

5th International Conference on Corpus Linguistics (CILC2013)

The Compilation of a Corpus of Business English: Syntactic Variation

María Luisa Carrió-Pastor^{a*}, Rut Muñiz-Calderón^b

^a*Universitat Politècnica de València, Departamento de Lingüística Aplicada, Camino de Vera, 14, 46022 Valencia, Spain*

^b*Universidad Católica de Valencia, Valencia, Facultad de ADE y Derecho, C/ Jorge Juan, 18, 46004 Valencia, Spain*

Abstract

This study focuses on the design of a real corpus extracted from a business environment. The compilation of a real corpus must be done in an objective way to extract samples of language used in an everyday context and with the minimum interference. The elaboration of a corpus based on texts from a business environment is somewhat problematic as companies are not used to providing information for linguistic research and the appropriate tagging of a real corpus is not an easy task. The tagging of a real corpus involves filtering the language selected, as it may contain some mistakes or variation. The objective of this paper is to describe the methodology followed in order to compile a real corpus and propose a tagging system of the syntactic variations found in a real corpus caused by the use of English as a second language. This proposal is made after the compilation of a corpus composed of one hundred and twenty e-mails written by Indian and Chinese employees who work in an international company. The corpus was tagged manually and several aspects were taken into account although in this paper we will focus on the tagging of variation.

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of CILC2013.

Keywords: variations; corpus; English as a second language

* Corresponding author. Tel.: +34963877530; fax: +34963877539.
E-mail address: lcarrio@idm.upv.es

1. Introduction

Nowadays, English is a global language used by millions of people in very different contexts. Due to this massive use of English as a lingua franca varieties of English in South East Asia, in Africa, in Europe, etc., have emerged. These different varieties have appeared over the years due to the coexistence of English alongside the local languages in multicultural contexts. These varieties show different linguistic and socio-cultural features which reflect their speakers' cultural background and should be taken into account in order to understand the evolution of English in particular and of language in general.

In this paper, we examine the issue of how to treat the variation found when tagging a corpus. Variation involves choosing a particular way to express reality in a foreign language that could be considered as peculiar by native speakers but is linguistically correct. The message is understood, but if native speakers are the addressees, they may view the manner of communication to be different. This happens not only when people with different language backgrounds communicate, but also when people from different parts of the same country explain or expound an idea using expressions that may be not common in other areas. We consider variation to comprise the different ways of communicating in one language where this does not interfere with the exchange of ideas.

Turning to the research on linguistic variation, it seems that the interest of some researchers has mainly centred on investigating variation through the association of particular discursive features with different linguistic backgrounds (Aijmer, 2001; Yli-Jokipii and Jorgensen, 2004; Schlee, 2009; Hyland, 2011). The main aim of these analyses is to describe, through contrastive rhetoric, differences in discourse patterns that sometimes operate as a barrier to effective communication.

We should also take into account that international business discourse, as with any type of discourse, is culturally-situated and therefore context-dependent, and that discourse, culture and context all play a key role in the communication process (Tognini-Bonelli, 2001; Bargiela-Chiapini, 2004: 31-34). International business communication in English is a field of research in which intercultural communication takes place between speakers of different nationalities. In the world of business and trade, the way the participants use English in intercultural communication has its own features that cannot be detached from their cultural background. In the case of business e-mails, some researchers such as Nickerson (1999) and Gimenez (2006) have shown that this kind of communication possesses specific characteristics in a business context. In this paper, our main purpose is to describe the tagging of linguistic variation that can be observed in e-mails written by non-native speakers of English. We focus our research on business English produced by Chinese and Indian writers in order to analyse the variations and the way these linguistic variations should be tagged.

As some researchers have pointed out (Gimenez, 2000; Rogerson-Revell, 2007; Cheung, 2011; Carrió and Muñiz, 2012; Weninger and Kan, 2013), language use tends to vary depending on the cultural, linguistic or social background of the speaker. Several researchers have observed that China and India make up the largest English speaking population in the world (Jiang, 2002; Jenkins, 2003; Bolton, 2003; Crystal, 2001; 2008) and this is the reason why some studies have focused on business communication in Asian countries (Bargiela-Chiapini and Gotti, 2005; Adamson, 2006; Salvi and Tanaka, 2011; Weninger and Kan, 2013). The above authors have shown in their studies that the obvious cultural differences between Asian countries and countries such as the USA and the UK is also reflected in how English is used there.

