

Clausal and phrasal complexity in research articles published in well-established and predatory journals: A case study of two journals in political science

Ying Wang, Karlstad University, Sweden
Josep Soler, Stockholm University, Sweden

Predatory publishing has attracted much scholarly attention recently, but little is known about the actual material published in predatory journals. In this paper, we address this gap focusing on syntactic complexity. Using both traditional syntactic complexity measures and more fine-grained indices of phrasal and clausal complexity, the study explores the similarities and differences between two corpora consisting of 220 research articles drawn from two comparable journals in the discipline of Political Science, one purportedly predatory and one top-ranking. The results show that the articles look similar in many respects (e.g., mean length of sentences/T-units, number of T-units per sentence). Differences are found in more fine-grained indices such as clausal complements, adverbial clauses, and noun phrases with noun pre-modifiers, which are associated with discipline-specific rhetorical and ideational functions. The study demonstrates the potential of linguistic analyses in contributing to our understanding of predatory publishing as a complex phenomenon.

Keywords: predatory publishing, research articles, syntactic complexity, political science, disciplinary writing

1 Introduction

Predatory publishing is a relatively recent phenomenon that has drawn increasing interest amongst researchers from different disciplines, most notably from scholarly communication sciences (e.g., Björk et al., 2020; Xia et al., 2015). Since 2008, when Beall first coined the term “predatory Open-Access journals” in the no-longer available *Scholarly Open Access* blog (Fazel & Heng Hartse, 2020, p. 184), much has been written about predatory publishing, both in the form of empirical studies, uncovering some of the key features of the phenomenon at a macro level (e.g., Shen & Björk, 2015), as well as in the form of letters to the editor in different journals (mostly within the medical sciences), warning authors about the risks and dangers of such journals (e.g., Cortegiani & Shafer, 2018). However, as a

Corresponding author’s email: ying.wang@kau.se

eISSN: 1457-9863

Publisher: University of Jyväskylä, Language Campus

© 2023: The authors

<https://apples.journal.fi>

<https://doi.org/10.47862/apples.120753>

much-debated and controversial issue, a widely agreed-upon definition of “predatory journals and publishers” was not made available until 2019. After much discussion by a group of 35 leading experts from different disciplines and countries, the following working definition was proposed:

Predatory journals and publishers are entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices (Grudniewicz et al., 2019, p. 211).

So far, empirical studies on predatory publishing have been focused largely on some of the most concerning and unethical aspects of the phenomenon, such as a lack of proper peer-review (Richtig et al., 2018) and use of plagiarism (Martin & Martin, 2016). Other studies have attempted to uncover the reasons why some scholars do end up selecting predatory journals as an outlet to have their work published (e.g., Cobey et al., 2019). Some of the reasons (e.g., authors’ self-reported lack of research proficiency, social identity threat as non-Western scholars, see Kurt, 2018) suggest predatory publishing is far more complex than being a matter of black or white. Until recently, little is known about the actual material published in predatory journals, probably due to an unquestioned presumption that whatever gets published in such journals must be of poor quality and therefore unworthy of serious consideration. However, it is increasingly recognised that knowing more about the actual material published in predatory journals has the potential to add fruitfully to our understanding of this complex phenomenon (Eriksson & Helgesson, 2017; Fazel & Heng Hartse, 2020). Indeed, in two of our recent studies addressing this gap (Soler & Wang, 2019; Wang & Soler, 2021), we have found that writers publishing in a selected predatory journal were based in higher education institutions all over the world. In other words, the phenomenon seems to affect scholars not only in developing countries in Africa and Asia, as suggested in the literature (e.g., Shen & Björk, 2015; Xia et al., 2015), but also those in developed countries in West Europe and North America. Additionally, while there are indeed articles that suggest unethical practice, there is also no shortage of what appear to be ‘serious’ papers, judging from some surface features such as length, structure and language use. A closer examination of the actual texts using corpus linguistics methods (e.g., keywords, lexical bundles), however, brought to light some key linguistic differences between the articles published in the predatory journal and those published in a top-ranking journal of the same field, which in turn indicated generally less sophisticated research methods and a lack of disciplinary writing skills in the former (Soler & Wang, 2019; Wang & Soler, 2021).

Building on our own previous work, in the present article, we turn our attention to syntactic complexity to dig deeper into the disciplinary nature of the texts in our corpora, hoping to contribute towards a better understanding of the shape of this selection of texts (e.g., what is present, what is missing or lacking) and what it can potentially inform us of the complexity of the phenomenon of predatory publishing.

We organise the article as follows: in the next section, we present a succinct literature review on syntactic complexity in academic writing research, followed

by a presentation of our material and methods in Section 3. Section 4 outlines the results and discussion, followed by the conclusions in the final section.

2 Syntactic complexity in academic writing research

Syntactic complexity is a concept widely used in the fields of second language acquisition and academic writing research, referring to the range of grammatical structures/units in language production and the degree of sophistication of such structures (Ortega, 2003). A large number of indices have been proposed to measure syntactic complexity. Traditionally, the indices have been focused on the clausal level, including those pertaining to the length of different production units (e.g., clauses, sentences, T-units) and the amount of clausal embedding and coordination. More recently, complexity at the phrasal level, such as the amount and range of modification in complex noun phrases, has started to garner attention, particularly in academic writing research (e.g., Biber et al., 2011; Kyle & Crossely, 2018).

Measures of syntactic complexity have proved to be important research tools in second language writing research to answer questions on the relationship between syntactic complexity and writing proficiency as well as the role of various variables in this relationship, including operationalisation of proficiency (e.g., school levels, program levels, holistic ratings), sampling condition (e.g., task type, timing of the writing, corpus size), and learner background (e.g., L1, L2 learning context) (Lu, 2010; Ortega, 2003).

The syntactic complexity measures are also used fruitfully in academic writing research. To begin with, Biber et al. (2011) demonstrated with empirical evidence that complex noun phrases and non-clausal phrases are characteristic of academic writing, whereas clausal subordination, which would increase the length of a T-unit, among others, is a feature of conversation (see also Biber et al., 1999). This observation was corroborated in a few other studies (e.g., Crossley & McNamara, 2014; Lu, 2011; Taguchi et al., 2013). Lu (2011), for instance, found that measures of complex nominals and coordinate phrases are stronger indices of quality in English as a Second Language (ESL) college-level writing than most other measures. As a consequence, the role of phrasal embedding has attracted increasing attention in academic writing research over the traditional indices of syntactic complexity that focus on embedded clauses (e.g., Staples et al., 2016). In terms of acquisition, finite dependent clauses are predicted to be acquired at earlier stages of writing development before non-finite dependent clauses and dependent phrases (Biber et al., 2011). In addition, the use of phrasal complexity features was found to increase as the writer's academic level increases (Staples et al., 2016).

