Learner Corpora and Cross-Linguistic Applications

Diana Oţăt[*], Titela Vîlceanu[**]

Abstract: Commonly featured as one of the most flexible analysis methodologies, corpus linguistics applications are an open class, facilitating the development of inter-and trans-disciplinary research studies. The extensive applicability of corpus linguistics syncs with the fast updating state-of-the-art technology, fostering multi-layered competences applied to design experiments, data collection and interpretation, lead to improved qualitative and statistical research tools. Driven by the requirement of multilayered competences, we embark upon the in-depth investigation of a learner corpus to identify the occurrence pattern and rate of mistranslation instances, based on a translation project carried out by 1st and 2nd year undergraduates in Translation and Interpretation.

Keywords: corpus-based study, cross-linguistic research, learner corpora, translation competence

1. Introduction

Defined as a systematized multidimensional and multi-purpose investigation, corpus linguistics has developed as an effective and resourceful methodology applied to almost any area of linguistics, and, in recent decades to translation theory and practice.

Nowadays we envisage of corpora as collections of texts delivered in a machine-readable form that can be constantly enlarged, at the touch of a button, to generate most varied and complex results that can be stored as research deliverables, or re-used as solid evidence for further applications. Resultantly, we highlight the key role of corpus linguistics as the interface of theoretical and experimental approaches to enhance first language, second and foreign language acquisition, translation practice and intercultural networking.

---

[*] Junior Lecturer, PhD, Department of Department of British, American and German Studies, Faculty of Letters, University of Craiova, Romania. E-mail: otatdiana@gmail.com
[**] Associate Professor, PhD, Department of Department of British, American and German Studies, Faculty of Letters, University of Craiova, Romania. E-mail: elavilceanu@yahoo.com

Alongside current trends in Corpus Linguistics and Translation Studies to open the doors for interdisciplinary and niche areas of expertise, mainstream literature pinpoints the emergence and evolution of learner corpora since the turn of the '90s. An early attempt to apply corpus linguistics to academic training was the *Collins Cobuild English Course* (*CCEC*), designed by Willis and Willis (1989) to make available the most frequent English words and phrases and their meanings.

The first international symposium on learner corpora, held in Hong Kong in 1991, brought together prominent scholars, members of the *ICLE project*, to join forces and open new horizons for modern-day learner corpora research areas (Tono: 2003). *ICLE* (*International Corpus of Learner English*) project encompasses argumentative essays written by higher intermediate and advanced learners of English belonging to different native language backgrounds. According to *ICLE* partners, over 400 published L2 papers have used *ICLE*, which amounts to more than "2.5 million words of argumentative essays written by university students with L2 English from several European countries, organized in different sub-corpora." (Granger *et al.* 2002: 17)

Acknowledging the highly adaptive and formative nature of corpus-based investigation, Biber and Reppen (1998: 157) are among the first to capitalize on the fertilizing ground provided by corpora analysis - a significant leap forward in our understanding of how "best to help students develop competence in the kinds of language they will encounter on a regular basis". Later on, contemporary scholars such as Granger (2002), Pravec (2002), Keck (2004), Myles (2005) and many others resorted to corpus linguistics and developed applied models for first and second language acquisition, research and practice.

In what follows we set out to frame current features and applications of corpus linguistics in order to carry out a corpus-based investigation aimed at identifying the nature and occurrence of mistranslation instances. Designed to face students with different translation tasks, the Translation Internship syllabusi for the 1st and 2nd academic year are aimed to develop trainees' translation competence and enhance self-reliance. The translation tasks assigned within these internship projects stand as reliable methodological tools, enabling tutors and supervisors to register, assess and quantify trainees' progress, as well as most frequent translation difficulties and instances of mistranslation exhibited at different linguistic levels. Also, the outcome of such corpus-based investigation may lead to further evaluation and implementation of corrective measures in translation training, hence contributing to translation teaching methodology and resources design.

2. Learner Corpora: theoretical insights and future prospects

Although initially developed to serve to two different research approaches and training goals, *Learner Corpus Research* (*LCR*) - intensively promoted by Granger (1993) and *Corpus-Based Translation Studies* (*CBTS*) - first introduced by Baker in 1993 - seem to display various key similarities, driven by a joint effort to improve trainees' communicative competence. Emerging almost simultaneously in

the early '90s, LCR and CBTS have set basically the same main objective, i.e. a multidimensional investigation of language transfer from trainees' L1 (in terms of LRC) or from the TL/SL for CBTS. Hence, the main analysis parametres for both corpora investigation approaches are set to register various aspects regarding interlanguage transfer and foreign language acquisition and/or translation. Admittedly, Granger (1998) acknowledges the transposable nature not only of the objectives, but also of the outcomes of such corpora investigations.

