# Syntactic Complexity and Lexical Diversity in English Conference Abstracts: Investigating Cross-Disciplinary Effects with Native Speaker Baseline (*)

**Amany Youssef**

**Associate Professor of Linguistics**

**Department of English Language and Literature**

**Faculty of Arts and Humanities**

**The British University in Egypt**

## Abstract

This study investigates potential cross-disciplinary effects on the extent of syntactic complexity and lexical diversity (SC & LD, respectively) in English conference abstracts authored by Egyptian (Arabic L1) researchers in two disciplines: Linguistics and Nuclear Science. The study establishes a native speaker baseline through parallel analysis of British-authored abstracts in the two disciplines under investigation. The data comprises 100 single-authored English conference abstracts, evenly divided over four contrastive categories: Eg(yptian)-Ling(uistics), Eg-N(uclear) Sc(ience), Br(itish)-Ling, and Br-NSc. Using two computational tools, L2 Syntactic Complexity Analyzer (L2SCA) and TEXT INSPECTOR, scores of SC and LD, respectively, have been extracted and managed in MS Excel through some statistical tools. The results have indicated significant and uniform cross-disciplinary effects in both the native and non-native groups in terms of SC, where the Ling abstracts have displayed longer and more complex production units. Furthermore, significant language nativity effects have been observed in terms of SC; English natives have been found to use more subordination, which is characteristic of more mature writing in their L1. Arabic natives, on the other hand, have made greater use of coordination which is the preferred structure-combining operation in their L1. In terms of LD, the native groups have outperformed the non-native groups across both disciplines.

## Keywords

Measuring linguistic complexity, syntactic complexity, lexical diversity, L2 writing, disciplinary effects

**الملخص:**

تبحث هذه الدراسة في تأثير اختلاف التخصصات العلمية على مدى التعقيد النحوي والتنوع المعجمي في ملخصات المؤتمرات باللغة الإنجليزية والتي ألفها الباحثون المصريون في مجالي

**(*)**Syntactic Complexity and Lexical Diversity in English Conference Abstracts, Vol. 8, Issue No.2, April 2019, pp.33-70.

اللغويات والعلوم النووية. تقارن الدراسة الأداء اللغوي للباحثين المصريين مع أداء متحدثي اللغة الإنجليزية من البريطانيين وذلك من خلال تحليل متوازي للملخصات التي كتبها باحثون بريطانيون في التخصصين قيد البحث. تشمل عينة البحث ١٠٠ ملخص مفرد للمؤتمرات باللغة الإنجليزية، مقسمة بالتساوي على أربع فئات متقابلة: مصريون-لغويات، مصريون-علوم نووية، بريطانيون-لغويات، بريطانيون-علوم نووية وقد استخدمت الدراسة اثنين من الأدوات الحاسوبية: محلل التعقيد النحوي L2 Syntactic (L2SCA) ومحلل التنوع المعجميText Inspector ، وقد تم استخراج النتائج وإدارتها في MS Excel من خلال عدد من الأدوات الإحصائية. وقد أشارت النتائج إلى تأثير اختلاف التخصصات على التعقيد النحوي، حيث استخدم باحثو اللغويات جملاً وعبارات أطول وأكثر تعقيدًا. وعلاوة على ذلك، لوحظ تأثير اللغة الأم فقد وجد أن الباحثين البريطانيين استخدموا مزيدًا من التراكيب التبعية subordination التي تتميز بها لغتهم الأولى في حين أن للباحثين المصريين قد استخدموا العطف اللغوي وهو ما يميز اللغة العربية. من حيث التنوع المعجمي، تفوق الباحثون البريطانيون على الباحثين المصريين بصرف النظر عن التخصص.

**الكلمات الدالة**

قياس التعقيد النحوي - التعقيد النحوي - والتنوع المعجمي – الكتابة باللغة الثانية - تأثير اختلاف التخصصات العلمية.

# 1. Introduction

Linguistic complexity may be viewed as "a dynamic property of the learner's L2 system at large …the degree of elaboration, the size, breadth, width, or richness of the learner's L2 system or 'repertoire', that is, … the number, range, variety or diversity of different structures and items that he [the learner] knows or uses" (Bulté & Housen[(1)], 2012, p. 25), as "[t]he extent to which language produced in performing a task is elaborate and varied" (Ellis, 2003, p. 340; as cited in Lu & Ai, 2015, p. 17), or as "the range and the sophistication of grammatical resources exhibited in language production" (Ortega, 2015, p. 82).

Linguistic complexity, however, is not the sole measure of L2 learner's performance; two other performance descriptors, accuracy and fluency, contribute to the Complexity-Accuracy-Fluency (CAF) triad of the L2 proficiency model proposed by Skehan (1989) "for the oral and written assessment of language learners as well as indicators of learners' proficiency underlying their performance" (as cited in Housen & Kuiken, 2009, p. 461). However, of the three components of the CAF, complexity has received the most attention in L2 research (Ansarifar, Shahriari, & Pishghadam, 2018).

Linguistic complexity has been extensively researched and

recognized in L2 research as a multidimensional construct operationalized through a wide range of mostly automatic measures (Lu, 2010; Crossley & McNamara, 2012, 2014; Kalantari & Gholami, 2017) targeting different language domains: lexical, morphological, syntactic/grammatical, and phonological (Bulté & Housen, 2012). In the literature, most studies have focused on the lexical and/or syntactic domains in the written production of L2 learners of different levels of proficiency (Wang & Slater, 2016), of different L1 backgrounds (Lu & Ai, 2015), of different levels of pragmatic competence (Youn, 2014), reading texts of varying degrees of complexity (Douglas & Miller, 2016), writing on different topics (Yang, Lu, & Weigle, 2015), and targeting different L2s (Bulon, Hendrikx, Meunier, & Van Goethem, 2017).

The genre that received most attention in the literature was learners' (argumentative) essays; very few studies have investigated linguistic complexity in research abstracts (Ansarifar et al., 2018). Furthermore, very few studies have attempted stronger validation of their results by comparing L2 performance to that of native speakers (Foster & Tavakoli, 2009; Lu & Ai, 2015). In fact, this is one of the significant aspects about the present study as explained in more detail in Section 2 (Research Statement and Hypotheses), Section 3 (Previous Studies), and Section 7 (Analysis). With regard to the tools of analysis, the present study has employed the L2 Syntactic Complexity Analyzer (L2SCA) and TextInpector for the syntactic and lexical domains, respectively (Sections 4 and 5). Statistical analysis has been carried out in MS Excel (Section 6, Data and Methodology). Notably, before carrying out the analysis, all abstracts have been screened for obvious grammaticality issues through MS Word grammar checker as shown in Section 7.

## 2. Research Statement and Hypotheses

The present study investigates potential cross-disciplinary (Linguistics vs. Nuclear Science) effects on both syntactic complexity and lexical diversity (SC & LD, respectively) in one of the under-researched genres, namely, research abstracts. The study establishes a native-speaker (British) baseline against which non-native (Egyptian) performance is compared. 100 recently published conference abstracts are evenly divided over four contrastive categories: Eg(yptian)-Ling(uistics), Eg-N(uclear) Sc(ience), Br(itish)-Ling,

and Br-NSc. Cross-group comparisons have been carried out to test three initial hypotheses illustrated in Table 1 below:

**Table 1: The three initial hypotheses of the present research**

| | |
|---|---|
| H1 | **Disciplinary Effects:** In both native and non-native groups, linguistics authors, whose discipline entails greater exposure to linguistic scrutiny, would display significantly higher levels of both SC and LD in their writing than their nuclear science counterparts. |
| H2 | **Language Nativity Effects:** In each discipline, the native authors would outperform the non-native authors in terms of SC and LD. |
| H3 | **Correlation between syntactic complexity and lexical diversity** scores: Since both are reflections of grammatical richness, it is expected that scores for SC and LD would correlate across all groups. |

The first two hypotheses predict greater complexity by certain respective groups, but neither of them specifies which, if any, of the measures of SC or LD would be affected. The third hypothesis, which predicts no difference between the syntactic and lexical domains in the complexity profile across all groups, is supported by an earlier study by Douglas & Miller (2016; see Section 3). The results of testing these hypotheses are presented in Section 7 (Analysis) and discussed in Section 8 (Discussion).

## 3. Previous Studies on Syntactic and Lexical Complexity

Syntactic and lexical complexity have been investigated either combined (e.g. Douglas & Miller, 2016; Bulon et al., 2017) or independently (e.g. Lu, 2010, 2011 and 2017; Ortega, 2015 for SC; and Salazar, 2011; Kyle & Crossley, 2015 for LC) in L2 production, with more focus on writing rather than on speaking (Biber & Gray, 2010; Chen & Zechner, 2011).