More recently, the construct has been employed to help uncover characteristics of the genre of research articles (RA) as produced by various writer populations. Ansarifard et al. (2018), for example, revealed that PhD and expert writers used a greater number and range of noun modifiers (e.g., pre-modifying nouns, *-ed* participle as post-modifiers, adjective-noun sequences as pre-modifiers, and prepositional phrases as post-modifiers) than MA students in RA abstracts, providing empirical evidence for the development of academic writing, which becomes more complex with experience. Through a comparison of RA manuscripts produced by English as a Lingua Franca (ELF) writers against those

authored by L1 American English speakers, Wu et al. (2020) revealed that ELF writers tended to use longer sentences and more coordinate phrases and complex nominal phrases than their L1 counterparts. Additionally, they provided interesting insights into the motivations of ELF writers to produce such complex structures. Long sentences seemed to grow out of the need to improve communication efficiency and coordinate phrases and complex nominal phrases were used to enhance clarity. In fact, there has been an increasing appreciation of the role of syntactic structures in realising various discursal, rhetorical, and communicative functions in RA writing practices among different writer populations in the most recent research in the field (see, e.g., Casal et al., 2021; Lu et al., 2020). Writers publishing their work in predatory journals can be regarded as an additional population that has been under-explored. Using syntactic complexity measures, the present study is an attempt to explore syntactic structures that characterise their writing in the hope of contributing to our understanding of academic writing development and discussing possible pedagogical implications.

Discipline, which is a key factor that affects academic writing practices (e.g., Hyland, 2008, 2012), has been established as an important parameter when it comes to syntactic complexity. Disciplinary variation at both levels (phrasal vs. clausal) has been observed in the literature. Broadly speaking, RAs in Humanities were found to use phrasal complexity features less extensively than those in Natural Sciences (Biber, 2006; Biber & Gray, 2016), which can be explained partly by the “more explicitly interpretative and less empiricist” nature of the Humanities domain (Hyland, 2008, p. 16). In addition, Staples et al. (2016) reported that premodifying nouns are used more frequently in Life and Physical Sciences, whereas nouns, nominalisations, and attributive adjectives occur more frequently in Social Sciences, and Arts and Humanities feature in particular *of* genitives and prepositional phrases. With regard to clausal features, finite clauses were found occurring most frequently in Arts and Humanities and least in Life and Physical Sciences, with Social Sciences lying in-between. These studies promise great potential in broadening the disciplinary focus in the study of syntactically complex structures. As Casal et al. (2021) highlighted, considerable degrees of disciplinary variation can exist even within the same broad disciplinary domain, reinforcing the importance of discipline specificity in academic writing research. The present study, focusing on one specific discipline (Political Science), which has not been studied previously in terms of syntactic complexity, will therefore contribute to the expansion of this line of research.

Despite all the advantages of syntactic complexity measures, as Ortega (2003) rightly cautioned, what they provide are quantitative results, or in other words, a starting point for the researcher to search for evidence in the data in order to gain a more nuanced understanding of the data. Using a selection of clausal and phrasal complexity measures, in combination with a qualitative analysis, the present study compares two corpora of RAs drawn from two journals in Political Science, one alleged predatory and one well-established, in order to answer the following research questions:

- 1) What are the similarities and differences between the two corpora in terms of the selected clausal and phrasal complexity measures?

- 2) How are different syntactic structures used in the two corpora in terms of frequency and function?

3 Data and methods

3.1 Corpora

In this study, we employed the same two corpora as used in Soler and Wang (2019) and Wang and Soler (2021): PJPS (Predatory Journal in Political Science) and TRJPS (Top-Ranking Journal in Political Science). Apart from the discipline, the two journals involved in the corpora are comparable in terms of aims and scope. The journal selected for the PJPS corpus is regarded as predatory for the following considerations. To begin with, it is not included in the relevant indexing database (e.g., Directory of Open Access Journals), and its publisher was included in the latest list of alleged predatory publishers, developed by Beall (2017) on the basis of the recommendations from the Committee on Publication Ethics (COPE). In addition, the journal also satisfies 14 out of 25 features of predatory journals listed by Eriksson and Helgesson (2017, p. 165), including providing an unrealistic impact factor calculation (based on Google Scholar), and engaging in email invitations sent out to authors with an expertise outside the journal's remit, with unrealistic time-frames leading to publication. The comparison journal involved in the TRJPS corpus is published by a well-known and prestigious academic publishing house and has been consistently ranked among the top journals in the field (ISI Journal Citation Reports). The status of the two journals was also confirmed by two expert informants in the field.

Considering changes over time in academic writing (e.g., Hyland & Jiang, 2018), the selection of articles was limited to the most recent time period, namely, the 2000s. The 110 RAs drawn from each journal were evenly distributed across the period. To build up the two corpora for the current analysis, the PDF files of the selected articles were converted to plain text files, which were then cleaned to exclude author names and affiliations, tables, figures, and formulas that occur in the original articles. The occurrences of noisy data as a result of data conversion were also dealt with in the cleaning process. Table 1 presents the word counts of the two corpora under investigation.

Table 1. Data used for the study

Corpora	Years	No. of articles	No. of words	Average length of articles
PJPS	2011-2018	110	653,997	5,665
TRJPS	2001-2018*	110	1,065,960	8,352

* The TRJPS corpus covers a slightly longer time span than does the PJPS corpus as the latter is a newer journal.

Ideally, more journals should be included in the study to allow generalisation. However, it is difficult to find journals with more or less the same scope and coverage on each end and the inclusion of more journals would complicate matters by introducing variables (e.g., topics and genres, author diversity and demographics) that may affect the comparability of the samples. We therefore

opted for a one-to-one journal comparison to ensure comparability to the greatest extent.

