Katalin Doró

University of Szeged, Department of English Language Teacher Education and Applied Linguistics

The lexical richness of undergraduate student essays: the influence of writing experience and essay writing strategies

L2 writing is a complex cognitive task, which requires and is facilitated by practice, familiarity with the task and appropriate writing strategies. This study investigates three lexical richness measures of argumentative essays written by Hungarian undergraduate students of English (n=89) and the correlation they have with students' self-reported writing behaviors and strategies. Surprisingly, no statistically significant differences were found between the reported writing behaviors of students and the lexical richness measures of their essays. This can be attributed partly to the limited strategy data that are gained through a questionnaire, but can also be due to the complex nature of writing as a process and to the variety of factors that contribute to writing as a product. The pedagogical implications of writing strategy instruction are also discussed.

Keywords: lexical richness, essay writing, strategies, EFL undergraduates

1   Introduction

Essay writing, especially in an academic context, requires great cognitive demand and much practice. Writing messages longer than a few short sentences needs planning, drafting, decision making concerning grammatical and lexical complexity, coherence and the message content. Second and foreign language (L2) students often have problems with vocabulary, grammar, spelling, organization (or a combination of these), and they need to constantly improve their text production by reading and writing a lot in the target language. In this study, some factors are investigated which are expected to have a direct influence on the lexical parameters of essays written by non-native undergraduate students of English. The first aspect to be explored is students' opportunities for producing longer texts in English, as it is hoped that practice opportunities help students in writing better texts. The second one is their text writing strategies, such as the avoidance of vocabulary items not fully known and the language (native vs. target) used for essay planning and drafting. It is also investigated how their view on the importance of lexical choice (compared to grammar, organization and spelling) is reflected in the lexical profile of their essays. In other words, it is explored whether those who claim to pay the most attention to lexical choices during writing do use a wider range of vocabulary or not. Answers to these questions may offer researchers and instructors valuable information to be incorporated in essay writing instruction and exam task elaboration.

2   Background

This section provides a general introduction to, and an overview of, the measures of lexical richness of written texts and writing strategies.

2.1   Measuring the lexical richness of essays

Lexical richness refers to the variety of words used in a text and usually measures how many different words the given text employs (Laufer 2005). The oldest and still most frequently used method is to count the different words in a text (types) and compare this number with the total number of words that appear in the text (tokens). Silva and Matsuda (2001: 98–100) list the following five ways to measure the lexical richness of texts:

    a. lexical variation – type/token ratio
    b. lexical density – content words/function words

    c. lexical sophistication – use of common words and rare words
    d. lexical errors – number of lexical errors or lexical errors/total errors
    e. lexical individuality – ratio of words unique to the writer

Granger and Wynne (2000) call attention to the fact that the use of the crude type/token ratio in large learner corpora presents researchers with considerable difficulty. First, due to the frequent typing or spelling errors that are in these texts, all the spelling variants of a word are counted as different types, unless corpus designers undertake the complex task of manually correcting the errors. The authors conclude that "it is not safe to use crude type/token or lemma/token ratios with learner corpora" (Granger & Wynne 2000: 7). The problems concerning spelling variants or errors, however, are ruled out in the present study, as it does not use a large learner corpus, but shorter single student essays which were corrected for spelling problems (see section 3.3). The criticism of the simple type/token ratio formula has led to the development of slightly different formulae, such as the S, R, C, U, D, P_Lex and vocd-D measures (for an overview see Kojima & Yamashita 2014 and Šišková 2012). McCarthy and Jarvis (2007, 2010) discuss the problem of text length and its effect on the measures of lexical richness and conclude that none of the above formulae could completely eliminate problems coming from text length. The authors point out that either a time limit should be set when collecting data (although this still does not result in texts of equal length), or texts have to be cut, in which case, considerable amount of valuable data may be lost. In our data collection giving participants a time and length limit minimized these problems. This meant that only very little text proportions had to be cut to be able to work with texts of the same length.

Laufer and Nation (1995) and Laufer (2005), while discussing some of the shortcomings of the measures, such as lexical variation and lexical density, offer an alternative method to capture the lexical sophistication in texts, namely the **lexical frequency profile** (LFP). The LFP measures the ratio of the first 2,000 word families in texts. If the calculation gives a high percentage, it means that the text used a high proportion of frequent English words.