The language used in business e-mails has been studied by researchers such as Barson, Frommer & Schwartz (1993), Warshauer (1995), Giménez (2000; 2006), Biesenbach-Lucas (2005) and Evans (2010, 2012). Although there are standard guidelines in this genre, variation exists in the linguistic characteristics of e-mails produced by speakers of different cultural backgrounds and this variation has not been widely studied. Mainly the characteristics of the language of e-mails have been the centre of the study of several researchers but the variation caused by communication produced by native and non-native speakers has not been one of the most popular issues to study. In this paper, we believe that variations in international communication such as e-mails should be detected, tagged and classified to observe the synchronic evolution of language in a global world.

In this study, we focus on the tagging of the syntactic variation of the English language when used by Chinese and Indian writers in order to analyse the evolution of the language and the internal mechanisms that make a language change until a linguistic variety appears. In the same line, the main hypothesis of this paper is that variation can be found in English when spoken by non-native speakers of English. The objectives of this paper are

first, to describe the methodology followed in order to compile a real corpus and, second, to propose a tagging system of the syntactic variations that may be found in a real corpus produced by the use of English as a second language.

2. Methodology

The present study describes the process of elaborating a corpus of e-mails from a business context. Therefore, in order to extract our results, we analysed a corpus of one hundred and twenty e-mails written by non-native English speakers with different linguistic backgrounds; sixty e-mails were written by Indian writers and sixty e-mails were written by Chinese writers. In this way, the first group of writers speak English as a second language because they live in a former British colony and the second group speak English as a foreign language, having had to learn the language *ab initio* at school. These e-mails were written during 2009 to 2011 by the employees of a Spanish company which exports its goods across the world. The total amount of words included in the corpus analysed was 13,320. The e-mails written by Indian employees included a total of 6780 words and the emails written by Chinese employees comprised 6540 words.

2.1. Materials

The Spanish company is located on the eastern Mediterranean coast of Spain, in Valencia. The enterprise specializes in the manufacture of high technology laser machines, which are used for the finishing of different fabrics. Moreover, one of the main activities of this enterprise is selling and exporting this equipment worldwide and it is considered to be one of the market leaders. The global trading situation has caused many American and European manufacturing companies to outsource production to the Southeast of Asia. Therefore, this company's major customers are in this part of the world, with Indian and Chinese buyers chief among them.

English is the language used for all the business interactions and the most usual internal and external medium of communication is e-mail. The use of English is compulsory among all interlocutors even those who share the same mother tongue. The e-mails compiled in the corpus were usually sent or received by the Sales Department and all the participants were non-native speakers from India and China.

The English language proficiency of the employees is high (B2 or higher, according to the levels of the Common European Framework) which is a compulsory condition for the personnel recruited. Indian employees have a good command of English as this country was a former British colony and belongs to the outer circle (Kachru, 2005). However it is difficult to find proficient Chinese employees due to the fact that English in China is learnt as a second or third language and it does not have the institutional presence it possesses in India, as it is part of the expanding circle (Kachru, 2005).

2.2. Method

When elaborating the corpus, we considered using a computer corpus analysis tool to detect the variations, but finally we considered that a manual analysis should be carried out in order to label the different variations detected. Thus, the raters detected the variations and the context, as well as carrying out the exact tagging of the variation.

The analysis had several steps; firstly, all the e-mails were carefully read and were classified into the two different groups mentioned above, according to the writers' origins. Secondly, the e-mails were numbered in order to classify them and find the occurrences easily. Thirdly, proper names were deleted and the name of the company was changed to a different name, *Laser Company*, for confidential reasons as this was requested by the company. Finally, the raters detected variations and tagged them in the corpus, inserting them into a spreadsheet. Figure 1 shows an example of an e-mail from an Indian employee:

