Exploitation of a Learner Corpus: Analysing Openings and Endings in Academic Forums

Universidad del Zulia
Maracaibo, Venezuela

Available in: http://www.redalyc.org/articulo.oa?id=31045571012
Exploitation of a Learner Corpus: Analysing Openings and Endings in Academic Forums*

Cristina Castillo Rodríguez, José María Díaz Lage

Universidad Internacional de La Rioja, España
cristina.castillo@unir.net, josem.diaz@unir.net

Abstract

The main aim of this paper is to analyse a learner corpus containing original written productions in the compulsory forums created in a concrete undergraduate subject. The reason for choosing the forum is that it is a space in which students can discuss in a more spontaneous way. The corpus compiled for this study will be managed with the corpus management software program AntConc. The study will be focused on the way these students have contributed to the forums but paying special attention to the expressions used in their openings and endings in written participations.

Keywords: Corpus linguistics, learner corpus, forum, opening expressions, ending expressions.

* This work has been carried out in the frame of the emerging research project “Análisis de patrones de errores gramaticales, estilísticos y de léxico en lengua inglesa a partir de producciones escritas de inmigrantes digitales”, (Universidad Internacional de La Rioja, 2014-2015).
Explotación de un corpus de estudiantes: análisis de saludos y despedidas en foros académicos

Resumen

El objetivo principal de este trabajo es analizar un corpus de estudiantes formado por textos escritos originalmente por estudiantes de grado en varios foros puntuables. La razón por la cual se han elegido los foros puntuables reside principalmente en que estos son espacios en los que los alumnos pueden debatir de una forma espontánea. El corpus compilado para este estudio se explotará por medio del programa de gestión de corpus llamado AntConc. El estudio se centrará en la forma en la que los alumnos participan en los foros con sus contribuciones prestando especial atención a las expresiones utilizadas en los saludos y despedidas.

Palabras clave: Lingüística de corpus, corpus de estudiantes, foro, expresiones de saludo, expresiones de despedida.

1. INTRODUCTION

The last few decades have witnessed an unprecedented increase in the use of information and communication technology in every area of everyday life. Tertiary education has not been alien to this trend, and one of the most noticeable fields of development in this realm has been e-learning.

Leaving aside the widespread adoption of online learning platforms¹ (or virtual learning environments, VLE), used to store course contents in combination with face-to-face lectures and seminars, which by now is almost an essential element of tertiary studies, the area of online learning is perhaps the best illustration of this tendency. The reasons for it are not only of economic efficiency, although this does certainly play a part and has been used by ICT companies to promote their educational products. Rather, the use of ICT, and specially Internet, for education has been presented as coherent with constructivist approaches to education (McMahon, 1997).
In this paper we will be concerning ourselves with one of the most crucial elements of such a methodology of learning: virtual forums, with the aim of analysing some expressions used by students in their openings and endings.

2. VIRTUAL FORUMS: MORE THAN A TOOL FOR COLLABORATIVE LEARNING

Virtual forums can be defined as virtual spaces set up for the purpose of asynchronous communication between the participants of a given course or programme. According to Arbey (2009:30), a virtual forum is “una comunicación grupal, dialógica, temática, asincrónica y argumentativa”. Forums have been seen as valid learning strategies (Arango, 2003) and as a way of developing critical thought (Fedorov, 2007). It goes without saying that for this to be true critical thought must be defined in a rather lax manner (Arango, 2003, and Fedorov).

Nevertheless, if we do define this thus, then certainly virtual forums can foster the development of the cognitive and attitudinal dimensions of critical thought.

Teaching practice seems to suggest that virtual forums serve more than one purpose. There is, first of all, a relevant distinction to be made between forums which are assessed as part of the students’ final grade and forums which are intended for communication with the tutor and operate therefore as a space for comments, inquiries and troubleshooting; but there are also several functional dimensions at work in one of these virtual forums. The most obvious one is, of course, offering a space where questions may be posted and answers obtained, although simultaneously virtual forums provide a sense of community and dialogue which can otherwise be missing from online learning. This contributes to a kind of collaborative learning (also called c-learning) that is currently perceived as highly desirable: the students’ interactions among themselves increase their involvement with the contents of a concrete subject or module.