3.2 Syntactic complexity measures under investigation

Two computational systems were used to measure a diversity of syntactic structures: the syntactic complexity analyser (SCA; Lu, 2010) for traditional measures of syntactic complexity, and the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016) for more fine-grained indices of phrasal and clausal complexity.

Table 2 summarises the 14 measures incorporated in SCA, which can be categorised into five types that are concerned with: length of production unit (MLC, MLS, MLT), amount of subordination (C/T, CT/T, DC/C, DC/T), amount of coordination (CP/C, CP/T, T/S), degree of phrasal sophistication (CN/C, CN/T, VP/T), and overall sentence complexity (C/S). The system adopts the most commonly used definitions for relevant production units and syntactic structures; see Table 3 for the definitions.

Table 2. Syntactic complexity measures included in SCA (adapted from Lu, 2017, p. 503)

Label	Description
MLC	Mean length of clause
MLS	Mean length of sentence
MLT	Mean length of T-unit
C/T	Number of clauses per T-unit
CT/T	Number of complex T-units per T-unit
DC/C	Number of dependent clauses per clause
DC/T	Number of dependent clauses per T-unit
CP/C	Number of coordinate phrases per clause
CP/T	Number of coordinate phrases per T-unit
T/S	Number of T-units per sentence
CN/C	Number of complex nominals per clause
CN/T	Number of complex nominals per T-unit
VP/T	Number of verb phrases per T-unit
C/S	Number of clauses per sentence

Table 3. Definitions for relevant production units and syntactic structures counted in SCA (adapted from Kyle & Crossley, 2018, p. 338)

Structure	Definition	Examples
Verb phrase	a finite or non-finite verb phrase (or reduced clause)	<i>ate pizza</i> <i>was hungry</i>
Complex nominal	nouns with modifiers, nominal clauses, gerunds and infinitives in subject position	<i>red car</i> I know <i>that she is hungry</i> <i>Running</i> is invigorating
Coordinate phrase	adjective, adverb, noun and verb phrases connected by a coordinating conjunction	She <i>eats pizza and smiles</i>
Clause	a structure with a subject and a finite verb ¹	<i>I ate pizza</i> <i>Because I was hungry</i>
Dependent clause	a finite clause that is a nominal, adverbial, or adjective clause	<i>I ate pizza because I was hungry</i>
T-unit	an independent clause and any clauses or non-clausal structures attached to or embedded in it	<i>I ate pizza</i> <i>I ate pizza because I was hungry</i>
Complex T-unit	a T-unit that contains a dependent clause	<i>I ate pizza because I was hungry</i>
Sentence	a group words bounded by sentence-ending punctuation (., ?, !,...)	<i>I went running today.</i>

Based on the results of SCA, a selection of more fine-grained clausal (e.g., number of adverbials per clause) and phrasal indices (e.g., number of adjectives per noun phrase) included in TAASSC were further analysed; see Tables 4 and 5 for an overview of the relevant indices.

¹ Sentence fragments punctuated by the writer that contain no overt verbs are also classified as clauses. Non-finite verb phrases (or reduced clauses) are excluded in the definition of clauses but are included in the definition of verb phrases.

Table 4. Selected clausal indices analysed by TAASSC (adapted from Kyle & Crossley, 2018, pp. 339–340)

Label	Structure	Description	Example
acomp	Adjective complement	an adjective that functions as a complement in a copular clause	She looks <i>beautiful</i> .
advcl	Adverbial clause	a clause modifying a verb phrase	The accident happened <i>as night fell</i> .
agent	Agent	the conceptual subject in a passive clause	The man has been killed by <i>the police</i> .
cc	Clausal coordination	clauses joined by a coordinating conjunction	Jill runs and <i>Jack jumps</i> .
ccomp	Clausal complement	a dependent clause serving as a complement	I am certain <i>that he did it</i> .
pcomp	Clausal prepositional complement	a prepositional phrase including a clausal prepositional object	They heard <i>about you missing classes</i> .
csubj	Clausal subject	a clause functioning as the subject of another clause	<i>What she said</i> is not true.
xcomp	Open clausal complement	a non-finite clausal complement	I am ready <i>to leave</i> .
parataxis	Parataxis	clauses or phrases inserted into a clause with no explicit markers of coordination or subordination	That man, <i>Jack continued</i> , is dangerous.
csubjpass	Passive clausal subject	a clause that serves as the syntactic subject of a passive clause	<i>That she lied</i> was suspected by everyone.
mark	Subordinating conjunction	a conjunction marking a subordinate clause	Forces engaged in fighting <i>after insurgents attacked</i> .

Table 5. Selected phrasal indices analysed by TAASSC (adapted from Kyle & Crossley, 2018, p. 341)

Label	Structure	Description	Example
amod	Adjectival modifiers	an adjective that modifies a noun or noun phrase	The man in the <i>red</i> hat gave that <i>tall</i> man some money.
prep	Prepositional phrases	a prepositional phrase that modifies a noun or noun phrase	The man <i>in the red hat</i> gave that tall man some money.
vmod	Verbal modifiers	a non-finite verb or verb phrase that modifies a noun or noun phrase	I don't have anything <i>to say</i> to you.
nn	Noun modifiers	a noun that modifies a noun or noun phrase	<i>Oil</i> prices are rising.
rmod	Relative clause modifiers	a relative clause that modifies a noun or noun phrase	I saw the person <i>you love</i> .
advmod	Adverbial modifiers	an adverb that modifies a noun or noun phrase	Today was a <i>really</i> hot day.
conj_and	Conjunction "and"	the conjunction "and" when used to join two nouns or noun phrases	Jack <i>and</i> Jill

4 Results and discussion

In this section, we first present the results of the comparison in respect of the 14 SCA measures, which will then lead to a discussion of a selection of TAASSC measures in the two corpora. Table 6 presents the syntactic complexity values analysed by SCA, together with the results of one-way ANOVA tests that were conducted to explore statistical significance.