When combined, syntactic and lexical complexity were often found to correlate with one another as reflections of grammatical richness and/or development in L2 writing (Douglas & Miller, 2016). Consistent syntactic and lexical short (Storch & Tapper, 2009; Bulté & Housen, 2014; Schenker, 2016) and long-term (Bulon et al., 2017) gains in L2 writing proficiency

were reported. Storch and Tapper (2009), for example, found that upon completion of an English for Academic Purposes (EAP) course in an Australian university, graduate students' writings showed improvement in terms of accuracy, use of academic vocabulary, and structure. On the other hand, lack of significant improvement in both lexical and grammatical complexity (despite improvement in fluency) has been reported by Knoch, Rouhshad, Oon, & Storch (2015) in international students' writing after three years of study in an English-medium Australian university. Commenting on the nature of their immersion experience, the students in Knoch et al.'s study reported in the interviews that "they were not required to do much writing in their degree studies and when they did, their lecturers almost exclusively commented on the content of their writing" (Knoch et al., 2015, p. 39). This suggests that the development of grammatical richness in learners' writing requires conscious attention and practice. Differential findings on the correlation between syntactic and lexical complexity have been reported by Xudong, Cheng, Varaprasad, & Leng (2010) who investigated the impact of an EAP course on the development of academic writing abilities of ESL/EFL graduate students at the National University of Singapore. They found that while students' post-course essays contained more academic vocabulary, they did not progress in terms of grammatical accuracy and fluency. The findings concerning language complexity development remain inconclusive perhaps due to the multi-factorial nature of this process.

As a dependent variable, sensitivity of syntactic and/or lexical complexity in L2 writing has been explored in relation to L1 background (Lu & Ai, 2015), the specific L2 being learned (Bulon et al., 2017), level of pragmatic competence (Youn, 2014), complexity of texts that learners most frequently read (Douglas & Miller, 2016), writing topic and writing quality (Yang et al., 2015), and cumulative experience (Ansarifar et al., 2018). Lu and Ai (2015) contrasted all 14 measures in the L2SCA (see Section 4) in 1400 EFL argumentative essays evenly representing seven L1s from four different language families: Sino-Tibetan (Chinese), Japonic (Japanese), Niger-Congo (Tswana), and Indo-European (Bulgarian, English, French, German, Russian)) extracted from the International Corpus of Learner English. They found that, when all learner groups were collapsed into one NN group, only 3 of 14 measures of syntactic complexity showed difference

from the (N)ative group. Yet, when the NN groups were disaggregated by L1 and compared against the N group, all 14 measures exhibited differences. They concluded that learners with different L1 backgrounds may not develop in the same ways in all dimensions of SC. Bulon et al. (2017) reported L2-sensitive impact of Content and Language Integrated Learning (CLIL) education on L2 proficiency of secondary school French-speaking Belgian pupils learning English and Dutch. Nearly all complexity measures significantly improved in the Dutch texts written by CLIL pupils while only half of such measures showed significant development in the English texts. Douglas and Miller (2016) reported strong correlation between syntactic and lexical complexity of 65 graduate students' most frequently leisurely-read texts on the one hand and of those students' writing on the other. Yang et al. (2015) investigated topic effect in 100 ESL argumentative essays by graduate students with non-homogenous L1s: Chinese, Korean, and Japanese. They found that one of the topics (their future plans), which, according to Yang et al. (2015) would naturally demand causal reasoning in task performance, invited a higher amount of subordination (finite and non-finite) and greater global sentence complexity measured via mean length of sentence (MLS; see Section 4). On the other hand, the other topic (appearance) elicited more elaboration at the finite clause level as reflected in more coordinate phrases and complex noun phrases. Furthermore, Yang et al. investigated the relation between SC and the writing quality as judged by human raters using the TOEFL iBT Independent writing scoring guide. Higher scores significantly correlated with global sentence complexity and T-unit complexity measured via mean length of T-unit (MLTU; see Section 4). Ansarifar et al. (2018) compared research abstracts by MA-level L1 Persian writers, PhD-level L1 Persian writers, and published writers from the field of applied linguistics in terms of phrasal modification features. They found that while the (less experienced) MA-level writers significantly differed from the expert writers, the (more experienced) PhD-level did not. Ansarifar et al. concluded that academic writing became more complex with experience.

Few studies in L2 performance have paid attention to establishing a native-speaker baseline for interpreting nonnative-speaker performance (Foster & Tavakoli, 2009; Lu & Ai, 2015 (discussed above)). Foster & Tavakoli (2009) subjected the performance of both native and non-native

speakers to equal scrutiny because they believed that "[i]f we investigate how learners perform language tasks, we should distinguish what performance features are due to their processing an L2 and which are due to their performing a particular task" (p. 866). They studied the writing of 100 learners of English: 40 learners from a variety of L1 backgrounds who were based in London (i.e. within the target language community) and 60 native speakers of Persian who were based in Tehran (i.e. outside the target language community). Foster & Tavakoli explored the influence of two narrative design features – namely, complexity of storyline and tightness of narrative structure – on the complexity, fluency, accuracy, and lexical diversity in the language of both native and non-native speakers. They found that storyline complexity correlated with more subordinate (i.e. complex) language by both native and nonnative speakers. With regard to narrative structure, on the other hand, a tight structure (as opposed to a loose structure) correlated with almost equally higher lexical diversity in the writing of both the native speakers and the London-based learners. The Tehran-based learners were lagging behind.

Benchmarking native/proficient speakers' performance has further enabled Wang and Slater (2016) to reveal the specific grammatical features where L2 performance diverged. Wang and Slater contrasted 38 written personal statements by Chinese non-English major college students with 15 personal statements by English proficient users (probably native speakers of English) extracted from the websites of a number of Canadian and American universities. The results indicated that Chinese EFL students used significantly fewer complex nominals, and shorter clauses and sentences than did the more proficient users (or rather, native speakers) of English.

In terms of data, most studies on complexity in written production have targeted academic (often argumentative) essays produced by L2 learners of different levels of proficiency (Lu, 2010, 2011 and 2017; Lu and Ai, 2015; Yang, Lu, &Weigle, 2015). Very few studies have targeted other genres like cover letters for job applications written by adult graduate students (Douglas and Miller, 2016), e-mail exchanges (Schenker, 2016), personal statements by students as they enter college (Wang and Slater, 2016), and research abstracts (Ansarifar et al., 2018).

In the present study, written research abstracts by both native and nonnative speakers across two disciplines (Linguistics and Nuclear Science) have been subjected to equal scrutiny. If L1 performance is shown to be

influenced by disciplinary effects in the same way as L2 performance is, then this would give stronger validity to disciplinary effects.

## 4. Measuring Syntactic Complexity

Traditionally, SC has been linked to clausal rather than to phrasal complexity (Diessel, 2004 and Ravid & Berman, 2010 as cited in Yang et al., 2015; Givo´n, 2009). However, several L1 and L2 developmental studies have taken phrasal complexity, especially that of noun phrases (NPs) as a significant indicator of syntactic complexity (see Ansarifar, et al., 2018) in terms of the complexity of pre- and post- phrasal modification).

In the literature, SC indicators range from simple measures such as the number of words before the main verb, where simple sentences like "She laughs" or "The girl left" with only one and two words before the main verb, respectively, would be contrasted with a sentence like "Thus, in syntactically simple English sentences there are few words before the main verb," with seven words before the main verb (McNamara, Crossley, & McCarthy, 2010, p. 69), to more elaborate sets of measures that target both clausal and phrasal complexity in terms of length of production units at all syntactic levels (phrase, clause, sentence, and T-unitc[(2)]), ratios of certain syntactic structures as well as the amount of clausal coordination and subordination (Wolfe-Quintero, Inagaki, & Kim, 1998; Bulte´ & Housen, 2012, 2014, 2015; Ortega, 2015;). Bulte´ & Housen (2014), for example, have compiled a set of measures targeting sentential complexity via mean length of sentence (MLS), mean length of T-unit (MLTU), sentence types (simple, compound, complex), and coordination; clausal complexity via mean length of clause (MLC); and phrasal complexity through mean length of noun phrase (MLNP). Interestingly, Bulte´ & Housen (2014) refrained from using automatic computer-assisted tools like *Coh-Metrix* (see Note **Error! Bookmark not defined.**) and the *L2 Syntactis Complexity Analyzer* (L2SCA; this section) claiming that they created 'computational noise,' and instead segmented the sentences in their data manually and managed them in MS Excel sheets. Martinez (2018), on the other hand, adopted B&H's compiled set of measures and used *Coh-Metrix*, which, in its public versions, analyzes texts on over 200 measures of cohesion, language, and readability.

In the present study, the L2SCA,which has been described as

providing "the most granular and comprehensive identification of writing samples" (Douglas and Miller, 2016, p. 4), is used. It is a web-based tool that was developed by Lu (2010, 2011) at Pennsylvania State University based on an extensive review of the literature on SC. This software analyzes the data using *Stanford Parser* and *Treegex*, producing results for 14 syntactic complexity indicators, including length and density of several syntactic structures, as well as the amount of coordination and subordination (see Table 2). L2SCA is open to public use in its single mode at https://aihaiyang.com/software/l2sca/single/. Single mode enables concurrent independent processing of up to two texts, with a maximum of 1000 word each.

