Linguistic Variation in EFL Students-Composed Virtual Texts in Different Registers

Zigrida Vincela
Faculty of Humanities
University of Latvia

Abstract

Linguistic variation has received increasing attention since the multidimensional analysis (MDA) was proposed by Biber (1988) and applied (Xiao, McEnery, 2005) to investigate the texts of different registers. Biber advocated that it is accurate to view register differences as continuous dimension of variation distinguishing texts. The purpose of the paper is to present linguistic variation found in EFL student texts grouped according to communicative purpose. EFL students at the University of Latvia continue to experience difficulties in register-determined choice of linguistic features in different virtual writing situations, so they have been involved in the composition of texts for various communicative purposes. These texts have been compiled into a corpus and examined for relevance of linguistic variation that may be found within.

Introduction

The researchers (Crystal, 2011; Herring, 2001; Kern 2006; Warschauer, 1996 and Whittaker, 2003), who have explored the texts of electronic discourse, have all concluded that this discourse comprises the registers that draw on linguistic features used in written as well as in spoken communication. This conclusion correlates with the theories on register perspective of discourse – the functional relationship of the situational context (field/topic, tenor/participants relationships and mode/organization) – and on the variation of linguistic features. A theoretical framework has been provided by the Systemic Functional Linguists Halliday (1970, 1989), Ferguson (1994) and Martin (2001) and these theories are reinforced by the multi-dimensional analysis (MDA) put forward and also implemented by Biber (1988) Biber, Coenrad (2009), Xiao and McEnery (2005). According to MDA, register differences in texts i.e. the texts of electronic discourse, are to be viewed as a continuous dimension of variation distinguishing themselves according to their communicative purpose. Such a register-based variation of linguistic features in modern electronic discourse can be a source of frustration for the EFL students at the University of Latvia, who are non-native speakers of English (Vincela, 2008). The purpose of this study, therefore, has been to compile a corpus of student-composed virtual texts over seven registers and to undertake an investigation of the lexical and grammatical features to identify their statistically computable frequency differences by applying MDA. The analysis would then reveal how far the students have

varied the lexical and grammatical features in their virtual texts according to the communicative purpose of a particular register.

1 Methodology: Corpus Compilation and Retrieval of Linguistic Features

The texts, included in the corpus, were written during the periods of student participation in the *IDEELS* (*Intercultural Dynamics in European Education through Online Simulation*) project and during e-course *English Academic Writing* virtual writing activities.

During their active involvement in the *IDEELS* Project simulations, the students collaboratively developed a series of virtual documents in order to create a simulation on the Parliament of the *Eutropian Federation* (*EF*) – and so as to in representing one imaginary country of the virtual *Federation*. During these simulations the students composed texts within various registers: statements, letters, and messages to synchronous online conferences. As a part of the active involvement in their writing activities as envisaged by the e-course *English Academic Writing*, each individual student drafted asynchronous online communication messages in which they investigated and analysed the communicative situations of online simulations and the resulting texts created during these simulations.

The corpus of data for the study comprises 1412 electronic texts (220, 012 tokens) created by the students during their prescribed writing activities. First of all, in order to compile the corpus data, the texts were extracted from both of the online environments – from Project *IDEELS* and from the e-course *English Academic Writing*. Secondly, the corpus data type was defined and a specialized electronic text filing system was designed to hold the complete texts, which were then arranged according to their communicative purpose in registers. All texts have been grouped into collaboratively and/or individually written texts. The collaborative texts had all been written during online simulations and, according to their communicative purpose were then sorted into statements, letters and simulation messages and post-simulation messages. Individual texts were written during the e-course *English Academic Writing* activities and, according to their communicative purpose, were then grouped into pre-essays written before and post-essays written after student participation in the simulations and asynchronous discussion messages. This arrangement of the data then allowed for the investigation and comparative analysis of the lexical and grammatical features of these virtual texts.

For the automated analysis of linguistic features, the corpus of student texts was part-of-speech annotated. This was performed by the tagger CLAWS (Constituent Likelihood Automatic Word-tagging System), developed at *UCREL* (University Centre for Computer Corpus Research) within Lancaster University, and which provided each token with a part-of-speech tag, which was then manually edited to check the selection of tags containing relevant word class information. In order to perform a detailed investigation of linguistic variation, seven previously mentioned registers of student electronic texts were selected for analysis. Thirty eight lexical and grammatical features belonging to each of the three following dimensions were identified: *Dimension A: Involved Versus Informational Production,* along

which more involved texts display higher dimension scores*; Dimension B: Explicit Versus Situation Dependent Reference,* along which more explicit texts display higher dimension scores, and *Dimension C: Abstract Versus Non-abstract Information,* along which higher dimension scores are displayed by more abstract texts. .

