Syntactic complexity in written English of incoming university students

Rebecca Present-Thomas
Bert Weltens
John H.A.L. de Jong
VU University Amsterdam

Abstract
Despite its increasing influence on institutional and even governmental policy, the Common European Framework of Reference (CEF) is currently lacking in comprehensive description of proficiency in writing, especially with regard to the higher levels. This paper reports on non-native English written data collected from 123 incoming university students at a Dutch university. Measures of syntactic complexity exhibited in the data will be compared against student proficiency self-assessments.

1. Introduction

Though widely adopted across a range of institutional and even governmental settings, many have expressed concerns that the CEF levels and descriptors are not suitable for direct application in particular contexts (Alderson, 2007; Fulcher, 2004; Hulstijn, 2007). Empirically validated descriptors for writing production, particularly at the upper levels, are especially lacking in the 2001 CEF publication. In fact, current descriptors have largely been reconstructed from scales describing different skills entirely (Council of Europe, 2001; 61), with the result that the descriptors do not seem to fit the skill they are supposedly referencing.

Despite its shortcomings, the CEF has gained traction across Europe and, indeed, around the world as the fashionable way to compare and contrast actual and target language proficiency across users, courses, examinations, institutions, etc. This paper reports on ongoing research concerned with constructing detailed and empirically sound descriptors of academic English writing. In other words, this project aims primarily at localizing the CEF with regard to academic writing, supplementing it with detailed descriptors of advanced written English proficiency. For the purposes of this project, academic writing is concerned with the texts and styles associated with academic settings, including typical university and post-graduate products such as essays, term papers, research reports and scientific articles.

Since the intention of this research program is to extend the CEF descriptors through the analysis of student writing samples, it was important to determine at an early stage how the participants fit into the framework with regards to the particular skill in question. The long-term research design allows for collection and analysis of data from students representing various stages in a university bachelor program that aims to produce C2-level English proficient graduates. This initial study only includes data from incoming 1st-year students and is concerned with describing the English writing proficiencies with which the students enter university.

Previous research has shown that students entering university in the Netherlands most often represent CEF level B2, though a significant number may also represent levels B1 and C1, as measured by CEF-benchmarked standardized tests and student self-assessment (Noijons and Kuiper, 2006; Present-Thomas, forthcoming). Additionally, previous research has suggested that the lexical range of this group of students is narrower than exhibited in published academic texts (Present-Thomas, forthcoming). This paper builds on these findings with continued exploration into the following research question:

What is the linguistic reality underlying the writing samples provided by incoming university students?

The specific element of proficiency addressed in this paper concerns syntactic complexity, which is given surprisingly little attention in the current CEF descriptors. Highly proficient writers are expected to display mastery of a number of complex syntactic strategies, such as subordination and coordination. Additionally, the increase in available syntactic and rhetorical strategies a learner has available leads to greater variety in the syntactic constructions produced. Accordingly, it can be expected that measures of syntactic complexity, which have been defined in many different ways by previous L2 researchers, will increase both in mean and in variance as proficiency.

Ortega (2003) conducted a research synthesis of 21 cross-sectional studies of syntactic complexity in L2 writing proficiency. The studies were classified and compared according to instructional setting (ESL vs. FL) and the proficiency criterion employed (program level vs. holistic rating). She found that learners in ESL settings tend to employ a higher degree of syntactic complexity in their writing than those in FL settings, most likely due to differences in entry level and pace of acquisition in the two distinct settings, such as target language access and use outside of the classroom. Additionally, studies that had employed a holistic rating of ability, rather than grouping learners by program level, resulted in measurements with less variance (i.e. a smaller standard deviation and narrower range of exhibited syntactic complexity measurements). This finding is presumably because a range of proficiencies is likely to be represented by a single program; the within-group variance would therefore be higher for studies in which learners are classified by program. Finally, Ortega (2003) proposed the following "critical magnitudes" for between group syntactic complexity comparisons:

a) Mean words per sentence (MLS): 4.5 or more words difference
b) Mean length of T-unit (MLT): 2 or more words difference
c) Mean length of clause (MLC): more than 1 word per clause difference
d) T-unit complexity ratio (C/T): +/- 0.20 or more clauses per T-unit difference

This study contributes to the field of L2 syntactic complexity research by adding a cross-sectional comparison between holistically defined CEF levels represented by students of a single program level (university entrance) in an FL setting. Although cross-sectional research is well represented in Ortega (2003), FL settings and holistically defined proficiency groups have thus far received less attention than their counterparts.