**Table 2: Syntactic Complexity Measures**

| | | | |
|---|---|---|---|
| Length of production unit | 1. | MLC | Mean length of clause |
| | 2. | MLS | Mean length of sentence |
| | 3. | MLT | Mean length of T-unit |
| Amount of subordination | 4. | C/T | Number of clauses per T-unit |
| | 5. | CT/T | Complex T-unit ratio |
| | 6. | DC/C | Number of dependent clauses per clause |
| | 7. | DC/T | Number of dependent clauses per T-unit |
| Amount of coordination | 8. | CP/C | Number of coordinate phrases per clause |
| | 9. | CP/T | Number of coordinate phrases per T-unit |
| | 10. | T/S | Number of T-units per sentence |
| Amount of phrasal sophistication | 11. | CN/C | Number of complex nominals per clause |
| | 12. | CN/T | Number of complex nominals per T-unit |
| | 13. | VP/T | Number of verb phrases per T-unit |
| Overall sentence complexity | 14. | C/S | Number of clauses per sentence |

(Lu, 2017, p. 503)

The L2SCA with its 14 measures was used by Lu and Ai (2015),

Wang and Slater (2016), and with minor adaptations by Yang et al. (2015) who selectively computed six measures—MLS, MLTU, MLC, TU/S, DC/TU, and CP/C—with the original version of the L2SCA. Yet, with regard to NP complexity, they claimed to have modified the pattern used in the L2SCA to identify complex NPs with a more inclusive characterization based on Biber, Gray, & Poonpon (2011) as noun phrases that contain one or more of the following: pre-modifying adjectives, post-modifying prepositional phrases, and post-modifying appositives. In addition, Yang et al. (2015) aimed to calculate a new SC index, namely, non-finite elements per clause (NFE/C) through subtracting 1 from the measure of verb phrases per clause (VP/C) since a clause should contain one finite VP, and hence the other VPs would be non-finite.

## 5. Measuring Lexical Diversity

Lexical diversity (LD) which counts how many different words are used in a text is one of two popular parameters of lexical complexity. The other is lexical density which gives the ratio of content words (i.e. nouns, verbs, adjectives and some adverbs) in a text. Both parameters have been used to characterize later lexical development in written production as argued by Johansson (2008) who compared these two parameters and concluded that they could be used interchangeably.

The traditional lexical density measure is the type-token ratio (TTR) calculated as the ratio of different words (types) to the total number of words (tokens). It follows that TTR would be at its maximum value when the number of word types is equal to the total number of words (tokens), i.e. when every word type occurs just once across the whole text. In that case, the text is either very low in cohesion or very short. Hence, a shorter text would generally have a higher TTR value than a longer text. The conclusion is that TTR is substantially affected by text length, which would render it reliable only when comparing (longer) texts of equal length (Johnsson, 2008; Koizumi &In'nami, 2012). It follows that, since abstracts in general are fairly short texts, and since the abstracts under study display considerable variation in mean length (indicated by high SD values in Table 3 and statistically significant difference in mean length in Table 4, Section 6), TTR would be unreliable, and is hence avoided in the present study.

Two popular measures of lexical diversity are VocD and MTLD.

Both measures are generally claimed to overcome the problem of sensitivity to text length (Jarvis, 2013; McCarthy & Jarvis, 2010). The former, originally referred to as the *D* measure, was developed by Brian Richards and David Malvern (Richards & Malvern 1997, as cited in Johansson, 2008). It is calculated by plotting the predicted decline of the TTR as texts become longer, then comparing the resulting mathematical curve with empirical data from a text sample. MTLD was developed by McCarthy (2005; as cited in Koizumi & In'nami, 2012) and is calculated as "the mean length of word strings that maintain a criterion level of lexical variation" (McCarthy & Jarvis, 2010, p. 381). Since each of these two measures targets unique lexical information, researchers are advised to use them in combination, rather than using any single index (McCarthy & Jarvis, 2010).

TEXT INSPECTOR[3] (https://textinspector.com/), the tool used in the current study, is a tagger and text complexity analyzer which was initially proposed and prototyped by Professor Stephen Bax and further developed by the software team at *Versantus*. In its free on-line version, it analyzes written texts individually offering, among other things, the two popular measures of lexical diversity: VocD and MTLD.

## 6. Data and Methodology

The data is composed of 100 single-authored English research abstracts by native (British) and non-native (Egyptian) researchers. The abstracts have been randomly extracted from books of abstracts published from 2012 to 2018 by reputable national and international conferences in two disciplines: Linguistics and Nuclear Science. Extraction of abstracts from full published papers has been avoided since they would likely have been subjected to editorial modification, and hence would not offer a truer reflection of a researcher's linguistic production. In addition, to evade an undesirable overlap of individual and collective competencies, only single-authored abstracts have been targeted. This presented quite a challenge when collecting Nuclear Science abstracts, where collaborative, rather than individual, research is the common practice. As a screening procedure, all abstracts have been checked for obvious grammatical errors using MS Word grammar checker to filter out any abstract with a clustering of structural accuracy issues (see Table 5, Section 7 for the results of this screening).

Table 3 displays the details of the four evenly-represented datasets:

Br(itish)-Ling(uistics), Eg(yptian)-Ling, Br-N(uclear) Sc(ience), and Eg-NSc.

**Table 3: Details of datasets**

| | | **No. of Abstracts** | | **Mean Length** | **SD** | **WC** |
|---|---|---|---|---|---|---|
| **NN** | Eg-Ling | 25 | | 219.2 | 53.80598 | 5480 |
| | Eg-NSc | 25 | | 262.95 | 82.40595 | 5259 |
| | | | | | | |
| **N** | Br-Ling | 25 | | 298.2 | 71.20861 | 7455 |
| | Br-NSc | 25 | | 284.35 | 64.10026 | 5687 |
| **All Grps** | | 100 | | | | 23881 |

Variation in the mean length of the abstracts under study has led to the avoidance of the TTR measure of lexical density as mentioned in Section 5. It can be observed from Table 3 and Table 4 that Br-Ling wrote significantly longer abstracts than did their Br-NSc counterparts. Significant ≤0.05 *p* values are shaded in Table 4. Within the N(on-)N(ative) groups, however, the situation is reversed. Eg-NSc wrote longer abstracts than Eg-Ling, but not significantly so. Examination of the Eg-NSc abstracts revealed that they often included reference material and data lines, which must have affected the word count. The N groups (disaggregated by discipline; Br-Ling and Br-NSc) wrote longer abstracts than did their NN (Eg-Ling and Eg-NSc, respectively) counterparts. However, the difference is statistically significant only in the linguistics discipline.

**Table 4: Significance of mean differences in abstract length**

(a) (b)

| | | **Means** | ***p*** | | | | **Means** | ***p*** |
|---|---|---|---|---|---|---|---|---|
| **Ling** | Br | 298.2 | 0.000002 | | **Br** | Ling | 298.2 | 0.000285 |
| | Eg | 219.2 | | | | NSc | 284.35 | |
| | | | | | | | | |
| **NSc** | Br | 284.35 | 0.208158 | | **Eg** | Ling | 219.2 | 0.327686 |
| | Eg | 262.95 | | | | NSc | 262.95 | |

Lexical diversity has been analyzed through the web-based TEXT INSPECTOR (https://textinspector.com/) which provides, among other things, measures for two measures of lexical diversity: VocD and MTLD (see Section 5). Syntactic complexity, on the other hand, has been measured using the web-based L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010; http://aihaiyang.com/software/l2sca/single/) with 14 well-defined measures of mean length of production units as well as the amount of subordination and coordination in a clause (see Section 4).

Each of the 100 abstracts (see Table 3) in the datasets was analyzed using L2SCA and then TEXT INSPECTOR. After scores for syntactic complexity and lexical diversity measures have been obtained for each abstract in the datasets, they have been compiled manually in Excel sheets and a set of independent samples *t* tests have been run to detect significance of differences across disciplines and language nativity. Correlation tests have been run to compare the lexical to the syntactic complexity measures.

## 7. Analysis

Presentation of the analysis of the results obtained for syntactic complexity and lexical diversity measures is organized according to the three initial hypotheses referred to in the Introduction section (Table 1). The first hypothesis (H1) predicting disciplinary effects is motivated by the fact that linguists would naturally be more language-focused, and hence are expected to display greater grammatical richness in their writing. The second hypothesis (H2) predicts language nativity effects. The third hypothesis (H3) predicts correlation between scores for syntactic complexity and lexical diversity within the same group since both are reflections of grammatical richness.