These linguistic features were retrieved from the text corpus with the help of two functions of *WordSmith Tools*: *Wordlist* and *Concord*. *Wordlist* was implemented for the identification of average word length and for the standardized type/token ratio (TTR). *Concord* served as tool to extract the frequency of all the other linguistic features.

The student text corpus contains complete texts of diverse magnitude. Thus to obtain an accurate assessment of the frequency distribution of linguistic features across the corpus, statistical procedures such as those implemented by Biber (1988) were used. Firstly, a dispersion analysis of linguistic features across the entire corpus was performed by calculating the normalized frequency and mean frequency of each linguistic feature. Then the standard deviations, which are a measure of the dispersion of linguistic features in a text corpus were calculated. The obtained information revealed that some of the lexical and grammatical features are distributed considerably evenly across the whole corpus. However, there are linguistics features that show an appreciable difference in dispersion, for example, first person pronouns, other nouns, and prepositions. The dispersion variations of the selected linguistic features in the student composed text corpus revealed that their use can show significant differences once the texts are arranged in registers according to their communicative purpose. As a result of this a more detailed comparison of the text groups was performed. In order to explore the correlation of linguistic features across text groups, all three dimension (*A, B, C*) scores for each text group were calculated by summing up the scores for each linguistic feature belonging to a particular dimension. To identify the significance level of the variation of linguistic features, Pearson's Correlation was applied. Test results confirmed, with a 95% confidence level, the significance of the differences of the texts written by the students. The obtained *p* scores ($p = 0.01$) are smaller than 0.05 and confirm that the frequency variation of linguistic features in the text groups are significant in all three dimensions across all seven registers. However, it is also important to analyse weather in all seven register groups the students' variation of linguistic features was adequate to the situational context of that register (i.e. if the students' variation of lexical and grammatical features correlates with the variation of communicative purpose of the texts grouped in registers)

2 Results and Discussion

MDA of linguistic features reveals significant frequency variations of these features in all three dimensions (*A, B, C*) in the collaboratively written texts during Project *IDEELS* online simulations and in the individually written texts envisaged by the e-course *English Academic Writing*. Figure 1 shows the results for all three dimension scores of the collaborative texts (statements, e-letters, simulation and post-simulation messages), whereas Figure 2 reveals the results for the three dimension scores of the individually written texts (pre-essays, post-essays and discussion messages).

2.1 Collaboratively Composed Texts

As it may be observed in Figure 1, in *Dimension A: Informational Versus Involved Production*, the collaboratively written texts show the most marked differences. The greatest dimension is displayed by post-simulation messages (Example 4 in Appendix). These messages were exchanged by the students after each simulation session and they were aimed towards the discussion of gained insights during the simulation. These messages, as it was envisaged by the communicative context, were highly interactional, resembling conversation. Collaborative statements (Example 1in Appendix), e-letters (Example 2 in Appendix) and simulation messages (Example 3 in Appendix) display considerably lower scores. Surprisingly, simulation messages, which were written during synchronous online communication, show an even lower score than letters, whereas statements, as it is to be expected, display the lowest score, as they are the most transactional of texts.