An important variable that has to do with these issues is the presence or absence of a moderator, generally the faculty member in charge of the virtual forum: whereas forums intended for the resolution of doubts obviously require a lecturer to take part in the discussion, assessable forums are generally devoid of this presence, aside from an introductory message explaining their objectives and regulations. Especially
in the cases where there is no moderator, the very communicative function of virtual forums requires participants to share a common discourse or, at the very least, a common set of assumptions about what constitutes proper etiquette in online conversations. As Ackerley points out:

Participants in a discourse community should not only be able to talk about themselves and their common interests, but also be able to produce the right kind of language to be identified as members. Not only may learners struggle to produce correct English, they may also lack an awareness of exactly how to write (or speak) in a discourse community which has specific norms, expectations and conventions (2013: 1-2).

In social forums, i.e. in those designed to foster a sense of community, one of the most frequent shortcomings is the participants’ inability or unwillingness to engage with their fellow students, so that each contribution is detached from the others (see García et al. 2010). However, this potential problem is not limited to social forums: as we will see, assessable forums can suffer from the same. In those cases, failure to engage in a dialogue with the other participants is usually associated to below par contributions. This is not without the importance for our topic –even though it falls beyond the scope of the present paper–, as it might prove enlightening to search for a possible relationship between unsatisfactory contributions and breaches of the accepted etiquette of online conversations.

Having said this, it must be acknowledged that there is no universal agreement as to what constitutes proper etiquette in conversations conducted via a virtual forum. One example will have to suffice. García et al. (2010) argue that, in the virtual forum that they analyse, “es obvio que los alumnos se vieron influidos por otros géneros digitales de comunicación interpersonal, como el correo electrónico (Ancarno, 2005), al introducir en sus intervenciones fórmulas de saludo, de despedida y firma”. This statement seems to suggest that, for the authors, there is not or should not be place in forums for greetings (openings) and goodbyes (endings). For the purposes of this paper, on the contrary, openings and endings are a valid part of the students’ contributions, and they constitute an important area of inquiry. It is worth pointing out that the object of our study is of a slightly more formal nature than that of Garcia et al. (2010), which may well account for the discrepancy.
Indeed, our study is concerned with academic forums, i.e., forums which aim at the exchange of thoughts and ideas regarding the topic at hand (Arango, 2003). Furthermore, the academic forums with which we are dealing are assessable forums where each student’s contribution will be awarded a mark that will be added to his/her final grade. This suggests from the outset that pragmatic dialogue will not be as relevant as in other, non-assessable types of academic forums, as for many of the students the perspective of a grade will inevitably introduce a makeshift approach to participation in the discussion. Taken as a whole, these forums will display a typical pattern: a dialogue between contributors who engage with each other’s posts, punctuated by a large number of unconnected, perfunctory messages that seem to exist in a vaccum.

This in turn seems to contradict a point which several authors make: that the asynchronous nature of virtual forums favours thoughtfulness and reflection. Thus, for instance, as Shana (2009: 221) states: “since students know that others will read their posts, they often take their time, and put significant thought and effort into them before posting. The web based exchange of ideas results in higher quality work than if students were left on their own”. García et al. (2010) recognise that the use of forums, as asynchronous tools, fosters in students a deeper reflection than other synchronous tools, for example, chats. That also coincides with the view of Arango (2003), who defends that it is the asynchronous feature of forums that promotes reflection in the content of messages.

Leaving aside the benefits of discussion forums for the participants, the posts contained therein constitute an adequate source for the compilation of a learner corpus, the analysis and study of which would give us important insights into the most salient mistakes made by students. This in turn would provide us with the opportunity to design strategies for the amelioration of those mistakes, including the development of specifically designed pedagogical materials. The use of learner corpora for the improvement of L2 students is of course nothing new (see Granger, 1994), but the use of learner corpora as a classroom resource is a crucial area that has not always been given the attention it deserves (Granger, 2009). Work in this field will doubtlessly yield valuable results in the foreseeable future.