Illustrative examples are extracted from the data and included in the analysis in sub-sections 7.1, 7.2, and 7.3. Each example is presented twice. In the first presentation of the example, the relevant tokens are marked via distinctive styles of underlining. In the second, a generative syntactic analysis of the example is presented via labelled bracketing (Carnie, 2013) to reveal the structure.

Initial grammar check screening of all abstracts through MS Word tools revealed very few grammatical errors across the four data sub-sets as

shown in Table 5.

**Table 5: Grammatical errors identified by MS Word grammar check in all abstracts**

| | | Type of error | No. of occurrences | Actual Occurrences |
|---|---|---|---|---|
| **N groups** | Br-Ling | None | | |
| | Br-NSc | 1. Number agreement on phrasal level | 1 | *"…enabling realization of* ****a scalable quantum devices*** *…"* (Abstract 12) |
| | | 2. Missing hyphenation that affects the structure | 1 | *"These heavy* ****element forbidden*** *lines are routinely used to determine electron temperatures and densities, …"*(Abstract 1) |
| | | 3. Two tensed verbs in one clause | 1 | *"It* ****is has been*** *shown that …"* (Abstract 19) |
| **NN groups** | Eg-Ling | 1. Subject-Verb Agreement | 1 | *"The corpus of the study* ****are*** *28 Text-Image Showcases …"* (Abstract 6) |
| | | 2. Number agreement on phrasal level | 1 | *"… from a highbrow to* ****a lowbrow values*** *through the analysis"*(Abstract 12) |
| | Eg-NSc | 1. Subject-Verb Agreement | 2 | *"The present study*****describe*** *the use of FO technique for water desalination …"* *"Different advanced* ****apparatus are*** *used for suchmeasurements."* (Abstract 4) |
| | | 2. Noun/Verb form | 1 | *"How the geometry of the radiation sources can* ****effect*** *the shielding design."* (Abstract 6) |
| | | 3. Verb form | 1 | *"… and these will* ***overcomes****."* (Abstract 4) |

This limited range of grammatical errors in the data is in line with the fact that the abstracts have been produced by mature writers who, even if non-native, have at least obtained a PhD degree in an English-medium discipline. Furthermore, the task of writing a research abstract is generally expected to involve careful revision on the part of authors to weed out any obvious grammatical errors.

## 7.1 First Hypothesis: Disciplinary Effects

Cross-disciplinary comparisons in syntactic complexity and lexical diversity are reported in this section for the N and NN groups. Table 6 displays the means, SD, and statistical *p* values for each of the 14 syntactic complexity measures in the NN groups. Shaded cells signal significant ≤0.05 *p* values.

**Table 6: Cross-disciplinary syntactic complexity comparison in the NN groups**

| | | | Eg-Ling | | Eg-NSc | | *p* |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | 1. | MLC | 15.35 | 3.67 | 17.54 | 4.76 | 0.03717 |
| | 2. | MLS | 27.54 | 8.28 | 24.95 | 4.77 | 0.09123 |
| | 3. | MLT | 25.78 | 6.52 | 23.18 | 5.30 | 0.06392 |
| **Amount of subordination** | 4. | C/T | 1.70 | 0.29 | 1.37 | 0.38 | 0.00057 |
| | 5. | CT/T | 0.55 | 0.21 | 0.28 | 0.20 | 0.00001 |
| | 6. | DC/C | 0.39 | 0.11 | 0.22 | 0.14 | 0.00002 |
| | 7. | DC/T | 0.68 | 0.28 | 0.35 | 0.41 | 0.00078 |
| **Amount of coordination** | 8. | CP/C | 0.58 | 0.23 | 0.56 | 0.30 | 0.39738 |
| | 9. | CP/T | 0.98 | 0.41 | 0.74 | 0.41 | 0.02293 |
| | 10. | T/S | 1.06 | 0.12 | 1.10 | 0.17 | 0.20353 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.29 | 0.77 | 2.55 | 0.87 | 0.13877 |
| | 12. | CN/T | 3.86 | 1.31 | 3.39 | 1.07 | 0.08436 |
| | 13. | VP/T | 2.49 | 0.49 | 1.90 | 0.51 | 0.00005 |
| **Overall sentence complexity** | 14. | C/S | 1.81 | 0.35 | 1.49 | 0.39 | 0.00182 |

It can be observed from Table 6 that, compared to Eg-NSc, Eg-Ling showed significantly higher use of complex T-units reflected through higher subordination in all four related measures (# 4-7: C/T, CT/T, DC/C, DC/T) and one related measure of coordination (CP/T, # 9).The complex T-Unit in example (1) contains three clauses: one main clause (Tense Phrase, TP) and two subordinate relative clauses (Complementizer Phrases, $CP_1$ and $CP_2$; marked in (1a) by double and dotted underlining, respectively) in a nested structure of successive post-modification. $CP_1$ post-modifies the NP *an Egyptian female media figure* while $CP_2$ post-modifies the NP *a field*.

(1) Eg-Ling, Abstract #18

a.

*Doaa Farouk is an example of an Egyptian female media figure who has entered a field which is well-occupied by male satirical writers of different ages and different backgrounds.*

b.

[TP Doaa Farouk is [NP an example of an Egyptian female media figure [$CP_1$ who has entered [NP a field [$CP_2$ which is well-occupied by male satirical writers of different ages and different backgrounds].

Complex T-units with subordinate clauses are more syntactically complex than simple or chained/coordinated constructions (Givo´n, 2009), and naturally include more than one VP. This explains the pattern that whenever subordination was high, it was accompanied with significant increase in phrasal sophistication via increased verb phrases per T-unit (VP/T, # 13) and overall sentence complexity (# 14).

Eg-NSc made greater resort to complex nominals (CN/C, # 11) which would enable writers to compress information through stacking modifiers before a head noun or through post-modification where a prepositional phrase or a relative clause follows the head noun. According to Price (1974), CNs have been noted to appear more frequently in technical writing (as cited in Douglas & Miller, 2016, p. 6). In the 41-word long example sentence in (2), there are two complex nominals. The first, $NP_1$, marked by single underlining in (2a) is headed by the N *interrelation* which is post-modified by the prepositional phrase, $PP_1$. The second CN, $NP_2$, marked by double underlining in (2a) is headed by the N *behavior* which is doubly pre-modified by the Adj(ective) *optical* and N *absorption*. In addition, the N *behavior* is post-modified by $PP_2$,where the N *samples* is post-modified by the reduced relative clause[(4)] (CP) with two adjunctive PPs ($PP_2$ and $PP_3$).

(**2**

(2) Eg-NSc, Abstract #23

a.

*In addition, we focused our attention on the interrelation between these changes and photoluminescence (PL) and optical absorption behavior of annealed and UV irradiated samples synthesized by chemical precipitation method at ambient temperature with a crystallite size (DSch.) ≈ 2 nm.*

b.

In addition, we focused our attention on [$NP_1$ the interrelation [$PP_1$ between these changes and photoluminescence (PL)]] and [$NP_2$ optical absorption behavior [$PP_2$ of annealed and UV irradiated samples [CP synthesized by chemical precipitation method [$PP_3$ at ambient temperature [$PP_4$ with a crystallite size (DSch.) ≈ 2 nm]]]]].

While Eg-NSc's greater use of CNs fell short of statistical significance, yet it might have contributed to significantly longer clauses MLC (# 1, $p$ = 0.03717).

Almost identical strong disciplinary effects can be observed in the N groups as shown in Table 7. Br-Ling displayed significantly greater amount of subordination in all four related measures (# 4-7), related highs in VP/T (# 13) and in overall sentence complexity (# 14). Example (3) below is a 50-word sentence which features a complex T-Unit. The subject, $NP_1$, marked by single underlining in (3a), is a CN headed by the N *project* and post-modified through subordination via the reduced relative clause,$CP_1$. The object, $NP_2$, marked by double underlining in (3a) is a CN headed by the N *map* and post-modified by the PP, *for English*. The complex T-Unit in (3) also contains a non-restrictive subordinated relative clause, $CP_2$ marked with wiggly underlining in (3a), which further contains another yet reduced relative clause, $CP_3$.

(3) Br-Ling, Abstract #15

a.

*The 'Mapping Metaphor with the Historical Thesaurus'*

*project, currently being undertaken at the University of Glasgow, is creating an interactive 'Metaphor Map' for English, which will show the 9 metaphorical connections between semantic domains made by speakers and writers of English from the Old English period to the present day.*

b.

[TP [$NP_1$ The 'Mapping Metaphor with the Historical Thesaurus' project, [$CP_1$ currently [$VP_1$ being undertaken at the University of Glasgow]]]], [T' is [$VP_2$ creating [$NP_2$ an interactive 'Metaphor Map' [PP for English]]]], [$CP_2$ which will [$VP_3$ show the 9 metaphorical connections between semantic domains [$CP_3$ [$VP_4$ made by speakers and writers of English from the Old English period to the present day]]]].

Hence, in addition to the main clause, the complex sentence in example (3) contains three subordination clauses (3 CPs) and four VPs.

