SYNTACTIC COMPLEXITY AT THE INTERMEDIATE LEVEL IN EFL WRITING BY EARLY BALANCED BILINGUALS

Melisa Grabovac
melisa.dervisevic89@gmail.com
Stockholm University, Sweden

Oleksandr Kapranov
oleksandr.kapranov@uib.no
University of Bergen, Norway

Received June, 15, 2016; Revised June, 26, 2016; Accepted June, 30, 2016

Abstract. The present article involves an empirical psycholinguistic study aimed at examining syntactic complexity in English as a Foreign Language (EFL) by early balanced Bosnian/Swedish bilingual EFL learners. 15 early balanced bilingual Bosnian/Swedish EFL learners were recruited for the study and matched with their respective control groups of intermediate EFL learners (15 speakers of Bosnian as their first language (L1) and 15 speakers of Swedish as their L1). The experimental task involved an unprepared writing assignment in English about the most significant invention of the 20th century. The corpus of the participants’ written assignments was analysed in L2 Syntactic Complexity Analyzer and SPSS software programs respectively. Data analysis involved measures of syntactical complexity. It has been found that the participants’ written assignments are characterised by statistically significant number of T-units scores in comparison with the Swedish L1 monolingual controls. These findings are further presented and discussed in the article.

Keywords: early balanced bilinguals, EFL writing skills, psycholinguistics, syntactic complexity, T-Unit.

1. Introduction

The present article involves an empirical psycholinguistic study which seeks to elucidate syntactic complexity in English as a Foreign Language (EFL) identified in an EFL written task produced by early balanced Bosnian/Swedish bilinguals. Following Ortega (2003), syntactic complexity in this study is understood as

variation and sophistication of forms that appear in written EFL language production. Specifically, syntactic complexity in the present research is investigated by means of identification of dependent and independent clauses in EFL writing. Presumably, syntactic complexity in EFL writing is associated with a host of variables, one of which is the level of bilingualism (Inoue 2016). Hence, this research is based upon a contention that the level of syntactic complexity in EFL writing produced by early balanced bilinguals would be quantitatively different from a control group of monolingual EFL learners.

Syntactic complexity is regarded as an important construct in EFL/ESL research involving written skills (Ahmadian & Tavakoli 2011; Alotaibi 2016; Byrnes 2009; Gaies 1979; Grodner et al. 2002; Kobayashi & Rinnert 1992; Macnamara & Conway 2014; Sotillo 2000). It is posited that syntactic complexity is associated with the growth of syntactic repertoire in the learner’s development in the target language (Ortega 2003). The growth of the syntactic repertoire is thought to involve a variety of features (Gaies 1980), such as the learner’s usage of multiclausal sentences exhibiting subordination and/or coordination, an increase in the number of clauses per T-unit, a varied and sophisticated usage of verb phrases, etc. (Hunt 1965; Ho-Peng 1983). Traditionally, the T-unit has been regarded as an index of syntactic complexity (Daiute 1981; Hunt 1965; Wolff 2000; Youn 2014).

The T-unit, or minimal terminable unit of language, is intended to measure the smallest word group that is considered a grammatical sentence, regardless of how it was punctuated (Hunt 1965). The T-unit is defined as “a single main clause (or independent clause) plus whatever other subordinate clauses or non-clauses are attached to, or embedded within, that one main clause” (Hunt 1965:93). T-units are low-level inference categories which are easily identifiable, thus providing an objective means of assessing sentence-level complexity in written texts (Sotillo 2000).

Previous research indicates that the standard measures of syntactic complexity involve number of words per T-unit, number of words per clause, clauses per T-unit, error-free T-units, the number of S(entence)-nodes per T-unit, error-free clauses, average lengths per syntactic unit, total number of subordinate clauses, total number of embedded subordinate clauses, clauses in terms of occurrences, the proportion of clauses per sentence, ratio of nodes per sentence, ratio of combined clauses per sentence (Bardovi-Harlig & Bofman 1989; Ben-Zev 1977; Byrnes 2009; Cummins 2000; Inoue 2016; Kobayashi & Rinnert 1992; Navés et al. 2003; Sotillo 2000). It should be noted that S-nodes do not take into account nontensed clauses and T-units do not consider the existence of multiple embeddings in an S-node (Bachmann et al. 1988:145; Dekydtspotter & Renaud 2014; Wolfe-Quintero et al. 1998). Despite these shortcomings, the measures of syntactic complexity are thought to be practical means of normative comparison across samples, populations and contexts (Ortega 2003:493), including child language acquisition, language impairment, language and aging, second language acquisition and EFL respectively (Yau & Belanger 1984). In EFL studies involving writing, the above-mentioned measures of syntactic complexity can be employed in conjunction with measures of syntactic variety, e.g.
the total number of different grammatical verb forms, such as tense and modality respectively (Ahmadian & Tavakoli 2011).

