). More evidence for a continuum between phonological and deep dyslexia: Novel data from three measures of direct orthography-to-phonology translation. *Aphasiology*, 25(5), 615-641.
- De Luca, M., Barca, L., Burani, C., & Zoccolotti, P.** (2008). The effect of word length and other sublexical, lexical, and semantic variables on developmental reading deficits. *Cognitive and Behavioral Neurology*, 21(4), 227-235.
- Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G.** (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, 93(2), 185-238.
- Vitevitch, M.S. & Luce, P.A.** (2004) A web-based interface to calculate phonotactic probability for words

and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, 36, 481-487.

States of the text: fixation, legitimation and multimodal publication

Fauchié, Antoine

antoine.fauchie@umontreal.ca

Chaire de recherche du Canada sur les écritures numériques, Canada

Publications (books, articles, etc.) are artifacts that carry texts, and their production requires complex editing processes, calling upon different types of skills and tools. Since the beginnings of computing and following a long technical evolution, digital technology has introduced mechanisms that modify both the editorial work and the workings necessary for the production of different forms and formats (Mourat, 2020). These technological *innovations*, which are numerous and sometimes reproduce pre-digital gestures identically (Ludovico and Cramer, 2012), are created and adopted in different stages or phases (Epron and Vitali-Rosati, 2018): theoretical principles are conceptualized, imagined and then implemented in prototypes, and finally industrialized for restricted or broad uses.

We are interested here in two technical principles derived from digital publishing technologies (Blanc and Haute, 2018), namely multimodal publishing and single source publishing. On the one hand, it is about being able to generate several different formats within the same editorial project — such as a printed book on paper, a digital book in EPUB format and an XML file —, rearranging the contents according to the form of the artifact, whether in terms of structure or graphic rendering (Haute, 2019). On the other hand, it is a question of producing these versions with a single source, limiting the usually repeated interventions on several source files, and making the editorial follow-up more fluid (Hyde, 2021). These two principles — both theoretical, implemented and proven — aim to simplify work on a text in the context of an editing process: to facilitate multiple interventions by the different people who act on a text, and thus to allow interventions at different levels and moments in the text production chain. These *states* of the text, whether reading, correcting, rereading, formatting, structuring, composing, transforming, publishing, can thus be considered and composed in various ways, without the risk of losing information or creating unnecessary conflicts between the participants.

If technological evolutions allow us to imagine new ways of working in publishing, what about their real implementation? The principles mentioned are implemented through methods, software or computer programs, often put forward to the detriment of questioning the organizations of humans who use these technologies (Gelgon, 2018). Can the legitimization of content (Vitali-Rosati, 2012) be integrated into a *single source publishing* chain? If multiple people can be involved in an editorial project simultaneously, text legitimization poses a significant challenge. At what point is a text fixed? If the contents can be modified at any time, and even technically after publication, the fixation of the text is questioned, as well as its citability (the text is moving) and its durability (several versions and states of the text coexist without systematic archiving). Finally, if the technical processes are numerous to realize this type of publication chain (Maxwell, 2019), perhaps they are not all desirable. We need to take a critical look at technical feasibility. Isn't a multimodal publishing chain based on a single source a utopia in the publishing ecosystem?

Our presentation focuses on the states of the text in relation to human organizations and technical solutions. We wish to analyze the mechanisms of fixation, legitimization and publication in the perspective of editing a text in several versions from a single source. Through several examples of publication chains, we will question the technical solutions at work, and we will draw up a panorama of the theoretical questions that are part of the digital humanities approach.

Bibliography

- Blanc, J. and Haute, L.** (2018). Technologies de l'édition numérique. *Sciences Du Design*, 8(2): 11–17 <https://www.cairn.info/revue-sciences-du-design-2018-2-page-11.htm> (accessed 23 March 2019).
- Epron, B. and Vitali-Rosati, M.** (2018). *L'édition à L'ère Numérique*. (Repères). Paris, France: La Découverte <https://www.cairn.info/l-edition-a-l-ere-numerique--9782707199355.htm>.
- Gelgon, A.** (2018). Un dialogue à réaliser : Design et technique. In, *.Txt 3*. Paris, France: Éditions B42, p. 158.
- Haute, L.** (2019). Livres mécaniques et chimères numériques. *Back Office*(3) http://www.revue-backoffice.com/numeros/03-ecrire-lecran/06_haute.
- Hyde, A.** (2021). Single Source Publishing *Coko* <https://coko.foundation/single-source-publishing/> (accessed 17 November 2021).
- Ludovico, A. and Cramer, F.** (2012). *Post-Digital Print: The Mutation of Publishing Since 1894*. Eindhoven, Pays-Bas: Onomatopée.
- Maxwell, J. W.** (2019). *Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms*.

Cambridge, Massachusetts, États-Unis d'Amérique: The MIT Press <https://mindthegap.pubpub.org/> (accessed 30 December 2019).

Mourat, R. de (2020). Le vacillement des formats. Matérialité, écriture et enquête : Le design des publications en Sciences Humaines et Sociales Université Rennes 2 Thèse <http://www.these.robindemourat.com/> (accessed 22 October 2020).

Vitali-Rosati, M. (2012). Auteur ou acteur du web ? *Implications Philosophiques* <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/12980>.

An experimental attempt to use Transfer Learning for Named Entity Recognition in letters from the 19th and 20th century

Flüh, Marie

marie.flueh@uni-hamburg.de
Universität Hamburg, Germany

Lemke, Marc

marc.lemke@uni-rostock.de
Universität Rostock, Germany

In this contribution, we investigate to what extent data from one digital scholarly edition project (*Dehmel digital 1*) can be used to fine-tune a large-scale language model, which was pre-trained for the purposes of another project (*The Complete Works of Uwe Johnson 2*). We discuss the opportunities and practical limitations of such an attempt dealing with differences in the language properties of the material that was used to pre-train the transferred model and the material that is applied to fine-tune this model for the specific NER task. As part of this, we investigate to what extent the quantity and quality of training data affects the performance of NER models. This places our contribution in the line of approaches dealing with computational analysis of textual material from different periods of time (Labusch et al., 2019; Schmidt et al., 2021; Ehrmann et al. 2022).

Technical prerequisites

We use the software *NEISS TEI Entity Enricher* (NTEE).³ Following a Transfer Learning (Kamath et al., 2019) approach the tool provides access to large-scale language models in a Bidirectional Encoder Representations from Transformers (BERT) architecture (Devlin et al., 2019; Underwood, 2019) for different languages, which can be