Higher use of complex nominals (CN/C, # 11) by NSc authors achieved statistical significance in the N group. The underlined string in example (4) below is an example of a CN. The head N *analysis* is post-modified by the PP *of incidents*, where the N *incidents* is further post-modified by a reduced relative clause, CP.

(4) Br-NSc, Abstract #11

a.

*An analysis of incidents reported by States to the Incident and Trafficking Database (ITDB) during 2007-2012 has recently been completed by the IAEA's Division of Nuclear Security.*

b.

[NP An analysis [PP of incidents [CP reported by States to the Incident and Trafficking Database (ITDB) during 2007-2012]]] has recently been completed by the IAEA's Division of Nuclear Security.

Like the pattern observed in the NN group, Br-NSc abstracts showed a higher (than Br-Ling abstracts) mean in MLC (# 1), but it fell short of

statistical significance. The Br-Ling were significantly higher regarding the other measures of length of production unit (MLS & MLT, # 2 and 3, respectively).

**Table 7: Cross-disciplinary syntactic complexity comparison in the N groups**

| | | | **Br-Ling** | | **Br-NSc** | | ***p*** |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | 1. | MLC | 15.834 | 5.346 | 16.206 | 3.866 | 0.390 |
| | 2. | MLS | 30.364 | 9.600 | 26.138 | 5.701 | 0.032 |
| | 3. | MLT | 27.872 | 8.465 | 24.169 | 5.637 | 0.037 |
| **Amount of subordination** | 4. | C/T | 1.794 | 0.370 | 1.526 | 0.308 | 0.004 |
| | 5. | CT/T | 0.482 | 0.146 | 0.392 | 0.174 | 0.027 |
| | 6. | DC/C | 0.395 | 0.094 | 0.316 | 0.125 | 0.007 |
| | 7. | DC/T | 0.736 | 0.314 | 0.517 | 0.281 | 0.006 |
| **Amount of coordination** | 8. | CP/C | 0.425 | 0.182 | 0.469 | 0.257 | 0.243 |
| | 9. | CP/T | 0.739 | 0.306 | 0.698 | 0.367 | 0.337 |
| | 10. | T/S | 1.096 | 0.126 | 1.090 | 0.110 | 0.434 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.260 | 0.583 | 2.538 | 0.768 | 0.078 |
| | 12. | CN/T | 4.057 | 1.279 | 3.779 | 1.139 | 0.211 |
| | 13. | VP/T | 2.632 | 0.546 | 2.180 | 0.505 | 0.002 |
| **Overall sentence complexity** | 14. | C/S | 1.949 | 0.367 | 1.660 | 0.345 | 0.003 |

Overlooking language nativity variation, and collapsing the four groups into one Ling group and one NSc group, disciplinary effects become even more pronounced as shown in Table 8. The All Ling group displayed significantly higher length of sentences (MLS, # 2), T-units (MLT, # 3), amount of subordination (measures # 4-7), amount of coordination (CP/T, # 9), amount of phrasal sophistication (VP/T, # 13), as well as overall sentence complexity (C/S, # 14). A well-motivated significantly higher use of complex nominals (CN/C, # 11) was made by the All NSc group who seem to compress information through CNs.

**Table 8: Cross-disciplinary syntactic complexity comparison**

| | | | All Ling | | All NSc | | *p* |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | | MLC | 15.592 | 4.545 | 16.875 | 4.346 | 0.076 |
| | 2. | MLS | 28.951 | 8.986 | 25.546 | 5.235 | 0.011 |
| | 3. | MLT | 26.826 | 7.551 | 23.673 | 5.440 | 0.009 |
| **Amount of subordination** | 4. | C/T | 1.748 | 0.333 | 1.447 | 0.354 | 0.000 |
| | 5. | CT/T | 0.515 | 0.183 | 0.335 | 0.196 | 0.000 |
| | 6. | DC/C | 0.391 | 0.099 | 0.271 | 0.141 | 0.000 |
| | 7. | DC/T | 0.710 | 0.295 | 0.435 | 0.356 | 0.000 |
| **Amount of coordination** | 8. | CP/C | 0.503 | 0.218 | 0.515 | 0.283 | 0.404 |
| | 9. | CP/T | 0.859 | 0.376 | 0.721 | 0.383 | 0.036 |
| | 10. | T/S | 1.080 | 0.122 | 1.094 | 0.141 | 0.292 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.278 | 0.675 | 2.544 | 0.812 | 0.039 |
| | 12. | CN/T | 3.958 | 1.287 | 3.582 | 1.114 | 0.061 |
| | 13. | VP/T | 2.561 | 0.517 | 2.039 | 0.520 | 0.000 |
| **Overall sentence complexity** | 14. | C/S | 1.878 | 0.362 | 1.573 | 0.375 | 0.000 |

With regard to lexical diversity, each of the 100 abstracts was analyzed via TEXT INSPECTOR and received one VocD score and one MTLD score. Scores of the two measures of lexical diversity (VocD and MTLD) were compared for each group and were found to strongly correlate with one another in all groups. In Table 9, the values of the correlation coefficient *r* is higher than +0.70 and close to +1.

**Table 9: Correlation results of *VocD* and *MTLD***

| | ***r*** |
|---|---|
| **Br-Ling** | 0.825293 |
| **Br-NSc** | 0.873854 |
| **Eg-Ling** | 0.860985 |
| **Eg-NSc** | 0.94713 |

Both VocD and MTLD scores indicate no statistically significant cross-disciplinary effects as shown in Table 10 with >0.5 *p* values at 0.168758 and

0.206921 for the N and NN groups, respectively. In terms of mean values, though, there is a tendency for NSc abstracts to show higher lexical diversity than Ling abstracts. While these results contradict the predicted pattern in the first initial hypothesis, i.e. the expectation that Ling authors would display higher lexical diversity, yet, it might be argued that a more concrete discipline (NSc) would involve higher use of diverse lexical currency to refer to and describe entities and processes in tests and experiments.

**Table 10: Cross-disciplinary lexical diversity comparisons in N & NN groups**

| | **MTLD** | | | ***VocD*** | |
|---|---|---|---|---|---|
| | Means | *p* | | Means | *p* |
| **Br-Ling** | 90.5164 | 0.145507 | | 98.7068 | 0.168758 |
| **Br-NSc** | 97.7888 | | | 93.474 | |
| | | | | | |
| **Eg-Ling** | 72.894 | 0.235957 | | 75.5144 | 0.206921 |
| **Eg-NSc** | 77.6248 | | | 81.3544 | |

It can be concluded that the first initial hypothesis was verified in terms of syntactic complexity, yet rejected in terms of lexical diversity.

## 7.2 Second Hypothesis: Language Nativity Effects

Comparisons based on language nativity were made in the Ling (

Table 11) and then in the NSc (Table 12) abstracts. Interestingly, it was found that, in Ling, there are very slight (statistically insignificant) differences between the mean values of the N and NN groups with the exception of measures # 8 and 9. CP/C and CP/T indicating phrasal coordination in clauses and in T-units, respectively, have been more significantly used by the NN group, i.e. those with Arabic L1.

**Table 11: Language nativity comparison in terms of syntactic complexity in linguistics**

| | | | **Br-Ling** | | **Eg-Ling** | | ***p*** |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | 1. | MLC | 15.834 | 5.346 | 15.350 | 3.669 | 0.355 |
| | 2. | MLS | 30.364 | 9.600 | 27.537 | 8.278 | 0.135 |
| | 3. | MLT | 27.872 | 8.465 | 25.780 | 6.517 | 0.166 |
| **Amount of subordination** | 4. | C/T | 1.794 | 0.370 | 1.701 | 0.290 | 0.163 |
| | 5. | CT/T | 0.482 | 0.146 | 0.549 | 0.212 | 0.099 |
| | 6. | DC/C | 0.395 | 0.094 | 0.386 | 0.105 | 0.373 |
| | 7. | DC/T | 0.736 | 0.314 | 0.684 | 0.278 | 0.270 |
| **Amount of coordination** | 8. | CP/C | 0.425 | 0.182 | 0.581 | 0.226 | 0.005 |
| | 9. | CP/T | 0.739 | 0.306 | 0.979 | 0.406 | 0.011 |
| | 10. | T/S | 1.096 | 0.126 | 1.064 | 0.118 | 0.176 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.260 | 0.583 | 2.295 | 0.768 | 0.429 |
| | 12. | CN/T | 4.057 | 1.279 | 3.860 | 1.313 | 0.297 |
| | 13. | VP/T | 2.632 | 0.546 | 2.491 | 0.487 | 0.169 |
| **Overall sentence complexity** | 14. | C/S | 1.949 | 0.367 | 1.806 | 0.349 | 0.082 |

Three instances of co-ordination on the clausal as well as phrasal level can be observed in the example sentence in (5), where the single underlined coordinating conjunction *and* heading the Conj(unction) Phrase, $ConjP_1$, joins two reduced subordinate relative clauses, $CP_1$ and $CP_2$. The two incidents of the double underlined coordinating conjunction *and* signal NP coordination as shown in (5b). $NP_4$ is a reduced NP; the unreduced version of it would be: *cognitive sociolinguistics theories.*

(5) Eg-Ling, Abstract #19

a.

