Syntactic complexity in applied linguistics research article abstracts: A corpus-based comparative analysis between MENA and international writers

Basim Alamri & Assem Alqarni
King Abdulaziz University (Saudi Arabia)
bmalamri@kau.edu.sa, amalqarni1@kau.edu.sa

Abstract

This study explored syntactic complexity variations in English research article (RA) abstracts written by native speakers of Arabic residing in the Arab world and expert international publication (IP) authors from various contexts. Three specialized corpora of English abstracts were compiled, each comprising 200 abstracts by Middle Eastern Arabs, North African Arabs (MENA), and expert IP authors, amounting to a total of 111,645 words. Using Lu's (2010) L2SCA, the data from the three corpora were analyzed using several procedures. Fourteen measures of syntactic complexity grouped into five categories were implemented. Findings across these five categories revealed significant differences among the three corpora with respect to syntactic complexity. The expert IP abstracts exhibited higher sentence complexity than did the MENA abstracts. Practical implications for Arabic learners and others with respect to writing pedagogy in English for research publication purposes will be discussed.

Keywords: English for research and publication purposes (ERPP), Middle East and North Africa (MENA), RA abstracts, syntactic complexity.

Resumen

Complejidad sintáctica en los resúmenes de artículos de investigación de lingüística aplicada: análisis comparativo basado en corpus entre árabes de Oriente Medio y Norte de África y escritores internacionales

Este estudio explora las variaciones de complejidad sintáctica en los resúmenes

de artículos de investigación en inglés escritos por hablantes nativos de árabe residentes en el mundo árabe y autores expertos en publicaciones internacionales de diversos contextos. Se recopilaron tres corpus especializados de resúmenes en inglés, cada uno de ellos compuesto por 200 resúmenes escritos por árabes de Oriente Medio y árabes del Norte de África, por un lado, y autores expertos en publicaciones internacionales, por otro, que sumaban un total de 111.645 palabras. Utilizando el L2SCA de Lu (2010), se analizaron los datos de los tres corpus mediante varios procedimientos. Se aplicaron catorce medidas de complejidad sintáctica agrupadas en cinco categorías. Los resultados de estas cinco categorías revelaron diferencias significativas entre los tres corpus con respecto a la complejidad sintáctica. Los resúmenes de expertos en publicaciones internacionales presentaban una mayor complejidad oracional que los resúmenes escritos por árabes de Oriente Medio y del Norte de África. Se discuten las implicaciones prácticas para, entre otros, los estudiantes árabes con respecto a la pedagogía de la escritura en inglés con fines de publicación de investigaciones.

Palabras clave: Inglés con fines de investigación y publicación, Oriente Medio y Norte de África, resúmenes de artículos de investigación, complejidad sintáctica.

1. Introduction

Scholarly interest in English for research and publication purposes (ERPP) has intensified over the last decade, as the field has emerged as a distinct and key research area (Flowerdew & Habibie, 2022). This paper explores ERPP in relation to peer-reviewed international journals (Cargill & Burgess, 2008) for native anglophone authors and those for whom English is an additional language (EAL). It further compares the syntactic complexity of English research article (RA) abstracts published in international peer-reviewed journals that have more publications from residents of the Arab world with those published in leading journals in the applied linguistics field. Syntactic complexity herein denotes "the degree of variety, sophistication, and elaboration of the structures that surface in language production" (Yin et al., 2021, p. 2).

Research on ERPP has attracted scholarly attention since the "publish or perish" phenomenon emerged in academia (Li & Flowerdew, 2020). Publication in scholarly journals is particularly advantageous for EAL scholars, allowing them to write for a wider audience, given that English is the most widely spoken second language worldwide (Hyland, 2016).

However, Flowerdew and Habibie (2022) asserted that written EAL texts exhibit key differences when compared with Anglophone texts, and EAL texts tend to be slower in being accepted for publication. While English-language publishing "poses challenges to [EAL authors'] visibility and participation in global scholarship" compared to Anglophone authors (Flowerdew & Habibie, 2022, p. 4), it is important to emphasize that "EAL writers may be as equally proficient in English as their Anglophone counterparts, or even better" (p. 18). The present study seeks to evaluate the accuracy of this assumption and yield further insights into ERPP practices and EAL academic writing research, particularly in MENA regions.

Although it is an under-researched region, MENA spans a vast geographical area, encompassing 22 countries with over 400 million people whose first language is Arabic. The region includes three non-Arab countries with different languages: Turkey (Turkish language), Iran (Farsi), and Israel (Hebrew). According to Kachru (1992), MENA countries are located within the Expanding Circle, where the English language is regarded as a foreign language. English in academia has been likened to "a tyrannosaurus rex … a powerful carnivore gobbling up the other denizens of the academic linguistic grazing grounds" (Swales, 1997, p. 374) and is the primary language for discipline-specific publication in MENA regions (Abdeljaoued & Labassi, 2020; Elyas & Mahboob, 2020). Researchers have a greater opportunity to attain international recognition by publishing in English. It is thus recommended that international scholarly journals recognize EAL, which is used by numerous authors, as an acceptable form of ERPP.

1.1. Syntactic complexity in academic writing

Syntactic structural complexity is recognized as a measure of sophistication (and variety). L2 ERPP writing may be analyzed to better understand its syntactic complexity, gauge its quality and proficiency, and measure syntactic structures (Ai & Lu, 2013; Casal & Lee, 2019; Lu, 2010; Yin et al., 2021). While research has focused primarily on quality assessments and correlations, syntactic complexity research has been used extensively as a benchmark for proficiency development, often as part of the complexity, accuracy, and fluency framework (Bulté & Housen, 2012; Casal & Lee, 2019; Lu, 2010; Ortega, 2003). Recent studies have investigated syntactic complexity features in RA part-genres, such as abstracts (Demir, 2021; Tovar-Viera, 2022; Wu et al., 2020).

Several indexes and measures have been developed to evaluate syntactic complexity and writing quality, particularly in comparisons of L1 and L2 writing (Ai & Lu, 2013; Bi & Jiang, 2020; Breeze, 2008; Kuiken et al., 2019; Ortega, 2003; Wu et al., 2020). Syntactic complexity can indicate some language users' proficiency (Crossley & McNamara, 2014), while for others, syntactic complexity can offer a means of testing pedagogical intervention effectiveness (Ellis, 2009; Ong & Zhang, 2010). Ai and Lu (2013) developed five syntactic complexity categories, adopted in the present study: length of production unit, amount of subordination, amount of coordination, phrasal complexity, and overall sentence complexity. These categories incorporated 14 indices (Table 3). For a comprehensive review of the development of syntactic complexity measures, see Cheung and Kemper (1992), Larsen-Freeman (1978, 2009), Ortega (2003), and Wolfe-Quintero et al. (1998).