This paper focuses on parting greetings or the absence thereof in the contributions to a learner corpus compiled from assessable, aca-
Conversational routines are difficult to teach because they are culture-bound and because their formal description and their situational frames are extremely complex (Yorio 1980). Learners of the language are usually not aware that there are rules governing these expressions but use phrases in a way which deviates from the native norm or which causes communicative misunderstandings (1996: 30).

The mistakes to which EFL students are liable when greeting their peers in virtual forums are many and go beyond grammatical or orthographic errors: these are of course possible, but the most salient ones are probably cases of non-idiomatic usage, including the likely possibility of linguistic interference between L1 and L2. As Ackerley (2013: 2) argues: “there are numerous reasons why learner language may be considered inappropriate to a genre (apart from the presence of errors)”. The author also quotes Stubbs in the sense that the latter claims that “words rather than phrases are typically considered the traditional units of language. Students tend to piece them together, often making direct translations from their L1” (2002: 215).

The point made by Stubbs (2002) merits careful consideration, as it seems to suggest that there are two stumbling blocks which we must take into account in our study: first, the possible errors derived from a deficient knowledge of the L2 might make it difficult to achieve a systematic overview of all the parting greeting formulae that the contributions employ, although the small size of our corpus renders this somewhat less problematic. Second, and this is the point that we specifically owe to Stubbs, if there is a deep-seated tendency to consider words rather than phrases the basic units of language, this will make the contributors struggle more with
greetings, and it will make it more difficult for our study to isolate the possible greeting problems they face in their virtual forum contributions.

All in all, the use of virtual forums in academic contexts is widely seen as beneficial but it is dependent upon a set of conditions, among which we may find the need for contributors to respect the conventions of the genre of discourse of which they are partaking. As already mentioned, we assume that greetings are part and parcel of those conventions; and, because they are subject to the specific features of formulaic expressions, greetings require an approach that takes their specificity into account.

The compilation and analysis of a learner corpus seems an ideal procedure to address the above points. Not only can it yield relevant results which may enhance our understanding of the problems that EFL students face in their virtual forum interactions: it can also suggest precise measures that can be taken to address those problems, most importantly in the shape of the design of specific teaching materials conceived to palliate those very shortcomings. In other words, the study of a learner corpus can be an outstanding tool for the diagnosis of the errors to which we should subsequently turn our attention.

3. METHODOLOGY

The corpus object of our study has been compiled from all the contributions found in the two compulsory forums proposed for the undergraduate subject ‘ICT tools applied to the learning of English language”, of the degree in Early Years Education.

A total of 201 participations have been compiled during two academic years, that is, 2013-2014 and 2014-2015. These contributions belong to 129 students studying this undergraduate subject.

The process for compiling the contributions has been the following. First of all, we downloaded every contribution and saved them in txt format according to a specific coding protocol. For instance, let us describe item 16TOENICTA: 16 corresponds to student number 16; TO means ‘texto original’, that is to say, source text; EN means English, according to codes proposed in ISO 3166 –alpha 2 for naming countries and their subdivisions; ICT is the abbreviation of Information and Communications Technologies, used for indicating the subject involved in this concrete corpus study; and A corresponds to the first compulsory forum of the aforementioned subject. According to this protocol, for coding the items belonging to the second
compulsory forum of the subject, we will follow the same system of numbers and abbreviations, except for the last letter, which will be B instead of A. For example, student 16 had also participated in the second forum, so the code for that will be 16TOENICTB.

Some special (and very few) cases of students retaking the subject—because they might have failed in a previous period or just because they could not attend the exam—also occur in the corpus. That is the case, for example, of student 20, who only participated in the academic year 2013-2014 in the second compulsory forum, and did the same in the next academic year 2014-2015. For this case, the coding protocol has been the following: 20TOENICTB1 and 20TOENICTB2. The distinguishing part of the code is the last number, i.e., 1 corresponds to the first participation in the first academic year, and 2 for the second contribution in the same forum the next academic year.