Table 6. Syntactic complexity (SCA) values in the two corpora

Measure	TRJPS		PJPS		One-way ANOVA		
	Mean	SD	Mean	SD	F value	Sig.	η_p^2
MLC	13.4	1.38	14.5	2.84	13.52	< .001	.058
MLS	26.3	3.57	25.3	5.03	2.811	.095	.013
MLT	24.1	3.16	23.5	4.44	1.01	.316	.005
C/S	1.96	.235	1.76	.298	31.13	< .001	.125
C/T	1.80	.191	1.63	.218	34.18	< .001	.136
CT/T	.517	.0719	.435	.110	42.56	< .001	.163
DC/C	.433	.052	.365	.0779	57.11	< .001	.208
DC/T	.786	.178	.613	.206	44.91	< .001	.171
CP/C	.371	.108	.482	.203	25.5	< .001	.105
CP/T	.660	.185	.771	.296	11.15	< .001	.049
T/S	1.09	.0504	1.08	.0795	3.839	.0514	.017
CN/C	2.01	.295	2.16	.509	6.948	< .01	.031
CN/T	3.61	.611	3.50	.807	1.204	.274	.005
VP/T	2.46	.302	2.19	.345	38.71	< .001	.151

A few points can be made from Table 6. First of all, the values of all measures (except for T/S) are higher than those reported in Lu (2011, p. 490) on college-level second language writing data (timed argumentative essays), thereby supporting the view that syntactic complexity increases as the writer becomes more experienced in writing. Secondly, the standard deviation (SD) values in TRJPS are all, without exception, lower than the PJPS counterparts, which means that there is less variation in the articles published in the top-ranking journal in terms of these syntactic complexity measures. As a type of “socially situated practice” (Hyland, 2012, p. 60), every academic discipline has its own conventional ways of making sense of and communicating specialist knowledge, which are shared by members of the community. This view is supported by mounting evidence from formulaic language research (e.g., Hyland, 2008; Wang, 2018). Wang (2018), for instance, found that expert writing demonstrates a lower degree of formal variability in what seem to be semantic transparent and syntactically flexible expressions such as *as can be seen* than does novice writing, indicating that even with these seemingly freely-generated expressions, there are ‘conventional’ ways that help distinguish members of the discourse community from new-comers. The lower SD values in the TRJPS corpus in Table 6 may therefore be taken as further evidence for the conventionality of disciplinary writing, manifest in syntactic complexity as well. At the same time, given the wider variation in the syntactic complexity scores among the articles published in the predatory journal, there is reason to argue that such conventionality is yet to be established in the writing of this particular group of writers.

Out of the 14 measures, four demonstrate no significant differences between the two corpora: mean length of sentences/T-units (MLS, MLT), number of T-

units per sentence (MLT), and number of complex nominals per T-unit (CN/T). In the remaining measures, the TRJPS corpus yielded a higher score for C/S, C/T, CT/T, DC/C, DC/T, VP/T (the effect size is interpreted as large with $\eta_p^2 > 0.1$ for all measures), meaning the use of a larger number of clauses and dependent clauses in general. Note that reduced clauses are counted as verb phrases (non-finite) by the system; therefore, the higher mean of verb phrases per T-unit in the TRJPS corpus may be associated with the same trend in terms of clause embedding.

By contrast, the predatory journal articles see a greater amount of coordination with more coordinated phrases per clause or T-unit (CP/C, CP/T); the effect size is interpreted as medium for CP/C but small for CP/T. Indeed, a search for one of the coordinating conjunctions (*and*) revealed a significant overuse in the PJPS corpus (3.25/100 words vs. 2.74/100 words in TRJPS, Log-likelihood = 348.79, $p < .0001$). Table 7 presents the results of a few measures involving coordinating conjunctions included in TAASSC.

Table 7. Values of relevant coordination measures included in TAASSC

Measure	TRJPS		PJPS		One-way ANOVA		
	Mean	SD	Mean	SD	F value	Sig.	η_p^2
conj_and per nominal	.0587	.0158	.0783	.0286	39.62	< .001	.154
conj per cl	.0689	.0191	.0901	.0304	38.35	< .001	.150
cc per cl	.00543	.00406	.00674	.00762	2.546	.112	.012

As shown in Table 7, coordination occurs least frequently at the clausal level (cc per cl), on which there is also no significant difference between the two corpora, while coordinated verb phrases (conj per cl) and noun phrases (conj_and per nominal) are more frequently seen in PJPS than in TRJPS (the effect size is interpreted as large for both). Examples (1) and (2) illustrate how *and* is used to combine noun phrases (e.g., *the goods and services*) or pre-modifiers (e.g., *early-morning and late-night timeslots*), and verb phrases (e.g., *became...and soon thereafter became, undergo...and invade*), resulting in fairly long clauses. On average, the clauses (MLC) in the PJPS corpus are significantly longer than those in the TRJPS, which may be partly attributable to the extensive use of phrasal coordination in the former.

- 1) While talk radio matured over a period of decades, moving from early-morning *and* late-night timeslots into primetime *and* all-talk formats (REF), the political blogosphere became an important organizing *and* fundraising tool after just a few years, *and* soon thereafter became an important part of elite American political discourse. (PJPS64)
- 2) The markets undergo an expansion of the goods *and* services they regulate (*and* therefore subject to the logic of short-term profit, to inequality *and* to neglect of “external diseconomies”), *and* invade functions that corresponded to the public sector *and* its democratic criteria of accountability. (PJPS11)

Table 8 presents the results of measures pertaining to clausal embedding in TAASSC.

Table 8. Values of relevant clausal complexity measures included in TAASSC

Measure	TRJPS		PJPS		One-way ANOVA		
	Mean	SD	Mean	SD	F value	Sig.	η_p^2
mark per cl	.153	.0300	.127	.0386	31.62	< .001	.127
ccomp per cl	.0955	.0205	.0832	.0300	12.63	< .001	.055
advcl per cl	.0666	.0142	.0519	.0184	43.8	< .001	.167
xcomp per cl	.0643	.0151	.0772	.0218	26.44	< .001	.108
rcmod all nominal	.0360	.00829	.0346	.0104	1.228	.269	.006
parataxis per cl	.0178	.00779	.0176	.0113	.027	.869	.0001
csubj per cl	.00883	.00417	.00663	.00462	13.66	< .001	.059
pcomp per cl	.00131	.00131	.00150	.00210	.633	.427	.003
csubjpass per cl	.000520	.000733	.000789	.00144	3.057	.0818	.014

We can see from Table 8 that, overall, there is a greater amount of subordination (mark per cl) in TRJPS than in PJPS (η_p^2 is > 0.1), which is consistent with what can be inferred from the SCA results. No significant differences were found between the two corpora regarding the following measures: pcomp (clausal complement that consists of a prepositional phrase), parataxis (inserted clauses with no explicit markers of coordination or subordination), and csubjpass (clauses serving as the syntactic subject of a passive clause), rcmod (relative clauses modifying a noun phrase). These types, together with csubj (clausal subjects), are also less commonly used than, for instance, ccomp (clausal complements; medium effect size: $.01 < \eta_p^2 < .06$), advcl (adverbial clauses; large effect size: $\eta_p^2 > .14$), xcomp (non-finite clausal complements; large effect size: $\eta_p^2 > .1$), wherein lie significant differences between the two corpora: more clausal complements and adverbial clauses in TRJPS and more non-finite clausal complements in PJPS.