Earlier studies have compared native and non-native speakers' writing using all or some of the five measures mentioned above (Ai & Lu, 2013; Foster & Tavakoli, 2009; Hinkel, 2003; Mancilla et al., 2017). Hinkel (2003), for example, found that advanced non-native speakers in U.S. universities were more likely than native speakers to favor simplified syntactic constructions. Mancilla et al. (2017) examined the writing of native and non-native speakers in online discussions and noted that non-native speakers used more coordination and complex phrases but less subordination, while more advanced non-native speakers used subordination to an extent close to that of native speakers. Similarly, Ai and Lu (2013) studied three corpora comprising 600 essays in total (200 essays per corpus): low-level non-native speakers; high-level non-native speakers; and native speakers. They observed significant differences between non-native and native authors' length of production unit, coordination, degree of sophistication, and subordination ratio.

Wu et al. (2020) investigated the syntactic constructions of unedited expert-level RAs from the SciELF corpus ((Written English as a Lingua Franca in Academic settings) (WrELFA, 2015) and American English RAs and observed several significant differences. English as lingua franca (ELF) authors used longer sentences, more coordinating phrases, and complex nominals. Moreover, they were more likely than American English RAs to favor nominal phrases. A comprehensive review of these studies' findings highlights three influencing factors by which the syntactic complexity

indexes vary: non-native writers' proficiency level (i.e., high or low); genre (e.g., sections of research articles); and whether the author is a native speaker or belongs to ELF regions. Nonetheless, specific sub-genres (e.g., abstracts) in published RAs within particular disciplines have yet to be examined in terms of comparison between Arab authors and international authors in light of ERPP research.

Computational software has the potential to render the adoption of multiple syntactic complexity measures to analyze numerous texts more objective and practical (Ai & Lu, 2013), as manual analyses may overlook several fundamental aspects. Lu's (2010) L2SCA, employed in the present study, is among the most reliable software programs, capable of automatically identifying 14 syntactic complexity indices for English-based texts.

1.2. Studies on RA Abstracts in MENA

The American National Standards Institute (ANSI) defines the abstract as "an abbreviated, accurate representation of the contents of a document, preferably prepared by its author(s) for publication with it" (Bhatia, 1993, p. 78). RA abstracts have attracted increased scholarly interest by virtue of the information that they provide on four aspects: what the author did, how the author did it, what the author found, and what the author concluded (Bhatia, 1993). Lorés-Sanz (2004) demonstrated that abstracts "constitute the gateway that leads readers to take up an article, journals to select contributions, or organizers of conferences to accept or reject papers" (p. 281).

Studies have analyzed the rhetorical patterns and linguistic features of RA abstracts across different disciplines, including applied linguistics (Lorés-Sanz, 2004), and to compare the writing of native versus non-native speakers. Other studies have analyzed RA abstract writing styles (El-Dakhs, 2018) and compared English RAs with those in other languages, such as Arabic (Alharbi & Swales, 2011; Alqinai, 2013; Dickins, 2017). These studies have reported various similarities and differences that may be attributable to L1 background or cultural influence.

RA abstracts from Arab writers remain underexplored and warrant further investigation from various perspectives, particularly in terms of their linguistic features (Alharbi & Swales, 2011; Fallatah, 2016). Alharbi and Swales (2011) compared 28 paired Arabic and English abstracts written by Arab authors for publication in Arab language science journals that require

translated versions of the abstracts and identified rhetorical and linguistic differences between Arabic and English pronoun use, promotional features, and number and order of rhetorical moves. The authors related these differences to scholastic traditions in the Arab world. Fallatah (2016) conducted a contrastive genre analysis on three sets of RA abstracts to investigate the differences and similarities in genre structure between abstracts written by Saudi and international researchers and found that Saudi English RA abstracts exhibited "more move presence fluctuation; verbosity; move cyclicality; excessive use of citation, acronyms, and listings; and multi-paragraphing" (p. 368).

In the Iraqi context, Friginal and Mustafa (2017) explored the linguistic features of English-language abstracts by U.S.-based and Iraqi scholars. Using Biber's (1988) multi-dimensional analytical (MDA) approach, the analysis involved eight parallel sub-corpora of RA abstracts in four disciplines (agriculture, nursing, engineering, and languages). The findings indicated certain similarities and differences in Iraqi writers' approach to structuring their abstracts, including the rhetorical preferences of Iraqi authors that differentiate them from other abstract writers across the four disciplines. Dickins (2017) explored how Arabic and English writers differ in their use of coordination based on different types of texts (e.g., political and cultural books, newspaper and magazine articles) derived from Dickins et al. (2016). The results showed that Arabic writers were more likely to employ coordination than their English counterparts, who used more subordination. Finally, Youssef (2019) investigated the syntactic complexity and lexical diversity of 100 single-authored English RA abstracts by Egyptian and British writers in linguistics and nuclear science and found that native Anglophone writers used more subordination, whereas Egyptian writers employed more coordination.

Collectively, these studies offer insight into the rhetorical structure employed by Arab scholarly writers. However, several gaps remain. For example, most studies have focused on genre structure by analyzing either only English abstracts between Arab writers and English native speakers (Fallatah, 2016) or paired Arabic–English abstracts (Alharbi & Swales, 2011; Alotaibi, 2013; El-Dakhs, 2018). The present study responds to the call from ERPP to investigate how L2 ERPP writers engage with syntactic complexity in scholarly writing. The few studies that have responded to this call (Ansarifar et al., 2018; Wu et al., 2020; Yin et al., 2021) were conducted in contexts other than the Middle East or North Africa. Moreover, few studies have

explored the linguistic characteristics of Arab writers from a multidimensional analytical perspective (e.g., Friginal & Mustafa, 2017). Youssef (2019) attempted to explore syntactic complexity among Egyptian writers using a small corpus of English conference abstracts. Our study, by contrast, includes writers from throughout the Arab world, which may offer a comprehensive perspective on MENA writers' syntactic complexity. Specifically, comparison of part-genres (e.g., abstracts) across the two author groups may facilitate observation of similarities and differences and delineate characteristics unique to each group.

The present study thus aims to contribute to the literature on the syntactic complexity of RA abstracts by focusing on an under-investigated L2 writing group—Arab authors residing in the Arab world—to provide insights into how these writers' RA abstracts resemble or differ from abstracts published in top journals in terms of SC. It is hoped that the study will enrich ERPP literature and promote awareness of EAL among novice Arab ERPP writers with respect to RA abstracts. A further objective is to examine SC in RA abstracts in the field of applied linguistics specifically to investigate how SC differs among MENA and expert international publication corpora. The study is guided by the following research questions:

> **RQ1.** What are the features of syntactic complexity in MENA and expert international publication corpora of RA abstracts?
>
> **RQ2.** Do any significant differences emerge across MENA and expert international publication (IP) corpora with respect to syntactic complexity in RA abstracts, and if so, what are these differences?

2. Methods

2.1. Corpora

Three specialized corpora comprising 600 English RA abstracts (approximately 111,645 words) were purposively selected from 10 peer-reviewed applied linguistic journals. To ensure a greater degree of comparability among the three corpora, the inclusion criteria for abstracts included publication during the last ten years (2010–2019); topics covering applied linguistics and English language-teaching disciplines; written by a maximum of two authors; and involving empirical studies. Furthermore, the expert IP corpus did not include abstracts written by MENA authors.