Aiming to charter the importance of corpus-based investigation as to enhance language acquisition, Granger (1998: 7) defines *learner corpora* as electronic collections of authentic L1 and L2 textual data complied in compliance with valid design criteria for a particular purpose. The author focuses on the application of corpus linguistics to L2 acquisition and highlights the fruitful outcomes generated by interlanguage corpus-based contrastive analyses. In thiscontext, Granger (1998) designs the *Integrated Contrastive Model*, successfully applicable to both research perspectives as well to hybrid investigation methodologies. To test the model, the author applies the *Contrastive Interlanguage Analysis* (*CIA*) arguing that corpus-based comparisons between native and non-native speakers would provide specific details to profile L1 and L2 users and related errors. Also, such comparisons enable both researchers and trainers to better identify particular L1 background features of the L2 learners. Designed to compare L2 learners' transfer performances in relation to learner's native language, CIA has brought significant contribution to error analysis. Though traditional views highly recommended to "treat causes of error very cautiously, for in many cases, what we see happening, however, is just the reverse" (Schachter and Celce-Murcia 1977: 67), contemporary corpus-based interlanguage research studies aim at assessing trainees' performance "in its own right rather than in respect of merely decontextualized errors" (Granger 1998: 34).

Granger's *CIA model* has been shared by prominent linguists and translation theorists such as Mason and Uzar (2000), Chesterman (2007) and Gilquin *et al*. (2008). Departing from Granger's *Integrated Contrastive Model*, Gilquin (2000) designs an analysis model of English and French causative constructions to validate the hypothesis that corpus-based interlanguage contrastive analyses provide some valuable results on L1 and L2 learners' transfer management. Also, Mason and Uzar (2000) seek to develop some natural language processing techniques applied to identify L2 learners' omissions of zero articles in an interlanguage corpus. Vanderbauwhede (2012) sets out to test the effectiveness of Granger's CIA applied to a corpus investigation of French *vs*. Dutch demonstrative determiner systems in L1 and their precise impact on written L2 productions as to establish some error prediction patterns in L2 transfers.

Tono (2003: 804) considers that Granger's "research avenue" may lead to successful results, though it might raise some methodological issues as well, and, argues that trainees' instruction year or age, listed by Granger as external selection criteria, would not necessarily validate that the subjects selected are comparable in terms of language proficiency. Concerned with error identification and improvement via corpora investigation, Tono (2003: 808) postulates that even though some

researchers may seek to implement large-scale manual tagging of all lexical expressions in learner corpora, various aspects regarding validity and reliability are not completely solved, as error tagging and error taxonomy still remain a thorny issue. Applicable to both language acquisition and translation practice, we share Tono's recommendations that corpus-based error tagging and categorization need to be purpose-oriented. Moreover, to secure some valid outcomes, the author suggests that a tagging scheme should include at least two aspects (a) linguistic category classification and (b) target modification taxonomy. Then, in terms of validity we would focus on error tagging and error assessment in the light of the specific research goals of each corpora investigation. As far as reliability is concerned, Tono claims that due to a lack of solid evidence, uncertainty of error type may stand as a serious problem and considers that the development of tagging schemes, which allow for alternative possibilities in terms of target forms, may be an efficient solution.

Extending the landscape, Aston *et al* (2004) cast light on the multifaceted perspective on the contemporary developments and tendencies of corpus-based research study applied to language training. Special attention is paid to the added value corpora investigations have brought to L2 learners, the didactic input via novel approaches and the inextricable relationship between corpora, their users and dedicated software.

In the same climate of opinion, Myles (2005: 376) argues that the key role of L2 acquisition is to develop applicable models meant to assist learners at a particular stage, while enhancing mental processes. According to Myles (2005: 377), the development of learner corpora to meet the above mentioned objectives has been constantly redesigned via inductive (bottom-up) and deductive (top- down) approaches. The author postulates that while top-down approaches use corpora investigation as tool to validate hypotheses, the bottom-up approaches exploit corpora as quantitative and qualitative research tools to formulate a hypothesis. Sharing the same perspective, Barlow (2005: 344) considers that bottom-up approaches in learner corpora are mainly applicable, if the research objective(s) is/are aimed at investigating particular issues concerning the trainees' language via introspection. On the other hand, top-down approaches develop trainees' skills to identify and describe certain linguistic patterns.