As part of the methodology used for this study with corpus, all the information related to the items of the corpus was registered in an excel file. The information gathered per item has been divided into seven columns:
2. Number of student.
3. Number of contributions.
4. Degree.
5. Subject.
6. Academic year.
7. Words in the ST.

The following figure shows the seven columns fulfilled with the corresponding information for each case:

![Figure 1. Information about the learner corpus](image-url)
As the texts of the corpus were directly compiled from the contributions provided by the students, we did not have to ‘clean’ the texts, as it usually happens when compiling texts from the Internet or when converting from PDF to TXT, for example (Castillo, 2009).

In the previous figure we have observed that some students have participated with more contributions than others. In fact, the instructions for every forum indicated a minimum participation of two contributions (in the case of forum 1) and three contributions (in the case of forum 2).

With the aim of separating the content of every contribution we have included in the txt some distinguishing sequences: *Contribution1*; *Contribution2*; *Contribution3*; and so on.

On the other hand, a final procedure for completing the methodology and before starting analysing the whole corpus has been the elimination of the real name of the student who had contributed to the forums either for introducing him/herself or at the end of the contribution. We have substituted once again every case with some corresponding sequences, as we list below:

- *Studentname*: when the student only used his/her name.
- *Studentname-capitalletter*: when the name was written in capital letter totally.
- *Studentname-diminutivo*: if the student used a shorter version of his/her name (pet name).
- *Studentnameandsurname*: when the student added his/her first surname.
- *Studentnameand1stlettersurname*: similar to the previous case but by using only the first letter of the surname instead of using the complete form of the latter.
- *Studentnameandsurnames*: if the student used his/her two surnames.

Once we have described the methodology in detail, we will focus on the results thrown after analysing the corpus with the corpus management software program AntConc.
4. RESULTS

In this section we will offer an analysis of our corpus focusing concretely on the opening and ending expressions used by these students in spontaneous written productions encountered in the forums of the subject.

The software used for this analysis has been AntConc, which is a corpus management program that allows the user to integrate and subsequently exploit a corpus in a very straightforward way. Its simple interface allows the user to add all the files of a concrete corpus and to navigate through the different tools offered by this program. Figure 2 shows the simple interface of AntConc with the files of our corpus integrated and ready to be exploited.

![Figure 2. Learner corpus files loaded in AntConc](image)

Once the corpus has been integrated we can start exploiting it and analysing the concordances according to the purposes we pursue in this study.

For our first search, we have typed the word ‘studentname*’ so that the system can throw not only all the ‘studentname’ cases, but also the
rest of the cases as the symbol [*] in this type of corpus management software is understood as ‘all the letters following the typed word’. If we sort the instances thrown 1L and 2L we will obtain the type of endings students have used for their contributions. Figure 3 shows some of these examples of endings.

Figure 3. Some examples of ending concordances in AntConc

Let us first analyse the endings by observing the concordances thrown on the left. There have been a total of 258 concordances for the aforementioned typed word\(^6\). From all these hits, the types of endings found in the corpus can be classified as we list below\(^7\):

1. Formal ending (147 hits).
   - [r/R]egards[./nothing]: 70 hits
   - [k/K]ind regards[./nothing]: 48 hits
   - Best regards[./.]: 20 hits
   - Best wishes[./.]: 9 hits
   - Best;: 5 hits
   - Sincerely [./nothing]: 5 hits
2. Expressing thanks (29 hits).

- Thank you very much[./nothing]: 5 hits
- [t/T]hanks[!./]: 5 hits
- Thanks (so much) for your attention [./nothing]: 4 hits
- Thanks[./!/nothing]: 4 hits
- Thank you[././nothing]: 4 hits
- Thanks Bye: 1 hit
- Thank you all,: 1 hit
- Thank you everybody for the contributions and for reading mine.: 1 hit
- Thanks for all your opinions.: 1 hit
- Thanks for your contribution girls,: 1 hit
- Thanks for reading,: 1 hit
- Thanks all,: 1 hit