*In the light of the historical context of these speeches, the results of using a hybrid model for discourse analysis developed by the researcher and based on a review of literature on humor, political discourse, and postulates of cognitive and interactional sociolinguistics theories, reveal interesting conclusions on political humor in the public discourse of the three Egyptian presidents.*

b.

In the light of the historical context of these speeches, the results of using a hybrid model for discourse analysis [ConjP$_1$ [CP$_1$ developed by the researcher] and [CP$_2$ based on a review of literature on [ConjP$_2$ [NP$_1$ humor], [NP$_2$ political discourse], and [NP$_3$ postulates of [ConjP$_3$ [NP$_4$ cognitive] and [NP$_5$ interactional sociolinguistics theories]]]]]], reveal interesting conclusions on political humor in the public discourse of the three Egyptian presidents.

Coordination on both the phrasal and clausal levels can also be observed in the Eg-NSc abstracts. In the example compound sentence in (**6**), the double underlined conjunction *and* in (**6**a) (head of ConjP$_1$ in (**6**b)) joins two TPs, i.e. two clauses. Two instances of phrasal NP co-ordination are shown in (**6**b) by the two Conjunction Phrases (ConjP$_2$ and ConjP$_3$) headed by the single underlined coordinating conjunction *and* shown in (**6**a).

(6) Eg-NSc, Abstract #10

a.

*Spark Plasma Sintering (SPS) has been used to produce high density pellets of TiNandZrN, and their microstructure and preliminary thermophysical properties have been characterized.*

b.

[ConjP$_1$ [TP$_1$ Spark Plasma Sintering (SPS) has been used to produce high density pellets of [ConjP$_2$ [NP$_1$TiN] and [NP$_2$ ZrN]]], and [TP$_2$[ConjP$_3$ [NP$_3$ their microstructure] and [NP$_4$ preliminary thermophysical properties]] have been characterized].

The pervasiveness of coordination in Arabic has been well documented (Othman, 2004; Alqinai, 2013; Dickins, 2017). Hence, Eg-Ling and Eg-NSc seem to be influenced by the writing style of their L1.

Language nativity effects are much clearer regarding the nuclear science groups (Table 12), where the N group shows significantly higher clausal, phrasal, and sentential sophistication. More specifically, Br-NSc scored significantly higher than Eg-NSc in almost all measures of clausal subordination, and in two measures that are related to clausal subordination, namely, VP/T phrasal sophistication and C/S overall sentence complexity.

**Table 12: Language nativity comparison in terms of syntactic complexity in Nuclear Science**

| | | | **Br-NSc** | | **Eg-NSc** | | ***p*** |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | 1. | MLC | 16.206 | 3.866 | 17.544 | 4.763 | 0.140 |
| | 2. | MLS | 26.138 | 5.701 | 24.953 | 4.768 | 0.214 |
| | 3. | MLT | 24.169 | 5.637 | 23.177 | 5.303 | 0.262 |
| **Amount of subordination** | 4. | C/T | 1.526 | 0.308 | 1.367 | 0.384 | 0.057 |
| | 5. | CT/T | 0.392 | 0.174 | 0.278 | 0.202 | 0.019 |
| | 6. | DC/C | 0.316 | 0.125 | 0.225 | 0.144 | 0.010 |
| | 7. | DC/T | 0.517 | 0.281 | 0.353 | 0.408 | 0.052 |
| **Amount of coordination** | 8. | CP/C | 0.469 | 0.257 | 0.561 | 0.304 | 0.127 |
| | 9. | CP/T | 0.698 | 0.367 | 0.744 | 0.405 | 0.340 |
| | 10. | T/S | 1.090 | 0.110 | 1.098 | 0.169 | 0.425 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.538 | 0.768 | 2.550 | 0.870 | 0.480 |
| | 12. | CN/T | 3.779 | 1.139 | 3.385 | 1.075 | 0.108 |
| | 13. | VP/T | 2.180 | 0.505 | 1.898 | 0.506 | 0.027 |
| **Overall sentence complexity** | 14. | C/S | 1.660 | 0.345 | 1.486 | 0.390 | 0.051 |

In the complex sentence in (7), there is a complex T-Unit with two clauses: one main, $TP_1$ in (7b), marked by single underlining in (7a) and one subordinate, $TP_2$ in (7b), marked by double underlining in (7a), with three VPs: two finite, $VP_1$ and $VP_2$, and one non-finite, $VP_3$. Notably, subordination has been reported to be favored in English, as a mark of

mature, interesting, and effective writing style (Oshima & Hogue, 1998).

(7) Br-NSc, Abstract #12

a.

*Strong Rydberg atom dipole-dipole interactions provide a mechanism for efficient single photon coupling to atomic ensembles [1], whilst entanglement is mediated via an off-resonant interaction with the superconducting microwave cavity to provide long distance (~mm scale) interaction lengths [2].*

b.

[$TP_1$ Strong Rydberg atom dipole-dipole interactions [T' [$VP_1$ provide a mechanism for efficient single photon coupling to atomic ensembles …]]], [CP whilst [$TP_2$ entanglement [T' is [$VP_2$ mediated via an off-resonant interaction with the superconducting microwave cavity [TP to [$VP_3$ provide long distance (~mm scale) interaction lengths …]]]]]]

When the four groups were collapsed into one N group and one NN group, masking disciplinary variation, an interesting pattern was revealed as shown in Table 13. The English native group (Ling+NSc) displayed a familiar significant resort to subordination, as well as the two measures related to it, namely, phrasal VP/T and overall sentence complexity (C/S). The NN group (Ling+NSc), on the other hand, showed significant use of coordination, which is characteristic of their L1, Arabic.

**Table 13: Language nativity comparison in terms of syntactic complexity**

| | | | All N | | All NN | | *p* |
|---|---|---|---|---|---|---|---|
| | | Measures | Means | *SD* | Means | *SD* | |
| **Length of production unit** | 1. | MLC | 16.020 | 4.621 | 16.447 | 4.351 | 0.318 |
| | 2. | MLS | 28.251 | 8.100 | 26.245 | 6.812 | 0.092 |
| | 3. | MLT | 26.021 | 7.359 | 24.479 | 6.025 | 0.127 |
| **Amount of subordination** | 4. | C/T | 1.660 | 0.363 | 1.534 | 0.377 | 0.046 |
| | 5. | CT/T | 0.437 | 0.165 | 0.413 | 0.247 | 0.291 |
| | 6. | DC/C | 0.356 | 0.116 | 0.306 | 0.149 | 0.031 |
| | 7. | DC/T | 0.626 | 0.315 | 0.518 | 0.384 | 0.064 |
| **Amount of coordination** | 8. | CP/C | 0.447 | 0.222 | 0.571 | 0.266 | 0.006 |
| | 9. | CP/T | 0.719 | 0.335 | 0.861 | 0.418 | 0.032 |
| | 10. | T/S | 1.093 | 0.117 | 1.081 | 0.145 | 0.320 |
| **Amount of phrasal sophistication** | 11. | CN/C | 2.399 | 0.689 | 2.422 | 0.822 | 0.440 |
| | 12. | CN/T | 3.918 | 1.207 | 3.623 | 1.211 | 0.113 |
| | 13. | VP/T | 2.406 | 0.568 | 2.194 | 0.576 | 0.034 |
| **Overall sentence complexity** | 14. | C/S | 1.805 | 0.382 | 1.646 | 0.400 | 0.023 |

With regard to lexical diversity, clear and statistically significant language nativity effects were observed in both disciplines, where the N groups outperformed the NN groups as shown in Table 14.

**Table 14: Language nativity comparison in terms of lexical diversity**

| | **MTLD** | | | ***VocD*** | |
|---|---|---|---|---|---|
| | Means | *p* | | Means | *p* |
| **Br-Ling** | 90.5164 | 0.002486 | | 98.7068 | 0.0001 |
| **Eg-Ling** | 72.894 | | | 75.5144 | |
| | | | | | |
| **Br-NSc** | 97.7888 | 0.004018 | | 93.474 | 0.040455 |
| **Eg-NSc** | 77.6248 | | | 81.3544 | |

Higher values of lexical diversity can be observed by the N groups in both disciplines in Table 14. The shaded values indicate statistical significance of

difference with $p$ <0.05. These results verify the second initial hypothesis in terms of lexical diversity.