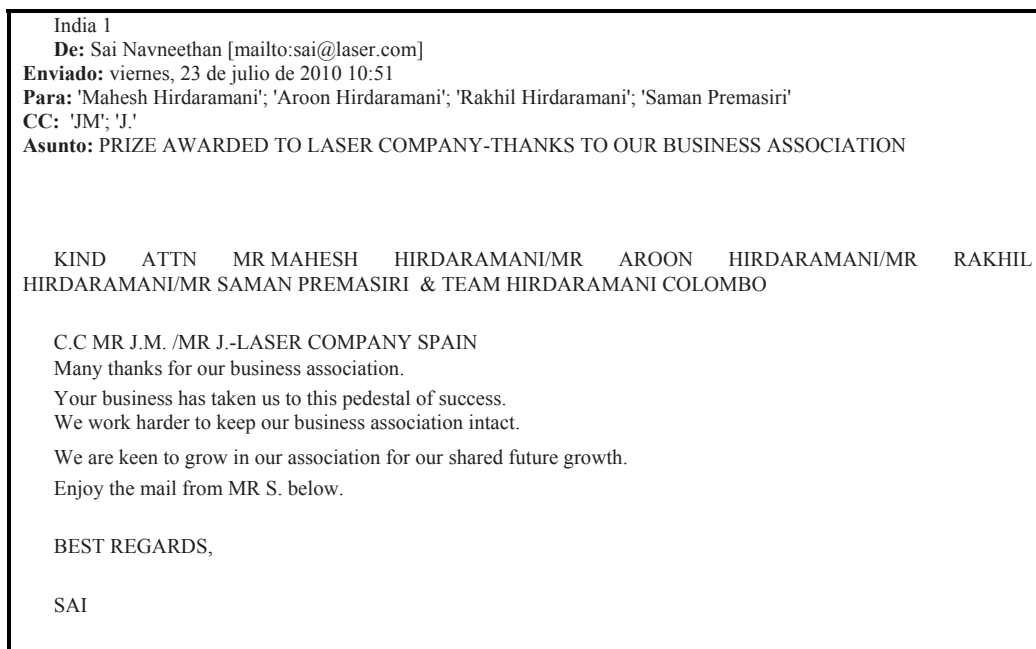


Figure 1. Example of an e-mail written by an Indian employee.

In order to carry out this analysis, some patterns were designed so the variations could be identified. These patterns were based on the syntactic characteristics of the English varieties studied by some linguists such as Bolton (2003), Kachru (2005), Kirpatrick (2007) and Sailaja (2009). In this way, in order to tag the corpus and the variations found, a tagging system was designed. The symbol * was manually placed next to the tag of the occurrence of variations to differentiate them from the standard use of the syntactic pattern. Then, the results were collated after raters had performed the manual tagging. The raters were four English language teachers, who were not native English speakers but were experts in grammar. The tagging system was designed to help raters identify the occurrences and it was based on the syntactic patterns that we observed showed variation when they were used by non-native English speakers. The tagging was divided into several syntactic patterns following the classification of errors proposed by James (1998) and Dagneaux, Dennes and Granger (1998).

The results were placed into tables according to the mother tongue of the authors and finally, we extracted the conclusions of this research. This study focused on the tagging of corpora of business e-mails written by speakers with different mother tongues but who use English to communicate. The tagging we show in this paper is just one part of a larger research study, which is focused on the analysis of the variation in different aspects of communication among non-native English speakers.

3. Results

Figures 2 and 3 show the two different tagging systems designed to detect the syntactic variation found in the corpus:

TAGGING OF SYNTACTIC VARIATIONS
Use of articles: definite<DA*>, indefinite<INDA*>
Use of pronouns: personal <PERP*>, possessive<POSP*>, demonstrative<DMP*>
Use of the verb tenses: present simple<PRS*>, present continuous <PRCONT*>, present perfect<PRPER*>, past continuous<PCONT*>, past simple<PAS*>, past perfect<PAP*>, future simple<FS*>, future perfect<FP*>

Use of adverbs: <ADV*>
Use of modal verbs: <MV*>
Use of passive voice: <PSV*>
Use of prepositions: <PP*>
Use of complex phrases: <NN*>, <NNN*>, <NNNN*>, <ADJN*>, <ADJNN*>, etc.
Use of connectors: <CN*>
Sentence structure: <SST*>

Fig. 2. Tagging of syntactic variation.