The authors in the two Arab corpora were native Arabic speakers affiliated with universities in the Arab world countries. Most authors in the international corpus were international authors—non-Arabs—affiliated with universities in Asia, Europe, and the United States. The term 'international writers/corpus' is problematic in this study because the authors represent different countries. To avoid such ambiguity, for the purposes of this study, we decided to use the term 'expert international publication' (IP) writers to denote applied linguistics scholars with publications in leading international journals. Following Wood's (2001) criteria (i.e., institution affiliation and authors' Arabic names), we also checked authors' biographies and websites to verify the Arabic authors' native status. While several previous studies compared native speakers and EAL writers from different parts of the world with various L1 backgrounds (e.g., Wu et al., 2020), we decided to restrict L1 background only to the Arab corpus and to trace comparisons with the international corpus. However, this restriction was not applied in the expert IP writers corpus in acknowledgment of the recent notion that high-level expertise in international publication and ERPP writing is not determined by the authors' L1 background (e.g., Flowerdew, 2015) because the authors are considered expert writers and thus publish in high-ranking international peer-reviewed journals (e.g., Thompson, 2005; Yin et al., 2021).

We selected 40 abstracts from each of the following scholarly journals to compile the expert IP corpus: *Journal of Second Language Writing (JSLW)*, *Applied Linguistics (AL)*, *English for Specific Purposes (ESP)*, *TESOL Quarterly (TQ)*, and *Journal of English for Academic Purposes (JEAP)*. These journals were selected as reputable and leading journals in the field of applied linguistics and English language teaching. No restrictions were applied to the authors' L1 background or affiliations in the expert IP corpus. However, it was difficult to select a balanced number of abstracts from each journal in the MENA corpora since some journals had more publications than others (as Table 1 shows). The abstracts in the other two corpora (i.e., MENA authors) were selected from the following peer-reviewed journals: *Arab World English Language (AWEJ)*, *English Language Teaching (ELT)*, *International Journal of Arabic-English Studies (IJAES)*, *International Journal of Linguistics (IJL)*, and *International Journal of English Linguistics (IJEL)*. These journals have more publications from authors residing in the Arab world. Tables 1 and 2 provide more detailed information about the three corpora.

Expert IP Corpus	No. of abstracts	Middle East Corpus	No. of abstracts	North Africa Corpus	No. of abstracts
AL	40	AWEJ	41	AWEJ	119
JEAP	40	ELT	45	ELT	43
JSLW	40	IJAES	36	IJAES	9
ESP	40	IJL	40	IJL	19
TQ	40	IJEL	38	IJEL	10
Total	200	**Total**	**200**	**Total**	200
Total no. of words	37,279 words	**Total no. of words**	36,246 words	**Total no. of words**	38,120 words

Table 1. Journal distributions of the three corpora.

Each corpus comprised 200 abstracts labeled as follows: Middle Eastern Arabs (ME_AR) (36,246 words), North Africa Arabs (NA_AR) (38,120 words), and expert IP writers (INT) (37,279 words). Table 2 presents the descriptive details of the three corpora. Note that there is a difference in the total number of words in the three corpora. However, the type of comparison being conducted in the present study should not be affected by this difference because "the syntactic complexity measures are all computed as ratios of one syntactic structure to another in complete texts" (Ai & Lu, 2013, p. 256). In addition, the effect of the total number of words in the three corpora was tested by an analysis of variance (ANOVA) test; it indicated no significant difference among the three corpora (i.e., F (df = 2, 597) = 2.114; p = 0.122). The syntactic complexity measures in the present study were computed based on the ratios of one syntactic structure to another in complete texts, following Ai and Lu (2013).

Register	No. of abstracts	Abstracts length		No. of sentences per abstract		Sentence length		Total no. of words
		Mean	SD	Mean	SD	Mean	SD	
Middle Eastern Authors	200	181.23	53.220	7.73	3.184	24.74	6.734	36,246
North African Authors	200	190.60	49.262	7.74	2.321	25.51	6.472	38,120
Expert IP Authors	200	186.39	31.517	6.66	1.817	29.82	9.587	37,279
Total	**600**	**186.08**	**45.738**	**7.38**	**2.553**	**26.69**	**8.032**	**111,645**

Table 2. Descriptive details of the three corpora.

2.2. Measures of syntactic complexity

All 600 RA abstracts were downloaded, manually added to a separate text file for each abstract, and checked for accuracy. We then deleted any components that were irrelevant to syntactic complexity analysis, such as section titles. To measure syntactic complexity, there are widespread automated text analysis

tools, such as L2 Syntactic Complexity Analyzer (L2SCA) and the Tool for the Automatic Analysis of Syntactic Complexity (TAASSC). Developed by Dr. Xiaofei Lu (Ai & Lu, 2013; Lu, 2010, 2011), we used the L2SCA to analyze the abstracts in the final dataset. We selected this tool because it is freely available, its accessibility and high reliability, and has been successfully applied in various studies to analyze and compare English RA abstracts (Tovar-Viera, 2022; Wu et al., 2020; Yin et al., 2021). The L2SCA was designed using Python to help researchers, instructors, and language teachers analyze the syntactic complexity of texts in English. As Table 3 illustrates, the tool provides comprehensive assessment of the sophistication and complexity of writing samples based on five main measures/categories comprising 14 different indices; these measures are (1) length of production units, (2) amount of coordination, (3) amount of subordination, (4) degree of phrasal sophistication, and (5) overall sentence complexity. Table 3 presents a summary of the indices. Regarding the tool's accuracy, Lu (2010) reported a higher correlation between human annotation and L2SCA—ranging between .834 and 1.000—on writing samples produced by college-level EFL learners.

Measure	Code	Definition
Length of production unit		
Mean length of clause	MLC	# of words/# of clauses
Mean length of sentence	MLS	# of words/# of sentences
Mean length of T-unit	MLT	# of words/# of T-units
Amount of subordination		
Clauses per T-unit	C/T	# of clauses/# of T-unit
Complex T-units per T-unit	CT/T	# of complex T-units/# of T-units
Dependent clauses per clause	DC/C	# of dependent clauses/# of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses/# of T-units
Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/# of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per sentences	T/S	# of T-units/# of sentences
Degree of phrasal sophistication		
Complex nominals per clause	CN/C	# of complex nominals/# of clauses
Complex nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/# of T-units
Overall sentence complexity		
Clauses per sentences	C/S	# of clauses/# of sentences

Table 3. Syntactic complexity measures used.

We calculated the 14 indices of syntactic complexity for the MENA and expert IP corpora using L2SCA after adding each abstract to a separate plain

text format file. The initial results were then entered into an Excel file, and SPSS software was used for further statistical analysis. To determine whether any possible significant difference was evident among the three corpora, an ANOVA test with post-hoc Bonferroni was implemented to account for multiple comparisons. Finally, we conducted a qualitative analysis of the identified syntactic complexity in light of the ERPP perspective—that is, we revealed on how Arab writers' writing resembles/differs from that of expert IP authors in terms of SC measures.

3. Results

This section presents the results of the two research questions with further discussion, illustrated by samples from the corpora.