It is suggested that syntactic complexity in EFL writing appears to interact with the use of cognitive strategies the learner employs (Kobayashi & Rinnert 1992). Psycholinguistic research indicates that cognitive strategies in several domain-specific areas are affected by bilingualism (Bialystok et al. 2003; Bialystok 1988; Molnar 2011). Current literature is suggestive of a contention that bilinguals may have an advantage over monolinguals in language acquisition and in language production respectively (Bialystok et al. 2003; Navés et al. 2003). In particular, research indicates that early balanced bilinguals enjoy positive outcomes in the additional language acquisition facilitated by the bilinguals’ linguistic awareness, i.e. the bilinguals’ linguistic knowledge (Bialystok 1988). Linguistic knowledge corresponds to the memory-based knowledge about a language, its syntactical and lexical composition. It is suggested that the bilingual’s linguistic awareness may lead to certain advantages attributed to the previously learnt languages (Bialystok et al. 2003). Presumably, these advantages involve the bilinguals’ awareness of formal linguistic features of language in general, and syntactic features in particular, the ability to reflect upon these features and their usage, the ability to intentionally monitor, plan and self-assess the effectiveness of linguistic processing (Bialystok et al. 2003; Norbert 2012).

An early balanced bilingual’s language learning is evocative of the observation that those learners who experience early exposure to an additional language ultimately surpass those learners who started their exposure as adults, in both pronunciation and syntax (Navés et al. 2003). However, the positive effects of bilingualism are inconclusive, since they have not been confirmed across various groups of bilinguals, in particular additive and subtractive bilinguals (Cenoz 1998; Cenoz & Valencia 1992). Whilst additive bilingualism implies no loss of an L1, i.e. the L1 is robustly used in socio-communicative contexts, subtractive bilinguals suffer from the loss of L1 skills due to the pressure to replace the socially non-dominant L1, (Lambert 1974). Slavoff and Johnson (1995) demonstrate that subtractive bilinguals exhibit no differences concerning initial morphosyntactic intuition. Given the inconclusiveness of empirical evidence, it can be assumed that syntactic complexity in EFL tasks executed by bilinguals will be in contrast with the EFL task execution by monolinguals. It remains to be investigated whether or not early balanced bilinguals would exhibit advantages in syntactic complexity over monolingual learners. To address this assumption, an empirical psycholinguistic experiment was conducted with early balanced Bosnian/Swedish bilingual EFL learners and contrasted with two groups of monolingual EFL learners. The experiment and its findings are further described in this article.

2. Hypothesis

Based upon previous research (Weissberg 2000), it was assumed that syntactic complexity would be robustly reflected in EFL writing tasks. Following previous research findings (Bialystok et al. 2003; Grosjean 2008; Herdina & Jessner 2000), it
was hypothesised that syntactic complexity in EFL writing tasks executed by early balanced bilinguals would be in contrast with syntactic complexity identified in the EFL writing tasks by their respective monolingual controls. Presumably, the contrast would be evident from measures of syntactic complexity involving T-Units and other associated variables. Arguably, the contrast in syntactic complexity would be associated with a contrast in lexical data identified in EFL writing tasks. Following the hypothesis, the main research questions were formulated as followed: i) Would early balanced Bosnian/Swedish bilinguals exhibit superior T-Unit values in their unprepared writing in the English language compared with their respective control groups? ii) Would early balanced Bosnian/Swedish bilinguals exhibit a contrast in their self-assessment of EFL skills in conjunction with the syntactic complexity data compared with their respective control groups?

