WCES 2012

Scientific articles in English in Economics Journals. A case study for Romania

Teodora Ghivirigă *

*a*Al. I. Cuza University of Iaşi, , B-dul Carol I, Nr. 11, Iaşi, RO-700506, Romania

Abstract

The large number of Economic journals publishing in English in Romania prompted an investigation into the language circulated to establish whether there are significant deviations from the language in articles by native English researchers. Two corpora were created and various linguistic features were compared: first reference pronouns, relative pronouns, passive forms, modal verbs, use of tense – aspect forms. With one exception, there were statistically significant differences between the two corpora. Further research can provide guidelines for more efficient reviewing of Romanian journals of Economics and improved didactics of academic prose, especially article writing, in terms of syllabi and teaching materials.

© 2012 Published by Elsevier Ltd. Selection and/or peer review under responsibility of Prof. Dr. Hüseyin Uzunboylu
Open access under CC BY-NC-ND license.

Keywords: corpus, research article, economics journal.

1. Introduction

One of the goals of the Bologna Process is to increase the attractiveness of the European academic system; one way to ensure this is an enhanced quality of the teaching staff, as assessed by various internal and/or external quality assurance organisms according to intensive or extensive criteria (or both). The evaluation of the faculty members' activity is done according primarily to the quantity/quality of research, as substantiated by material published in internationally recognized publications registered in various databases, number of citations, factor of impact, possibly derived work. This emphasis on research has resulted in increased pressure on the academic staff of all degrees to publish the results of their work in specialized journals. The Romanian academia is not different. Universities are striving to increase their number of scientific journals, but also to upgrade the existing ones by changing their publication policy. The top criterion is scientific content, however, another criterion, yielding a high score in the journal evaluation process, is the ratio of articles published in widely spoken languages, of which English is expectedly the first choice. It follows thus that the national language seems no longer an appropriate vehicle for the circulation of knowledge even within the national arena. The faculties of economic studies and business schools in Romanian universities are obviously going with the trend. A quick check of the major faculties' sites indicates that they produce at least one specialised journal each, whose texts, accessible on-line, are mainly – if not exclusively – in English. The natural outcome is a large amount of text in a foreign language which is now considered the *lingua franca* of publication. At a cursory glance the articles and the web-pages of the journals display an unquestionably professional design, following their reputed models, however, there are language lapses that may unfortunately suggest otherwise. Two questions suggest themselves at this point: what is the linguistic

---

* Corresponding Teodora Ghiviriga.
  E-mail address: teoghivi@gmail.com

quality of these texts? and how does it measure up against the language of comparable international publications of high standing?

2. Methodology

Multi-variable corpus analyses have been performed starting with Biber's seminal study (1988) mainly to create a linguistic "profile" of a certain genre and thus facilitating genre identification for a particular text. More recently, at the far end of the spectrum, Coniam (2004) considers the possibility of an individual profile resulting from the creation of a "personal corpus" of academic writing and what syntactic, lexical and textual features should be the subject of investigation. Other, more eclectic, methodologies have been adopted (Nishina, 2007), also applied across genres, for the comparison of literary and non-literary texts and the identification of distinctive language patterns occurring in a certain text type. In the present investigation, the style (academic writing), the type of discourse (specialised) and the genre (scientific article) are the major criteria for corpus creation; the present corpus-driven comparative analysis aims at establishing to what extent certain linguistic features are similar in the two corpora and where they diverge. The resulting profile can provide the background for a) more efficient language monitoring and reviewing for Romanian journals of Economics and b) improved didactics of academic prose and especially research article writing in terms of curricula and teaching materials.