Other recent developments within the field of learner corpora have set as focal point some in-depth interlanguage analyses of complex structures. Growing aware of the difficulties encountered even by advanced learners to properly use L2 collocations, Nesselhauf (2005) dwells on the importance of corpus-based analysis applied to investigate the manifestation of trainee's difficulties at different levels. In the same spirit, Fitzpatrick (2007) develops a corpus-based analysis model to investigate phrases and structures errors, while Lozano and Mendikoetxea (2010) embark upon corpora investigations in terms of word order alternations.

To keep up with the dynamics of breakthrough technologies in language acquisition and translation practice, Granger (2010: 14) opens up the horizon towards corpus-based applications in cross-linguistic research and highlights that "linguistics and translation studies now have a common resource". She identifies two main types

of corpora used in cross-linguistic research, i.e. corpora consisting of original texts in one language and their translations into one or more languages, labelled by the author as *translation corpora*, and corpora that encompass original texts in two or more languages, "matched by criteria such as the time of composition, text category, intended audience", i.e. *comparable corpora*. Aiming at establishing which type of corpora may generate more reliable outcomes, the author designs a highly useful checklist. According to Granger (2010: 17) translation corpora set as research parameters the degree of equivalence between L1 and L2, although, as far as comparable corpora are concerned, these parameters are less resourceful. However, it seems that "what constitutes an advantage for one type of corpus, constitutes a disadvantage for the other and vice versa", since translation corpora rely on a limited availability of texts, while comparable corpora bank on a wider availability of texts. Regarding the diversification of corpus-based cross-linguistic applications, Granger advocates that such applications may supply any field that rests on the analysis of two or more languages with valuable input. Under the circumstances, she mentions the current development of automated translation, "notably via the creation and gradual update of translation memories", the design and evolution of electronic lexicography and thematic maps, and the steady updating of pedagogical material and teaching methodology.

The collaborative efforts carried out by Granger and Lefer (2017) materialize into the *Multilingual Student Translation Corpus* aimed at bridging the gap between learner corpus research and translation studies. The two authors design and implement a new international corpus, i.e. the *Multilingual Student Translation* (*MUST*) project (Granger and Lefer 2017). According to the authors, the *MUST* project is aimed at revising some weaknesses of earlier collections and set as strategic objective the development of a multilingual corpus. The project partners are in charge with the investigation of language transfer from and into 25 languages (50 language pairs). The annotation system prescribed displays typologies used in both LCR and CBTS. In addition, it enables users to mark translation strategies (transposition, simplification or explicitation), hence facilitating cross-linguistic assessments (syntax-discourse, lexicon-syntax, syntax-phonology, etc.). For reliable outcomes, the corpus underpins expert translations that function as reference works for students' translations. *MUST* sought to design a hybrid corpus to meet the needs of both linguists and translation theorists, i.e. a standardized use of the metadata and annotation system applied to all the translations included in the database to secure comprehensive data comparisons and reliable interpretations.

3. Corpus-based investigation: sustainable development of translation competence

Based on the principles of cross-linguistic research applications and on the close connection between LRC and CBTS, we set out to design a translation corpus addressed to 1st and 2nd year undergraduates enrolled in the Translation and Interpretation programme at the Faculty of Letters, University of Craiova.

Mapped out as a win-win project, the compilation of our corpus, on the one hand, sought to improve trainees' multi-layered competence and increase self-reliance via translation corpus-based assignments, and, on the other hand, to carry out an in-depth investigation as to profile learners' L2 proficiency level and their interlanguage performance when transferring the message from the ST (L1) into the TT (L2). It is worth mentioning that the students selected for the project have different native-language backgrounds (Romanian, Moldavian, Serbian, and French) besides a solid command of Romanian.

The design parametres observed both external and internal criteria; hence the sample selection considered the communicative function of the texts and the degree of language difficultly as trainees were challenged to translate two chapters of a representative novel authored by a contemporary Romanian author (Mircea Cărtărescu, *De ce iubim femeile*).

Adopting and adapting some of the main principles recommended by Sinclair (2005: 2-9) for corpus design, our model was developed in compliance with:

- *the principle of representativeness* – the design of a corpus "as representative as possible of the language from which it is chosen" (Sinclair 2005: 2). Cross-sectional criteria were applied as we selected trainees with different L2 proficiency levels (1st and 2nd academic year). Furthermore, in line with the CIA principles, our corpus-based investigation of L2 learners' translation performance covered multilingual L1 subjects. Table 1 below indicates the rate of representativeness as related to L2 proficiency and native language variation.