## 7.3 Third Hypothesis: Relationship between Syntactic and Lexical Reflections

The results obtained from correlation tests between each of the syntactic complexity measures and each of the lexical diversity (LD) indicators (VocD and MTLD), lead to rejecting the third initial hypothesis. The results are presented in Table 15 for VocD:

**Table 15: Correlation between *VocD* and syntactic complexity measures**

| | | ***r*** | | | | |
|---|---|---|---|---|---|---|
| | | | Br-Ling | Br-NSc | Eg-Ling | Eg-NSc |
| **Length of production unit** | **1.** | MLC | -0.12 | 0.14 | -0.47 | 0.07 |
| | **2.** | MLS | 0.02 | -0.02 | -0.45 | -0.14 |
| | **3.** | MLT | -0.17 | 0.20 | -0.56 | -0.06 |
| **Amount of subordination** | **4.** | C/T | -0.12 | 0.08 | -0.21 | -0.18 |
| | **5.** | CT/T | -0.22 | 0.02 | -0.28 | -0.17 |
| | **6.** | DC/C | -0.28 | 0.01 | -0.23 | -0.22 |
| | **7.** | DC/T | -0.23 | 0.08 | -0.26 | -0.23 |
| **Amount of coordination** | **8.** | CP/C | -0.08 | 0.10 | -0.29 | 0.00 |
| | **9.** | CP/T | -0.12 | 0.14 | -0.36 | -0.08 |
| | **10.** | T/S | 0.36 | -0.43 | 0.07 | -0.20 |
| **Amount of phrasal sophistication** | **11.** | CN/C | -0.11 | 0.10 | -0.52 | 0.02 |
| | **12.** | CN/T | -0.14 | 0.16 | -0.58 | -0.12 |
| | **13.** | VP/T | -0.33 | 0.34 | -0.37 | -0.12 |
| **Overall sentence complexity** | **14.** | C/S | 0.11 | -0.14 | -0.14 | -0.26 |

Results for MTLD are presented in Table 16:

**Table 16: Correlation between *MTLD* and syntactic complexity measures**

| | | *r* | | | | |
|---|---|---|---|---|---|---|
| | | | Br-Ling | Br-NSc | Eg-Ling | Eg-NSc |
| **Length of production unit** | **1.** | MLC | -0.09 | 0.15 | -0.47 | 0.08 |
| | **2.** | MLS | -0.02 | 0.09 | -0.41 | -0.10 |
| | **3.** | MLT | -0.19 | 0.19 | -0.47 | -0.04 |
| **Amount of subordination** | **4.** | C/T | -0.18 | 0.07 | -0.08 | -0.21 |
| | **5.** | CT/T | -0.33 | 0.05 | -0.10 | -0.17 |
| | **6.** | DC/C | -0.29 | 0.07 | -0.04 | -0.23 |
| | **7.** | DC/T | -0.26 | 0.11 | -0.10 | -0.26 |
| **Amount of coordination** | **8.** | CP/C | 0.04 | 0.19 | -0.43 | 0.00 |
| | **9.** | CP/T | -0.06 | 0.20 | -0.43 | -0.06 |
| | **10.** | T/S | 0.34 | -0.18 | -0.01 | -0.17 |
| **Amount of phrasal sophistication** | **11.** | CN/C | -0.13 | 0.08 | -0.46 | 0.06 |
| | **12.** | CN/T | -0.21 | 0.15 | -0.47 | -0.09 |
| | **13.** | VP/T | -0.29 | 0.28 | -0.18 | -0.12 |
| **Overall sentence complexity** | **14.** | C/S | 0.02 | -0.03 | -0.07 | -0.27 |

A great deal of uniformity has been observed between the results in Table 15 and Table 16, which is in line with the strong correlation between VocD and MTLD values (see Table 9, section 7.1). The vast majority of the correlation co-efficient *r* values are close to zero, indicating a weak negative/downhill linear relationship between most of the SC measures and each of VocD and MTLD, respectively. In the N groups, however, moderate positive/uphill linear relationship (indicated via shading in Table 15 and Table 16 is shown to occur with Br-Ling in T/S (measure #10, coordination) and with Br-NSc in VP/T (measure #13, phrasal sophistication). An interesting pattern of moderate negative/downhill linear relationship occurs in the Eg-Ling group between LD measures (VocD and MTLD) and measures of length of production units and some of the measures of coordination and phrasal sophistication with *r* ranging from -0.36 to -0.58. This negative relationship could mean that when Eg-Ling wrote longer

syntactic units, they used less diversified lexis. In other words, they circulated a common set of lexis over longer production units, probably resulting in more coherent texts.

## 8. Discussion

Observed disciplinary effects, which materialized only in the syntactic domain with Ling(uistics) authors using elaborate structures, specifically subordinate constructions, more often than their N(uclear) Sc(ience) counterparts, come as an expected result. First, linguists are expected to be more language-focused than nuclear scientists. Secondly, based on the Language Expectancy Theory, syntactically simple language has been linked to thinking on concrete as opposed to abstract levels (Burgoon, 1995; Miller & Burgoon, 1979, as cited in Averbeck & Miller, 2014). NSc, as a discipline, would naturally be expected to involve more concrete thinking levels than would Ling. Yet, while NSc authors used simpler clauses with less subordination, they produced significantly more complex nominals (CN/C) on the phrasal level, possibly to compress information through pre- and post-modifiers. This might suggest some form of trade-off between certain measures of SC. In a different context, Lu & Ai (2015) cited several studies (Bardovi-Harlig, 1992; Ortega, 2003; Norris & Ortega, 2009) that found that as the learners' level of proficiency advanced, "the role of subordination is subdued as phrasal-level complexification becomes the most pervasive means of syntactic complexity" (Lu & Ai, 2015, p. 24).

Clear nativity effects, which were statistically verified across both syntactic and lexical measures, resulted in an interesting division between native (N) and non-native (NN) groups. The N (English L1) abstracts were characterized by higher use of subordination while NN (Arabic L1) abstracts by coordination. Interestingly, Othman (2004) reported that English "makes use of more subordination than coordination, while Arabic favors the use of coordination rather than subordination." (p. 12). Subordination is indeed favored in English since, according to Oshima & Hogue (1998), writing with "only short, simple sentences is boring and ineffective, as is writing that uses too many compound sentences. Writing with complex sentences and participial phrases, structures that use subordination, is generally considered more mature, interesting, and effective in style" (p. 163). Alqinai (2013), however, argues that Arabic preference for coordination is not to be

considered a drawback or a sign of immaturity in the Arabic discourse" (p. 8). It is rather a matter of preference. Hence, NN authors were more inclined to use the preferred structure-combining operation (coordination) of their L1 (Arabic).

With regard to the relationship between syntactic complexity and lexical diversity measures, the results reported in the literature were inconclusive. While some studies on ESL learners of unidentified age groups reported developmental increase in complexity in both domains (syntactic and lexical; Storch and Tapper, 2009; Bulté & Housen, 2014; Schenker, 2016; Bulon et al., 2017), a study on the development of academic writing abilities of ESL/EFL graduate students reported a differential pattern, where an increase in academic vocabulary (lexis) was not accompanied with a comparable development on the syntactic level (Xudong et al., 2010). In another study on the writing of graduate students (aged 23-42), the lexical measures drawn from Lexile® Framework correlated with three of the SC measures of the L2SCA. Two measures, MLC and MLS, showed moderate correlation at 0.52 and 0.47, respectively. The third, CN/T, showed rather strong correlation with the lexical measures at 0.92. In the present study, where the writing of more mature authors has been investigated, lexical measures (VocD and MTLD) showed weak negative/downhill linear relationship with most of the SC measures. This could suggest that a linear relationship between syntactic complexity and lexical diversity becomes weaker with more mature authors. Furthermore, while syntactically complex constructions (particularly complex sentences with subordination) characterize writing that is "generally considered more mature, interesting, and effective in style" (Oshima& Hogue, 1998, p. 163), coherence would be strengthened by the circulation of lexis, i.e. by less lexical diversity. It follows that research abstracts written in English would be expected to display greater use of syntactically complex structures (mainly subordination) for effective style and less varied lexis for coherence. This eschews the results of the present study, to a great extent.

## 9. Conclusion

This paper presented a thorough examination of both syntactic complexity and lexical diversity (SC & LD, respectively) in 100 English conference abstracts written by both native (British) and non-native (Egyptian) researchers in two disciplines: Linguistics (Ling) and Nuclear Science (NSc). Regarding SC, the study revealed clear disciplinary effects on certain dimensions of SC. Regardless of language nativity, Ling authors displayed significantly greater resort to subordination and related measures of clausal complexity than did their NSc counterparts. Regardless of discipline, a strong link could be established between the authors' choice of clause-combining operation and the preferred operation in their respective L1. English native authors resorted to subordination which was recognized in the literature to characterize mature, interesting, and effective writing style. Arabic native authors, on the other hand, made more significant use of coordination, which was acknowledged as a pervasive operation in Arabic. Our findings have useful implications for L2 writing pedagogy; Egyptian authors, especially in science, should be encouraged while writing in English to break away from the preferred choice(s) in their L1 (coordination), and target constructions (subordination) characteristic of L2.

