Register, Genre, Rhetorical Functions:
Variation in English Native-Speaker and Learner Writing

Ekaterina Zaytseva

Johannes Gutenberg-Universität Mainz, Department of English and Linguistics

Jakob-Welder-Weg 18, 55099 Mainz

E-mail: zaytseve@uni-mainz.de

Abstract

The present paper explores patterns and determinants of variation found in the writing of two groups of novice academic writers: advanced learners of English and English native speakers. It focuses on lexico-grammatical means for expressing the rhetorical function of contrast in academic and argumentative writing. The study's aim is to explore and to compare stocks of meaningful ways of expressing the rhetorical function of contrast employed by native and learner novice academic writers in two different written genres: argumentative essays and research papers. The following corpora are used for that purpose: the *Louvain Corpus of Native English Essays* (LOCNESS), the *Michigan Corpus of Upper-level Student Papers* (MICUSP), the *British Academic Written English* corpus (BAWE) and two corpora of learner English, i.e. the *International Corpus of Learner English* (ICLE) and the *Corpus of Academic Learner English* (CALE) – the latter being a corpus of advanced learner academic writing, currently being compiled at Johannes Gutenberg-Universität Mainz, Germany. The study adopts a variationist perspective and a functional-pedagogical perspective on learner writing, aiming at contributing to the field of second language acquisition (SLA), by focusing on advanced stages of acquisition and teaching English for academic purposes.

Keywords: novice academic writing, rhetorical function of contrast, variation, function-oriented annotation

1. Introduction

The branch of the SLA focusing on advanced levels of proficiency puts forward issues that are problematic for researchers, EAP teachers, and foreign language learners alike. Those include the need for an exhaustive description of language performance on an advanced level and a set of defining characteristics which could be further developed into assessment criteria.

One of the factors responsible for the problematic nature of "advancedness" is a somewhat narrow view of this stage of language acquisition as on the one hand, "no more than 'better than intermediate level' structural and lexical ability for use", as pointed out by Ortega and Byrnes (2008:283); and yet, on the other hand, as language performance, not "flawless" enough to be considered native-like.

2. Theoretical Background

Advanced learner writing has recently been the object of a number of corpus-based studies (cf. e.g. Callies, 2008; Gilquin & Paquot, 2008; Paquot, 2010). It has generally been analysed from a pedagogical perspective, i.e. against the yardstick of English native-speakers' writing, where features of learner writing have often been characterized as non-native-like. Among the areas identified as problematic for advanced learners are most notably accurate and appropriate use of lexis, register awareness, and information structure management. Yet, studies adopting a variationist perspective on advanced learners' output and considering a possible influence of different kinds of variables are still scarce (cf., however, Ädel, 2008; Paquot, 2010; Wulff & Römer, 2009). One of the reasons for this could be the lack of corpora representing advanced academic learner writing (Granger & Paquot, forthcoming), which makes it difficult, for example, to analyse the importance of genre and writer's genre (un)awareness as possible determinants of variation. The existing corpora include the following projects in progress: the '*Varieties of English for Specific Purposes*' database (VESPA) (cf. Granger, 2009), the

Corpus of Academic Learner English (CALE)[1], and the *Cologne-Hanover Advanced Learner Corpus* (CHALC) (Römer, 2007).

The pedagogical approach to learners' language production has brought forward particular kinds and methods of learner data analysis. One of them is annotating a learner corpus for errors (cf. Granger, 2004). Valuable as it is, this kind of corpus annotation, however, does not allow for a truly usage-based perspective on learner language production, where learners' experience with language in particular social settings is the focus of attention.

Corpus-based analyses of native English academic writing, meanwhile, have revealed that this register is characterised by a specific kind of vocabulary on the one hand (Biber et al., 1999; Coxhead, 2000; Paquot, 2010) and by certain kinds of grammatical structures on the other hand (e.g. Biber, 2006; Kertz & Haas, 2009). In addition, it has been pointed out that the register of native English academic writing displays a certain degree of variation as well, e.g. there is discipline- and genre-based variation in the form and use of lexico-grammatical structures used in written discourse (Hyland, 2008). However, there is little information on possible variation in different genres produced by novice native English academic writers (cf., however, Wulff & Römer, 2009).