For this research, two corpora were compiled, one consisting of articles authored by at least one native English economist and published in specialized journals (corpus Native English NC), and a second consisting of articles authored by non-native (Romanian) economists published in English in specialized journals in Romania (corpus non-native NNC). The two corpora created (NC and NNC) consist of 22 articles and 49 articles respectively, a little under 200,000 words each (190153 tokens, 10213 types for NC and 179946 tokens, 10262 types for NNC, respectively). In the case of NC, texts were selected from international journals, issues dating from 2006 to present, and native authorship was attributed by name and institutional affiliation. In the case of NNC the articles date as early as 2005 to the current year; an extra criterion was added in the sampling process, namely location (university): as far as possible – given the scarcity of sources in some instances – a comparable number of texts were extracted from journals published in the various universities in Romania (Bucharest, Craiova, Iași, Oradea, while data from Cluj were not available) and were organized into subcorpora accordingly; the text are all from the domain Economics and focus especially on the topic "economic crisis" (but not only). The corpora are raw, due to the preliminary nature of this investigation; texts were prepared manually and processed with a concordancer (AntConc, version AntConc3.2.4w), with some of the data (ex. Passive structures, aspectual forms) extracted by using *wild cards*.

The aim of this investigation, as stated above, is to assess to what extent certain features in the language used in articles authored by Romanian researchers (non-natives) are similar to those of the language produced by native writers. To that end, a set of features to be analyzed was established (selected from Biber's list of features with functional associations, 1998:145); given the limited space allowed for this paper, the results will refer only to: lexical specificity and general text statistics, pronouns and place-holders, passive forms, aspect, modals and subordination features (relative pronouns as indicators of relative clauses); other features have already been analyzed (elements of hedging, discourse markers, coordinators, prepositions, adverbs, articles) or will be part of further inspection and statistical processing (quoting strategies, factor analysis).

3. Results and discussion

3.1. General indicators

Before proceeding to the frequency counts of the features considered, a general analysis of the two corpora was done, to assess the collections of texts in terms of general indicators such as lexical richness and readability scores.

To this end, two versions of 1Mb were produced by randomized selection of the initial collections in order to be processed with *Compleat lexical tutor (*http://www.lextutor.ca/vp/eng/ 2011/06), a vocabulary profiler based on Laufer and Nation's Lexical Frequency Profiler; the resulting vocabulary profile offers data on lexical density, type-token ratio and lexical density, as well as the ratio of words of Anglo-Saxon origin (including function words) vs. Greco Latinate words. For both corpora, the figures are identical, or almost, for all the indicators: the type-token ratio (0.6) and the lexical density index (0.57) are the same for both corpora. Differences appear in two areas: in the percentage of off-list words, different by one unit (NC 9.84 > NNC 8.8), and in the Anglo-Saxon index, which is higher by two points in the case of the NC (NC 63.05% > NNC 61.28%); the latter indicates a higher ratio of words of Anglo-Saxon origin and consequently a higher proportion of Greco-Latinate i.e. supposedly scientific vocabulary in the NNC. These corroborate with other general indicators, such as average number of characters and syllables per word and average number of words per sentence: while the former show the corpora to be almost identical (average number of characters per word NC 5.26 ~ NNC 5.3), for the latter there is a difference of 4.39 words per sentence. Increased sentence length in the texts by Romanian authors is paralleled by the values of the readability scores – Gunning Fox, Coleman Liau index, Flesch Kincaid and SMOG (using the *Readability Calculator*), as shown in Table 1; the values are higher for NNC in all four cases. Such values suggest differences in the lexical choice of NNS, with heavier emphasis on Greek-Latinate elements, which is predictable, given Romanian's linguistic typology, and in instruments for nominalization.

Table 1. Readability values for the two corpora

Corpus	NC	NNC
Gunning Fox index	16.65	18.90
Coleman Liau index	13.83	14.28
Flesch Kincaid Grade level	14.35	16.40
SMOG	15.44	17.02