2. Methodology

2.1 Participants

Over 2 years, incoming bachelor students of English Linguistics at VU University Amsterdam were required to take an academic writing test shortly into their first year of study. Native speakers of English were removed from the dataset and the remaining participants represented the following native languages: Dutch (n=114; monolingual or bilingual with another non-English language counted here), Arabic (2), Berber (1), Lebanese (1), Papiamento (1), Persian (1), Portuguese (1), Slovak (1), and Turkish (1). The complete dataset included a total of 430 text samples collected from 123 students.

2.2 Test content and delivery

The academic writing test consisted of standardized items taken from an international test of academic English, the Pearson Test of English Academic (PTE). Two of the items required students to provide short summaries, one based on a 200-word written text (response length: 1 sentence) and one based on a 90-second excerpt of an audio-recorded lecture (response length: 50-70 words). Additionally, each student was presented with one or two essay prompts that elicited argumentative-style responses between 200 and 300 words in length.

The test was followed by a short demographics and self-assessment survey (DSA). In response to this survey, students provided self-assessment measures based on the CEF Common reference level descriptors for overall English language proficiency and for English writing proficiency (Council of Europe, 2001: 24, 26-27). Students who believed themselves to be at levels other than B1, B2, and C1 were removed from the analysis due to low numbers.

The test was administered in proctored sessions at the university using the online Qualtrics Survey Software (Qualtrics Labs Inc., Provo, UT, 2010, version 18xxx) for test delivery. This platform was selected for its relative ease of customization, which enabled replication of PTE testing conditions with limited resources.

2.3 Syntactic Complexity

Next, responses were analyzed for syntactic complexity using the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010). This program automatically analyzes text files, first calculating the frequencies of 9 distinct units of linguistic analysis: words, sentences, verb phrases, clauses, t-units, dependent clauses, complex t-units, coordinate phrases, and complex nominals. Based on these linguistic units, 14 measures are produced, all of which have been suggested by previous research as relevant in L2 writing proficiency. As Lu (2010) points out, each of the measures focuses on one of the following 5 categories: length of production unit, sentence complexity, subordination, coordination, and particular structures. An overview of the 14 syntactic complexity measures is provided in Table 1.

Category	Measure	Code	Definition
Length of production unit	Mean length of clause	MLC	# words / # clauses
	Mean length of sentence	MLS	# words / # sentences
	Mean length of T-unit	MLT	# words / # T-units
Sentence complexity	Sentence complexity ratio	C/S	# clauses / # sentences
Subordination	T-unit complexity ratio	C/T	# clauses / # T-units
	Complex T-unit ratio	CT/T	# complex T-units / # T-units
	Dependent clause ratio	DC/C	# dependent clauses / # clauses
	Dependent clauses per T-unit	DC/T	# dependent clauses / # T-units
Coordination	Coordinate phrases per clause	CP/C	# coordinate phrases / # clauses
	Coordinate phrases per T-unit	CP/T	# coordinate phrases / # T-units
	Sentence coordination ratio	T/S	# T-units / # sentences
Particular structures	Complex nominals per clause	CN/C	# complex nominals / # clauses
	Complex nominals per T-unit	CN/T	# complex nominals / # T-units
	Verb phrases per T-unit	VP/T	# verb phrases / # T-units

Table 1. Overview of syntactic complexity measures provided by L2SCA (Lu, 2010).

Descriptive statistics for the entire set of responses (n=430) were calculated in order to gain insight into the central tendency and dispersion among the entire set of written responses produced by the incoming students to the various test tasks. Responses were then grouped by

CEF level (based on self-assessment data) and task type in order to highlight differences between CEF levels and task types.

3. Results

3.1 Student Proficiency Levels

Following the CEF tradition of self-assessment using "can-do" statements, students were asked to select which one of the six writing descriptors from the self-assessment grid best described their abilities in English (Council of Europe, 2001, p. 24). Half of the students (51%) estimated themselves to be at CEF level B2, while the majority of the remaining students were split equally (24%) between levels B1 and C1 (Present-Thomas, forthcoming). Self-assessment data was used to divide students into proficiency-based groups for syntactic complexity analysis, as shown in Table 2.

CEF level	# students	# texts
B1	30	103
B2	63	222
C1	30	105

Table 2. Proficiency-based (self-assessment) division of students and texts.

3.2 Syntactic Complexity

3.2.1 Syntactic Complexity Across All Incoming Students

All 430 writing samples provided by the students during the academic writing test were run through the L2SCA program. Descriptive statistics for each of the 14 syntactic complexity measures calculated by the program are provided in Table 3.

