Nejla Kalajdžisalihović

University of Sarajevo

SOCIAL LANGUAGE PROCESSING AND READING ONLINE—ON WRITTEN ERRORS IN PREDATORY JOURNALS

Abstract

The present paper discusses linguistic features of potentially predatory academic journals and the rising trend of predatory publishing where several factors are concerned: a) the researcher is unfamiliar with all the aspects of digital literacy and the changing landscape of scholarly publishing, and/or b) the researcher is eager to publish and is not a native speaker of English (Nobes, 2017). In some cases, the researcher is proficient in English and attributes errors on the journal webpages to different factors, some of which may even be linked to social language processing and personality traits (Boland, 2016; Hancock, 2010). Since Beall's list of predatory journals is no longer available, researchers abroad are starting blogs to discuss predatory publishing. Therefore, one aim of this paper is to present the most frequent linguistic features in potentially predatory journals and discuss why authors choose not to detect grammatical errors on their websites. Furthermore, this paper invites scholars to pay more attention to written errors while reading online and links responses on written errors to personality traits of the researcher and readability scores of domain-specific texts.

Key words: readability, text processing, predatory journal, error, contemporary English

1. Introduction

The present paper discusses academic writing in predatory journals and the rising trend of predatory publishing. Other than the adjective 'predatory' used to premodify the noun 'journal', adverbs modifying the adjective 'predatory' used for academic journals in this context are: 'potentially', 'possibly' or 'probably'. It is difficult to establish the degree to which a certain journal is 'potentially predatory', 'possibly predatory' or 'probably predatory'. Furthermore, according to Beall (2017), predatory journals not only publish poor quality

research but also resort to 'template plagiarism' by means of which authors re-publish previously published data and findings.

In the research conducted so far, it has been concluded that the cues to deception are not universal or identical across research domains and are context and genre-specific as well (Markowitz et al., 2014, p. 5). Since the world's academic literature is based on trust, as stated in Beall's (2017) conference talk, one aim of this paper is to introduce new terms that have begun to be used in the world of publishing as they are relevant for the lexis of contemporary English, i.e. terms coined to refer to fraud and deceit. Furthermore, a relevant framework to mention in terms of authors' international background is that of Social Language Processing (SLP), an interdisciplinary framework that joins social and psychological theory with computational techniques for modelling the relationships between discourse and social dynamics (Hancock, 2010). SLP discusses issues relevant when determining the degree to which people and groups share a common perspective.

For instance, reading is sometimes observed as a process depending on one's personality and culture and deception as a function of perceiver's culture, deceiver's intention and deceiver-deceived relationship (Seiter et al., 2002). What this means is that, for instance, personality traits of the reader may have an effect on the assessment of the language used on the journal fronts of predatory journals. Depending on a number of different factors, the reader may opt to overlook grammatical errors on journal fronts. Also, since authors often ask for their papers to be translated into the English language, it is unlikely that the language of predatory journals will be recognized as deceit by authors themselves. Therefore, it is important to pay special attention to journal fronts and the English language used on the websites of potentially predatory journals which seek authors that need to be recognized, after conferences, or after their talks.

In this paper, the grammatical errors and other features of journal fronts which tend to be overlooked by authors who are not proficient users of English will be presented. For instance, a researcher may be unfamiliar with all aspects of digital literacy (i.e. trusts the websites and accepts to read their content). Another reason for deciding to submit a manuscript to a potentially predatory journal is that the researcher may be eager to publish and tends to attribute suspicion to the changing landscape of scholarly publishing and is not a native speaker of English. In some cases, the researcher is proficient in English and attributes errors on the journal webpages to different factors (e.g. the errors are temporary and will be fixed). Even with the number of published articles on predatory publishing becoming lower, this paper also invites scholars to pay

more attention to the English language used by potentially predatory journals, 25% of all open-access journals in total being 'predatory' (Perry, 2015).