3.2. Specific features

First person reference: pronouns. **Passives**. **Place-holder subjects**. The use of pronouns in general and of self reference pronouns in particular have been the subject of research with focus on the first pronoun either as an element creating writer identity (Hyland, 2002) or as an indicator of linguistic and pragmatic competence in a comparative analysis of a learner vs. native speaker corpus (Neff at al., 2004). While the NNC analyzed here consists of texts by professional writers not by learners, some findings of Neff's study mentioned above can be of relevance: there is a tendency to overuse first person forms, especially first person plural, which is to be attributed not to the convention of self effacement and objectivity in Anglo-Saxon academic prose but rather to the influence of L1 on Spanish NNS students (Neff, 2004:79) or to the presence of various oral features in essays by novice (Japanese) writers among which the overuse of 1st person pronouns (McCrostie, 2008:105); this may be corroborated with Hyman's findings (2002:1098, 1106), relative to the use of first person pronominal forms by Hong Kong undergraduates, whose texts (on Economics) reveal preference for the first person singular (while no comparison with native writers is intended) with a tendency, however, to avoid self mention.

The values in the NNC and NC indicate a clear difference in the use of 1st person pronouns with a preference of the Romanian NNS for the plural (121 occurrences of 1st person. sg. – 0.6726 per 1,000 compared to 300 occurrences of 1st person pl. – 1.6671 per 1,000 words). Both total values (NC 927 occurrences with 4.875 per 1,000 and NNC 563 occurrences with 3.128 per 1,000) and specific values for the singular (NC 359 / 1.8879 normalized value vs. NNC 121 / 0.6726 normalized value) and plural forms (NC 568 / 2.9870 normalized value vs. NNC 300 / 1.6671 normalized value) are shown to be relevant ($\chi^2 = 2.54$ df= 2 p<0.001). The figures for NC are also consistent with the figures in Nishina (2007:15) for the general academic register without any subject specialization (*I* 1.82/1,000 and *we* 2.79/1,000). Several hypotheses can be forwarded for this difference: an influence of L1 rhetoric, which encourages the use of introductory phrases such as *We can see, We notice* etc. (see Neff, 2004:79), an avoidance strategy of making direct statements ("author invisibility" in Hyland's terms) or lack of familiarity with

the recent change in trend in Anglo-Saxon style, i.e. reducing 1st person plural pronouns and Passive forms in academic writing. Further investigation will consider the various uses of *we* (inclusive vs. exclusive) or possible instances of ambiguous use, as well as the various functions of *I* in the text (as researcher, indefinite, biographical etc., see Vladimirou, 2007: 144) or other sets of discourse functions that *I* may have (for example, Hyland, 2002:1099).

The figures on the use of pronouns correlate with the pattern of use for the Passive structures in the two corpora. Among the conventions of academic prose, the Passive (often paired with the omission of the agent) has been regarded as one of the most important strategies to achieve the expression of objectivity, to the extent to which some consider it an essential feature of good academic prose. However, many studies mention the over-/under-/misuse of Passive forms, associated with impersonal subject place holder *it* or structures with existential *there*. Hyland & Milton (1997: 198) acknowledge a general background of difficulty in manipulating impersonal forms in scholarly writing, which may indicate limited use of passives in NNS texts; a similar find is offered by Reid (in Neff, 2002:75), according to which Spanish NNS use fewer Passive forms than NS. However, the number of Passive forms in NNC is (slightly) higher than that in NC (NC 1158 occurrences / 6.0893 per 1,000 words vs. NNC 1281 / 7.1188 per 1,000 words, and long/agent passives counts NC 554 / 2.9134 vs. NNC 654 / 3.6340), which is consistent with Hinkel's remark (1999:101) that Asian NNS use Passive constructions "significantly more frequently than natives", without, however, offering any quantitative data. This tendency can possibly be correlated with the lower rate of 1st person forms in the NNC discussed above.