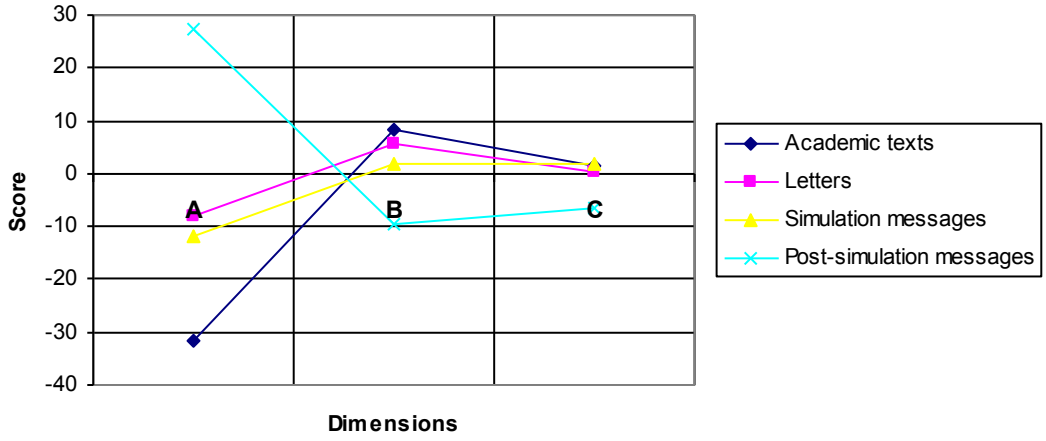

Figure 1: Comparison of dimension scores of collaboratively written texts

In *Dimension B* (Figure 1) *Explicit Versus Situation Dependent Reference*, which marks the degree of explicitness in discourse, there are noteworthy differences among the four registers. The highest score is displayed by statements (Example 1 in Appendix), so they are the most explicit. E-letters (Example 2 in Appendix) exhibit a lower score than statements, whereas simulation messages (Example 3 in Appendix) display a lower score than e-letters. Although both simulation and post-simulation messages display the lowest scores, there is a noteworthy difference in the scores between these two registers in this dimension. Simulation messages, which aimed at consensus building among the delegations of the simulated *Eutropian Federation,* on topical policy issues of this country, are considerably more explicit. In contrast, post simulation messages, which contain opinion exchange of simulation insights and thus resemble a conversation, are the least explicit.

The scores of *Dimension C* (Figure 1) *Abstract Versus Non-abstract Information* reveal surprising results. Although the statements (Example 1 in Appendix) that display the highest score in this dimension, prove to be abstract and thus transactional texts, the simulation conference messages (Example 3 in Appendix) display nearly just as high a

dimension score as do the statements. These messages were developed by using the messaging system for synchronous online communication that functioned similarly to a 'chat' system; however, the communicative context of conferences – negotiations of on specific issues considered by the students who simulated various Parliamentary members of the virtual Federation – were required to choose of the linguistic features that highlight abstract information. Post-simulation messages (Example 4 in Appendix), as to be been expected according to the communicative context, show considerably lower scores by comparison to statements, letters and simulation messages.

The obtained dimension scores reveal that the students had explored the situational context of each collaborative writing activity in order to achieve its communicative purpose.

2.2 Individually Composed Texts

The obtained results show that in *Dimension A: informational versus involved production* all three individually composed text groups reveal noteworthy differences (Figure 2). The dimension scores show that pre-essays (Example 6 in Appendix) are more involved texts than post-essays (Example 7 in Appendix) and messages (Example 5 in Appendix). This also confirms that before their participation in simulation writing activities the students experienced difficulties in recognition of the functional relationship between the situational context and linguistic features. Post-essays, according to this dimension score, are thus revealed to be more formational than pre-essays. This proves that the students have more relevantly recognised the functional relationship between situational context and linguistic features than in the case of pre-essays.

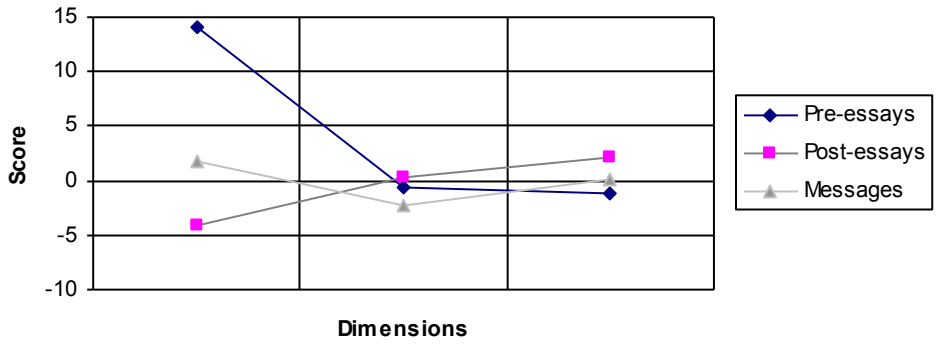

Figure 2: Comparison of dimension scores of individually written texts

Alongside *Dimension B:* (Figure 2), *Explicit Versus Situation-Dependent Reference*, there is a less noteworthy difference among the three text groups than along *Dimension A*. The highest score among the three text-groups is displayed by post-essays (Example 7 in Appendix), which confirms that the post-essays are more explicit than pre-essays and messages, however they are less explicit than simulation statements and letters.