Raters used this tagging system to detect and classify the variation found in the e-mails analysed. Examples [1], [2], [3] and [4] have been taken from the two groups of e-mails analysed:

[Example 1]. Tagging of the variation of complex phrases found in the Indian e-mails:

“Daughter wedding <NN*> will be celebrated next month...”

“Meeting time today<NNN*> is at seven in the meeting room”.

“Kind courtesy<ADJN*> to receive the visitors from Bombay”.

“Today we have a small dinner meeting<ADJNN*>to discuss the agreement with the Chinese counterparts”.

[Example 2]. Tagging of the variation of complex phrases found in the Chinese e-mails:

“Chinese New Year Holiday<NNNN*> is celebrated in the headquarters of the company”.

“China factory <NN*> will be visited by the Managers”.

“Hong Kong China clients<NNN*> are welcome in the headquarters of Laser Company”.

[Example 3]. Tagging of the variation found in the Indian e-mails:

“We are valuing <PRCONT*> our business association”

“We are keen to meet again and develop an<INDA*> opportunity given”

“Happy new year for business and it will be given <PSV*>family prosperity in 2010 and life too”

“Must <MV*>honestly complement for the store, music and ambience was world class<SST*>”

“We would be delighted <PSV*>to have your presence”

“We are fortunate to be associated with your esteemed group<ADJN*> in business<SST*>”

[Example 4]. Tagging of the variation found in the Chinese e-mails:

“We want all the parts sent us with the next Big M. shipment together<SST*>”.

“We have no room for Big m., it’s not make <PSV*>sense to have room for production”.

“Ok, we have tried everything now<ADV*>. I think it’s time to get some one here R., you are already in China, come to NINGBO and sort this out for us<PP*>”

“Need assistance urgently to find out why!!!!<SST*>”

“I <PERP*>my concept that is clear of matter who closes the deal that is NOT important for us<SST*>. The important is closed<PSV*>the deal”

“The visit of your technician or sales team must be free<MV*>.”

As can be seen in the examples provided here, the variations found in the use of English as a lingua franca show differences according to the group of writers analysed. The tagging system used demonstrates that speakers of English as a second language and speakers of English a foreign language employ the language in quite different ways. If we observe the use of language by the two groups in the whole corpus, Chinese writers use more direct language and Indian writers use more ornamented language. In this paper, some examples are shown, although sentences such as “We are fortunate to be associated with your esteemed group in business” are frequently found in

the corpus of e-mails written by Indian. A detailed explanation of these aspects has not been taken into account in this analysis, but we were able to make this observation while tagging complex phrases composed of adjectives not usually associated to certain nouns. In this sense, non-native speakers of English vary the foreign language they use to communicate in an international context, changing the patterns of the target language to adapt them to the source language parameters.

4. Conclusions

In this study, our aim was to show the need to tag not only the usual patterns or errors of language, but also the examples of variation which may exist, which perhaps may later lead to the establishment of different varieties of a language used internationally. The results shown here are a sample of the tagging system used to detect and classify variation found in the English language when used by Chinese and Indian writers, but our long term purpose is to tag the examples of variation in texts written by speakers of different linguistic backgrounds, and to analyse the evolution of language and the internal mechanisms that make a language change until a new variety of it appears. In the results of this paper we identified and tagged the syntactic variation found in the use of the English language by non-native speakers. We described the methodology followed in order to compile a real corpus and we proposed a tagging system for the syntactic variations found in a real corpus of texts involving the use of English as a second language.

The questions of synchronic syntactic variation are obviously closely tied to questions of syntactic change. This is the reason why theoretically oriented research on syntactic change has focused on questions regarding the relationship between acquisition and change, as well as grammar. We believe that a better understanding of synchronic variation is clearly a prerequisite for more general theoretical insights in the field of syntactic change. This is the reason why we have designed a tagging system for the examples of variation in the use of English by non-native speakers. In future research, we are going to analyse different aspects such as modality and lexical issues, etc., in order to tag and classify the variation produced in the use of English as a second language.