3. Expressing standard greetings (4 hits).

- Greetings[./,]: 2 hits
- Greetings to all.: 1 hit
- A greeting: 1 hit

4. Wishing something to the receivers (15 hits).

- Have a nice day[!/,]: 4 hits
- Have a nice evening[!/,]: 2 hits
- Have a nice Mother’s Day!: 1 hit
- With a final expression using look forward to: 2 hits
- Take care: 1 hit
- Enjoy,: 1 hit
- Enjoy your weekend,: 1 hit
- Regards and smiles from [place],: 1 hit
- Smiles and regards,: 1 hit
- I hope you find them useful!: 1 hit
5. Informal endings (8 hits).
- Good bye,: 2 hits
- See you soon [!/,:]: 2 hits
- See you soon. Bye: 1 hit
- Talk to you soon!: 1 hit
- Cheers[./nothing]: 2 hits

6. Sense mistakes in endings (3 hits).
- Kind rewards,: 2 hits
- Mindfully: 1 hit

7) No endings: 41 hits.

As we can see from these results, the most frequent ending expressions in our learner corpus have been the expressions collected in formal endings and expressing thanks. In the second section of endings, there has been a wide variety of thanks expressions as contribution endings, so hits have been more or less distributed among the aforementioned options.

The most outstanding use of ending has been within the framework of the first section, the most frequent ones being Regards, Kind regards, and Best regards –the total of hits have been 128–, all of them with their different variations (i.e. followed by commas or full stops or without punctuation). As seen with these results, we can state that our students have preferred a more formal ending rather than an informal one.

Nevertheless, opposite to these results, we have also gathered a great number of hits (41) in which no ending expressions have been provided by students in their contributions. With regard to the rest of endings, the results indicated before have not been quite significant if we compare them to the hits encountered in the corpus related to the first section expressions.

Openings in our learner corpus have also been varied, according to what we have observed in the concordances. For searching for the opening examples, we have typed ‘contribution*’ and then we have sorted the results using two other levels: 2R and 3R. Once the software has sorted all the concordances (607 hits) the results thrown and analysed have been the following:
1. Standard opening with dear + noun/pronoun (49 hits).
   - Dear [A/a]ll[./!/]: 29 hits
   - Dear studentname,: 11 hits
   - Dear classmates,: 3 hits
   - Dear ICT mates,: 2 hits
   - Dear colleag[u]es,9: 2 hits
   - Dear partners!: 1 hit
   - Dear all again!: 1 hit

2. Good + noun (128 hits).
   - Good morning[./!//:/nothing]: 56 hits
   - Good [A/a]fternoon[./!//:/nothing]: 52 hits
   - Good evening[./!/nothing]: 12 hits
   - Good night[./nothing]: 8 hits

3. Informal opening with Hello, Hi and Hey (231 hits).
   - Hello[./!/]: 51 hits
   - Hello again[!//:/nothing]: 22 hits
   - Hello everybody[./!/.../nothing]: 24 hits
   - Hello everyone[./!/]: 29 hits
   - Hello + studentname[./!//:]: 16 hits
   - Hi[./!//:/nothing]: 15 hits
   - Hi again[!/./everybody,/nothing]: 11 hits
   - Hi +studentname[!/./nothing]: 17 hits
   - Hi all[!/./nothing]: 3 hits
   - Hi everybody[!/./]: 10 hits
   - Hi everyone[./!/./]: 27 hits
   - Hi there!: 3 hits
   - Hi good afternoon: 1 hit
   - Hi partners::: 1 hit
   - Hey everybody,: 1 hit
Apart from these results, we have collected some other hits (136) concerning no openings at all in the messages posted by students in the forums.

The results thrown by AntConc have revealed to us that there is a high tendency to use a more informal opening while the tendency for endings is a more formal expression, according to the number of hits of these groups. On the other hand, taking into account the faculty member’s openings and endings in those forums, that is, *Dear all*, and *Kind regards*, respectively, we can check that there is also a slight and subtle tendency of linguistic imitation in students. Even as far as punctuation is concerned, if we analyse these two types of hits in detail, we can see that in the opening the use of comma is preferred (*Dear all*, has 27 hits; *Dear all.* 1 hit, and *Dear all!* 1 hit), and the same occurs with the ending, that is to say, comma is the punctuation most frequently used by students (*Kind regards*, 35 hits, *Kind regards.* 7 hits, and with no punctuation after the expression there was a total result of 6 hits).