3.1. Research Question 1: Comparison between MENA and expert IP corpora

The first research question seeks to identify different features of syntactic complexity in the three corpora and investigates any possible significant differences. Table 2 summarizes the descriptive statistics (means and standard deviations) of the 14 SC measures for our data. An independent sample *t*-test was conducted to investigate significant differences between the MENA corpus and the IP expert corpus in terms SC measures. Inspection of Q-Q Plots revealed that the 14 SC indices were normally distributed for both groups and that there was homogeneity of variance in two indices (i.e., Coordinate phrases per clause (CP/C) and Clauses per sentences (C/S)) as assessed by Levene's Test for Equality of Variances. The *t*-test results are summarized in the "MENA vs. INT" column in Table 4. Overall, the expert IP corpus exhibited higher means than the MENA corpus across all SC indices. Therefore, the overall sentence complexity (measured using C/S) revealed a significant difference (t = -3.953, p = .000 < .05) between the MENA and expert IP corpora, indicating that each corpus exhibits a unique syntactic complexity.

The independent sample *t*-test results revealed significant differences between the MENA and expert IP corpora across all syntactic complexity measures. Regarding the length of production unit, the mean lengths of clause, sentence, and T-unit (MLC, MLS, and MLT, respectively) are greater in the expert IP corpus—MLC (M= 16.95, SD= 6.167), MLS (M= 29.82,

SD= 9.587), and MLT (*M*= 27.77, *SD*= 9.940)—than in the MENA corpora—MLC (*M*= 15.39, *SD*= 3.657), MLS (*M*= 25.13, *SD*= 6.608), and MLT (*M*= 23.74, *SD*= 6.101), respectively.

With respect to the subordination ratio (measured by C/T, CT/T, DC/C, and DC/T), the MENA corpus uses a significantly smaller proportion of the four aspects of subordination than the expert IP corpus. The results show statistically significant differences in the amount of subordination as of C/T (t = -3.013, p = .003 < .05), CT/T (t = -2.531, p = .012 < .05), DC/C (t = -3.287, p = .001 < .05), and DC/T (t = -3.514, p = .001 < .05).

Regarding the amount of coordination, our results reveal that the MENA corpus differs significantly from the expert IP corpus in terms of phrasal coordination per unit and sentential coordination (measured by CP/T (t = -5.092, p = .000 < .05) and T/S (t = -2.256, p = .025 < .05), respectively) and in the amount of phrasal coordination per clause (CP/C) (t = -3.895, p = .000 < .05).

Finally, the MENA corpus uses a smaller proportion of phrasal sophistication aspects, including complex nominals (i.e., per unit and per clause) and phrasal verbs (measured using CN/C (t = -5.343, p = .000 < .05), CN/T (t = -7.728, p = .000 < .05), VP/T (t = -2.237, p = .026 < .05)).

Measure	Code	ME_AR (N=200)		NA_AR (N=200)		MENA (N=400)		INT (N=200)		MENA vs. INT (t-test)				
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Levene's Test for Equality of Variances F*	*P*	*t*	*Sig.*	*Cohen*
Length of production unit														
Mean length of clause	MLC	14.89	3.580	15.88	3.677	15.39	3.657	16.95	6.167	9.642	.002*	-3.317	.001*	-0.309
Mean length of sentence	MLS	24.74	6.734	25.51	6.472	25.13	6.608	29.82	9.587	6.002	.015*	-6.219	.000*	-0.570
Mean length of T-unit	MLT	23.22	5.800	24.27	6.360	23.74	6.101	27.77	9.940	13.569	.000*	-5.258	.000*	-0.489
Amount of subordination														
Clauses per T-unit	C/T	1.59	.343	1.56	.374	1.57	.358	1.68	.449	12.535	.000*	-3.013	.003*	-0.270
Complex T-units per T-unit	CT/T	.44	.195	.41	.212	.42	.204	.47	.227	5.529	.019*	-2.531	.012*	-0.223
Dependent clauses per clause	DC/C	.34	.121	.32	.132	.33	.127	.37	.143	5.327	.021*	-3.287	.001*	-0.290
Dependent clauses per T-unit	DC/T	.58	.322	.54	.324	.56	.323	.68	.424	14.292	.001*	-3.514	.001*	-0.317
Amount of coordination														
Coordinate phrases per clause	CP/C	.48	.273	.51	.270	.50	.272	.59	.291	2.458	.117	-3.895	.000*	-0.337
Coordinate phrases per T-unit	CP/T	.75	.427	.77	.380	.76	.403	.95	.467	4.043	.045*	-5.092	.000*	-0.452
T-units per sentences	T/S	1.07	.129	1.06	.114	1.06	.122	1.09	.149	11.386	.001*	-2.256	.025*	-0.202
Degree of phrasal sophistication														
Complex nominals per clause	CN/C	2.16	.629	2.33	.682	2.25	.661	2.60	.826	9.785	.002*	-5.343	.000*	-0.479
Complex nominals per T-unit	CN/T	3.36	.970	3.55	.980	3.45	.979	4.24	1.254	12.622	.000*	-7.728	.000*	-0.696
Verb phrases per T-unit	VP/T	2.40	.669	2.47	.657	2.43	.663	2.58	.801	7.058	.008*	-2.237	.026*	-0.200
Overall sentence complexity														
Clauses per sentences	C/S	1.70	.418	1.65	.406	1.67	.412	1.82	.452	2.756	.097	-3.953	.000*	-0.342

*Note: *statistical significance difference at the level p < .05.*

Table 4. Mean, standard deviations, and independent sample t-test of 14 syntactic complexity indices among the three corpora.

Measure	Code	ANOVA F	ANOVA Sig.	ME_AR vs. NA_AR	ME_AR vs. INT	NA_AR vs. INT	MENA vs. INT (*t*-test)
Length of production unit							
Mean length of clause	MLC	9.900	.000*	–	*	–	*
Mean length of sentence	MLS	25.047	.000*	–	*	*	*
Mean length of T-unit	MLT	19.733	.000*	–	*	*	*
Amount of Subordination							
Clauses per T-unit	C/T	5.452	.005*	–	*	*	*
Complex T-units per T-unit	CT/T	4.621	.010*	–	–	*	*
Dependent clauses per clause	DC/C	7.108	.001*	–	–	*	*
Dependent clauses per T-unit	DC/T	7.890	.000*	–	*	*	*
Amount of Coordination							
Coordinate phrases per clause	CP/C	8.088	.000*	–	*	*	*
Coordinate phrases per T-unit	CP/T	14.379	.000*	–	*	*	*
T-units per sentences	T/S	3.265	.039*	–	–	*	*
Degree of phrasal sophistication							
Complex nominals per clause	CN/C	19.706	.000*	*	*	*	*
Complex nominals per T-unit	CN/T	36.806	.000*	–	*	*	*
Verb phrases per T-unit	VP/T	3.230	.040*	–	*	–	*
Overall sentence complexity							
Clauses per sentences	C/S	8.480	.000*	–	*	*	*

Note:
**Indicates a statistically significant difference ($p < .05$).*
– indicates a non-significant difference ($p > .05$).

Table 5. Differences in the mean frequencies of 14 syntactic complexity indices among the three groups.