One other difference that deserves further inspection is the significant under-use of existential *there* in the Romanian NNC corpus; it was considered as one of the variables in the initial set on account of the fact that the canonical word order can be violated and subject inversion in Romanian is acceptable. Results from research on various NNS seem to converge: Palacio-Martinez and Martinez-Insua identify higher frequency of *there* constructions with L1 Spanish users of English, with a ratio of 24/10,000 NS vs. 39/10,000 for Spanish NNS (2006:219), while Callies & Zaytseva (2011:54) mention an overuse of impersonal subject placeholders (*it, there*) as an alternative strategy to suppress the agent where the NS favour other reporting strategies; also, Hinkel (1999:102) claims that Asian NNS use fewer non referential *it* sentences, but significantly more existential *there* sentences than NS. The finds in NNC set the Romanian writers on Economics apart from this general tendency: the number of *there* sentences in the NC (632 occurrences, 3.3236 /1,000 words) exceed by far those in NNC (186 occurrences, 1.0336/1,000 words). Further qualitative research of the corpus may reveal whether the low figure is related to incorrect subject-verb sequence, or to sentences with the subject doubled by a "false" anticipatory *it* where in Romanian an impersonal passive construction would be used as a mark of impersonality and objectivity. (ex. *It is spoken English and French here.)

Use of tense-aspect forms. One of the areas where differences were expected to be significant between the corpora was that of tense use and aspect use, given the typological dissimilarity between the two languages and intuitive personal and shared teaching and testing experience with Romanian learners of English. Use of tense and/or aspectual forms in the research article and in its parts ("moves", in Swales' terms) has been subject to extensive investigation in genre-related studies starting with Swales (1990, 2004) to the present. A number of contrastive studies show disparities in the use of aspectual forms in the various moves: e.g. texts by Asian students display lower rates of Present Perfect forms in the introductory section of the academic article than natives, indicating a tendency to avoid them (Hinkel, 2004:21), which applies equally to Progressive forms (ibid.). A comparatively recent quantitative study on technical texts by Spanish writers of English (Carrio-Pastor, 1997) claims that no significant differences existed in the use of tense-aspect combinations. These finds are more in keep with the results in NNC regarding aspect-tense combinations (Past/Present/Future Progressive and Past/Present/Future Perfect): the difference is moderately significant, with χ^2=3.21, df=1, p=0.073. This suggests that although Romanian does not have systematic aspectual oppositions, L1 has little influence on the use of aspect forms; further research is needed to verify their appropriateness and correctness in NNC.

Modals. Modality appears to be notoriously problematic for nonnative speakers of English. The richness of expression of modality in English plays a central role in the pragmatic proficiency of a NNS and is an important element of the conventions of academic writing; often it may turn as a source of confusion and misuse. Modality – and especially modal verbs – have been studied in its various contexts: in relation to topic or discipline (Piqué et al., 2001, 2002) or text types and genres (Biber, 1998), expression (modal auxiliaries, lexical verbs, adjectives or adverbs etc.), as well as in the stance adopted by the author (hedging, boosting, attitude markers etc.) (Hyland, 2008); some contrastive studies point at dissimilarities in the modals used to express a certain modal meaning: NS prefer *would* and *will*, followed by other modals to express epistemic modality, while (Hong Kong) NNS resort to *will* and *may* (Hyland & Milton, 1997:189); German and French learners show a clear preference for *can* / *could* to express possibility and *may* is selected by French learners, with a generalized tendency to overuse *might* in all NNS groups listed above (Aijmer, 2002:62), while other modals (*must, should*) have a variable pattern of usage depending on topic. Therefore, significant variations were anticipated: differences do not lie in the total raw number of modals (NC 11.617 per 1,000 words vs. NNC 12.237 per 1,000 words), but rather in the selection of the modals in the two corpora (highly significant statistically, $\chi^2 = 335$, df=8, p<0.000) as presented in Table 2.

Table 2. Modals in the corpora (per 1,000 words)

Corpus	must	may	might	can(not)	could	should	would	shall	will
NC	0.514	1.682	0.610	2.382	1.141	1.330	2.303	0.026	1.688
NNC	0.883	1.089	0.272	3.606	0.872	1.205	0.911	0.188	3.206

As anticipated based on the specialized literature, some modals are overused in NNC (*can, will, must*), while others are underused (*may, might, could* and *would*); however, the patterns in NNC do not integrally parallel any of the groups in the examples above. Further analysis of modals in NNC needs to establish whether and to what extent they are appropriately used, with focus on the highly significant disparities, and to refine analysis by taking into consideration modal meanings in order to establish a pattern characteristic of non-native Romanian writers of research articles in journals of Economics, a domain whose epistemological position may orientate selection of modal meaning.