3. Participants

15 early balanced Bosnian/Swedish bilingual participants were recruited for the study in Stockholm. The participants’ mean age was 16 y.o. at the time of the experiment. The participants identified themselves as early balanced Bosnian/Swedish bilinguals, being equally proficient in both Bosnian1 and Swedish respectively. The participants were secondary school students with the Swedish language as their medium of instruction. The participants had two classes of English per week. The participants were age-matched with the respective control groups. The first control group consisted of 15 Swedish L1 speakers who were intermediate-advanced EFL learners. Those controls resided in Stockholm, Sweden and were secondary school students whose language of instruction was Swedish. Those controls had two classes of English per week. The second control group involved 15 Bosnian L1 speakers who were intermediate-advanced EFL learners from Mostar, Bosnia-Herzegovina and were secondary school students with Bosnian as their language of instruction. Those controls were reported to have two classes of English per week. All the participants and their respective controls reported English to be their foreign language.

4. Procedure

The participants and their respective controls were accessed at their respective EFL classes by research assistants, one in Stockholm (Sweden) and another in Mostar (Bosnia-Herzegovina). First, the participants and their respective controls were asked to fill a questionnaire pertaining to their socio-linguistic background (e.g., age, gender, L1, L2, approximate amount of exposure to EFL, etc.) and rate their EFL proficiency on the Likert’s scale. Second, the participants and their respective control groups were instructed to write an unprepared essay in English titled “The most significant invention of the 21st century is…”.

5. Data analysis

The participants’ and their respective controls’ written data were analysed in software programs L2 Syntactical Complexity Analyzer (Xiaofei 2010) and in SPSS
respectively. The automatic segmentation into T-units, clauses, etc. by *L2 Syntactical Complexity Analyzer* (Xiaofei 2010) was manually checked by two linguists following the guidelines developed in Ortega et al. (2003).

### 6. Results and discussion

The application of the algorithm *L2 Syntactical Complexity Analyzer* (Xiaofei 2010) to the corpus of essays yielded the descriptive statistics summarised in Table 1 below:

**Table 1**

<table>
<thead>
<tr>
<th>Measure</th>
<th>S/B/E</th>
<th>B/E</th>
<th>S/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>M 143 (STD 73)</td>
<td>M 106 (STD 41)</td>
<td>M 65 (STD 40)</td>
</tr>
<tr>
<td>Sentences</td>
<td>M 9 (STD 4,4)</td>
<td>M 8 (STD 3)</td>
<td>M 4 (STD 2)</td>
</tr>
<tr>
<td>VP</td>
<td>M 20 (STD 10)</td>
<td>M 16 (STD 6,6)</td>
<td>M 9 (STD 5)</td>
</tr>
<tr>
<td>Complex Nominal</td>
<td>M 16 (STD 8)</td>
<td>M 13 (STD 7)</td>
<td>M 7 (STD 4)</td>
</tr>
<tr>
<td>Complex Nominal/T-Unit</td>
<td>M 2 (STD 0,8)</td>
<td>M 1,3 (STD 0,5)</td>
<td>M 1,7 (STD 0,6)</td>
</tr>
<tr>
<td>Coordinate Phrase/Clause</td>
<td>M 0,9 (STD 0)</td>
<td>M 0,2 (STD 0,1)</td>
<td>M 0,1 (STD 0,2)</td>
</tr>
<tr>
<td>Clauses</td>
<td>M 16,7 (STD 9)</td>
<td>M 13 (STD 5,3)</td>
<td>M 7 (STD 4,5)</td>
</tr>
<tr>
<td>T-Units</td>
<td>M 9,6 (STD 5)</td>
<td>M 9,6 (STD 4)</td>
<td>M 4,6 (STD 2,4)</td>
</tr>
<tr>
<td>Dependent clauses</td>
<td>M 7 (STD 4)</td>
<td>M 3 (STD 2,3)</td>
<td>M 3 (STD 2)</td>
</tr>
<tr>
<td>Complex T-Units</td>
<td>M 4,6 (STD 3)</td>
<td>M 2 (STD 1,3)</td>
<td>M 2 (STD 1,7)</td>
</tr>
<tr>
<td>Coordinate phrases</td>
<td>M 3 (STD 1,4)</td>
<td>M 2 (STD 1,4)</td>
<td>M 1 (STD 1,3)</td>
</tr>
<tr>
<td>MLS</td>
<td>M 16,7 (STD 4)</td>
<td>M 13,9 (STD 3,5)</td>
<td>M 15 (STD 5)</td>
</tr>
<tr>
<td>ML T-Units</td>
<td>M 15,7 (STD 4)</td>
<td>M 11,5 (STD 2,7)</td>
<td>M 14 (STD 3,5)</td>
</tr>
<tr>
<td>ML Clause</td>
<td>M 8,8 (1,6 STD)</td>
<td>M 8 (STD 1,5)</td>
<td>M 9 (STD 2,8)</td>
</tr>
<tr>
<td>Clause per sentence</td>
<td>M 1,8 (0,2 STD)</td>
<td>M 1,6 (STD 0,4)</td>
<td>M 2 (STD 0,4)</td>
</tr>
<tr>
<td>VP/T-Unit</td>
<td>M 4 (STD 0,4)</td>
<td>M 1,7 (STD 0,4)</td>
<td>M 2 (STD 0,5)</td>
</tr>
<tr>
<td>Clause/T-Unit</td>
<td>M 1,8 (0,2 STD)</td>
<td>M 2 (STD 0,3)</td>
<td>M 1 (STD 0,5)</td>
</tr>
<tr>
<td>Dependent Clause/Clause</td>
<td>M 0,4 (STD 0,7)</td>
<td>M 0,2 (STD 0,1)</td>
<td>M 0,3 (STD 0,1)</td>
</tr>
<tr>
<td>Dependent Clause/T-Unit</td>
<td>M 0,8 (0,2 STD)</td>
<td>M 0,3 (STD 0,2)</td>
<td>M 0,6 (STD 0,4)</td>
</tr>
<tr>
<td>T-Unit/Sentence</td>
<td>M 1 (0,1 STD)</td>
<td>M 1,2 (STD 0,2)</td>
<td>M 1 (STD 0,2)</td>
</tr>
</tbody>
</table>