In terms of *Dimension C: Abstract Versus Non-abstract Information,* the three text groups show noticeable differences (Figure 2). The most abstract are post-essays (Example 7 in Appendix). Surprisingly the pre-essays (Example 6 in Appendix) present a considerably lower score than messages.

The obtained dimension scores of the three groups of individually written texts reveal that the students have shifted the variation of linguistic features to a more adequate use in their post-essays thorough all three dimensions.

3 Conclusions and further research

The results thus for obtained allow for the conclusion that, apart from the pre-essays, the students tended to vary the linguistic features in the texts according to their communicative purpose. During simulations, the students practiced investigation of the situational context details for each text type in order to reach its communicative purpose and thus to advance the flow of the simulation activities. Moreover, the positive impact of simulation-based writing activities on overall students' competence of register variation is confirmed by their shifting of linguistic features from a less situationally adequate use in the pre-essays to a more situationally adequate use in post-essays. Further, it would be useful to explore the student-composed texts in comparison with the contextually similar authentic English texts that linguists have grouped in registers. Such a comparison would allow for a further identification of the linguistic features, the variation nuances of which seem to be the primary cause of difficulties in particular registers for the students.

References

Biber, D. (1988). *Variation across Speech and Writing.* Cambridge: Cambridge University press

Biber, D., Conrad, S. (2009). *Register Genre, Style*. Cambridge: Cambridge University Press

Crystal, D. (2011). *Internet Linguistics*. Routledge: Taylor and Francis Group.

Ferguson, Ch. (1994) Dialect, register and genre: working assumptions about conventualization. In: Biber and Finegan (eds.). pp. 315-347.

Halliday, M.A.K. (1970) Language structure and language function. In: (ed.) J. Lyons *New Horizons in Linguistics*. Baltimore: Penguin Books.

Halliday, M.A.K. (1989) *Spoken and Written Language.* Oxford and New York: Oxford University Press.

Herring, S.C. (2001) Computer-mediated Discourse. In: D.Schifrin, D. Tannen, H.E. Hamilton (eds.) *The Handbook of Discourse Analysis* Malden, MA: Blackwell, pp. 612-706.

Kern, R. (2006). Perspectives in technology in learning and teaching languages. *TESOL,* Quarterly, 40 (1) pp. 183-211.

Martin, J.R. (2001). Language, register and genre. In: (eds.) A.Burns and C.Coffin *Analysing English in Global Context* pp. 149-166, London: Routledge

Xiao, Z., McEnery,A. (2005). Two approaches to genre analysis: three genres in modern American English. *Journal of English Linguistics*, 33 (1), March, pp. 62-82.

Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom. *CALICO* journal, 13 (2) pp. 7-26.

Whittaker, S. (2003). Theories and methods in mediated communication. In: A.C.Graesser, M.A. Grensbacker, S.R. Goddman (eds.) *Handbook of Discourse Processes.* Mahwah.New Jersey, London: Lawrence Erlbaum Associate, Publishers, pp. 243-286.

Vincela, Z. (2008). Corpus analysis: application of data-driven language studies. *Proceedings of the Third Baltic Conference on Human Language Technologies.* Kaunas: Vytautas Magnus University and NGLST (Nordic Graduate School of Language Technologies) pp. 327-333.

Appendix

Example 1: Statement

Policy Statement
Table of contents
1 Executive summary
2 Goals
3 Data Protection in *Bardland*
4 Environmental Information Regulations in *Bardland*
5 Freedom of Information in *Bardland*
6 Bibliography

1Executive summary
By means of this policy statement we will try to find solutions to the existing problems connected with Data Access and Protection (*DAP*) in *Bardland*. Being a part of the *Eutropian Federation* we expect to get support from other *Eutropian* countries, and to come to common solutions. The solution of problems connected with data access and protection, undoubtedly, are of vital importance. The overall goals of Bardland are to develop and to improve the system of *DAP* in *Eutropian Federation*.
3 Freedom of Information in Bardland
The Freedom of Information Act (FoIA) was issued to give people the right to ask for information from public authorities, which do not fall under data protection or environmental information regulations as listed above. FoIA is intended to promote a culture of openness and accountability among public sector bodies and to facilitate better public understanding of how public authorities carry out their duties, why they make the decisions they do, and how they spend public money. The FoIA requires government agencies to make public their data access and protection.