3.2. Research Question 2: Comparison among the three corpora

As RQ 1 anticipated, several statistically significant differences in syntactic complexity were identified between the MENA and expert IP corpora. We further examined similarities and differences in authors' writing across the three corpora in terms of SC measures. ANOVA tests were employed to investigate possible significant differences among the three corpora.

As Table 5 illustrates, Middle East Arabs' (ME_AR) and North Africa Arabs' (NA_AR) writing shows a higher degree of similarity except in complex nominals per clause (measured using CN/C). Additionally, the NA_AR corpus exhibits higher values than the ME_AR corpus in almost all indices except the amount of coordination and verb phrases. Both the ME_AR and NA_AR corpora differ significantly from the expert IP corpus in most SC measures.

In the overall sentence complexity, the results showed significant differences [$F(2,597)=8.480$, $p<.05$] between each of the ME_AR and NA_AR corpora as a separate corpus against the expert IP corpus in terms of overall sentence

complexity (measured by C/S) results. As Table 4 illustrates, authors in the expert IP corpora (M = 1.82, SD = .452) tend to write more complex sentences, followed by Middle Eastern corpus (M = 1.70, SD = .418) and finally authors in the North African Arab authors (M = 1.65, SD = .406). However, ME_AR and NA_AR corpora do not exhibit any significant difference in the overall SC measure.

The three length of production unit measures show statistically significant differences among the three corpora; the mean lengths of clauses MLC [$F(2,597) = 9.900$, $p<.05$], sentences MLS [$F(2,597) = 25.047$, $p<.05$], T-units MLT [$F(2,597) = 19.733$, $p<.05$]. The length of production units in the expert IP corpus—MLC (M = 16.95, SD = 6.167), MLS (M = 29.82, SD = 9.587), and MLT (M = 27.77, SD = 9.940)—are significantly longer than those in both the MENA corpora. The NA_AR corpus—MLC (M = 15.88, SD = 3.677), MLS (M = 25.51, SD = 6.472), and MLT (M = 24.27, SD = 6.360)—shows a longer unit of production than the ME_AR—MLC (M = 14.89, SD = 3.580), MLS (M = 24.74, SD = 6.734), and MLT (M = 23.22, SD = 5.800)—in the three indices. Moreover, no significant within-MENA corpora differences were observed in the three length of production unit measures, indicating that Arab writers follow comparable writing styles in terms of length of production unit.

The findings regarding the mean values of the amount of subordination (i.e., clauses per T-unit (C/T) [$F(2,597) = 5.452$, $p<.05$] and dependent clauses per T-unit (DC/T) [$F(2,597) = 7.890$, $p<.05$]) reveal significant differences between both MENA corpora and the expert IP corpus. Unlike the ME_AR corpus, the NA_AR and expert IP corpora also show significant differences in complex T-units per T-unit (CT/T) [$F(2,597) = 4.621$, $p<.05$] and dependent clauses per clause (DC/C) [$F(2,597) = 7.108$, $p<.05$] measures. It thus appears that ME_AR writers use a subordination comparable to that used by the authors in the expert IP corpus. While the ME_AR corpus tends to include more subordinations than NA_AR, the expert IP corpus shows the most subordination.

Regarding the amount of coordination, the mean values of the coordinate phrases per clause (CP/C) [$F(2,597) = 8.088$, $p<.05$] and coordinate phrases per T-unit (CP/T) [$F(2,597) = 14.379$, $p<.05$] measures also showed significant differences between the ME_AR and NA_AR corpora on one hand and the expert IP one on the other hand. Regarding coordination at the sentential level T/S [$F(2,597) = 3.265$, $p<.05$], significant differences were

observed between the NA_AR (M = 1.06, SD = .114) and expert IP (M = 1.09, SD = .149) corpora but not between ME_AR and the other two corpora. Finally, ME_AR and NA_AR show comparable coordination use, with no significant differences in the mean values of the CP/C, CP/T, and T/S measures.

Finally, regarding phrasal complexity, the results reveal the only significant differences among the three corpora, which are found in the mean values of complex nominals per clause (CN/C) [$F(2,597) = 19.706$, $p<.05$]. Significant differences also emerged between the both MENA corpora and expert IP corpora in terms of complex nominals per T-unit (CN/T) [$F(2,597) = 36.806$, $p<.05$] measure but not between the ME_AR and NA_AR corpora. Regarding the verb phrases per T-unit (VP/T) measure [$F(2,597)=3.230$, $p<.05$], the only significant difference observed was between the ME_AR (M = 2.40, SD = .669) and expert IP (M = 2.58, SD = .801) corpora. The following subsection discusses these similarities and differences in detail, illustrated by samples.

4. Discussion

4.1. Comparison between MENA and expert IP corpora

As column (MENA vs. INT (*t*-test)) in Table 5 reveals, our results indicate significant differences between the MENA and expert IP corpora in all syntactic complexity. In terms of overall syntactic complexity, as presented in Table 2, the mean abstract length in the expert IP corpus appears higher than that in the MENA corpus (M = 186.39, SD = 31.517 and M = 185.92, SD = 51.430, respectively). However, the authors in the MENA corpus wrote significantly more sentences per abstract (M = 7.74, SD = 2.783) than the expert IP authors (M = 6.66, SD = 1.817).

A closer look at the mean length of sentence (MLS) reveals that although the expert IP corpus had fewer sentences per abstract, the mean sentence length (M = 29.82, SD = 9.587) is significantly higher than those in the MENA corpus (M = 25.13, SD = 6.608). This indicates that the expert IP corpus is syntactically more complex than the MENA corpus. The mean sentence length per abstract in the expert IP corpus is in line with Biber and Conrad (2009), who suggested that an average sentence length of 30 words is the standard. According to Biber and Gray (2010), academic writing is constructed using longer sentences and T-units)MLT(. However, the present

study's findings contradict those of Tovar-Viera (2022), who reported abstracts in Ecuadorian journals that had sentences consisting of 35–38 words. Significantly, excessive word counts may render abstracts more grammatically complex (Tovar-Viera, 2022).

The number of clauses is among the most widely employed measures of syntactic complexity in SLA research (Pallotti, 2015). Dependent clauses are particularly relevant to syntactic complexity. In the present study, we found that writers in the expert IP corpus employ more and longer dependent clauses (i.e., DC/C and DC/T) than their MENA counterparts. Accordingly, writers in the expert IP corpus appear to support Hyland's (2002) argument that higher subordination is widespread in academic writing.

Regarding the amount of coordination, expert IP corpus shows a significantly higher employment of coordination at both the sentential (T/S) and phrasal levels (CP/C, CP/T). Wu et al. (2020) identified greater use of coordination as a signature feature of ELF research articles. The use of subordination and coordination by MENA and expert IP writers contradicts Othman's (2004) assertion that English "makes use of more subordination than coordination, while Arabic favors the use of coordination rather than subordination" (p. 12). According to Oshima and Hogue (1999), complex sentences and participial phrases (e.g., subordination) are preferred and considered "more mature, interesting, and effective in style" (p. 163). The differences between their findings and those of the present study may be attributable to the part-genre given that our study examined RA abstracts, while Oshima and Hogue (1999) did not focus specifically on abstracts.