Explanation of the abbreviations in Table 1: S/B/E = early balanced Bosnian/Swedish bilinguals who participated in the study, participants; B/E = Bosnian L1 monolingual control group in Bosnia; S/E = Swedish monolingual L1 control group in Sweden; VP = verb phrase; MLS = mean length of the sentence; ML = mean length; VP/T-Unit = verb phrase per T-Unit

The subsequent analysis of those statistics in SPSS yielded measures of statistical significance. Paired t-tests revealed statistical significance in contrasting i) T-units of the participants’ with their respective Swedish L1 controls: p = 0.0024, t = 3.34, df = 28, standard error of difference = 1.5. ii) MLT (Mean Length of T-units) of the participants’ with their respective Bosnian L1 controls: p = 0.0051, t = 3.0354, df = 28, standard error of difference = 1.340; iii) MLT (Mean Length of T-units) of the participants’ with their respective Swedish L1 controls: p = 0.0141, t = 2.6192, df = 28, standard error of difference = 0.178; iv) Clause
per T-Unit of the participants’ with their Swedish L1 controls: \( p = 0.0011, t = 3.6299, \ df = 28, \) standard error of difference = 0.147.

The participants’ and the respective control groups’ self-assessment of their EFL skills involving listening, speaking and writing were computed in SPSS and presented in Table 2:

### Table 2

**Self-assessment of the participants’ and the control groups’ EFL skills on the Likert scale**

<table>
<thead>
<tr>
<th>Participants/Controls</th>
<th>EFL listening</th>
<th>EFL speaking</th>
<th>EFL writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/B/E</td>
<td>M 32 (STD 2)</td>
<td>M 30 (STD 3)</td>
<td>M 30 (STD 2)</td>
</tr>
<tr>
<td>B/E</td>
<td>M 27 (STD 4)</td>
<td>M 26 (STD 4)</td>
<td>M 14 (STD 2)</td>
</tr>
<tr>
<td>S/E</td>
<td>M 28 (STD 5)</td>
<td>M 27 (STD 6)</td>
<td>M 10 (STD 3)</td>
</tr>
</tbody>
</table>

It has been assumed in the hypothesis that early balanced bilinguals would have an advantage in acquisition of additional languages and that they would produce texts of higher syntactic complexity. Specifically, syntactic complexity would involve higher T-unit values compared with the analogous values of the monolingual groups. Data analysis indicates that the participants outperformed the control groups on the following measures: M T-Units, M Length of T-Units, and the number of clauses per T-Unit. These findings can be taken to indicate that early balanced bilingualism offers certain advantages in syntactic complexity. The results of the present data analysis seem to support a contention that T-Unit is a robust means of measuring overall syntactic complexity of writing samples, and as such, T-Units are indicative of the language learners’ syntactic maturity (Shaw & Liu 1998).