Table 1 The representatives rate of the selected subjects

Study year	Romanian	Serbian	Moldavian	French
1st year	8	0	8	1
2nd year	21	2	2	0

- *the principle of documentation* – which states that both corpus design and content "should be documented fully with information about the contents and arguments in justification of the decisions taken." (Sinclair 2005: 8). In this respect our monolingual corpus totals 2,905 words, and 42 students (1st and 2nd year) of different nationalities were assigned to translate from Romanian (SL/L1) into English (TL/L2) the first two chapters of one of the contemporary Romanian literature bestsellers.
- *the principle of topic or subject matter* – driven by the "use of external criteria" (Sinclair 2005: 10). In our case it is validated, as the texts sampled and their translation are aimed to provide a comprehensive representation of L2 command and translation performance at different linguistic levels, i.e. the transfer of different structures, phrases, lexical items, etc. from the SL into the TL.
- *the principle of homogeneity* – since "a corpus should aim for homogeneity in its components while maintaining adequate coverage, and rogue texts should be avoided." (Sinclair 2005: 14). For Sinclair *rogue*

10

texts are those texts "which stand out as radically different from the others in their putative category, and therefore [are] unrepresentative of the variety on intuitive grounds." Our corpus observes this principle entirely.

Having established the selection criteria and applied the design principles, the following step of our project was to assign the translation tasks. All the students were asked to provide a translation of the Romanian ST into English within a 3-week deadline. In terms of text editing, they were required to send the final version via e-mail, in in a word.doc format (TNR 12, 1, 5, justified) and a PDF format.

It is worth mentioning that all students complied with the deadline set and handed in their translated version of the selected texts in due time. After storing and organizing the texts according to the subjects' proficiency level (1st and 2nd academic year), we designed and applied an analysis model aiming to identify particular translation difficulties and/or instances of mistranslation while observing the external criteria set previously, i.e. the proficiency level and the linguistic and socio-cultural background of the subjects. The analysis model developed for our corpus-based investigation is illustrated in Diagram 1 below.
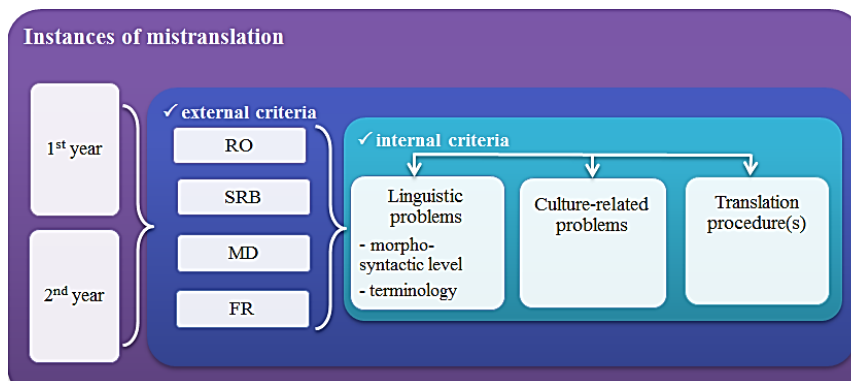

Diagram 1 Corpus-based analysis model

Based on Granger's (2002) *Contrastive Interlanguage Analysis* we carried out qualitative and quantitative corpus investigations, aiming at establishing the occurrence patterns and the frequency rates of translation difficulties and/or mistranslations based on the previously established internal and external criteria.

In terms of internal criteria, most of the mistranslation instances encountered were at the morpho-syntactic level, particularly related to the use of tenses and sequence of tenses. However, their representation and occurrence pattern tend to vary in terms of external criteria, i.e. the native language background of the subjects. Thus, if the Romanian students seem to encounter some difficulties in the use of Present Perfect and Past Tense, the Serbian and Moldavian students had difficulties at the syntactic level, i.e. word order and noun phrases. Also, the most frequently encountered morphological errors concern aspects of pronominal use (Serbian, Moldavian and French), whereas Romanian students wre faced with inversion issues and derivation.

11

e.g.

ST:	*De-a lungul timpului, am rămas acelaşi jerk căruia nu-i pasă cu ce se-mbracă…*
RO - EN	*Over time, I remained the same jerk who does not care about what he wears...*
SRB - EN	*Over time, I remained the same jerk who doesn't care what he wears…*
MD - EN	*Over time I have only remained same jerk, who think that no matter in what he dresses...*
FR - EN	*Over time, I have stayed the dame Jerk who doesn't care what he is wearing*