Words that tend to be followed by *that*-clausal complements in the TRJPS corpus include verbs such as *assume, imply, suggest, indicate, expect, find, argue, demonstrate, show, estimate*, adjectives such as *clear, evident, likely, plausible*, and nouns such as *belief, view, evidence, possibility*. Most of these words can be regarded as stance markers, used to convey the writer's attitude or evaluation. Looking at the same data from the perspective of lexical bundles (i.e., recurrent word combinations), Wang and Soler (2021) reported a higher proportion of stance making bundles in the TRJPS corpus, with a myriad of expressions featuring the same underlying syntactic pattern, such as *we find that, we show that, we expect that, and we argue that*, as in Example (3). The more frequent use of complement clauses in the top-ranking journal articles is likely to be motivated by the same need to establish authorial voice and stance. Examples (4) to (7) illustrate some additional patterns that have emerged from the same corpus. Another element most of these examples have in common is the presence of the writer in the form of first-person pronouns (*we, our*), by which the message is endowed with the quality of directness and unambiguity.

- 3) On a three-judge panel where W represents a single judge, *we argue that* a is positive due to the fact that some costs of reversal fall on the entire court. (TRJPS10)
- 4) *Our findings suggest that* a single dimension works well. (TRJPS44)

- 5) *It is plausible that* a party gains votes under nonproportional rules but does not gain to the same degree as another party in the election. (TRJPS69)
- 6) Second, *we find it striking that* citizens' coalition-based inferences about parties' positions on Europe are not supported by expert judgments or by content analyses of party manifestos, especially given the growing salience of Europe as displayed in the bitter public debates over the financial assistance packages offered to distressed economies in Greece, Spain, Ireland, and Portugal, and the growth of populist, anti-European integration parties such as Golden Dawn in Greece, the French National Front, Italy's Five Star movement, and the Dutch Party for Freedom. (TRJPS101)
- 7) *It is our view that* calculations of relative advantage for the two largest parties under the given sequential allocation rules determined the decision to have 10 ministries. (TRJPS67)

Moving on to adverbial clauses, the following examples (8-13) should give a sense of how adverbial clauses led by *as* are used in the TRJPS corpus. As the examples demonstrate, *as*-adverbials fulfil a multitude of functions, including structuring discourse (i.e., referring to information in other parts of the text) as in (8) and (9), introducing a stance (10), citing sources (11), making comparisons (12), and denoting conditions (13). Again, some of these functions, such as text-structuring signals associated with expressions *as can be seen* and *as described* (somewhere), were found to stand out in the top-ranking journal articles under investigation (Wang & Soler, 2021). The results of the present study underline the need for broadening the scope of investigation from individual expressions that occur frequently to syntactic structures and their functions which may also help distinguish different disciplines and writer groups.

- 8) *As can be seen from the figure*, contemporary governments have had more consistently unified support in the legislature than in the post-War. (TRJPS87)
- 9) *As described above*, comparing our control group (marginal eligibles) to our treatment group (marginal ineligibles) within a narrow range around the treatment cutoff (ineligibility to vote) allows us to look at the impact of exogenous variation in preregistration on turnout. (TRJPS01)
- 10) Furthermore, *as expected*, we found the effect of casualties to be most pronounced among those least interested in politics. (TRJPS05)
- 11) It may indeed be the case, *as Fearon (1997) suggests*, that hand-tying is more commonly used as a commitment strategy during international crises rather than beforehand, but our analysis suggests that tying hands by engaging one's reputation can prevent crises as well. (TRJPS89)
- 12) Respondents who rate the candidates equally are treated as missing (*as are those who cannot decide later in the sample*). (TRJPS26)
- 13) *As casualties increase*, low-attention respondents become increasingly likely to vote, whereas voting among those who pay greater attention to politics is relatively insensitive to casualties. (TRJPS05)

The PJPS corpus, in contrast, yielded more non-finite clausal complements on average, most typically in the form of *to*-infinitives and gerunds as in Examples (14) to (16). Such

structures are of course not unique to the PJPS corpus; they are merely more frequently used by comparison. As can be seen from the examples, these non-finite clausal complements are often controlled by evaluative adjectives or nouns, particularly in a syntactic structure involving an empty subject *it*. Unlike in Examples (3), (4), (6), (7) taken from the TRJPS corpus, where the authors do not hide away from the opinions expressed, the writers' opinions in Examples (14) and (15) are presented in a 'faceless' way, with the writer staying at a distance from the claim made. Objectivity is a key feature of academic style and the use of *it* as an empty subject is one strategy taught at academic writing courses to help the writer establish an objective distance from the proposition. However, it has emerged from the comparison in the present study, as well as that in Wang and Soler (2021), that writers publishing in the top-ranking journal actually prefer a more personal style, which may well be discipline-specific. Indeed, as recognised by the expert informants in the field, this more personal style allows the writers, as they are also expected, to take responsibility for what they claim.

- 14) The analysis has clearly shown the favouring of certain types of knowledge over others: while *it is difficult to* determine the exact impact of knowledge on the policy change process, the case-study demonstrates that different types and uses of knowledge are *essential to* achieve policy change outcomes. (PJPS101)
- 15) In this regard, *it is important to* note that Mar del Plata is distinguished by the historical reception of international migration flows, especially since 1890. (PJPS01)
- 16) The patron-client relationship within the Middle Eastern countries has achieved *success in* maintaining American and Allied security interests in the region. (PJPS09)

CN/C (complex nominals per clause), which is seen as a more accurate indicator of academic writing quality than the number of complex clauses (Biber et al., 2011; Staples et al., 2016), has a higher value in the PJPS corpus than in the TRJPS counterpart. Table 9 presents the results of a number of measures related to complex nominals incorporated in TAASSC: verbal modifiers (vmod), adjective modifiers (amod), adverbial modifiers (advmod), prepositional phrases (prep), and pre-modifying nouns (nn).