Relative pronouns and clauses. In order to investigate one pattern of subordination, namely the relative clause, the *wh-* forms were analyzed (*who, which* and also *that,* with further consideration of *that* deletion and *wh*-clauses which could not be integrated in the present study). In an article on the Euro-English project, Mollin (2006:48), referring to a previous study, contradicts its findings that this alleged variety (i.e. Euro-English) may move towards an interchangeability of the relative pronouns *who* and *which*, emphasizes the limited number of related errors in the corpus but points at the different distribution. A more comprehensive study focusing only on relative pronouns (Olson Flanigan & Inal, 1996:221) with NS speaking American English and NNS of various L1 indicate the natives' general preference for *that* (present and omitted) in writing and the non native's choice of *wh-* words (*who, which*). In the case Romanian NNS, *which* is by far the most frequent option (NC 620 / 3.260 per 1,000 words vs. NNC 1054 / 5.857 per 1,000 words), with a tendency to replace both animate an inanimate referents and to "translate" Romanian phraseology (ex. "the reason for which companies pay…" or "the way in which firms progress"), associated with slight underuse of *who* and flexional forms *whom* and *whose* (NC 214 / 1.125 per 1,000 words vs. NNC 134 / 0.744 per 1,000 words) and moderate underuse of *that* (NC 2934 / 15.429 per 1,000 words vs. 1929 / 10.719 per 1,000 words). The difference can be therefore considered significant ($\chi^2 =282$, df=4, p<.0.000); the disparity in total number of occurrences may suggest diverging patterns of subordination, to be further analysed.

4. Conclusions

The figures extracted at this stage seem to indicate significant differences between the two corpora and encourage further investigation in language, text structure, rhetorical conventions. The reference to research in the field on the same linguistic features may not be entirely relevant due to the dissimilar profile of the texts (authors: commonly learner vs. expert users in my case, type of discourse: general vs. specialized – on a restricted domain, i.e.

Economics, genre: essays or other types of texts vs. expert academic writing in one genre only – the research article), however, they may point to possible trends and patterns of deviation from the native use. The issue of "native command of the language" should be carefully addressed: it is often left unspecified and may become a fetish; nevertheless, different assessment standards do exist by native and non-native reviewers: Whitley, 1998 (in Steinman, 2003:85) found that non natives rated a group of papers higher while a group of English natives rated the same group lower. Therefore, the Economics journals in English published in Romania can but profit by careful monitoring and quality assessment of the language used, resulting from corpus investigation; an accurate description of the language they circulate can also provide efficient guidelines in the didactics of language for special purposes and in academic writing.

References

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In Granger, S., & Hung, J., & Petch-Tyson, S., (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching*. John Benjamins. (pp. 55-76)

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge University Press.

Björk, B., Roos , A. & Lauri, M. *Scientific journal publishing: yearly volume and open access availability.* At http://informationr.net/ir/14-1/paper391.html, 2011/10

Callies, M.,& Zaytseva, E.. (2011). *The Corpus of Academic Learner English* (CALE): A new resource for the study of lexico-grammatical variation in advanced learner varieties. In Hedeland, H., Schmidt, T., & Wörner ,K. (Eds.), *Multilingual Resources and Multilingual Applications* (pp. 51-56). Hamburg Working Papers in Multilingualism B 96.

Carrio-Pastor, M.L. (2009). Contrasting Specific English Corpora: Language Variation. *International Journal of English Studies. Special Issue*, 221-233. www.um.es/ijes 2011/10

Coniam, D. (2004). Concordancing oneself: Constructing individual textual profiles. *International Journal of Corpus Linguistics* 9(2), 271-298.