ST:	*Sigur că poate splendoarea ei se amestecă acum în mintea mea cu irealul caruselului de pe malul oceanului, cu leii de mare îngrămădiţi unii peste alţii lângă debarcader,cu omul-statuie încre menit pe postamentul său…*
RO - EN	*Surely her splendor now mingles in my mind with the unreal ocean carousel, sea lions piled over one another beside the pier, with the statue-man standing on his post…*
SRB- EN	*Of course it is possible that her prettiness now interferes in my mind with unrealistic ocean carousel, with sea lions piled up on top of each other next to the pier, with a statue of a frozen dumbfounded man on his stool…*
MD - EN	*Sure maybe her splendor is now mingling in my mind with the carousel on the ocean's shore, with sea lions piled up over each other beside the wharf, with the man-statue, dumbfounded on his pedestal...*
FR - EN	*Sure maybe as her splendour mixes now on my mind with an unreal ocean carousel, with the lions bordering over each other near the pier statue man confided on the post*

As expected, terminological issues were more challenging for the Serbian, Moldavian and French students, although some mistranslation instances were registered among Romanian students as well.

e.g.

ST:	*În loc să-ncep acum să explic de ce aceste flash-uri de frumuseţe pură sunt atât de minunate literar (pe cât de banale par la prima vedere), părăsesc locomotiva aşa cum e şi trec la vagoane.*
RO - EN	*Instead of trying to explain why these flashes of pure beauty are so literally wonderful (as trivial as they look at first glance), I'm leaving the locomotive as it is and head to the wagons.*
SRB - EN	*Instead of starting to explain why these flashes of pure beauty are so literally wonderful (however banal they may seem at first sight) I'm leaving the locomotive as it is and make my way to wagons*
MD - EN	*Instead start explaining why these flashes of pure beauty are so literally wonderful (still banal they may seem at first sight) I leave the locomotive as it is and move on to the wagons*
FR - EN	*Instead to start now and explain why this flash of pure beauty are more wonderful*

12

literary (and trivial at first glance), I leave the locomotive as is it and pass the wagon.

As far as culture-related problems are concerned, our corpus investigation revealed that although foreign students were expected to deal with cultural gaps when managing Romanian culture-specific items, the translation version of the Romanian students exhibited considerable instances of culture-bound mistranslations. A possible explanation is that that foreign students would have performed extensive research to find the most appropriate culture-related meanings (Romanian and foreign) before transferring them to English.

e.g.

ST:	*Vorbeam în citate nu din snobism, nici ca să mă dau mare (nu te puteai da mare decât cu muzica rock şi cu lista gagicilor pe care le-ai fi avut...*
RO - EN	*I was talking about in snobbery quotes, not be greater (you could be greater with the rock music and the list of the babes you had...*
SRB - EN	*I talked in quotes not out of snobbery, nor desire to become great (you couldn't become great without rock music or babes list which you would have…*
MD - EN	*I was talking in quotes not from snobbism, nor to do the grand (you couldn't do the grand only with rock music and the list of girls that you had had...*
FR - EN	*I was talking in quote not snobbery, not to brag (you couldn't brag more than rock music and with list of girlfriends you could have...*
ST:	*Şi s-a întâmplat, de asemenea, ca fata să aibă fire de walkman şerpuindu-i afară din urechi şi pierzându-se, subţiri şi duble, sub pânza sari-ului...*
RO - EN	*And it also happened that the girl had a Walkman strings winding out of her ears and disappearing, thin and double, beneath the sari...*
SRB - EN	*It also happened, that the girl had cables from the headphones of her Walkman which went in and out of the ears like a snake, thin but double, underneath the sari fabric…*
MD- EN	*And it is happened, also that the girl had the wires of Walkman, which were sneaking out of her ears and then were seeing even under the canvas of the sari...*
FR- EN	*it happened as well the girl to have the wires of the Walkman shaking them out from the ears and thin and double, under the jump suit...*

According to our investigation, subjects' mistranslations generated by applying unsuitable translation procedures are evenly distributed among the subjects, irrespective of their language background.

Further differences were recorded in relation to the other external criteria, i.e. language proficiency. In case of the 1st year subjects the occurrence of mistranslations is both more frequent and more seroius, considering the communicative purpose of the TT. Most of our 1st year students display translation difficulties at the syntactic level and in the management of culture-bound items. Since the theoretical framework of translation strategies and procedures is taught starting

13

with the 1st semester of the 2nd academic year, we could not assess their translation performance in terms of translation procedures adequacy. The students enrolled in the 2nd academic year showed a higher level of language proficiency both at linguistic and translation-related levels. Beyond the language background differences, some variations were recorded at the syntactic level. Furthermore, cognitive gaps (lack of knowledge) were registered in terms of translation procedures misuse. In both situations, a final conclusion would highlight a different involvement degree and individual study among 2nd year undergraduates.