Table 9. Values of relevant phrasal complexity measures (complex nominals) included in TAASSC

Measure	TRJPS		PJPS		One-way ANOVA		
	Mean	SD	Mean	SD	F value	Sig.	η_p^2
vmod all nominal	.0315	.00958	.0308	.0109	.308	.579	.001
advmod all nominal	.0186	.00459	.0161	.00697	9.835	< .01	.043
amod all nominal	.318	.0543	.306	.0663	2.299	.131	.010
prep all nominal	.265	.0369	.282	.0466	8.991	< .01	.040
nn all nominal	.179	.0496	.159	.0706	5.806	< .05	.026

As can be seen from Table 9, there is no significant difference between the two corpora in terms of verbal modifiers and adjective modifiers. Table 8 also shows no difference in respect of relative clause modifiers. Significant differences were found in the scores of *advmod*, *prep*, and *nn*, with more adverbial modifiers and noun modifiers, but fewer prepositional modifiers in TRJPS than in PJPS. The size of the difference between the two corpora is regarded as medium ($.01 < \eta_p^2 < .06$) for all the three measures. Therefore, the number of prepositional modifiers per nominal may be largely accountable for the higher CN/C ratio in the predatory journal articles.

A search of the preposition *of* in the two corpora revealed that it is indeed overused in the predatory journal articles: 4.34/100 words in PJPS vs. 3.62/100 words in TRJPS (Loglikelihood = 532.49, $p < .0001$). The main patterns of usage are demonstrated in Examples (17) to (19), including *of* genitive (e.g., *the role of, the form/size/scope/number of, the hypothesis/theory of*), *of* complement (e.g., *the idea of, the probability of, an example of*), nominalisation of (e.g., *the spread of, the development of*), and formulaic sequences (e.g., *as a consequence of, on the basis of, by means of, on behalf of*).

- 17) This is an important argument, but one that does not consider the historical experience *of* France and Frenchmen, and which leaves out the fact that it was Frenchmen, elected by other Frenchmen in a democratic nation, who voted on the destiny *of* France by rejecting the EDC. Perhaps the National Assembly vote was a bad choice in the eyes *of* the outside world; but it was nevertheless the choice *of* France. (PJPS97)
- 18) According to this theory, the appearance *of* a sophisticated floating voter can be seen as a consequence *of* the process *of* cognitive mobilization. That process encompasses two distinct aspects: on the one hand, the spread *of* education which increases citizens' cognitive skills, and on the other hand, the development *of* mass media which decreases the cost *of* acquiring political information... (PJPS30)
- 19) In other words, the fundamental idea *of* gradual social and economic convergence (and/or integration) has been substituted by the political notion *of* divergence. (PJPS112)

Examples (18) and (19) give a taste of the extensive use of nominalisations in the data, either as the head noun of a noun phrase (*appearance, spread*) or as a complement led by *of* (*mobilization, convergence, divergence*). Like the use of *it* as an empty subject together with an embedded non-finite clause to distance the writer from their assessment, nominalisation represents another common linguistic feature of academic writing in the interests of conciseness. By turning clauses into noun phrases, nominalisation is a useful tool for packing more information into a single sentence. Again, we can see that the writers contributing to the predatory journal seemed to incorporate some common features of academic writing to a greater extent than did the writers publishing in the top-ranking journal.

In contrast to the prominent status of *of*-phrases in the predatory journal articles, adverbial and noun modifiers stand out in the top-ranking counterparts.

Starting from adverbial modifiers, a quick scan of words ending with *-ly*² suggested, first, such words, most of which adverbs, occur significantly more frequently in the TRJPS corpus (1.75/100 words vs. 1.42/100 words in PJPS, Loglikelihood = 278.81, $p < .0001$). Secondly, a fair proportion of such adverbs pertain to the writer's comments on the proposition, including hedges or boosters (e.g., *highly, substantially, clearly, largely, seemingly*) and stance markers (e.g., *reasonably, unsurprisingly, strikingly, unfortunately*). Additionally, there are a set of adverbs associated with the manner in which something happened (e.g., *systematically, quickly, routinely, increasingly*). Adverbs are used to modify adjectives or verbs, and therefore when they occur in a noun phrase, multiple layers of modification are involved, enabling the writer to compress a great deal of information into one sentence and thus communicating knowledge concisely. The following examples (20-22) show how information is arranged into complex constructions, namely noun phrases containing multiple modifiers.

- 20) Collectively, these results suggest candidate evaluation that is constrained by partisanship, but that, within those constraints, varies from moment to moment in response to transient and thus *highly* accessible character and issue information. (TRJPS77)
- 21) We also show that the class of models Fey and Ramsay propose make the *substantively* unwarranted assumption that an actor can unilaterally impose peace on an opponent who strictly prefers war. (TRJPS70) (noun complement clause that)
- 22) Our identification of six levels of policy, and four patterns of historical policy development, helped uncover two much more *empirically and historically* accurate patterns of policy development on federal and private forest lands in the U.S. Pacific Northwest than would typically have been identified using orthodox models of policy dynamics. (TRJPS45)

The noun head in Example (20) (*information*) is modified by a defining noun phrase (*character and issue*) and two adjectives (*transient, accessible*), the latter being further modified by a booster (*highly*). In Example (21), the noun head (*assumption*) is modified by an adverb + *-ed* adjective premodifier as well as a complement clause headed by *that* as a postmodifier where there is a further embedded relative clause. The nominalisation allows two propositions (what someone assumes, what the writer thinks of the assumption) to be presented in the same clause. The highlighted noun phrase in Example (22) is even more complex with the use of a variety of phrasal features, including an attributive adjective (*accurate*) modified by two adverbs of manner (*empirically, historically*), multiple postmodifying prepositional phrases (*of policy development, on federal and private forest lands, in the U.S. Pacific Northwest*), and a postmodifying finite clause of comparison (*than...*) with an embedded non-finite clause (*using...*). As in Example (21), in this way, the writer manages to provide detailed information of the method/process as well as a result, together with an assessment, within one single sentence.