Hinkel, E. (1999). Objectivity and credibility in L1 and L2 academic writing. In Hinkel, E., (Ed.), *Culture in Second Language Teaching and Learning*. (pp. 90 – 108). Cambridge: Cambridge University Press

Hinkel, E.(2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research* 8, 1, 5-29

Hyland, K., & Milton, J. (1997). Qualification and Certainty in L1 and L2 Students' Writing. *Journal of Second Language Writing*, 6(2), 183 – 205.

Hyland, K. (2002) Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics* 34, 1091–1112.

Hyland, K. (2008). Disciplinary voices. Interactions in research writing. *English Text Construction* 1(1), 5-22. John Benjamins.

Kuo, C.H. (1999). The Use of Personal Pronouns: Role Relationships. *Scientific Journal Articles. English for Specific Purposes* 18 (2), 121-138.

Lozano, C., & Mendikoetxea, A.(2008). Postverbal subjects at the interfaces in Spanish and Italian learners. In G. Gilquin, S. Papp, & M. BelenDiez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 85 – 125). Rodopi: Amsterdam New York.

Mc Crostie, J. (2008) Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal* 32, 97-114.

McEnery, T., & Xiao, R. (2005) *Passive constructions in English and Chinese. A corpus-based contrastive study*. http://eprints.lancs.ac.uk/63/ 2011/11.

Mollin, S. (2006). English as a Lingua Franca: A New Variety in the New Expanding Circle? *Nordic Journal of English Studies*, vol. 5 no. 2 (2006), 41-57. http://hdl.handle.net/2077/3149 2011/11

Neff, J. et al. (2004). Formulating Writer Stance: A Contrastive Study of EFL Learner Corpora. In U. Connor, & T. Upton (Eds.), *Applied corpus linguistics: a multidimensional perspective* (pp. 73- 90). Amsterdam & Atlanta: Rodopi.

Noguchi, J., Orr, T., & Tono Y.(2006). Using a dedicated corpus to identify features of professional English usage: What do "we" do in science journal articles? In A. Wilson, D. Arcer, & P. Rayson (Eds.), *Corpus Linguistics Around the World* (pp. 155 – 166/12). Rodopi.

Olson Flanigan, B., & Inanl, E. (1996). Object relative pronoun use in native and non-native English: A variable rule analysis. *Language Varition and Change*. vol. 8 no. 2, pp. 203-226.

Palacios –Martinez, I., & Martinez-Insua, A. (2006). Connecting linguistic description and language teaching: native and learner use of existential *there. International Journal of Applied Linguistics*, 16 (2), 214- 232. Blackwell Publishing.

Piqué, J., Posteguillo, S., & Andreu-Besó, J. V. (2001). A pragmatic analysis framework for the description of modality usage in academic English contexts. http://institucional.us.es/revistas/elia/2/16.%20pique.pdf 10/2011

Piqué, J., Posteguillo, S., & Andreu-Besó, J. V.(2002). Epistemic and Deontic Modality: A LInguistic Indicator of Disciplinary Variation in Academic English. *LSP & Professional Communication* vol. 2, no. 2 (2002), pp. 49-65

Readability Calculator . http://www.online-utility.org/english/readability_test_and_improve.jsp 2011/09

Steinman, L. (2003) Cultural Collisions in L2 Academic Writing. *TESL Canada Journal*, vol. 20 (2), 80-91

Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press

Swales, J. (2004). *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press

Vladimirou, D. (2007). "I suggest that we need more research". Personal reference in linguistics journal articles. In C. Gabrielatos, R. Slessor & J.W. Unger (Eds.), *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching. Volume 1. Papers from LAEL PG 2006* (pp. 139 – 157). Department of Linguistics and English Language Lancaster University.

Yasunori, N. (2007). A Corpus-Driven Approach to Genre Analysis: The Reinvestigation of Academic, Newspaper and Literary Texts. *ELR Journal*, 1 (2). at http://ejournals.org.uk/ELR/article/2007/2 2011 /10