With regard to LD, VocD and MTLD scores indicated no statistically significant cross-disciplinary effects. On the other hand, clear and statistically significant language nativity effects were observed in both disciplines, where the N groups outperformed the NN groups in both VocD and MTLD scores. Weak and generally negative/downhill correlation was observed between SC measures and each of VocD and MTLD, respectively. This could mean that when the authors under study wrote longer and more complex syntactic units, they used less diversified lexis. In other words, they seem to have circulated a common set of lexis over longer production units, possibly resulting in more coherent texts.

Before closing, it is essential to acknowledge two limitations of the current study. The first is the relatively small sample size which may have resulted from the strict inclusion criteria adopted in data collection as discussed in section 6. The second limitation is the absolute absence of background information on the L2 proficiency level(s)as well as the length of professional career of the NN (Egyptian) researchers. Retrieval and

investigation of such background information is beyond the scope of the present study.

Future research is hoped to avert at least one of the limitations of the present study by aiming for a bigger sample size. Furthermore, the writing of authors with a variety of disciplinary and L1 backgrounds could be investigated to either consolidate the present findings or provide new insights into this area of research.

## Notes

1) In their taxonomic model of complexity, Bulté and Housen (2012) distinguish between *cognitive complexity* and *absolute complexity.* The former is concerned with the mental ease/difficulty by which learners acquire a certain language feature. The latter, *absolute complexity,* encompasses *linguistic complexity,* which is the focus of this research, along with propositional complexity and discourse-interactional complexity.
2) A T-unit or the "minimal terminable unit" was introduced by Hunt (1965) as a unit of syntactic measurement that consists of one independent clause with all of its dependent (subordinate) clauses. It differs from a sentence in the fact that a sentence may contain a set of coordinate independent clauses.
3) Other tools include Linguistic Inquiry and Word Count (LIWC; http://www.liwc.net/tryonline.php; Pennebaker, Booth, & Francis, 2007) which calculates the frequencies of self-reference, social and cognitive words as well as words denoting positive and negative emotions; Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Crossley, & McCarthy, 2010; McNamara &Graesser, 2012) which, in its public versions, analyzes texts on over 200 measures of cohesion, language, and readability; and the Tool for the Automatic Analysis of LExical Sophistication (TAALES; Kyle & Crossley, 2015) which calculates text scores for 135 lexical indices related to word frequency, range, academic language, and psycholinguistic information.
4) Reduced relative clauses feature a deleted relative pronoun and possibly a deleted linking verb (Carnie, 2013, pp. 16-17). In the example in(2), the head N *sample*s is post-modified by the reduced relative clause *synthesized by chemical precipitation method.* An unreduced version would be: *samples <u>which are</u> synthesized by chemical precipitation method.*

## Works Cited

Alqinai, J. (2013). Mediating punctuation in English Arabic translation. *LinguisticaAtlantica, 32,* 2-20. file:///C:/Users/tr/Downloads/22521-34072-1-PB.pdf.

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. *English for Academic Purposes, 31,* 58-71.

Averbeck, J. M., & Miller, C. (2014). Expanding Language Expectancy Theory: The Suasory effects of lexical complexity and syntactic complexity on effective message design. *Communication Studies, 65*(1), 72–95.

Bardovi-Harlig, K. (1992). A second look at T-Unit analysis: Reconsidering the sentence. *TESOL Quarterly, 26*(2), 390-395.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complex1ity, elaboration, explicitness.*Journal of English for Academic Purposes, 9,* 2-20.

Biber, D., Gray, B., &Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45,* 5–35. doi:10.5054/tq.2011.244483

Bulon, A., Hendrikx, I., Meunier, F., and Van Goethem, K. (2017). Using global complexity measures to assess second language proficiency: Comparing CLIL and non-CLIL learners of English and Dutch in French-speaking Belgium. *Travaux du CBL, 11*(1), 1-25. https://sites.uclouvain.be/bkl-cbl/wp-content/uploads/2017/04/Bulon_et_al_2017.pdf.

Bulte´, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency – Investigating complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.

————. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26,* 42-65.

————. (2015). Evaluating short-term changes in L2 complexity development. *Círculo de LingüísticaAplicada a la Comunicación, 63,* 42–76.

Burgoon, M. (1995). Language expectancy theory: Elaboration, explication, and extension. In C. R. Berger & M. Burgoon (Eds.), *Communication*

*and social influence processes* (pp. 29–52). East Lansing, MI: Michigan State University Press.

Carnie, A. (2013). *Syntax: A generative introduction*. UK, USA: Wiley Blackwell Publishers Ltd.

Chen, M., & Zechner, K. (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous non-native speech. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, June 19-24, pp. 722–731. www.aclweb.org/anthology/P11-1073.

Crossley, S. A., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In S. Jarvis, & S. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106–126). Bristol: Multilingual Matters.

_____ (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66–79.

Dickins, J. (2017). The pervasiveness of coordination in Arabic, with reference to Arabic>English translation. *Languages in Contrast, 17*(2), 229-254.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge, England: Cambridge University Press.

Douglas, Y., & Miller, S. (2016). Syntactic and lexical complexity of reading correlates with complexity of writing in adults. *International Journal of Business Administration, 7*(4), 1-10. https://doi.org/10.5430/ijba.v7n4p1

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Foster, P., &Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*(4), 866–896.

Givo´n, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution.* Amsterdam: John Benjamins.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193–202. https://pdfs.semanticscholar.org/cb42/704113df4cbff874806cdf4a9f05d5f8065f.pdf.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics, 30*, 461–473.

Hunt, K. (1965). Grammatical structures written at three grade levels. *National Council of Teachers of English (NCTE)* Research report No. 3. Champaign, IL, USA: NCTE. ftp://128.91.234.106/papers/faculty/beatrice_santorini/hunt-1965.pdf.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning, 63*, 87–106.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund University Working Papers, 53*, 61-79. http://journals.lub.lu.se/index.php/LWPL/article/download/2273/1848.

Kalantari, R. and Gholami, J. (2017). Lexical complexity development from dynamic systems theory perspective: Lexical density, diversity, and sophistication. *International Journal of Instruction, 10*(4), 1-18. https://doi.org/10.12973/iji.2017.1041a.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal ofSecond Language Writing, 28*, 39–52.

Koizumi, R., &In'nami, Y. (2012). Effects of text length on lexical diversity measures. *System, 40*, 522–532. doi: http://dx.doi.org/10.1016/j.system.2012.10.017.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757-786.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15, 474–496.

_____ (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*(1), 36-62.

_____ (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing, 34*(4), 493-511.

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.

Martínez, A. C. L. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing, 35,* 1–11. https://www.sciencedirect.com/science/article/pii/S107529351730051X

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*(1), 57–86. https://alsl.gsu.edu/files/2014/03/The-linguistic-features-of-quality-writing.pdf.

McNamara, D. S., &Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 188–205). Hershey, PA: IGI Global. ftp://129.219.222.66/Publish/pdf/McNamara_Graesser_Coh-Metrix.pdf.

Miller, M. D., & Burgoon, M. (1979). The relationship between violations of expectations and the induction of resistance to persuasion. *Human Communication Research, 5*, 301–313.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555–578.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518.

_____ (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing, 29*, 82-94.

Oshima, A., & Hogue, A. (1998). Writing academic English. Third Edition. London: Longman.

Othman, W. (2004). Subordination and coordination in English-Arabic translation. *Al-Basaer, 8*(2), 12-33. http://www.translationdirectory.com/article899.htm.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC2007: Linguistic inquiry and word count*. Retrieved from http://www.gruberpeplab.com/teaching/psych231_fall2013/documents/231_Pennebaker2007.pdf.

Price, B. D. (1974). Noun overuse phenomenon article. *The Language Quarterly, 2*(4), 29-37.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language, 30*, 3–26.

Richards, B. J. & Malvern, D. (1997). *Quantifying lexical diversity in the study of language development.* Reading: Faculty of Education and Community Studies.

Salazar, D. J. L. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing. PhD dissertation. University of Barcelona.

https://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf

Schenker, T. (2016). Syntactic complexity in a cross-cultural E-mail exchange. *System, 63*, 40-50.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes, 8*, 207-223.

Wang, S., & Slater, T. (2016). Syntactic complexity of EFL Chinese students' writing. *English Language and Literature Studies, 6*(1), 81-86.

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

Xudong, D., Cheng, L. K., Varaprasad, C., &Leng, L.M. (2010). Academic writing development to ESL/EFL graduate students in NUS. *Reflections on English Language Teaching, 9*(2), 119–138.

Yang, W., Lu, X., &Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53-67.

Youn, S. J. (2014). Measuring syntactic complexity in L2 pragmatic production: Investigating relationships among pragmatics, grammar, and proficiency. *System, 42*, 270–287.