Measure	Mean	SD
MLC	10.00	4.40
MLS	21.61	11.33
MLT	18.92	9.89
C/S	2.23	1.01
C/T	1.94	0.84
CT/T	0.57	0.30
DC/C	0.40	0.18
DC/T	0.87	0.75
CP/C	0.25	0.29
CP/T	0.45	0.49
T/S	1.18	0.37
CN/C	1.13	0.81
CN/T	2.17	1.69
VP/T	2.76	1.26

Table 3. Descriptive statistics for the syntactic complexity measures across all responses (n=430).

While the means may provide insight into the central tendencies of incoming students' proficiencies with regards to syntactic complexity, the relatively large standard deviations indicate that the responses were widely dispersed in terms of the syntactic complexity measures. The most likely factors contributing to this dispersion are differences in proficiency, differences in task type, individual differences, or a combination of these factors.

3.2.2 Syntactic Complexity Across CEF Levels

Setting aside for a moment the issue of task type, which is addressed in 3.2.3, responses were grouped according to the students' self-assessed CEF level for written production. Table 4 shows the descriptive statistics for the measures calculated for each proficiency-based group of responses. Five of the 14 measures correlate positively with the self-assessment data: mean length of clause, mean length of T-unit, dependent clause ratio (DC/C), complex nominals per clause, and complex nominals per T-unit. Most of the remaining measures exhibit an increase between two levels (i.e., from B1 to B2 or from B2 to C1), but this increase in complexity is not held through all three levels. Interestingly, the measure of clauses per T-unit actually decreases slightly from B1 to B2 and again from B2 to C1. However, two of these measures, coordinate phrases per clause and coordinate phrases per T-unit, did not exhibit the steady increase from levels B1 to B2 to C1 that would indicate that the differences were due to proficiency. The third measure, complex nominals, did exhibit a steady increase, but the increase was only found to be significant between levels B1 and C1.

Measure	**B1 (n=103)**		**B2 (n=222)**		**C1 (n=105)**		**ANOVA**	
	Mean	SD	Mean	SD	Mean	SD	F	Sig.
MLC	9.31	3.47	10.03	4.42	10.64	5.08	2.41	.09
MLS	20.10	9.66	22.26	12.56	21.72	9.99	1.29	.28
MLT	18.15	9.68	18.97	9.94	19.55	10.03	.53	.59
C/S	2.20	0.95	2.29	1.14	2.14	0.75	.85	.43
C/T	1.96	0.87	1.94	0.87	1.90	0.74	.14	.87
CT/T	0.55	0.31	0.57	0.30	0.57	0.30	.18	.84
DC/C	0.38	0.20	0.40	0.17	0.41	0.19	.60	.55
DC/T	0.86	0.75	0.89	0.80	0.86	0.62	.08	.92
CP/C	0.19	0.18	0.29	0.32	0.25	0.30	4.03	.02
CP/T	0.34	0.32	0.51	0.52	0.45	0.53	4.25	.02
T/S	1.15	0.33	1.20	0.42	1.16	0.31	.65	.53
CN/C	**1.00**	**0.65**	**1.13**	**0.75**	**1.28**	**1.03**	**3.06**	**.05**
CN/T	1.97	1.50	2.18	1.73	2.33	1.76	1.18	.31
VP/T	2.78	1.32	2.76	1.29	2.76	1.14	.01	.99

Table 4. Comparison of means for syntactic complexity measures, grouped by CEF level (n=430).

Despite the trends noted above, most of these differences are not significant. Table 4 also shows the results of one-way ANOVAs of the means with CEF level as the factor. The Bonferroni correction was employed to counteract effects of simultaneous comparisons. Despite the observed increases in five measures as proficiency levels increased, only three of the 14 measures were found to exhibit significant differences between levels (α=0.05). However, two of these measures, coordinate phrases per clause and coordinate phrases per T-unit, did not exhibit the steady, monotonic increase from levels B1 to B2 to C1 that would indicate that the differences were due to proficiency. The third measure, complex nominals,

did exhibit a steady increase, but the increase was only found to be significant between levels B1 and C1.

3.2.3 Syntactic Complexity Across Task Types

It is also likely that any significant differences between CEF levels concerning syntactic complexity may be masked by differences in task. For example, one of the three response types included in the writing test was a 1-sentence summary of a 200-word written passage. It is not difficult to imagine that students instructed to limit their response to a single sentence would be more inclined to pack as much information as possible into this sentence than they might be for the average sentence included in their longer (50-70 words without restriction on the number of sentences) summary response or the essay response.