It is important to note that the above-discussed formulae provide only surface-level information about the lexicon of texts, as they base their calculation on single words rather than on lexical bundles (Biber & Barbieri 2007), sentences or larger sections of texts. Therefore, they are unable to replace trained raters who base their judgments on holistic or analytic scales.

2.2   The influence of writing strategies

The measures of lexical richness usually look at the writing as a product rather than as a process. Traditionally, writing was considered more of a product, and the writing

process itself was neglected (Sun & Feng 2009; Paran 2012). Nunan (1999: 272) also claims that "one of the most controversial aspects of writing pedagogy has been the tension between process and product approaches to the teaching of writing". Writing involves several skills, steps and strategies, and a successful piece of writing requires a good combination of many factors. The first group of these factors is advanced L2 language proficiency in general and writing skills in particular. The second group involves strategies used during writing and practice with writing.

There has been a debate among researchers to what extent native language (L1) writing skills and strategies are transferable to an L2. Silva (1993), in her empirical study involving adult native speakers of more than 20 different languages, pointed to more differences than similarities between L1 and L2 writing. She found that her undergraduate student participants did less planning in L2 both on the global and the local levels (while planning and writing their texts). They had more difficulties organizing their thoughts and concentrated the most on grammatical accuracy. On the contrary, a year later Berman (1994) concluded, based on data gathered from secondary school Icelandic students of English, that there are close similarities between L1 and L2 writing. He found that "many learners transfer their writing skills between languages, and their success in doing so is assisted by the grammatical proficiency in the target language" (Berman 1994: 29). In other words, proficient bilinguals are able to transfer L1 writing skills more successfully than beginners. The same idea of a proficiency threshold for writing transfer was confirmed by Ito (2009) in his study that involved over 300 Japanese EFL university students, and by Beare (2000), who investigated the writing strategies of English-Spanish bilinguals.

L2 student writers with low levels of proficiency in the target language may employ different writing strategies depending on their aims. Uzawa and Cumming (1989), for instance, differentiated between what they called "keeping the standards" and "lowering the standards" strategies. The first group of strategies refers to the learners' effort to balance for their insufficient language skills by asking for help, using sources, revising and also needing and using considerable time for text production. The second group of strategies is typical of students who do not try to produce excellent texts; they rather put little mental effort into the process and take as little time as necessary to finish the task.

Writing strategies can be classified into four main groups of categories: a) **rhetorical strategies**, such as organizing ideas into a coherent text and keeping genre conventions in mind, b) **metacognitive strategies**, involving the planning, evaluating and self-monitoring steps, c) **cognitive strategies**, such as generating ideas, imitating models and revising; and finally d) **social/affective strategies**, which involve asking for help and feedback, reducing anxiety, keeping up motivation and drawing on

previous experience (Riazi 1997; Mu & Carrington 2007). Some researches argue for the usefulness of direct strategy instruction in L2 writing. From the large group of strategies Cao (2011) pointed out four that are typical of English learners. The first one is the **strategy of avoidance**, namely the simplification of the syntactic structure or the use of familiar vocabulary in order to avoid errors. The second one is the so-called **strategy of preparation**, which refers to the memorization of set phrases and sentences to be used during writing. The **strategy of language switch** means the mechanical translation of ideas and phrases from the native language to the target language, which often happens under time constraints. The last one is the **strategy of spelling**, which is nothing else but the illegible writing of parts of words when a student is unsure about the spelling of the target word.

We can conclude that research into the writing process and strategies used by second and foreign language learners has produced a range of different results, depending on the focus, the participants and the methods of the studies. Among others, language use, planning, avoidance and the concentration on certain aspects of the writing process are rhetoric, metacognitive and cognitive strategies that seem to be the most relevant ones in the case of a controlled exam situation; therefore, they are investigated in this study.

3   Methods

3.1   Research questions

The literature reviewed in the previous sections and my experience in teaching L2 writing suggest that certain writing strategies may result in better student essays. Therefore, this study addresses the following main research questions:

1. Do the essays of students who often produce longer texts in English have better lexical richness measures than the essays of those who do not engage in frequent text production?
2. Do students' self-reported text writing strategies influence the lexical profile of their written texts?
3. Is there a significant difference in the results between subject groups tested in different academic years?