Regarding phrasal sophistication, the expert IP corpus exhibited more complex nominals per clause, T-unit, and verb phrases per T-unit than the MENA corpus. It may be that the main goal of phrasal sophistication is to compress information using pre- and post-modifiers. In another context, Youssef (2019) reported higher usage of complex nominals per T-unit and verb phrases per T-unit in linguistics RA abstracts written by English native speakers from the United Kingdom than among Egyptian writers, who exhibited a slightly higher usage of complex nominals per clause.

4.2. Comparison among the three corpora

The results indicated that the expert IP corpus differed from ME_AR and NA_AR with respect to overall syntactic complexity. The results also indicated a higher degree of similarity between ME_AR and NA_AR. These

outcomes may be attributed to the fact that writers from homogeneous L1 backgrounds (e.g., Arabic) may share some linguistic aspects regardless of their geographical location. Wu et al. (2020) reported no significant differences between ELF writers, who belonged to several countries and had different L1 backgrounds, and American English writers in terms of overall sentence complexity. Similar to writers in the MENA corpora, Yin et al. (2021) observed lower overall sentence complexity in writing produced by emerging Chinese writers than that of expert international publication writers. Our findings may be interpreted to indicate that MENA writers may have attained the mastery of sentence complexity, such as those found in the international corpus (Biber & Gray, 2010; Yin et al., 2021).

Regarding length of production unit, the expert IP corpus exhibited greater sentence length in sentence, clause, and T-unit, unlike the ME_AR and NA_AR, which exhibited similar results. The results also revealed no significant differences between NA_AR and expert IP corpora in MLC. Wu et al.'s (2020) findings also reported that MLC and MLS are significantly longer in the ELF RA corpus than in the American corpus. These findings support the claim that academic writing is typically characterized by longer clauses and sentences (Brown & Yule, 1983; O'Donnell, 1974; Wu et al., 2020). One possible interpretation may be that some authors employ longer sentences embedded with clauses, phrases, and T-units to facilitate communication (Wu et al., 2020) and presentation of meaning (Li & Ge, 2009).

The expert IP corpus shows more subordinations (i.e., clauses per T-unit, complex T-units per T-unit, dependent clauses per clause, dependent clauses per T-unit), followed by ME_AR and then NA_AR. supporting the widespread perception that academic writing includes more subordination (Hyland, 2009). Hinkel (2003) found that non-native speakers (NNS) writers used less subordination than their native speaker (NS) counterparts, given the fact that the MENA corpus includes L2 writers. Finally, the mean values of all subordination measures revealed no significant differences between ME_AR and NA_AR, suggesting that Arab writers across both MENA corpora may use a similar degree of subordination. Yin et al. (2021) similarly reported that IP expert writers employed more subordination than emerging Chinese writers in all part-genres other than the abstract.

Authors in the expert IP corpus used significantly more coordination than their counterparts in the MENA corpora at both the phrasal and sentential

levels, corroborating Wu et al.'s (2020) findings regarding ELF RAs (i.e., ELF writers use more coordination to significantly enhance clarity). It appears that expert IP writers' main purpose in employing more coordination in RA abstracts is to condense important information within the word limit, particularly in sections with limited word counts, such as abstracts (Ansarifar et al., 2018; Yin et al., 2021). Within the MENA corpora, ME_AR writers use more subordination, while their NA_AR counterparts employ more coordination. Interestingly, several studies have documented the use of coordination in Arabic (Alqinai, 2013; Dickins, 2017; Othman, 2004). ME_AR writers thus appear to be more influenced by their L1's writing style. Another possible interpretation may be related to writers' preferences, since no significant differences were identified between the corpora in terms of subordination and coordination usage.

The results reveal that the expert IP corpus used more complex nominal and verb phrases than the MENA corpora. These results are in partial agreement with Wu et al.'s (2020) findings that ELF writers employ more complex nominals per clause and T-unit but fewer verb phrases than American English writers. Ansarifar et al. (2018) also reported greater phrasal complexity in RA abstracts written by expert writers.

The significant differences across the three corpora in the complex nominals per clause (CN/C) measure merits further discussion. The means of the three corpora are as follows: ME_AR (M = 2.16, SD = .629) NA_AR (M = 2.33, SD = .682), and expert IP (M = 2.60, SD = .826). According to Lu (2010), the term 'complex nominal' refers to "(1) noun phrases with one or more of the following pre- or post-modifiers: adjective, possessive, prepositional phrase, adjective clause, participle, or appositive; (2) nominal clauses, and 3) gerunds and infinitives in subject position" (Ai & Lu, 2013, p. 255). First, in line with Wu et al. (2020), the differences confirm that ELF writers have a tendency to produce longer complex nominals and, in turn, longer sentences, while ELF writers rely mostly on complex nominals postmodified by prepositional phrases. This also corroborates Biber and Gray's (2010) findings that nominal/phrasal structures are deemed preferable to clausal structures.

Second, for both ELF and expert IP writers, writing for different journals and/or disciplines will impact how they write, as each journal has its own style and requirements. This will ultimately impact the use of nominals (CN/C). This impression was partially fueled by McCambridge (2015), who

stated that writing for ELF journals is inherently problematic because ELF authors write in English for international audiences but are not native speakers. Therefore, ELF writers rely heavily on frequent, well-regarded nominals in EAL journals. Wu et al. (2020) also asserted that postmodified prepositional phrases demonstrated explicitness in meaning, which ELF authors largely adopted.

Finally, one of the primary differences between the ME_AR and NA_AR authors is their L1 background: the former speak Arabic while the latter speak Arabic and French in most regions in North Africa. We assumed that authors from North Africa would use more complex nominals than authors from the Middle East owing to their French background. Our assumption was supported by Lu and Ai (2015), who determined the influence of the French language on English syntactic complexity. In particular, authors from French backgrounds might be expected to outperform authors from other backgrounds, such as German and Russian.

Close examination of a random sample from the corpora indicated that authors in the expert IP corpus employ complex nominal clauses to convey more complex information, using longer sentences to report studies' findings and implications, as Excerpts (1), (2), and (3) indicate, with several pre-/post-modifiers (**in bold**) to noun phrases and clauses (*in italics*).

(1) "The findings shed real-time conceptions ***of (un)successful academic stance and engagement in group oral contexts****, as well as confirm the usefulness of* ***verbal*** *protocols in revealing previously* ***hidden*** *complications for group* ***oral*** *assessments in an academic context, with* ***accompanying*** *suggestions for resolving such complications."* (INT_40)

(2) "Findings **of the study** show *that* ***the proposed intercultural*** *approach stimulates students' thinking, helps them better comprehend how to immerse themselves in* ***diverse*** *perspectives on* ***complicated international*** *issues, and helps them become* ***global*** *citizens able to deal effectively with multiculturalism in the work environment*." (NA_AR74)

(3) The study concluded *that the students committed* ***more transfer*** *errors in their use of English negation and the definite article than other types of errors in the same* ***syntactic*** *areas as a result of the effect of CJA*. (ME_AR18)

5. Conclusion

The present study has aimed to a) describe the features of syntactic complexity across the RA abstracts of MENA and expert IP corpora and b) determine the significant differences among the corpora. The MENA and expert IP corpora were inspected in terms of five syntactic complexity categories—length of production unit, amount of subordination, amount of coordination, phrasal complexity, and overall sentence complexity—incorporating 14 indices (Table 3).