3. Project Aims and Objectives

The present paper reports on work in progress exploring patterns and determinants of variation found in the writing of two groups of novice academic writers: advanced learners of English and English native speakers. It focuses on lexico-grammatical ways for expressing the rhetorical function of contrast in academic and argumentative writing. The study's aim is to explore and subsequently to compare stocks of meaningful ways of expressing contrast employed by native and learner novice academic writers in two different written genres: argumentative essays and research papers. For that purpose the following corpora are used: three corpora of native English corpora: the *Louvain Corpus of Native English Essays* (LOCNESS) (Granger, 1996), the *Michigan Corpus of Upper-level Student Papers*

(MICUSP)[2], the *British Academic Written English* corpus (BAWE) (Nesi, 2008) as well as two corpora of learner English, i.e. the *International Corpus of Learner English* (ICLE) (Granger, 2003) and the *Corpus of Academic Learner English* (CALE)[3] - a corpus of advanced learner academic writing, currently being compiled at Johannes-Gutenberg-Universität Mainz, Germany.

Another aim of the study is to investigate to what extent the influence of the variable 'genre' is a possible determinant of variation in the written production of various groups of academic writers. In this respect, it is important to address the issue of novice writers' genre awareness and to discuss the question of native-speaker norm. In addition, the paper explores the existence of interlanguage (IL)-specific strategies used by advanced learners to express rhetorical functions in writing.

The latter will be achieved by annotating both corpora of advanced learner writing for the rhetorical function of contrast. This kind of function-oriented annotation, though still rare in English learner corpus research, presents researchers with a valuable opportunity to view learners as active language users, rather than learners demonstrating deficient knowledge of the target language. In addition, the potential of multidimensional corpus analysis (Biber & Conrad, 2001) is currently being considered as a highly useful method of distinguishing between different registers and genres.

The study, thus, adopts a variationist perspective to novice academic writing, considering advanced interlanguage as a variety in its own right. At the same time, a functional-pedagogical perspective allows for a further analysis of those areas of language use that are still problematic for advanced learners, and reveals meaningful ways in which learners cope with writing-related tasks.

4. Function-oriented annotation

The advantage of adding a function-driven annotation is that it makes it possible to generally identify contrast in learner writing and to pin down an extensive stock of language means, treated as writers' lexico-grammatical preferences for signaling this rhetorical function in written discourse.

---

[1] http://www.advanced-learner-varieties.info

[2] http://micusp.elicorpora.info/www.micusp.org

[3] http://www.advanced-learner-varieties.info

Further on, the encoded information allows for function-driven, together with form-driven searches in learner writing, resulting in a comprehensive and accurate picture of the variety of lexico-grammatical means for expressing contrast used by two groups of (advanced) German learners in their writing. In addition, a subsequent quantitative analysis can provide valuable insights into general and individual preferences of learners in terms of which items are particularly favoured in the context of a specific writing-related task set in a specific situation of language use. Moreover, its combination with a qualitative analysis of patterns and determinants of variation in the ways of expressing contrast in writing promises to shed more light on general written argumentation strategies employed by (advanced) German learners.

In order for this kind of annotation to be reliable, several conditions have to be met, which when applied to the present project, imply clarification of the concept of a rhetorical function and a clear definition of the rhetorical function of contrast in terms of its aim and distinctive characteristics, complemented by a list of possible language items for its realization in writing.

The next step involves annotating each instance of contrast being expressed in written discourse in both corpora of (advanced) German learner writing (i.e. CALE-GE and ICLE-GE). This stage is followed by a detailed description and categorization of the lexico-grammatical means for expressing contrast in learner writing. Subsequently, comparative analyses, quantitative as well as qualitative, are carried out, in order to reveal possible patterns and determinants of variation that exist in the novice academic writing.

Preliminary findings reveal a slight degree of genre-induced variation in German learners' writing in terms of sentence placement of the contrastive item *however*, see Table 1 below.

Corpus	Corpus size, N of tokens	Initial	Non-initial	Total
ICLE-GE	234.423	103	125	228
%		45	55	
CALE-GE	55.000	49	27	76
%		64	36	

Table 1: Position of the contrastive item *however*

As the table shows, German learners seem to prefer the initial sentence positioning of *however* in academic (CALE-GE), rather than in argumentative (ICLE-GE) writing. Thus, the item *however* found in the sentence initial position is almost 1,5 times more frequent in term papers than in argumentative essays. This seems to tie in well with one of the findings recently reported by Wagner (2011). In her empirical study, she points out a tendency for *however* to take up the initial sentence position in literature and cultural studies texts, rather than in linguistic texts and general corpora (2011:43). Due to a modest number of words contained in the version of the CALE corpus used at the time of analysis (see Table 1), the preliminary finding reported in the current paper should be treated with caution. A further analysis of a greater number of occurrences in a bigger corpus is needed in order to provide more empirical evidence for supporting and accounting for this finding.