3.2 Participants

The participants involved in this study were 89 third-year students at the undergraduate level English studies program at a large Hungarian university, participating in a comprehensive study concerning the lexical proficiency and text writing practices of non-native students of English. The participants formed two subgroups, Groups A (n=41) and B (n=48), that were tested during two consecutive academic years. This was done in order to insure a better generalizability of the results.

Participants were all Hungarian native speakers and English as a foreign language (EFL) users. They had an average of twelve years of previous English studies, and were expected to have a C1 level of English proficiency. Twenty of them had a one to five month stay, while twelve of them a longer than six month stay in an English native speaking environment. There was no statistically significant difference between the two subgroups in terms of the years of English studies or the time spent in an English-speaking country.

3.3 Data collection instruments and data analysis

Each participant wrote an argumentative essay during their end-of-the-third-year English proficiency exam. This assured that all students within the same study group had the same circumstances for writing. Students were given ninety minutes to respond to prompts and write an argumentative essay of 300–350 words in length, without using a dictionary. Group A was provided with a choice of three prompts, all having two argumentative sides to choose from. This meant altogether six options. Students in Group B, those who took the comprehensive language exam a year later, had a choice of two, instead of three prompts due to changes in the exam policy. Topics included the brain drain phenomenon, e-book storage in libraries, renewable sources of power, gender and age inequality at jobs and healthy lifestyle. My experience in teaching writing to these groups of undergraduate students shows that homework assignments would not be easily comparable as the time spent on and the effort put into writing assignments greatly vary not only from student to student, but also the same students may work very differently on various assignments due to time constraints, availability of sources (such as dictionaries and the internet), type of assignment, their tiredness level, to mention a few.

The essays had to be standardized in terms of length and spelling due to the sensitivity of the lexical measure to text length (see section 2.1), but with the goal of leaving out the least number of words possible. Since most students produced a text within the given 300–350-word limit, the cut-off point was 300 words. This length restriction meant usually the loss of the final one or two sentences only. Evident spelling

errors were corrected so that the computer program would recognize these words instead of wrongly categorizing them as non-existent or rare words. However, unlike in Laufer and Nation (1995), lexical errors that contained existing English words (such as wrong phraseology or phrasal verbs) were not excluded from the texts. The main reason for this was the problematic nature of deciding what should be classified as an error and how much of the text around it should be deleted. Another reason was that the texts themselves were too short to allow for extensive deletion. On the contrary, non-existing English words (for example, *embetter* instead of *improve*) and proper names were excluded from the analysis, as they would significantly increase the number of words classified as above the first 2,000 English words.

Three lexical richness measures were used for this study, namely the lexical frequency profile (LFP), the type/token ratio and the lexical density. The lexical profile of the texts, following Laufer (1998), indicates the proportion of the first 2,000 most frequent words and can be quickly calculated with the help of text tools, such as the *Complete Lexical Tutor* (Cobb 2000). The condensed lexical profile argued for by Laufer (1998) does not further analyze academic and off-list (less frequent) words. The type/token value indicates the ratio of different words and all words in a text, while lexical density figures show the ratio of content words and function words.

Questionnaire data referring to the frequency of writing longer texts in the target language and a selected group of writing strategies were included in the analysis and collected in the participants' native language a few weeks earlier than the essays. Questions were close-ended with options to choose from, except for the first two questions (A and B below). The instrument had been piloted prior to the data collection and options were formulated during this pilot phase. In the case of question C, participants had to indicate a first, second, third and fourth choice (organization, appropriate vocabulary, grammar, and spelling), based on their importance during writing. For questions D, E and F, two or three possible answers were provided. The questions are listed in Table 1 below.

TABLE 1. Questions used in the questionnaire.