Table 5 shows the descriptive statistics and one-way ANOVAs for the same set of responses when task type, instead of CEF level, is the factor. The difference between this task-based analysis and the previous CEF-based analysis is clear: every one of the 14 measures shows significant differences in syntactic complexity between tasks. More specifically, the reading summary task, which requires a single-sentence response is significantly more complex (α=.01) than both of the other response types across all 14 measures with only one exception: the difference between the reading summary responses and the essay responses with regards to the dependent clause ratio is not significant. Additionally, the 200-300 word essay responses also seem to be more complex than the 50-70 word listening summary responses, with significant (α=.01) results for all but 4 of the syntactic complexity measures: the 3 coordination measures and complex nominals per clause.

Measure	**RS (n=123)**		**LS (n=123)**		**Essay (n=184)**		**ANOVA**	
	Mean	SD	Mean	SD	Mean	SD	F	Sig.
MLC	14.01	6.14	7.33	1.47	9.11	1.58	120.77	.00
MLS	34.08	12.42	12.79	5.16	19.16	4.19	251.75	.00
MLT	29.19	11.43	11.58	5.20	16.95	3.62	200.31	.00
C/S	2.83	1.45	1.78	0.75	2.14	0.47	40.91	.00
C/T	2.35	1.20	1.60	0.72	1.88	0.38	28.80	.00
CT/T	0.74	0.41	0.40	0.23	0.56	0.15	50.44	.00
DC/C	0.43	0.26	0.31	0.15	0.44	0.10	24.00	.00
DC/T	1.20	1.06	0.57	0.69	0.86	0.33	24.40	.00
CP/C	0.38	0.48	0.18	0.16	0.22	0.12	18.66	.00
CP/T	0.73	0.77	0.27	0.24	0.39	0.21	34.98	.00
T/S	1.28	0.61	1.13	0.24	1.14	0.15	6.95	.00
CN/C	1.81	1.20	0.73	0.30	0.96	0.28	86.80	.00
CN/T	3.70	2.18	1.23	1.13	1.78	0.56	114.54	.00
VP/T	3.74	1.56	1.80	0.89	2.76	0.57	109.84	.00

Table 5. Comparison of means for syntactic complexity measures, grouped by task (n=430).

These results show the important role that task plays in the syntactic complexity of a writing sample, regardless of student proficiency. It can therefore be argued that task differences, if not properly accounted for, may lead to validity issues in any cross-sectional proficiency-based syntactic complexity analysis.

3.2.4 Syntactic Complexity Across CEF Levels Revisited

In order to provide a more accurate analysis of proficiency-based differences in syntactic complexity, responses were grouped by task type and ANOVAs were performed for each set of responses. The results of these analyses are reported in Table 6, which indicates for each task type which of the syntactic complexity measures showed a steady increase or decrease from CEF levels B1 to B2 to C1, and whether or not these differences were found to be significant. Significance results are only reported when a steady (monotonic) increase or decrease in means was displayed.

Measure	**RS (n=123)**		**LS (n=123)**		**ES (n=184)**	
	Trend	**Sig?**	**Trend**	**Sig?**	**Trend**	**Sig?**
MLC	Increase	No	Increase	No	Increase	Yes
MLS	None		Increase	No	None	
MLT	Increase	No	Increase	No	Increase	No
C/S	None		Increase	No	Decrease	No
C/T	Decrease	No	Increase	No	Decrease	No
CT/T	None		Increase	No	None	
DC/C	None		Increase	No	None	
DC/T	None		Increase	No	None	
CP/C	None		None		None	
CP/T	None		Increase	No	None	
T/S	None		Increase	No	Decrease	No
CN/C	Increase	No	Increase	No	Increase	Yes
CN/T	Increase	No	Increase	No	Increase	No
VP/T	Decrease	No	Increase	No	Increase	No

Table 6. Mean trends and ANOVA results for task-separated texts with CEF self-assessment as the factor.

Neither summary task exhibited significant differences for any of the 14 measures. For the essay responses, however, significant (α=.01) monotonic differences between CEF levels were found for two of the syntactic complexity measures. Post-hoc (Bonferroni) test results for these significant measures are provided in Table 7.

Dependent Variable	**(I) CEF**	**(J) CEF**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**
MLC	**B1**	**B2**	**-.90**	**.28**	**.01**
	B1	**C1**	**-1.14**	**.33**	**.00**
	B2	C1	-.24	.28	1.00
CN/C	B1	B2	-.12	.05	.05
	B1	**C1**	**-.18**	**.06**	**.01**
	B2	C1	-.06	.05	.69

Table 7. Post-hoc test results for significant differences found in essay responses (n=184).