In comparison, formulation in Example (22) obviously requires more careful planning and editing to be specific but also more concise than that in Example

² The sheer number of such words (18,640 in the TRJPS corpus alone) means it is impossible to carry out a systematic analysis within the scope of the present study.

(23). In that sense, the use of adverbial modifiers, rather than prepositional postmodifiers, in complex nominals may be more predictive of the writers' expertise level in disciplinary writing.

- 23) Further, Lois Bryson suggests that the word “welfare” should be changed to “elfare” or “ill-being” and contends that these new names would more *accurately* describe the realities of the programs in practice... (PJPS75)

Noun modifiers in TRJPS are often associated with specialist terminology, as in Examples (24) to (26), representing economical and conventional ways of packaging knowledge of the field.

- 24) These *outsiders activists, interest groups, and party bosses* - use their control over *party nominations*, conditioned on institutional rules, to ensure ideological behavior among officeholders. (TRJPS47)
- 25) I reevaluate the links between *candidate race, district composition*, and turnout by leveraging a nationwide database of over 185 million individual *registration records*, including estimates for the race of every voter. (TRJPS99)
- 26) Two new national surveys address *the measurement problem* directly by asking respondents how they would vote and how they think their representatives voted on key *roll-call votes*. (TRJPS22)

As revealed in Soler and Wang (2019) through a keyword analysis, the top-ranking journal articles contain more disciplinary and research-related keywords than those in the predatory journal articles, which consist of more general academic vocabulary. The significant difference between the two corpora in terms of the amount of noun modifiers may thus be a further indication of the difference in relation to the type as well as amount of discipline-specific knowledge shared in the two sets of RAs. It would be interesting to find out how these discipline-specific concepts are expressed in the PJPS corpus - whether they simply occur less often in favour of other concepts or they are expressed in other ways. In the former case, the concepts would be informative of the nature of research involved in the predatory journal; in the latter case, such a comparison as performed in the present study would certainly have considerable pedagogical implications for less experienced writers of the field.

5 Conclusion

A few words of caution are in order before we draw any conclusions from this study, namely that the comparison was based on one journal in each domain in one single discipline. While this kind of one-to-one journal comparison helped us rule out certain factors (e.g., scope and aims) that could complicate our interpretation of the results, should more journals be involved in the study, it goes without saying that we need to be cautious in making any generalisations. That said, the study did provide us with some insights into predatory publishing and academic writing that would warrant further investigation with more data. To begin with, no significant difference was found in many measures (such as mean length of sentences/T-units, number of T-units per sentence, number of complex nominals per T-unit, number of coordination at the clausal level, number of

relative clauses per nominal, number of adjective modifiers per nominal), suggesting that the writing quality in the two corpora may look much alike, if only on the surface.

Secondly, the top-ranking journal articles yielded a greater amount of clausal embedding, and the predatory counterparts exploited more complex nominals and coordinated phrases per clause. According to Biber et al. (1999), finite dependent clauses, particularly *that*-complement clauses (e.g., *I think that...*) and finite adverbial clauses (e.g., *Because I think...*), are extensively used in spoken language. This is why they argued against the traditional syntactic complexity measures focusing on clausal embedding to gauge the quality of academic writing and proposed measures at the phrasal level (such as complex nominals and coordinated phrases) as stronger indices of quality (see also Lu, 2011). This view seems to be challenged by our findings based on the traditional complexity measures. One explanation could be that previous studies such as Lu (2011) tend to focus on the development of university-level academic writing, which is more general by nature than disciplinary writing. Even in those studies which involve professional disciplinary writing, there can be a great deal of disciplinary variation, and the results of the present study may well be specific to the discipline of Political Science.

Indeed, the analysis of fine-grained clausal and phrasal complexity measures indicated that the overall results of the traditional measures were most likely attributable to a small set of specific structures. For instance, the higher score of clausal embedding in the TRJPS corpus may be related to complement and adverbial clauses in particular, which, as the qualitative analysis demonstrated, were associated with specific rhetorical or discoursal functions. The higher score of complex nominals per clause in the PJPS corpus was associated with a larger number of noun phrases with prepositional modifiers such as *of*-phrases. In terms of complexity, such phrases are obviously not at the same level as noun phrases with adverbial modifiers, for instance, which stood out in the top-ranking journal articles. These findings confirm Lu's (2017) observation that the relationship between syntactic complexity and writing quality can vary across different task variables, to which discipline is clearly relevant. In fact, we would argue that it may be simplistic to associate quantitative results of syntactic complexity measures directly with levels of quality in disciplinary writing. To echo Casal et al.'s (2021) suggestion, a closer examination of the function-specific usage patterns of syntactically complex structures may produce more interesting and useful insights into the role of such structures, regardless of their position in the hierarchy of 'complexity', in disciplinary knowledge-making practices. In the present study, the quantitative results were indeed useful in identifying areas for a qualitative analysis, which, unfortunately, was not performed in a systematic way; the number of relevant instances was too large to be manageable within the scope of this study. Future research may want to look further into the functions of clausal complements and adverbial clauses and noun phrases with pre-modifying nouns, to mention a few.

In conclusion, the results of the present study suggest that articles published in predatory journals can be difficult to identify with a quick glance at the language used. On the face of it, they may look very similar to those published in more reputable outlets in terms of syntactic complexity. This means that it is probably not lack of language proficiency that has prompted the authors to

publish in predatory journals. However, through a more fine-grained analysis, some subtle differences could still be found. In the present study, such differences were associated specifically with rhetorical techniques (as reflected in the use of clausal complements and adverbial clauses) and the packaging of specialist knowledge (noun phrases with pre-modifying nouns), among other things. The differences with regard to rhetorical techniques may suggest a lack of awareness of the need and/or the way of effectively constructing authorial voice and stance in the discipline among the authors of predatory journal articles. The differences in connection with specialist terminology may be more concerning, as it raises doubts about the nature and quality of research that gets published in predatory journals. As emerged from Kurt's (2018) survey study, authors' self-reported lack of knowledge of research is one of the main factors influencing their decision to select predatory journals, which publish everything that is claimed to be scientific – along with authors' fees – without quality checks. The results of the present study may thus be taken as providing empirical support for this account.