2. Corpus

The corpus used for the purpose of this paper is the website content published by Lukić et al. (2014). After browsing the content of each website and after analyzing e-mails received from potentially predatory journals, the following features have been extracted as indicators that the journal might be predatory:

2.1. Grammatical errors

Across journal fronts, i.e. websites of potentially predatory journals a consistency has been detected in the grammatical features described below:

– Errors in third person singular: (1) Each accepted paper range between 2000-5000 words.
– Missing plural: (2) data analysis in all discipline
– Incorrect past participle: (3) *charges will be pay*

2.2. Tone

In regards to the social language processing mentioned in the introduction of the present paper, the following features have been extracted to demonstrate a language of 'immediacy', i.e. a tone that places authors in an inferior position asking them to react immediately, press the 'pay' button and have their paper published, e.g.:

– (4) (authors) will be charged; each additional word will be charged
– (5) we ask authors to pay the article processing charge
– (6) (authors) can now pay by the Pay Now button

The usage of capital letters may also be an indicator that a journal is possibly a predatory one, e.g.:

– (7) ONLINE PAYMENT!!!
– (8) IMPORTANT: If you do not receive e-mail in your 'inbox', check your 'bulk mail' or 'junk mail' folders.
– (9) IMPORTANT: Sometimes you will not recieve response because emails don't go to your mail box. Always check your Junk/spam email folder.

2.3 Impersonalised title of the paper

Once sent, the paper is usually impersonalized and a reference number given for identification, e.g.:

– (10) A reference number will be mailed to the corresponding author
– (11) Paper ID 255400 - Review Report

2.4 Generic reviews applicable to any paper

Another feature in terms of the style of peer review is that in fake journals, the review may be applied to any topic, e.g.:

– (12) Discussion is a section in which the author discusses the findings of the study by relating them to 1) Aims of the research 2) Research questions raised 3) Literature review. That is, s/he should discuss, CRITICALLY, how the findings have achieved/failed to achieve the aims of the study, have answered the research question, which the study aimed to answer and how the findings of the study are 1) in line with the view of the scholars/findings of the studies surveyed in literature review 2) are against /different from the view of the scholars/findings of the studies surveyed in literature review, etc.

Interestingly, there have been several cases where a fake journal congratulated authors on their papers being accepted although the papers were not written by authors but rather— by random text generators (Sergan, 2015; Stromberg, 2014).

3. Discussion

A question arises as to what makes the author trust the journal front of a fake journal. It is proposed that, because trust is in-built in academia when it comes to publishing (or at least it has been until recently), the reader rarely notices the features described above. To test the narrativity scores of journals, ten journal fronts were analysed through Coh-Metrix, a piece of software for computing coherence and cohesion in text that has been tuned for the English language. It may be used for multiple purposes, one of them being to allow readers, writers, educators and researchers to measure the difficulty of text for a target audience. It may be used in corpus analysis, computational linguistics, education, and cognitive science.

For instance, in the ten journal fronts analysed to illustrate how Coh-Met works, the 'referential cohesion' average for the texts analysed is 37.75%. This

means that cohesion gaps require the reader to make inferences, as explained in the analysis following the result, which can be challenging and unsuccessful without prior knowledge. In Coh-Met, 'coherence' is defined as a linguistic category that contributes to the coherence of the mental representation of text and is measured by means of the repetition score, the verb phrase density, noun incidence, pronoun incidence, etc. (see: Appendix).

A low narrativity score (7.25%) indicates that the content is less story-like and such texts are more difficult to comprehend. In other words, the author "browses, reads, clicks, and submits the manuscript" because of having assumptions about the context and processes linguistic strings, thereby avoiding more complex inferencing (Žegarac, 1998, pp. 343-344).

To support this claim, there are also important results from eye-tracking studies in terms of reading online:

- "To complete their tasks efficiently, people have learned to pay attention to elements that typically are helpful (e.g. navigation bars, search boxes, headlines) and ignore those which are usually void of information" (Pernice, 2018).
- "Online readers read in an F pattern (Nielsen, 2006) and give up reading the content after the two screenfuls of content when browsing" (Fessenden, 2018).

It can be, therefore, concluded that the very nature of the multimodal platform of online journals is, at least when it comes to readability scores, a convenient platform for manipulating trust by decreasing attention (see: Galir, 2017). The author is, in that way, distracted from text containing errors pointing to discriminators which are relatively constant among journal fronts indicating a potentially common authorship.

Other content appearing at the top of journals' websites intended to distract the reader are flashing images, e-mail addresses, editorial boards that have not agreed to be affiliated with the journals, fake impact factors, acronyms, or flashing 'pay now' buttons, all appearing in the very top of the fold of the website where fixation occurs, according to eye tracking studies conducted recently (Pernice, 2018; Nielsen, 2006).