These results indicate that significant differences in mean clause length exist between CEF levels B1 and B2. Additionally, mean clause length and complex nominals per clause exhibited significant differences between levels B1 and C1. However, no significant differences were found between levels B2 and C1 in any of the syntactic complexity measures. Particularly when taken together with the (non-significant) positive trends exhibited for these same two measures in the summary tasks, it seems that there is a positive correlation between writing proficiency and both mean length of clause and complex nominals per clause.

4. Discussion

Section 3.2.1 outlines the results of an ANOVA between student self-assessed CEF levels of 430 written responses to tasks on an academic writing test provided by incoming students at a Dutch university. Although this analysis ignores differences in the response tasks included in the dataset, results indicate that differences in syntactic complexity exist between CEF levels B1 and B2 with specific respect to coordinate phrases per clause and coordinate phrases per T-unit. Significant differences between levels B1 and C1 are also found for the measure of complex nominals per clause. Although the results presented in section 3.2.2 show that significant differences in just about every measure of syntactic complexity occur between different types of writing tasks, particularly when students are limited to a single-sentence response, section 3.2.3 confirms for essay responses that the differences identified in section 3.2.1 exist even when the confounding variable of task type is removed from the analysis. Section 3.2.3 further identifies mean clause length as an additional significant difference between CEF level B1 and both levels B2 and C1 (though not between level B2 and C1).

In more practical terms, these results seem to indicate that as students become more proficient in written English, they tend to produce longer clauses. Additionally, differences in syntactic coordination exist between CEF levels B1 and B2, with B2 writers coordinating more phrases per clause and T-unit than B1 writers. Finally, differences in complex nominal usage seem to exist between the B and C levels; at a clausal level, complex nominal usage increases as students reach CEF level C1.

When the results of this study are viewed in light of Ortega's (2003) critical magnitudes, it should be noted that although statistically significant differences were found in CN/C (when all task types were analyzed together, as well as for essays exclusively), and MLC (when the essay task was analyzed exclusively). The only comparison to display differences of the magnitude identified in Ortega (2003) was between task types (see section 3.2.3). This feature, independent of instructional setting, proficiency or proficiency criterion, was not addressed in the research synthesis other than to indicate that tasks varied considerably from one study to the next (Ortega, 2003: 499). However, the clear interaction found in this study between task type and syntactic complexity shows that this is a variable worthy of further exploration.

Finally, it must be noted that this study is not without limitations. For example, the CEF level assignment is based on self-assessment data provided by students who were untrained in the use of the CEF scales. There is, therefore, no guarantee that each of the students involved in the study used the same standard to choose their CEF level. Furthermore, even if there was a way to ensure that they had effectively all used the same standard the problem remains that the difference between a low B2 and a high B2 is greater than between a high B1 and a low B2.

Additionally, the dataset for this study is particularly small in size. As with all studies that utilize corpus-based methodology, small sample sizes may lead to the effect of individual author differences masking or even falsely identifying more general differences between the groups in question, proficiency-based response groups in this case. Therefore, continued research with a larger corpus containing text samples from students at different staged in a bachelor program, as well as native and known C2-level texts, is in progress. The long-term data collection plan for this project allows for not only cross-sectional comparison across CEF levels and different stages in an English-language bachelor program, but also longitudinal comparisons as these incoming students progress through the program.

References

Alderson, J. C. (2007). The CEFR and the Need for More Research, *91*(4), 659-663. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1540-4781.2007.00627_4.x/full.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. (Council of Europe, Ed.). Cambridge: Cambridge University Press.

Fulcher, G. (2004). Deluded by Artifices? The Common European Framework and Harmonization. *Language Assessment Quarterly*, *1*(4), 253-266.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *Modern Language Journal*, *91*(4), 663-667. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1540-4781.2007.00627_5.x/full.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496.

Noijons H., J. & K. (2006). Mapping the Dutch foreign language state examinations onto the common European framework of reference. Arnhem: Cito.

Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency : A Research Synthesis of College-level L2 Writing. *Applied Linguistics*, *24*(4), 492-518.

Pearson Education Ltd. (2011). Pearson Test of English Academic : Automated Scoring. Retrieved September 28, 2011, from http://www.pearsonpte.com/research/Pages/AutomatedScoring.aspx

Qualtrics Labs Inc. (2010). Qualtrics Survey Software. Provo, UT, version 18xxx