Currently, predatory publishing seems to pose additional challenges for scholars worldwide, particularly for those in (semi-)peripheral contexts. Although authors who choose to publish in predatory outlets may be difficult to target as a group, unlike university students or novice writers, we would argue that the results from studies like ours can potentially add to pedagogical interventions aimed to assist authors when navigating the complexities of the current publishing landscape (Fazel & Heng Hartse, 2020). In addition to focusing on how to select potential publication outlets, the apparently minuscule micro-level textual differences that set apart the top-ranking journal articles from the predatory ones might make an important difference for authors to develop an authorial voice of their own, for instance, which might be a crucial aspect shaping readers' (and particularly reviewers') interpretations of the texts and of their authors' expert status (Matsuda & Tardy, 2007). We therefore end with a call for further research on micro-level linguistic features with larger corpora from a wider variety of specialisations, on the one hand, and on the other hand, for applied interventions for (novice) writers in the (semi-)periphery that do not end at the level of awareness raising about predatory publishing as a phenomenon.

Disclosure statement

The author declared no conflict of interest.

References

- Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*, 31, 58–71.
<https://doi.org/10.1016/j.jeap.2017.12.008>
- Beall, J. (2017). Beall's list of predatory publishers. <https://scholarlyoa.com/bealls-list-of-predatory-publishers-2017/>
- Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. John Benjamins. <https://doi.org/10.1075/scl.23>
- Biber, D., Johansson, S., Leech, J., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.

- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35. <https://doi.org/10.5054/tq.2011.244483>
- Biber, D., & Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge University Press.
- Björk, B-C B. C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How frequently are articles in predatory open access journals cited? *Publications*, 8(2), 17. <https://doi.org/10.48550/arXiv.1912.10228>
- Casal, J. E., Lu, X., Qiu, X., Wang, Y., & Zhang, G. (2021). Syntactic complexity across academic research article part-genres: A cross-disciplinary perspective. *Journal of English for Academic Purposes*, 52. <https://doi.org/10.1016/j.jeap.2021.100996>
- Cobey K.D., Grudniewicz A, Lalu M.M., Rice, D.B., Raffoul, H., & Moher, D. (2019). Knowledge and motivations of researchers publishing in presumed predatory journals: a survey. *BMJ Open* 2019;9:e026516.
- Cortegiani, A., & Shafer, S. (2018). “Think. Check. Submit” to avoid predatory publishing. *Critical Care*, 22, 300. <https://doi.org/10.1186/s13054-018-2244-1>
- Crossley, S., & McNamara, D. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learner. *Journal of Second Language Writing*, 26, 66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Eriksson, S., & Helgesson, G. (2017.) The false academy: Predatory publishing in science and bioethics. *Medicine, Health Care, and Philosophy*, 20, 163–170. <https://doi.org/10.1007/s11019-016-9740-3>
- Fazel, I., & Heng Hartse, J. (2020). Gray areas of academic publishing: ‘predatory journals’ under the lens. *Journal of English for Research Publication Purposes*, 1(2), 184–194. <https://doi.org/10.1075/jerpp.20008.faz>
- Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G., Cukier, S., Allen, K., Ardern, C., Belcom, L., Barros, T., Berger, M., Ciro, J., Cugusi, L., Donaldson, L., Egger, M., Graham, I., Hodgkinson, M. Khan, K., Mabizela, M., Manca, A., Milzow, K., Mouton, J., Muchenje, M., Olijhoek, T., Ommaya, A., Patwardhan, B., Poff, D., Proulx, L., Rodger, M., Severin, A., Strinzel, M., Sylos-Labini, M., Tamblyn, R., van Niekerk, M., Wicherts, J., & Lalu, M. (2019). Predatory journals: No definition, no defence. *Nature*, 576, 210–212.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. (2012). *Disciplinary Identities: Individuality and Community in Academic Discourse*. Cambridge University Press.
- Hyland, K., & Jiang, F. (2018). “In this paper we suggest”: Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, 51: 18–30. <https://doi.org/10.1016/j.esp.2018.02.001>
- Kurt, S. (2018). Why do authors publish in predatory journals? *Learned Publishing*, 31(2), 141–147. <https://doi.org/10.1002/leap.1150>
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Doctoral Dissertation, Georgia State University]. <https://doi.org/10.57709/8501051>
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349. <https://doi.org/10.1111/modl.12468>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers’ language development. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511. <https://doi.org/10.1177/0265532217710675>

- Lu, X, Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes*, 44, 1–16. <https://doi.org/10.1016/j.jeap.2019.100832>
- Martin, A., & Martin, T. (2016). A not-so-harmless experiment in predatory open access journals. *Learned Publishing*, 29(4), 301–305. <https://doi.org/10.1002/leap.1060>
- Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, 26, 235–249. <https://doi.org/10.1016/j.esp.2006.10.001>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>
- Richtig, G., Berger, M., Lange-Asschenfeldt, B., Aberer, W., & Richtig, E. (2018). Problems and challenges of predatory journals. *Journal of the European Academy of Dermatology and Venereology*, 32(9), 1441–1449. <https://doi.org/10.1111%2Fjd.15039>
- Shen, C., & Björk, B.-C. (2015). ‘Predatory’ open access: a longitudinal study of article volumes and market characteristics. *BMC Medicine*, 13(1), 230. <https://doi.org/10.1186/s12916-015-0469-2>
- Soler, J., & Wang, Y. (2019). Linguistic differences between well-established and predatory journals: A keyword analysis of two journals in political science. *Learned Publishing*, 32, 259–269. <https://doi.org/10.1002/leap.1244>
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149–183. <https://doi.org/10.1177/0741088316631527>
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, 47(2), 420–430. <https://doi.org/10.1002/tesq.91>
- Wang, Y. (2018). *As Hill seems to suggest*: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *Journal of English for Academic Purposes*, 33, 12–23. <https://doi.org/10.1016/j.jeap.2018.01.003>
- Wang, Y., & Soler, J. (2021). Investigating predatory publishing in political science: a corpus linguistics approach. *Applied Corpus Linguistics*, 1(1), 1–15. <https://doi.org/10.1016/j.acorp.2021.100001>
- Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, 43. <https://doi.org/10.1016/j.jeap.2019.100798>
- Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2015). Who publishes in “predatory” journals? *Journal of the Association for Information Science and Technology*, 66(7), 1406–1417. <https://doi.org/10.1002/asi.23265>

Received August 4, 2022

Revision received November 16, 2022

Accepted November 30, 2022