5. Conclusion

The project presented in the present paper sets out to explore advanced IL-specific strategies for coping with a writing-related task in the context of English academic and argumentative writing. This is achieved by combining a functional-pedagogical view with a variationist perspective on learner writing and annotating the rhetorical function of contrast in the two corpora of learner writing. At the same time, the findings of the project will contribute to the area of variation in novice native English academic writing and will further a definition of the native speaker norm, which advanced learners are generally expected to aim at.

6. References

Ädel, A. (2008): Involvement features in writing: do time and interaction trump register awareness? In G. Gilquin, S. Papp. & M. B. Díez-Bedmar (Eds.), Linking up Contrastive and Learner Corpus Research. Amsterdam, Atlanta: Rodopi, pp. 35-53.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999): Longman Grammar of Spoken and Written English. Harlow: Pearson Education.

Biber, D., Conrad, S. (2001): Introduction: Multidimensional analysis and the study of register variation. In S. Conrad & D. Biber (Eds.), Variation in English: Multidimensional Studies. London: Longman,

pp. 3-13.

Biber, D. (2006): University Language: A Corpus-Based Study of Spoken and Written Registers. Amsterdam: John Benjamins.

Callies, M. (2008): Easy to understand but difficult to use? Raising constructions and information packaging in the advanced learner variety. In G. Gilquin, S. Papp. & M. B. Díez-Bedmar (Eds.), Linking up Contrastive and Learner Corpus Research. Amsterdam, Atlanta: Rodopi, pp. 201-226.

Coxhead, A. (2000): A new academic word list. TESOL Quarterly, 34(2), pp. 213-238.

Gilquin, G., Paquot, M. (2008): Too chatty: Learner academic writing and register variation. English Text Construction, 1(1), pp. 41-61.

Granger, S. (1996): From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), Languages in Contrast. Text-Based Cross-Linguistic Studies. Lund Studies in English 88. Lund: Lund University Press, pp. 37-51.

Granger, S. (2003): The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly, 37(3), pp. 538-546.

Granger, S. (2004): Computer learner corpus research: Current status and future prospects. In U. Connor & T. Upton (Eds.), Applied Corpus Linguistics: A Multidimensional Perspective. Amsterdam, Atlanta: Rodopi, pp. 123-145.

Granger, S. (2009): In search of a general academic vocabulary: A corpus-driven study. Paper Presented at the International Conference 'Options and Practices of L.S.A.P Practitioners', 7-8 February 2009. University of Crete, Heraklion, Crete.

Granger, S., Paquot, M. (Forthcoming): Language for Specific Purposes. Retrieved from http://sites.uclouvain.be/cecl/archives/GRANGER_P AQUOT_Forthcoming_Language_for_Specific_Purp oses_Learner_Corpora.pdf , 17.12.2010.

Hyland, K. (2008): As can be seen: lexical bundles and disciplinary variation. English for Specific Purposes, 27(1), pp. 4-21.

Kerz, E., Haas, F. (2009): The aim is to analyse NP: the function of prefabricated chunks in academic texts. In R. Corrigan, E. Moravcsik, H. Ouali & K. Wheatley

(Eds.), Formulaic Language: Volume 1. Distribution and historical change. Amsterdam, Philadelphia: John Benjamins, pp. 97-117.

Nesi, H. (2008): BAWE: An introduction to a new resource. In A. Frankenberg-Garcia, T. Rkibi, M. Braga da Cruz, R. Carvalho, C. Direito & D. Santos-Rosa (Eds.), Proceedings of the 8th Teaching and Language Corpora Conference. Held 4-6 July 2008 at the Instituto Superior de Línguas e Administração. Lisbon, Portugal: ISLA, pp. 239-246.

Ortega, L., Byrnes, H. (2008): Theorizing advancedness, setting up the longitudinal research agenda. In L. Ortega & H. Byrnes (Eds.), The Longitudinal Study of Advanced L2 Capacities. New York: Routledge/Taylor & Francis, pp. 3-20.

Paquot, M. (2010): Academic Vocabulary in Learner Writing: From Extraction to Analysis. United States: Continuum Publishing Corporation.

Römer, U. (2007): Learner language and the norms in native corpora and EFL teaching materials: a case study of English conditionals. In: S. Volk-Birke & J. Lippert (Eds.), Anglistentag 2006 Halle. Proceedings. Trier: Wissenschaftlicher Verlag, pp. 355–63.

Wagner, S. (2011): Concessives and contrastives in student writing: L1, L2 and genre differences. In J. Schmied (Ed.), Academic Writing in Europe: Empirical Perspectives. Göttingen: Cuvillier, pp. 23-49.

Wulff, S. & Römer, U. (2009): Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. Corpora, 4(2), pp. 115-133.