e.g

ST: *Mi-e cu neputinţă să spun dacă era doar un obiect estetic lipsit cu desăvârşire de psihologie sau dacă, dimpotrivă, era numai psihologie, derealizată, proiecţie a privirilor fascinate ale celor din jur.*					
1st Academic year			2nd Academic year		
RO	MD	FR	RO	SRB	MD
I find it impossible to say whether it was just an aesthetic object completely lacking of psychology, or whether, on the contrary, it was only psychology, derealized, a projection of the fascinated looks of those around.	*I find it impossible to say whether it was just an aesthetic object completely lacking in psychology, or whether, on the contrary, it was only psychology, derealized, a projection of the fascinated looks of those around.*	*I can't say if it was an esthetical object with a complete psychology or if, against was psychology, unrealistically project of fascinate view of people*	*I couldn't tell if she was just an aesthetic object completely devoid of psychology, or on the contrary, if she was only psychology, derealization, a projection of the fascinated looks of those around her*	*I am helpless to say if it was only an aesthetic object completely devoid of psychology or if it was, on the contrary, only psychology, a derealization of one, projection of fascinated looks of those around.*	*It's impossible for me to say if it was an aesthetic object that was totally devoid of psychology or if, on the contrary, it was only psychology being not handled, a projection of people's excited eyes*

4. Computer-assisted analysis and results interpretation

To shed some light on the frequency rate of the mistranslation instances encountered following our qualitative analysis of the corpus, we resorted to dedicated software. Our quantitative analysis was carried out via MAXQDA 12 - Software for Qualitative and Mixed Methods Research.

First we imported the translated corpus into the software and organised the documents according to the previously established external criteria, i.e. subjects'

14

language background and L2 proficiency level. The next step was to encode the encountered instances of mistranslation according to the internal criteria set.

We selected the code list option and labelled each internal criterion as follows:

- ■ morpho-syntactic mistranslations - yellow
    - ■ at morphologic level - purple
    - ■ at syntactic level with orange
- ■ culture-related instances of mistranslation – green
- ■ misused translation procedures – blue

The codes were then organized on code sets according to the external criteria set (1st Academic Year/2nd Academic Year).

Selecting the Visual Tools option from the Tool Bar, we could generate, at a touch of a button, some code-based document portraits to compare the frequency of mistranslation instance in terms of L2 proficiency level or in relation to the language background of the students. Figures 1 and 2 illustrate the results obtained by applying a contrastive corpus investigation in terms of users' L2 proficiency level vs. internal criteria
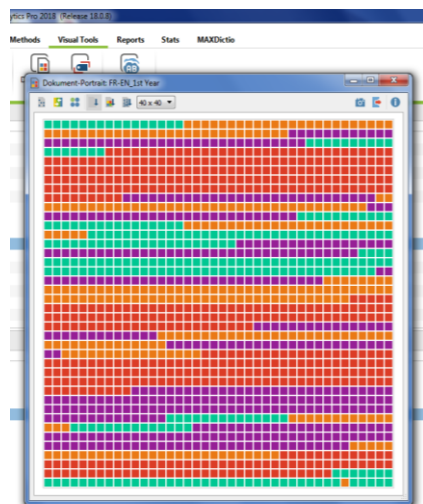

Figure 1 *Document Portrait – 1st year students' mistranslation frequency on internal criteria*
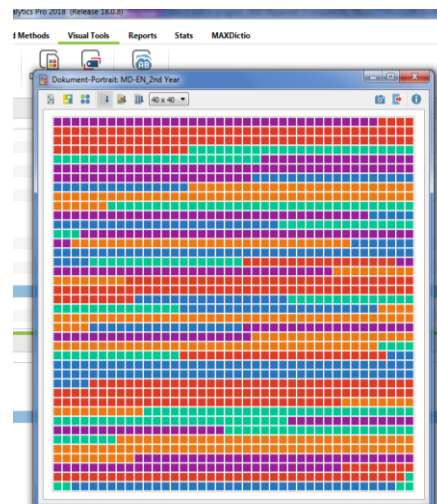

Figure 2 *Document Portrait – 2nd year students' mistranslation frequency on internal criteria*