A	Number of times you write longer texts per week in English (including homework assignments, e-mails, letters, etc., containing a minimum of 10 sentences).
B	Number of times you write longer texts per week in Hungarian (including home-work assignments, e-mails, letters, etc., containing a minimum of 10 sentences).
C	During essay writing what aspect do you focus on the most?
D	What is the way you plan and write your essay most often?
E	When the closest equivalent of a Hungarian word/expression does not come to mind in English, what do you do most often?
F	If you are not sure of the spelling of a word you want to use and cannot check it, which strategy do you use more often?

4   Results and discussion

4.1   Lexical richness measures in the essays

The lexical frequency profiles of the essays produced by the 89 third-year students ranged between 77.67 and 95.37 (mean=86.53, SD=9.44), the type/token ratios between 0.43 and 0.62 (mean=0.53, SD=0.04) and the lexical density figures between 0.43 and 0.58 (mean=0.5, SD=0.03). These numbers indicate individual variability among the essays. The reasons behind this can be numerous, for instance, general language proficiency and vocabulary proficiency differences among the students, topic choices, attention paid to the task, the tiredness and stress levels of the students during the exam, experience with the task type, general writing strategies and strategies chosen for the task (such as producing a full draft or doing key words/ideas brainstorming only) and motivation (for example, targeted best grade possible, therefore, trying to do their best or targeted bare pass, which means trying to achieve the minimum required). The present study does not investigate most of these factors; for this reason it is not possible to give conclusive answers. Furthermore, it is assumed that all the factors mentioned above interact to some extent during task completion. The influence of general writing experience and essay writing strategies are discussed in the following subsections.

4.2   The influence of the frequency of writing longer texts

As students with more experience writing longer texts in English were expected to produce lexically richer essays, participants were asked about their opportunities for producing longer texts in English. Results indicate that students, although being in the

third year of target language medium education, produce very few texts consisting of a minimum of ten sentences. Students reported to write between one and 25 such texts per week, with an average of 3.45 texts.

As a control question (question B), students also had to report on their writing experience in their native language (Hungarian), as it was assumed that their text production behavior may show similar patterns in the two languages. In other words, a student who writes very little in the L1 may do the same in the target language; similarly, those who write a lot may do so in both languages. On average, students produced 3.61 longer texts in Hungarian per week. This is slightly above the English average, but is still a very low figure. Altogether, these values could hardly be considered a desirable amount of extensive writing practice in either of the two languages.

In order to answer research question number one, the reported writing experience was compared to the lexical measures of the essays produced under exam circumstances. Surprisingly, the Pearson correlation found weak and statistically insignificant differences on the group level. This result could be interpreted in various ways. First, the majority of the students had very limited consecutive writing experience, which makes the groups not very heterogeneous in this sense, therefore, a difference between the more and less experience reported is very small. Second, students with more experience in producing longer texts in the target language do not necessarily write essays that are richer in vocabulary. They may use the same restricted vocabulary, have little topic related specific lexicon or may not feel motivated to produce lexically rich texts if they are able to do the task with less varied vocabulary. Third, the interpretation of the data needs to be taken cautiously, as the analysis is based on the students' stated overall writing experience without knowing the type of writing, the purpose, the length and the genre of the texts they produce regularly.

4.3   The influence of overall writing strategies

In order to answer research question number two, answers provided to questions C to F in the questionnaire were examined. Question C referred to the attention paid to sophisticated vocabulary choice during writing in the target language. It was expected that those who claim to dedicate more attention to the aspect of carefully chosen, appropriate vocabulary use in essays (versus grammar, organization and spelling), produce lexically richer texts. Table 2 reports on the attention paid to these different aspects while writing target language texts. Data show that there is some visible difference between the two study groups in terms of the perceived importance of the various factors during writing, but tendencies are the same. It is interesting to note that vocabulary is not a popular first choice; indeed, very few students in both groups

indicated this as the most important factor to consider while producing essays. For many students, it is only a second or third aspect, while for slightly over one-fourth of the students, vocabulary remains the least important factor. Spelling, which may influence vocabulary choice, is ranked last by a considerable number of students (42% and 46% in the two groups). In terms of the differences between the two groups, it is seen that students in Group A had a more grammar oriented writing, while students in Group B were paying more attention to vocabulary and organization.

To conclude, vocabulary does not seem to be the priority of choice among the several factors involved in writing, of which only four were asked to be rated. Appropriate vocabulary choice seems to be the third or fourth among these for most students, preceded by grammar, organization or even spelling.