4. Conclusion

Beall (2016) defined predatory journals as "counterfeit journals [used] to exploit the open-access model in which the author pays. These predatory publishers are dishonest and lack transparency". In this paper, the focus of research

has not been on determining whether a particular journal is 'predatory', 'potentially predatory' or 'possibly predatory'.

On the contrary, the names of the journal fronts are not given in this paper as the aim was to analyse the website content by means of Coh-Met to determine and assess how the reader might be processing the English language on the journal fronts in terms of coherence, narrativity and text readability given that journal fronts are online.

It has been concluded that the text of the websites analysed, grammatical errors aside, is low in narrativity and cohesion. At the same time, the website content contains numerous distractions such as images, acronyms, e-mail addresses and names of scholars who have not agreed to be affiliated with these journals. A reader cannot know this, especially if he or she comes from a culture where scholarly publishing is based on trust.

At the same time, the features extracted may be seen to have high repetition scores, indicating a common authorship behind journals. At this moment, the area of predatory publishing is still insufficiently researched and scholars publishing in English as a second language are advised to consult other colleagues and journal evaluation tools (Reele et al. 2017) while, at the same time, awareness needs to be raised regarding the eye tracking studies and how an average reader selects what content to process while scrolling and deciding when to stop reading.

References

Beall, J. (2017). Plagiarism is everywhere: detecting and reporting plagiarism in predatory journals [video file]. Retrieved on Oct 28 2018 from https://www.youtube.com/watch?v=_X6r0b_Ivxs.

Beall, J. (2016). Essential information about predatory publishers and journals. *International Higher Education*. Number 86. Summer 2016. Retrieved on August 16 2018 from https://ejournals.bc.edu/ojs/index.php/ihe/article/viewFile/9358/8368.

Boland, J. E., Queen, R. (2016). If you're house is still available, send me an email: personality influences reactions to written errors in email messages. *PLoS ONE* 11(3). Retrieved on Oct 28 2018 from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149885.

Coh-Metrix Common Core (2012). Retrieved on Nov 4 2018 from http://129.219.222.70:8084/Coh-MetrixResearch.aspx.

Fessenden, T. (2018). Scrolling and attention. Nielsen Norman Group. Retrieved on Nov 7 2018 from https://www.nngroup.com/articles/scrolling-and-attention/.

Galir, P. (2017). Usporedba čitljivosti teksta na različitim medijima [MA thesis]. Varaždin: Sveučilište Sjever.

Lukić, T. (et al.) (2014). Predatory and fake scientific journals/publishers—A global outbreak with rising trend. A Review. Available at http://www.dgt.uns.ac.rs/pannonica/papers/volume18_3_3.pdf.

Markowitz, D.M., Powell, J.H., Hancock (2014). The writing style of predatory publishers. *American Society for Engineering Education*, 2013. ASEE Annual Conference and Exposition, Conference Proceedings.

Nielsen, J. (2006). F-shaped pattern for reading web content (original study). Nielsen Norman Group. Retrieved on Nov 7 2018 from https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/.

Nobes, A. (2017). Critical thinking in a post-Beall vacuum. Retrieved on Feb 11 2018 from https://www.researchinformation.info/feature/critical-thinking-post-beall-vacuum.

Perry, S. (2015). 'Plagiarism, fraud, and predatory publishing' are polluting science. *MinnPost*. Retrieved on Feb 11 2018 from https://www.minnpost.com/second-opinion/2015/04/plagiarism-fraud-and-predatory-publishing-are-polluting-science-says-bioethic.

Pernice, K. (2018). Banner blindness revisited: users dodge ads on mobile and desktop. Nielsen Norman Group.Retrieved on Nov 6 2018 from https://www.nngroup.com/articles/banner-blindness-old-and-new-findings/.

Reele, S. et al. (2017). Journal Evaluation Tool. *LMU Librarian Publications & Presentations*. Retrieved from https://digitalcommons.lmu.edu/librarian_pubs/40.

Seiter, J. S., Bruschke, J., & Bai, C. (2002). The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. *Western Journal of Communication,* 66(2), pp. 158-180.

Sergan, E. (2015). Why a fake article titled "Cuckoo for Cocoa Puffs?" was accepted by 17 medical journals. Retrieved on Oct 28 2018 from https://www.fastcompany.com/3041493/why-a-fake-article-cuckoo-for-cocoa-puffs-was-accepted-by-17-medical-journals.

Stromberg, J. (2014). A paper by Maggie Simpson and Edna Krabappel was accepted by two scientific journals. *Vox*. Retrieved on Oct 25 2018 from https://www.vox.com/2014/12/7/7339587/simpsons-science-paper.