Also, by activating simultaneously the document sets of our corpus and their corresponding code sets, the software generates a distribution chart of the codes in relation to both external criteria – Code-Matrix-Browser option in the Main Menu. Figure 3 indicates the overall results obtained via our corpus-based investigation.
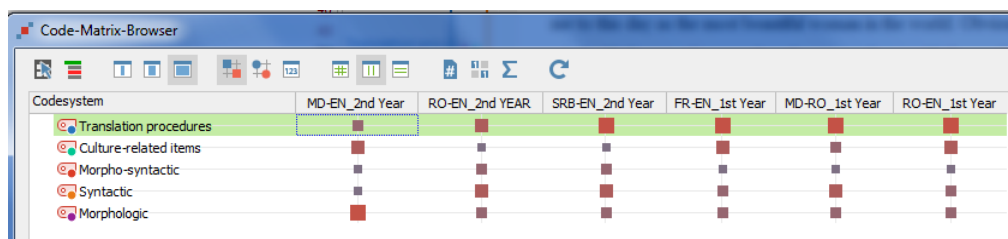
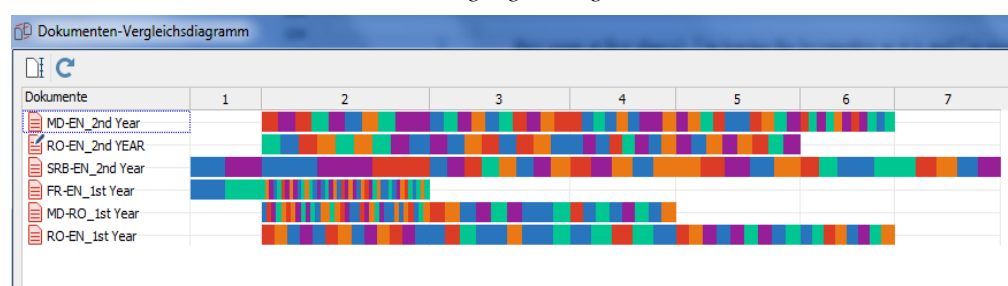Figure 3 *Code-Matrix-Browser – mistranslation distribution on L2 proficiency level and native-language background*

Figure 4 *Documents Comparison Chart – contrastive quantitative analysis in relation to on L2 proficiency level and native-language background*

Our computer-based corpus investigation validates the results obtained previously, namely that the most frequent instances of mistranslation are encountered in terms of morpho-syntactic choices, irrespective of the L2 proficiency level of the subjects (68%). At the syntactic level, a mild rising trend is registered among the 1st year students (39%), while the 2nd year students seem to have difficulty at the morphological level. Some differences have been recorded when comparing the frequency rate of morpho-syntactic mistranslation to the native language background of the students. Hence, Romanian students seem to have tense-related difficulties, while Serbian and Moldavian students show some difficulty in terms of word order (syntactic level) and the use of the article and other determiners at the morphological level. In contrast, for both 1st year and 2nd year Romanian students, the issue of word formation and derivation is still a challenge.

Culture-bound instances of mistranslation were more frequently encountered among the 1st year students 48% *vs*. 16% - in the 2nd year, who also experienced difficulty in the use of loan words and naturalization. Contrary to our expectations, Romanian students had some problems when dealing with the transfer of culture-bound items, 17% for the 1st year and 8% for the 2nd year.

Even though encoded for both groups of subjects, as previously mentioned, the occurrence of mistranslations due to poor knowledge of translation procedures was investigated only among 2nd year students. The findings indicate that the frequency rate of these instances is not influenced by the subjects' native language background.

5. Conclusion

Departing from the hypothesis that learner corpora contribute to the development of intercultural communication competence of students directly via translation tasks and indirectly as methodological tools to assess and improve students' language performance, the outcomes of our theoretical framework and the hands-on application are validated.

The added value of corpus-based investigation as integrated research and training method is secured by its cross-linguistic applicability, as LCR and CBTS target a mutual goal, i.e. purpose-oriented language performance improvement. In terms of corpora design and research developments, similar criteria and principles may be applied to the two main types of corpora (translation and comparable corpora), highlighting once again the joint effort towards L2 acquisition and improvement to meet the contemporary labour market demands.

As sustainable input, the Contrastive Interlanguage Analysis Method applied to Learner Corpora and Corpus-Based Translation Studies may contribute to an effective updating of academic curricula and syllabi, hence enriching the research community with the expertise gained. Such good practice examples would then apply and branch out to other academic disciplines connected to Translation Studies, such as pragmalinguistics, intercultural communication, cultural studies, teaching methodology, etc.