TABLE 2.   Student ranking of the importance of the four factors in essay writing.

	Order of choice	Group A (n=41)	Group B (n=48)
Grammar	First	44.4%	37.5%
	Second	42.2%	37.5%
	Third	13.3%	16.7%
	Fourth	0%	8.7%
Organization	First	37.8%	37.5%
	Second	13.3%	25%
	Third	17.8%	18.8%
	Fourth	31.1%	18.8%
Spelling	First	15.6%	16.7%
	Second	24.4%	12.5%
	Third	17.8%	25%
	Fourth	42.2%	45.8%
Vocabulary	First	2.2%	8.3%
	Second	20.5%	27.1%
	Third	51.1%	37.5%
	Fourth	26.7%	27.1%

As a next step in the investigation of overall writing strategies, students were asked in question D to report on the strategies they use during written text production under controlled circumstances when no help is allowed to be used (such as dictionaries, sample essays, topic-related reading, friends or relatives). Exams are very similar in the sense that no sources can be used, unlike in the case of take-home assignments. Subjects first reported on their strategy of L1 versus L2 use during essay planning. The use of both languages during the planning phase was reported by 77.4% of the participants (84.4% of students in Group A and 66.7% of Group B). The option, according to which they write down only ideas that come to their mind directly in English, was chosen by the remaining 22.6% (15.6% of Group A and 33.3% of Group B). Although these results

show a visible difference between individuals within and among the subgroups, the paired-samples t-test could not identify statistically significant differences in the lexical richness measures of the essays of students who indicated different language choices.

Similarly, statistically significant differences were not found in the case of the other two strategies related to lexical choices and avoidance during essay writing. The overwhelming majority of students in both groups (95.6% in Group A and 95.8% in Group B) reported using another lexical item close in meaning to the target item if this one did not come to their mind. Only the remaining four percent claimed to leave the information out when they have problems accessing the English target word. The last strategy, related to possible spelling problems, revealed that 70% of the students use a lexical item close in meaning if the spelling of the target word is problematic for them. The remaining 30% of the participants guess the right form. These results indicate avoidance and guessing strategies in writing. It is likely that these strategies result in the use of fewer lexical items in writing compared to what students would use if they were asked to talk about the given topic instead of writing about it. Nevertheless, writers have the advantage of a longer timeframe to think their thoughts over, plan and adjust their language in writing. Again, these replacement strategies had no statistically significant bearing on the lexical profile of student essays written under controlled circumstances.

A limitation of the results is that the selected strategies and the alternatives given to students in the questionnaire could only provide a partial picture of the strategies and aims of student writers. Also, it needs to be underlined again that the strategies in this study referred to self-reported general writing strategies that were not connected to a specific writing task. A think-aloud protocol used in other studies (e.g. Beare 2000), however, could not be employed under exam conditions and post-testing interviews could also reveal only a fraction of the cognitive strategies used during writing. Analyzing both the draft and the final versions of essays and asking students to reflect on the steps and changes made during the writing process are hoped to provide a fuller picture in future research on the strategies and their influence on the final product.

5   Conclusion

This study investigated the lexical richness of essays written by undergraduate non-native students of English as part of a comprehensive English language exam. The lexical frequency profile, the type/token ratio and the lexical density measures were calculated. As it was assumed that writing practice and certain writing strategies may influence the lexical measures of these essays, questionnaire data were collected referring to some general rhetorical, cognitive and metacognitive writing strategies.