Following previous research literature (Iwashita et al. 2008), it should be noted that in the present data set, syntactic complexity is particularly evident from the participants’ increased use of the number of clauses per T-unit, i.e. the T-unit complexity ratio, the ratio of dependent clauses to the total number of clauses, the number of verb phrases and the mean length of T-Units. However, these findings should be taken with caution, since, for instance, Shaw & Liu (1998) indicate that syntactic complexity is associated with the decrease of the total number of verb phrases and the increase in nominalisation. Nevertheless, the present results seem to support the contention that the average length of the T-unit correlates closely with the maturity of a writer, i.e. the more proficient students write fewer and longer T-units in their compositions than the less proficient students (Hunt 1970).

In the present corpus of EFL essays, the participants’ writing is characterised by an increased number of verb phrases (M 20, STD 10) compared with both the control groups, e.g. M 16 (STD 6.6) by the control group from Bosnia and M 9 (STD 5) by the control group comprised of Swedish L1 monolinguals. The participants’ EFL writing exhibits the presence of nominalization, which is evident from the mean number of complex nominals and the ratio of complex nominals per T-Unit.
Whilst the measures of complex nominals and the ratio of complex nominal per T-Unit are not statistically significant, it is observed in Figure 1 above that the participants exhibit a tendency to outperform the two control groups in terms of nominalisation. In concert with the studies by Wolfe-Quintero et al. (1998) and Iwashita et al. (2001), this measure is deemed to capture grammatical complexity. It has been assumed in the hypothesis that early balanced bilinguals would exhibit a contrast in their self-assessment of EFL skills in conjunction with the syntactic complexity data compared with their respective control groups. The results of the questionnaire involving the participants’ and the controls’ self-assessment of the EFL skills appear to support the hypothesis.

**Fig. 1. Nominalisation across the groups**

![Nominalisation across the groups](image)

**Fig. 2. The participants’ and the controls’ self-assessment of the EFL skills on the Likert scale**

![The participants’ and the controls’ self-assessment of the EFL skills on the Likert scale](image)
As evident from Figure 2 above, the participants self-assess their EFL writing skills higher in comparison with both the control groups. These findings are specifically noteworthy within the context of the T-units scores exhibited by the participants (M 9.6; STD 5) and the control group of Bosnian L1 monolingual EFL learners (M 9.6; STD 4). It is observed from the self-assessment data that the participants and the control group of Bosnian L1 monolingual EFL learners tend to evaluate their EFL writing skills higher in comparison with the control group comprised of Swedish L1 monolinguals. Whilst the Swedish L1 monolingual controls have produced less T-Units in their EFL writing task (M 4.6; STD 2.4), their self-assessment of the EFL writing skills appears to be lower than both the participants’ and Bosnian L1 monolinguals’ self-assessment. A possible explanation of the difference in self-assessment can be provided by the examination of multiple variables which are involved in the early bilinguals’ acquisition of their additional language (Kapranov 2013; 2015). However, it is beyond the scope of the present article to offer a plausible explanation of these variables and to shed light on whether or not there is a correlation between a higher self-assessment of the EFL writing skills and the actual higher number of T-Units in a piece of the EFL writing.

7. Conclusions

The participants in the present study have exhibited significantly higher results in the EFL writing task compared with the control groups. Specifically, paired t-tests have revealed statistical significance in contrasting i) T-units of the participants’ with their respective Swedish L1 controls; ii) Mean Length of T-units of the participants’ with their respective Bosnian L1 controls; iii) Mean Length of T-units of the participants’ with their respective Swedish L1 controls; iv) Clause per T-Unit of the participants’ with their Swedish L1 controls. These findings are suggestive of the bilinguals’ advantage in the study of a foreign language, in particular EFL. Additionally, it has been found that the early balanced participants in this study appear to self-assess their EFL writing skills higher on the Likert scale in comparison with the control groups. It is assumed that further empirical studies will elucidate whether or not there is a correlation between a higher self-assessment of the EFL writing skills and the actual higher number of T-Units in EFL written tasks executed by the early balanced bilingual EFL learners.

References