References

- Adamson, J. (2006). The globalization debate in business English: Exploiting the literature through matrices. *The Asian ESP Journal*, 1, 51-57.
- Aijmer, K. (2001). Modal adverbs of certainty and uncertainty in an English-Swedish perspective. *Language and Computers*, 39, 97-112.
- Bargiela-Chiappini, F. (2004). Intercultural business discourse. In N. C. Candlin and M. Gotti (eds.), *Intercultural aspects of specialized communication* (pp. 29-51). Bern: Peter Lang.
- Bargiela-Chiappini, F. and M. Gotti. (2005). *Asian Business Discourse*. Bern: Peter Lang.
- Barson, J.; J. Frommer & M. Schwartz. (1993). Foreign language learning using e-mail in a task-oriented perspective: Interuniversity experiments in communication and collaboration. *Journal of Science Education and Technology*, 2-4, 565-584.
- Biesenbach-Lucas, S. (2005). Communication topics and strategies in e-mail consultation: Comparison between American and international university students. *Language Learning & Technology*, 9-2, 24-46.
- Bolton, K. (2003). *Chinese Englishes. A Sociolinguistic History*. Cambridge: Cambridge University Press.
- Carrió, M. L. and R. Muñiz. (2012). Lexical variations in business e-mails written by non-native speakers of English. *LSP Professional Communication, Knowledge Management and Cognition*, 3-1, 4-13.
- Cheung, M. (2011). Sales promotion communication in Chinese and English: A thematic analysis. *Journal of Pragmatics*, 43, 1061-1079.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. (2008). Two thousand million? *English Today*, 24, 3-4.
- Dagneaux, E., S. Dennes and S. Granger. (1998). Computer-aided error analysis. *System*, 26, 163-174.
- Evans, E. (2010). Business as usual: The use of English in the professional world in Hong Kong. *English for Specific Purposes*, 29, 153-167.
- Evans, S. (2012). Designing email tasks for the Business English classroom: Implications from a study of Hong Kong's key industries. *English for Specific Purposes*, 31, 202-212.
- Gimenez, J. C. (2006). Embedded business emails: Meeting new demands in international business communication. *English for Specific Purposes*, 25, 154-172.
- Gimenez, J. C. (2000). Business e-mail communication: some emerging tendencies in register. *English for Specific Purposes*, 19, 237-251.
- Hyland, K. (2011). Projecting an academic identity in some reflective genres. *Ibérica*, 21, 9-30.
- James, C. (1998). *Errors in Language Learning and Use*. London: Longman.
- Jenkins, J. (2003). *World Englishes. A resource book for students*. New York: Routledge English Language Introductions.
- Jiang, Y. (2002). China English: issues, studies and features. *Asian Englishes*, 5, 4-23.
- Kachru, B. (2005). *Asian Englishes today. Asian Englishes beyond the canon*. Hong Kong: Hong Kong University Press.

- Kirkpatrick, A. (2007). *World Englishes. Implications for International Communication and English Language Teaching*. Cambridge: Cambridge University Press.
- Nickerson, C. (1999). The use of English in electronic mail in a multinational corporation. En Bargiela-Chiappini y Nickerson (eds.), *Writing business: Genres, media, and discourses*. New York: Pearson Education. 35-56.
- Rogerson-Revell, P. (2007). Using English for International Business: A European case study. *English for Specific Purposes*, 26, 103-120.
- Sailaja, P. (2009). *Indian English*. Edinburgh: Edinburgh University Press.
- Salvi, R. and H. Tanaka. (eds). (2011). *Intercultural interactions in business and management*. Bernd: Peter Lang.
- Schleef, E. (2009). A cross-cultural investigation of German and American academic style. *Journal of Pragmatics*, 41, 1104-1124.
- Tongnini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Warshauer, M. (1995). *Email for English teachers: Bringing the Internet and computer learning networks into the language classroom*. Alexandria, VA: Teachers of English to speakers of other languages.
- Weninger, C. and K. H.-Y. Kan. (2013). (Critical) Language awareness in business communication. *English for Specific Purposes*, 32, 59-71.
- Yli-Jokipii, H. and P. E. F. Jorgensen. (2004). Academic journalese for the Internet: A study of native English-speaking editors' changes to texts written by Danish and Finish professionals. *Journal of English for Academic Purposes*, 3, 341-359.