References

1. Aston, G., Bernadini, S., and Stewart, D. (eds.), *Corpora and Language Learners* (Studies in Corpus Linguistics. Amsterdam: John Benjamins, 2004.
2. Baker, M*., Corpora in Translation Studies. An Overview and Some Suggestions for Future Research.* Amsterdam and Philadelphia: John Benjamins, 1993.
3. Barlow, M., "Computer-based analysis of learner language", in Ellis R. and Barkhuizen G.P. *Analysing Learner Language*, Oxford: Oxford University Press, pp. 335–357, 2005.
4. Biber, D. and Reppen, R., "Comparing native and learner perspectives on English grammar: a study of complement clauses", in Granger, S. and Hung, J.(eds) *A bird's eye view of computer learner corpus research*. Oxford: Oxford University Press, pp. 145-158, 1998.
5. Chesterman A., *Contrastive Functional Analysis*. Amsterdam and Philadelphia: John Benjamins, 2007.
6. Fitzpatrick, E., *Language and computers, corpus linguistics beyond the word: Corpus Research from phrase to discourse*. Amsterdam, Netherlands: Rodopi, 2007.
7. Granger, S., "The International Corpus of Learner English", in Aarts, J., P. de Haan and N. Oostdijk (eds.) *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi, pp. 57-69, 1993.
8. Granger, S., "The computer learner corpus: a versatile new source of data for SLA research", in Granger, S. (ed.) *Learner English on Computer*. London and New York: Longman, pp. 3-18, 1998.
9. Granger, S., "A Bird's-eye View of Computer Learner Corpus Research", in Granger, S., Hung, J. and Petch-Tyson, S. (eds.) *Computer Learner Corpora*, Second

Language Acquisition and Foreign Language Teaching. Amsterdam and Philadelphia: John Benjamins, 2002.

10. Granger, S. and Lefer, M. A., "Bridging the gap between learner corpus research and translation studies: The Multilingual Student Translation corpus". Paper presented at the 4th Learner Corpus Research Conference LCR 2017, Bolzano Italy, https://www.researchgate.net/project/MUST-Multilingual-Student-Translation-corpus, 2017.

11. Gilquin, G., "The Integrated Contrastive Model: Spicing up your data", in *Languages in Contrast*, Vol. 3, Issue 1, pp. 95 –123, 2000.

12. Gilquin, G. and Paquot, P. 2008. "Too chatty: Learner academic writing and register variation", in Thompson P. (ed.) *Corpus-based EAP Pedagogy. Special issue of Journal of English for Academic Purposes*, 2008.

13. Keck, C., "Corpus linguistics and language teaching research: bridging the gap", https://www.researchgate.net/publication/237970784, in *Language Teaching Research*, 8 (1), pp.83-109, 2004.

14. Lozano, C. and Mendikoetxea, A.,"Postverbal subjects in L2 English: A corpus-based study", https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article/interface-conditions-on-postverbal-subjects-a-corpus-study-of-l2-english, in Bilingualism: Language and Cognition 13(4), pp, 475–497, 2010.

15. Mason, O. and Uzar R., "NLP meets TEFL: Tracing the zero article", in Lewandowska-Tomaszczyk, B. and Melia, J.P. (eds.) *PALC '99: Practical Applications in Language Corpora: Papers from the International Conference at the University of Lódz*, 15-18 April 1999, Frankfurt am Main: Peter Lang, pp.105-116, 2000.

16. Myles, F., *Interlanguage corpora and second language acquisition research*, http://journals.sagepub.com/doi/abs/10.1191/0267658305sr252oa?journalCode=slrb, in Second Language Research, 21(4), pp. 373-391, 2005.

17. Nesselhauf, N., *Collocations in a Learner Corpus*. Amsterdam: John Benjamins, 2005.

18. Schachter, J., and Celce-Murcia, M., "Some reservations concerning error analysis", in *TESOL Quarterly*, 11(4), pp. 441-451, 1977.

19. Sinclair, J., "Corpus and Text - Basic Principles", http://ota.ox.ac.uk/documents/creating/dlc/, Oxford: Oxbow Books, http://ota.ox.ac.uk/documents/creating/dlc/, in Wynne, M., (ed.) Developing Linguistic Corpora: a Guide to Good Practice, pp. 1-16, 2005.

20. Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm*, 4-8 August 1991, Berlin and New York: Mouton de Gruyter, pp.379-397, 1991.

21. Pravec, N., "Survey of learner corpora", https://www.researchgate.net/publication/247685282_Survey_of_Learner_Corpora, in *ICAME Journal*, 26, pp. 81-114, 2002.

22. Tono, Y., "Learner corpora: design, development and applications", in Archer, D. et al (eds.) *Proceedings of Corpus Linguistics 2003*, Lancaster University, pp. 800-809, 2003.

23. Vanderbauwhede, G., "The Integrated Contrastive Model evaluated: The French and Dutch demonstrative determiner in L1 and L2", https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1473-4192.2012.00321.x, in International Journal of Applied Linguistics, Vol. 22, Issue 3, pp. 392-413, 2012.

24. Wills, J., and Wills, D., *Collins COBUILD English Course*. London: Collins, 1989.