Our findings revealed significant differences in all syntactic complexity categories among the MENA and expert IP RA abstracts, including overall sentence complexity (C/S). First, expert IP writers employed greater length of production unit and amount of subordination than the MENA writers; this result is in line with Hinkel's (2003) findings regarding international students with different L1 backgrounds (Japanese, Chinese, Indonesian, Korean, and Arabic). However, Wu et al. (2020) reported that ELF authors from ten different countries (China, Brazil, Finland, Czech, France, Spanish, Russian, Swedish, Italian, Romania) use longer sentences and more coordinate phrases and complex nominals in the interest of clarity and explicitness. The same holds true for coordinations, as the expert IP writers showed more coordinations, more complex phrases, and sentences than the MENA corpora; this result was also asserted by Ai and Lu (2013).

Furthermore, the MENA corpora, ME_AR and NA_AR, revealed almost no significant differences except in using complex nominals per clause (CN/C). This suggests that EAL authors from similar language backgrounds adopted similar SC features, distinguishing the quality of Arab authors' writing from that of expert IP authors. Nonetheless, one of this study's most striking findings was the significant difference between ME_AR and NA_AR authors in their use of complex nominals per clause CN/C). Moreover, the NA_AR corpus exhibits higher values than the ME_AR corpus across almost all indices except in coordination and verb phrases. We assumed that this dichotomy could be a result of the authors' various L1 backgrounds since the authors from North Africa mostly speak Arabic and French, while writers from the Middle East speak only Arabic. This influence was supported by Lu and Ai (2015), but this issue is worth exploring further. Overall, the similarity of syntactic complexity between Middle East and North Africa corpora could be linked to the close relation between orthography and writing. In other words, it is known that orthography plays

important roles in writing (Thompson-Panos & Thomas-Ru□ i□, 1983). Therefore, the orthography of Arabic language (MENA authors' L1) could have contributed to the similarity between Middle East and North Africa. In the same vein, we assume that the differences in the majority of syntactic complexity features found in MENA and expert IP corpora are due to the alphabetical difference the writers use.

When comparing the expert IP corpus with either the ME_AR or NA_AR independently, the MENA corpora exhibited similar findings to the expert IP writers corpus, yet the expert IP corpus revealed higher means in all indices. More specifically, comparison of the expert IP writers corpus with each MENA corpus revealed that, of the 14 indices, the ME_AR showed no significant differences in two subordination indices (CT/T, DC/C) and one amount of coordination index (T/S). Meanwhile, the NA_AR showed no significant differences in the following two indices: length of production unit (MLC) and verb phrases per unit (VP/T). These findings may be attributed to the MENA writers' multiple attempts over various production stages to attain a higher syntactic complexity. They may also be related to the fact that the abstracts were edited by a native or a native-like English speaker (Tovar-Viera, 2022). Considering the minor dichotomy, we may conclude that expert IP writers employ their own writing schema in prestigious journals despite their various similarities.

Genre analysis literature indicates that RA abstracts published in more prestigious journals indexed in the Web of Science and Scopus differ from those in less prestigious publications that are not indexed in either of these two highly ranked databases (El-Dakhs, 2018). The present study's corpus includes both more (i.e., expert IP writers corpus) and fewer (i.e., MENA corpus) prestigious journals, as we might assume that the publications in the two journal groups differ with respect to quality or impact. This factor may significantly affect the writing in the publications. That is, the differences found in the study may be attributed not only to the authors' L1 backgrounds but also to the differences in the aforementioned journal factors. It would be interesting to investigate the differences between local and international journals in terms of their different audiences and writers' tendencies to accommodate their readers.

Our findings have several key implications for L2 writing pedagogy. First, EAL authors in general—and MENA authors in particular—may be regarded as unique language users given that they exhibit distinct

characteristics with respect to syntactic complexity. For this reason, it is important to enhance the awareness of parties interested in EAP regarding this issue in terms of designing or teaching learning materials. Meanwhile, it is important to note that this study demonstrated L1's influence on ERPP writing, even when EAL authors write for peer-reviewed journals, because writing in academia is challenging. Thus, Biber and Gray (2010) asserted that academic writing requires specific characteristics—for example, special attention should be paid to degree of phrasal sophistication and complex nominals per clause, since these features showed statistical differences across all three corpora. It will also be beneficial if genre-writing teachers design specific activities around analysis of particular syntactic features or enhancing the use of longer sentences/clauses in RA abstracts. Finally, ERPP researchers must pass EAL findings on to practitioners to facilitate further action research (Jenkins, 2015).

This study has several limitations, highlighting avenues for future research. First, the present study focuses exclusively on applied linguistics. Future studies may complement the present study by comparing MENA and expert IP authors writing for the applied linguistics discipline in comparison with another discipline. The exploration of different disciplines would yield further insights into any influence resulting from different majors. For instance, comparisons among disciplines in hard sciences and social sciences would uncover a possible influence the disciplines have on syntactic complexity, especially complex nominals per clause which displayed a significant difference among the three corpora. Similarly, comparing Middle East with North Africa authors who write for the same journals in particular fields would reveal the extent to which the syntactic complexity is influenced by the authors L1 culture and/or educational system. Second, the study analyzed a small corpus and involved abstracts exclusively which may affect the generalizability of our findings. Future investigations should examine how other sections of scholarly RAs differ (i.e., introduction, methods, and results) and how international journals might resemble one another or differ in this regard. Further, analyzing larger corpora would help in providing a comprehensive picture of the study findings and understanding deeply the MENA usage of syntactic complexity. Another future avenue would be to investigate the syntactic complexity of MENA writers who have succeeded in publishing in the scholarly journals used in the present study. Lastly, we adopted Lu's (2010) 14 measures for our analysis, and it would be interesting to employ more comprehensive and fine-grained indices that take into

account specific subtypes of clausal or phrasal structures for comparisons between MENA and expert IP writers. The reason behind this suggestion is that syntactic complexity includes various sorts of linguistic features, such as discourse features, which can be investigated in more detail by using fine-grained analysis.

Article history:
Received 21 May 2022
Received in revised form 17 August 2023
Accepted 3 March 2024

References

Abdeljaoued, M., & Labassi, T. (2020). English as the lingua franca of academic publishing in Tunisia. *World Englishes*, *40*(2), 245-258. https://doi.org/10.1111/weng.12511

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). John Benjamins. https://doi.org/10.1075/scl.59.15ai

Alharbi, L. M., & Swales, J. M. (2011). Arabic and English abstracts in bilingual language science journals: Same or different? *Languages in Contrast*, *11*(1), 70-86. https://doi.org/10.1075/lic.11.1.06alh

Alotaibi, H. (2013). *Research article abstracts and introductions: A comparative genre-based study of Arabic and English in the fields of educational psychology and sociology.* PhD dissertation. Texas A&M University-Commerce.

Alqinai, J. (2013). Mediating punctuation in English Arabic translation. *Linguistica Atlantica*, *32*, 2-20.