Results suggest that the writing opportunities and the strategies students were asked to report on do not directly influence the lexical richness measures of their essays written during an exam. This is understandable in light of the fact that a combination of a variety of strategies is used during essay writing, as has been pointed out by Riazi (1997) and Mu and Carrington (2007). Moreover, the exam circumstances may not push students as much to perform their best as we may think. Some may employ avoidance strategies and strive for a bare pass. Also, many other factors, such as time limit, anxiety level, employed effort, environment, topic choice, and language proficiency do have a role in the outcome of the writing task. Students were asked to indicate in the questionnaire what they believed to be the most typical strategies they use; nevertheless, there is considerable variation as to what specific strategies students employ in given situations or in a given stage of writing. A think-aloud protocol or a post-writing interview may reveal more closely what the factors are that lead to certain strategy choices. However, these methods were not available for the present study due to the controlled nature of the essay writing. Further research with a smaller number of participants, but with a closer look at the process of writing at various stages, may inform us about more specific strategies and the reason behind their choices. Although positive correlations were not found in this study between general self-reported strategies and actual language use in essays, the importance of strategy training and process writing seem to be evident. As Negari (2011: 303) concluded in his study, "students can become better learners if they become more aware of their learning processes and then decide to act on that awareness." Data also revealed that students in this study group, when they were asked about the four important aspects of text construction (grammar, organization, spelling and vocabulary), did not consider advanced vocabulary as the most important factor to keep in mind; therefore, awareness raising about the importance of sophisticated language use should also be a crucial part of writing instruction. Besides class-related writing tasks, students should also be suggested opportunities to practice extensive writing, as the majority of them lack the necessary practice.

References

Beare, S. 2000. *Differences in content generating and planning processes of adult L1 and L2 proficient writers*. An unpublished doctoral dissertation. University of Ottawa, Ottawa, Canada. Available: http://www.ruor.uottawa.ca/fr/handle/10393/8515?show=full.

Berman, R. 1994. Learner's transfer of writing skills between languages. *TESL Canada Journal*, *12* (1), 29–46.

Biber, D. & F. Barbieri 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26* (3), 263–286.

Cao, X. 2011. Learners' strategies in English writing. *Sino-US English Teaching*, *8* (4), 221–226.

Cobb, T. 2000. *Complete lexical tutor*. Available at http://www.lextutor.ca.

Granger, S. & M. Wynne 2000. Optimizing measures of lexical variation in EFL learner corpora. In J.M. Kirk (ed.) *Corpora galore*. Amsterdam: Rodopi, 249–258.

Ito, F. 2009. Threshold to transfer writing skills from L1 to L2. *Kanto-koshinetsu Association of Teachers of English Bulletin*, *23*, 1–10.

Kojima, M. & J. Yamashita 2014. Reliability of lexical richness measures based on word lists in short second language productions. *System*, *42*, 23–33.

Laufer, B. 1998. The development of passive and active vocabulary in a second language. Same or different? *Applied Linguistics*, *19* (2), 255–271.

Laufer, B. 2005. Lexical frequency profile. From Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics*, *26* (4), 582–588.

Laufer, B. & P. Nation 1995. Lexical richness in L2 written production: can it be measured? *Applied Linguistics*, *16* (3), 307–312.

McCarthy, P. & S. Jarvis 2007. Vocd: a theoretical and empirical evaluation. *Language Testing*, *24*, 459–488.

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

Mu, C. & S. Carrington 2007. An investigation of three Chinese students' English writing strategies. *TESL-EJ*, *11* (1), 1–23.

Negari, G. M. 2011. A study on strategy instruction and EFL learners' writing skill. *International Journal of English Linguistics*, *1* (2), 299–307.

Nunan. D. 1999. *Second language teaching and learning*. Boston: Heinle & Heinle.

Paran, A. 2012. Language skills: questions for teaching and learning. *ELT Journal*, *66* (4), 450–458.

Riazi, A. 1997. Acquiring disciplinary literacy: a social-cognitive analysis of text production and learning among Iranian graduate students of education. *Journal of Second Language Writing*, *6* (2), 105–137.

Silva, T. 1993. Toward an understanding of the distinct nature of L2 writing. *TESOL Quarterly*, *27* (4): 657–677.

Silva, T. & P. K. Matsuda (eds) 2001. *On second language writing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Šišková, Z. 2012. Lexical richness in EFL students' narratives In C. Ciarlo & D. S. Giannoni (eds) *Language studies working papers. Vol. 4*. Reading: University of Reading, 26–36.

Sun, C. & G. Feng 2009. Process approach to teaching writing applied in different teaching models. *English Language Teaching*, *2* (1), 150–155.

Uzawa, K. & A. Cumming 1989. Writing strategies in Japanese as a foreign language: lowering or keeping up the standards. *The Canadian Modern Language Review*, *46*, 178–194.