Žegarac, V. (1998). What is phatic communication? In V. Rouchota and A.H. Jucker (eds.), *Current Issues in Relevance Theory* (pp. 327-361). Amsterdam: John Benjamins.

Appendix

Components of text analysed by Coh-Met applications

Narrativity: Narrativity seems intuitive: the more story-like a text the higher the narrativity score, and the easier the text. Though this is true, some texts will score high on narrativity and not seem very story-like.

Syntactic Simplicity: Syntactic simplicity is measured through several indices such as average number of clauses per sentence, the number of words per sentence, and the number of words before the main verb of the main clause. Texts with fewer clauses, fewer words per sentence, and fewer words before the main verb will give a text a higher score for syntactic simplicity. T.E.R.A. also measures the similarity of the sentences within each paragraph. Paragraphs that contain sentences with similar structures and verb tenses are easier to read.

Word Concreteness: Concrete words (mask, spoon, forest, ammunition) are words that refer to things you can see, hear, taste, touch, feel, or smell. Abstract words (democracy, appear, success, joy) cannot easily be seen, heard, touched, felt or smelled. A text with relatively high numbers of concrete words is easier to read and will have a high word concreteness score.

Referential Cohesion: Referential cohesion is the overlap between words, word stems, or concepts from one sentence to another. When sentences and paragraphs have similar words or conceptual ideas, it is easier for the reader to make connections between those ideas. Sometimes, however, low cohesion is desirable if you want the reader to create connections to understand the text.

Deep Cohesion: Deep cohesion measures how well the events, ideas and information of the whole text are tied together. T.E.R.A. does this by measuring the different types of words that connect different parts of a text. These words are called connectives. There are different types of these connectives: time connectives such as *after, earlier, before, during, while, later*; causal connectives such as *because, consequently, thus*. Then there are additive connectives such as *both, additionally, furthermore, moreover, what is more*. There are also logical connectives; *actually, as a result, due to*. Finally, adversative connectives are words that connect two phrases or notions that on some level conflict with each other, such as "My favorite sport is baseball. **However**, I watch more football" or "Whales are not fish **yet** they spend their life in the water." Some examples of adversative connectives are: but, yet, however, although, nevertheless.

All of these connectives help to tie the events, ideas and information in the text together for the reader (Coh-Metrix Common Core, 2012).

DRUŠTVENI FAKTORI U PROCESIRANJU JEZIKA I ČITANJU U ONLAJN OKRUŽENJU — PRAVOPISNE GREŠKE U PREDATORSKIM ČASOPISIMA

Sažetak

U radu se govori o jeziku potencijalno predatorskih akademskih časopisa čiji broj se u posljednje vrijeme povećava usljed nekoliko faktora: a) istraživač nije upoznat sa svim aspektima digitalne pismenosti zbog činjenice da se priroda izdavaštva akademskih radova mijenja i/ili je b) istraživač želi objaviti rad u časopisu na engleskom jeziku pri čemu istraživač nije izvorni govornik engleskog jezika (Nobes, 2017). U nekim slučajevima, istraživač odlično poznaje engleski jezik ali pripisuje greške na internet stranicama predatorskih časopisa različitim faktorima, pri čemu neki od njih mogu biti povezani sa društvenim faktorima u procesiranju jezika ili ličnim stavovima istraživača (Boland, 2016; Hancock, 2010). S obzirom na to da Beallova lista predatorskih časopisa više nije dostupna, mnogi autori naučnih radova u inostranstvu počinju pisati blogove na temu predatorskih časopisa. S tim u vezi, jedan od ciljeva ovog rada je predstaviti najfrekventniji jezični sadržaj koji upućuje na to da se radi o potencijalno predatorskom časopisu i razmotriti iz kojeg razloga autori ne obraćaju pažnju na gramatičke i pravopisne greške na internet stranicama predatorskih časopisa. Ovaj rad također poziva autore da detaljnije analiziraju gramatičke i pravopisne greške dok čitaju sadržaj internet stranica naučnih i stručnih časopisa na engleskom jeziku i, s tim u vezi, povezuje stav prema gramatičkim i drugim greškama s ličnim stavom istraživača i čitljivošću teksta.

Ključne riječi: čitljivost, procesiranje teksta, predatorski časopis, greška, savremeni engleski jezik