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

Bhatia, V. (1993). *Analyzing genre: Language use in professional settings*. Longman

Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, *91*, 102248. https://doi.org/10.1016/j.system.2020.102248

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2-20. https://doi.org/10.1016/j.jeap.2010.01.001

Breeze, R. (2008). Researching simplicity and sophistication in student writing. *International Journal of English Studies, 8*(1), 77-104. https://doi.org/10.1016/j.jeap.2010.01.001

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins. https://doi.org/10.1075/lllt.32.02bul

Cargill, M., & Burgess, S. (2008). Introduction to the special issue: English for research publication purposes. *Journal of English for Academic Purposes*, *2*(7), 75-76. https://doi.org/10.1016/J.JEAP.2008.02.006

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, *44*, 51-62. https://doi.org/10.1016/j.jslw.2019.03.005

Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics*, *13*(1), 53-76. https://doi.org/10.1017/S0142716400005427

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*, 66-79. https://doi.org/10.1016/j.jslw.2014.09.006

Demir, C. (2021). A prelude to determine 'datum point' for MA writing in English: Comparing syntactic complexity of inner-circles and expanding-circles. *Arab Journal of Applied Linguistics*, *6*(2), 31-70.

Dickins, J. (2017). The pervasiveness of coordination in Arabic, with reference to Arabic>English translation. *Languages in Contrast*, *17*(2), 229-254. https://doi.org/10.1075/lic.17.2.04dic

Dickins, J., Hervey, S. G. J., & Higgins, I. (2016). *Thinking Arabic translation* (2nd Ed.). Routlege.

El-Dakhs, D. A. S. (2018). Comparative genre analysis of research article abstracts in more and less prestigious journals: Linguistics journals in focus. *Research in Language*, *16*(1), 47-63. https://doi.org/10.2478/rela-2018-0002

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474-509. https://doi.org/10.1093/applin/amp042

Elyas, T., & Mahboob, A. (2020). World Englishes in the Middle East and North Africa (MENA). *World Englishes*, *40*(2), 154-158. https://doi.org/10.1111/weng.12504

Fallatah, W. (2016). Features of Saudi English research articles abstracts. *Arab World English Journal (AWEJ)*, *7*(2), 368-379. https://dx.doi.org/10.24093/awej/vol7no2.25

Flowerdew, J. (2015). Some thoughts on English for Research Publication Purposes (ERPP) and related issues. *Language Teaching*, *48*(2), 250-262. https://doi.org/10.1017/S0261444812000523

Flowerdew, J., & Habibie, P. (2022). *Introducing English for research publication purposes*. Routledge.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, *59*(4), 866-896. https://doi.org/10.1111/j.1467-9922.2009.00528.x

Friginal, E., & Mustafa, S. S. (2017). A comparison of U.S.-based and Iraqi English research article abstracts using corpora. *Journal of English for Academic Purposes*, *25*, 45-57. https://doi.org/10.1016/j.jeap.2016.11.004

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, *37*(2), 275-302. https://doi.org/10.2307/3588505

Hyland, K. (2002). Options of identity in academic writing. *ELT Journal*, *56*(4), 351-358. https://doi.org/10.1093/elt/56.4.351

Hyland, K. (2009). *Teaching and researching writing* (2nd ed.). Longman.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58-69. https://doi.org/10.1016/j.jslw.2016.01.005

Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice*, *2*(3), 49-85. https://doi.org/10.1515/eip-2015-0003

Kachru, B. B. (Ed.) (1992). *The other tongue: English across cultures* (2nd ed.). University of Illinois Press.

Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019). Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics*, *29*, 161-170. https://doi.org/10.1111/ijal.12255

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*(4), 439-448. https://doi.org/10.2307/3586142

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579-589. https://doi.org/10.1093/applin/amp043

Li, L.-J. & Ge, G.-C. (2009). Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985-2004). *English for Specific Purposes*, *28*(2), 93-104. https://doi.org/10.1016/j.esp.2008.12.004

Li, Y., & Flowerdew, J. (2020). Teaching English for Research Publication Purposes (ERPP): A review of language teachers' pedagogical initiatives. *English for Specific Purposes*, *59*, 29-41. https://doi.org/10.1016/j.esp.2020.03.002

Lorés-Sanz, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes*, *23*(3), 280-302. https://doi.org/10.1016/j.esp.2003.06.001

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., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*, 16-27. https://doi.org/10.1016/j.jslw.2015.06.003

Mancilla, R. L., Polat, N., & Akcay, A. O. (2017). An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions. *Applied Linguistics*, *38*(1), 112-134. https://doi.org/10.1093/applin/amv012

McCambridge, L. (2015). Academic writing in an ELF environment: Standardization, accommodation-or transformation? In T. Lillis, K. Harrington, M. R. Lea & S. Mitchell (Eds.), *Working with academic literacies: Case studies towards transformative practice* (pp. 185-193). Parlor Press.

O'Donnell, R. C. (1974). Syntactic differences between speech and writing. *American Speech*, *49*(1/2), 102-110. https://doi.org/10.2307/3087922

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218-233. https://doi.org/10.1016/j.jslw.2010.10.003

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

Oshima, A., & Hogue, A. (1999). *Writing academic English* (3rd ed.). Longman.

Othman, W. (2004). Subordination and coordination in English Arabic translation. *Al-Basaer Journal*, *8*(2), 1-21.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117-134. https://doi.org/10.1177/0267658314536435

Swales, J. M. (1997). English as Tyrannosaurus rex. *World Englishes*, *16*(3), 373-382. https://doi.org/10.1111/1467-971X.00071

Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, *4*(4), 307-323. https://doi.org/10.1016/j.jeap.2005.07.006

Thompson-Panos, K., & Thomas-Ružić, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly*, *17*(4), 609-623. https://doi.org/10.2307/3586616

Tovar-Viera, R. (2022). Syntactic complexity in journal research article abstracts written in English. *MEXTESOL Journal*, *46*(2), 1-12. https://doi.org/10.61871/mj.v46n2-2

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity*. University of Hawaii at Manoa.

Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 71-83). Cambridge University Press.

WrELFA (2015). *WrELFA corpus*. https://www.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/wrelfa-corpus

Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, *43*, 100798. https://doi.org/10.1016/j.jeap.2019.100798

Yin, S., Gao, Y., & Lu, X. (2021). Syntactic complexity of research article part-genres: Differences between emerging and expert international publication writers. *System*, *97*, 102427. https://doi.org/10.1016/j.system.2020.102427

Youssef, A. (2019). Syntactic complexity and lexical diversity in English conference abstracts: Investigating cross-disciplinary effects with native speaker baseline. *HERMS*, *8*(2), 33-70.

Basim Alamri is an Associate Professor in the English Language Institute at King Abdulaziz University, Saudi Arabia. His research interests revolve around topics related to L2 academic writing for publication, including genre

studies and corpus linguistics, EAP/ESP, discourse analysis, and technology in the classroom.

Assem Alqarni is an Assistant Professor in the Applied College at King Abdulaziz University, Saudi Arabia. He is interested in formulaic language, corpus linguistics, and academic writing. His research investigates L1 and L2 use of language in light of corpus and formulaic language.