Example 2: E-letter

Subject: RE: GOV: Agencies of the Government.
To the Delegation of Northland

We got acquainted with the formulation of the functions of the Special Agency proposed by your Delegation, and we would like to express our general agreement with all the ideas. Your idea of dividing the responsibilities of the agency is really very useful as it could make the Data operation process more accurate and elaborate.
However. we would like to state our strong conviction that the separation of the functions of the Agency, according to the storage of different kinds of data, could cause a kind of overlapping responsibilities and even could be a reason for raising the bureaucratic tendencies within this sphere.
Concerning all other points our political views are quite similar.

We look forward to further negotiations between our Governments,
The Government Delegation from the Republic of Bardland

Example 3: Simulation conference messages

<Message 89> Our delegation also agrees with the majority of the participants, and we would like to speak over this issue in the afternoon.
<Message 61> This institution is based on the platform of the Governmental Service and is a kind of independent organisation with all the actions regarded as the part of State Restricted Information Service. The access to this Data should be in hands of experts.
<Message 97> There are certain legal requirements to disclose information, for example, in relation to infectious deseases. Doctors are obliged to report incidences of specified infectious diseases to health boards and the Infectious Diseases Surveillance Centre. This information must be disclosed without identifying the patient. The confidentiality of personal information such as medical records must be protected by the data protection legislation. beyond doubt.
<Message 99> We would like to make some proposals about Data Access and Protection in financial field. The access to the financial history of the individual is available to: (1) data subject; (2) financial institution; (3) Government (in exceptional cases). By exceptional cases we mean if the data subject is involved in criminal affairs. Then the Government is allowed to access the individual's financial history.

Example 4: Post-simulation conference messages

All the topics were quite interesting bacause they are actually the everyday problems that affect all the countries. aren't they? But we are FOR SURE more informed about them now!!
Why do you think so?
We think that we also learned a lot of new things.
What do you mean?
It was great and we gained new experience!
Our feelings are the same!
Who are you?
Where are you from?
We're actually not form Spain.

Example 5: Individual message

It is possible to use personal pronouns in the essay. But the writer should avoid expressing his personal opinion too straightforwardly by, for example, writing. *I guess*. Instead of this phrase, another sentence should be used, and it should be more formal, and it means that the first person pronoun, most probably, will disappear. However, in my opinion, it is necessary to use personal pronouns in quotations, for example, if the writer wants to make a quotation from a book and the quote contains first person pronoun it won' t be omitted, of course.

Example 6: Pre-essay

It has always been very important to know how formal and informal communication differ. Nowadays it's extremely important.

I can say that we must communicate in a different way with people who are older. I can say that for me informal communication is easier because I know that I can say everything that is on my mind. I don't have to use special words; I don't have to think about what to say. I can say everything what is on the tip of my tongue. But I understand that more formal communication is also necessary.

When do we use formal language? Do we use it only when we speak with teachers and business partners? I guess. But when I'm with my friends, relatives or family I speak absolutely differently. And, of course, this way of communication is much easier, because I don't have to think about the choice of a right word or intonation.

I can say that I use both informal and formal communication. When I communicate with people who are of my age, I use informal language, and when I communiacte with people older than me I use

formal language. I don't like formal language because I think it does not make the special connection between two or more people.

Example 7: Post-essay

The question of various kinds of institutions for punishing the criminals has always been a topical issue. Yet, more contradictory are discussions concerning the sentence of under-age law-breakers. As there is no clear evidence of the refinement of behaviour, and lifestyle of people leaving the institutions of punishment, one might doubt whether the system of sentence actually works. A new type of establishments aiming to take a highly disciplined and pedagogically tended care of under-age law-breakers should be introduced in Latvia.

Though there is a distinction between the punishment adapted to grownup criminals and teenage law-breakers. it is not providential enough. Under-age criminality should be controlled by special pedagogical approach. as teenagers are very flexible social group that can easily be influenced to change for better. Although under-age criminals often come from disadvantaged families, thus bringing the vice into society, there is a chance for them to be persuaded of a different way of living. By the help of new establishments, problematic teenagers would become aware of their own abilities to receive from society appreciation by giving positively tended attitude in return.