With regard to the participation of the students interested in the forum, some of them have contributed with more valuable-content messages showing a great interest in the debate established for the different forums. In the instructions of the forums, it was indicated that students had to participate twice in the first forum, and three times in the second one. According to these requirements, out of 103 students participating in the first forum, 55 students have strictly respected the instructions of the first forum, that is, with a total of two contributions (around 53%). On the other hand, 37 students (almost 36%) have participated more than twice in the first forum; however, the number of contributions of this group of students interested in fostering the debate of the first forum has been varied: 21 have contributed three times, 11 have contributed four times, 3 have contributed five times, and 2 have contributed six times. There has been very little participation of students contributing less than twice, i.e., only once: 11 students (around 10% out of the total from the first forum).

As far as the second forum is concerned, out of 98 students who participated in the forum, 68 students (around 70%) have participated with three contributions, which was the number of contributions indicated in the instructions. Less numerous were the participants with more contributions than required in this second forum (15 students, or around 15% of the participants) distributed as follows: 10 students with four
contributions, 3 students with five contributions, and 2 students with a total of 6 contributions. A similar percentage has been found for those participating with less than three contributions (around 15%): 11 students have contributed twice, and 4 students have contributed only once.

5. FINAL REMARKS AND FUTURE CONSIDERATIONS

In this paper we have analysed a very concrete aspect of the texts contained in a learner corpus compiled from the forums participation in an undergraduate subject. We refer to the use of openings and endings in this type of academic texts so as to check which types of expressions were the most frequent ones and the tendency followed by students.

Generally speaking, we have observed in this paper that student participation showed a high number of hits concerning the use of formal endings, while the opening expressions preferred by students were more informal. We suspect that the use of these expressions might be due to a desire of closeness with their partners, in the case of the openings; however, the tendency of endings has been towards the use of a formal expression. Besides, with regard to punctuation, the tendency of students has been the imitation of their faculty members in charge of the subject and, therefore, in charge of starting the debate with a post. That is to say, the most frequent punctuation following certain openings and endings has been the use of commas.

As future lines, we would want to undertake some experiments so as to check whether or not imitation is a practice found in some other linguistic aspects, especially choice of vocabulary but also use of idioms. That will undoubtedly guide us to conclude with some interesting outcomes in linguistic behaviour from our undergraduate ad postgraduate students.

Notes

1. For an overview of learning platforms, please see Graziadei et al. (1997).
2. Regarding the different types of forums we recommend the publications of Arango (2003), Álvarez and Morán (2010), and García et al. (2010).
3. In fact, Owen, Grant, Sayers and Facer (2006: 10) state that c-learning can refer to: “community learning, communicative learning
or collaborative learning”. For an overview of the tools that could be used for fostering c-learning, see Castillo and Ríos (2013).

4. Needless to say, Aijmer’s work (1996) focuses on spoken conversations rather on written exchanges. This difference should be taken into account when considering the present discussion: a virtual forum can certainly be conceptualised as a conversation, but its features are very different from those of a spoken conversation.

5. In order to keep privacy, every student has been assigned a number instead of using their real names.

6. There have been 10 other hits not corresponding to endings, but to openings, as students sometimes introduced themselves or named other partners.

7. In brackets we have included the possible options found in the corpus. That is to say, if a word was used with upper case or lower case, and if they were followed by commas, full stops or there was no punctuation after these endings.

8. Once again there have been other 69 hits not corresponding to openings, as the word ‘contribution’ corresponded to parts of the content of our students’ messages.

9. This item also shows a mistake because there was a hit which lacked the use of the letter ‘u’.

References


Exploitation of a Learner Corpus: Analysing Openings and Endings in Academic Forums


