V International Conference on Corpus Linguistics (CILC2013)

A proposal for the detection and classification of discourse errors

Eva M. Mestre Mestre[*], María Luisa Carrió-Pastor

Departamento de Lingüística Aplicada, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain

Abstract

Our interest lies in error from the point of view of language in context, therefore we will focus on errors produced at the discourse level. The main objective of this paper is to detect discourse competence errors and their implications through the analysis of a corpus of English written texts produced by Higher Education students with a B1 level (following the Common European Framework of Reference for Languages). Further objectives are to propose categories which could help us to classify the different errors in the interlanguage of second language learners and to identify the causes of these errors in the different stages of second language learning in order to offer opportunities for improvement. The methodology followed is a mixed research method which includes quantitative and qualitative analyses of the corpus.

Keywords: Second language acquisition; error; discourse competence; communicative approach

[*] * Eva Mestre Mestre Tel.: +34963877530; fax: +34963877539.
E-mail address: eva.mestre@upvnet.upv.es

1. Introduction

As is well known, learning an L2 (mostly English) has become a crucial skill worldwide, and has had some noticeable effects in the European context, economic and social factors playing a role. In this sense, finding new and better ways to teach English has been a matter of concern both for teachers and for researchers. Consequently, a considerable amount of literature is dedicated to the study of Second Language Acquisition (SLA), which has evolved from traditional to content and task-based methods, and finally to the view of language acquisition as a tool for communication (Sinclair and Coulthard, 1975; Canale and Swain, 1980; Canale, 1983; Bachman, 1990; Savignon, 2002). Thus, the aim in second language learning is now to attain communicative competence in a language different from one's own, preferably English.

Turning to second language approaches, the mainstream method in SLA is the Communicative Approach, which supports the idea that language is essentially communication, thus implying a mastery of the language beyond its particular grammar. The priority of this approach is to meet the communicative needs of learners, which implies the prioritisation of communicative competence over grammatical competence. Canale and Swain (1980) defined communicative competence in terms of four components: a) grammatical competence, which includes the use of words and knowledge of rules, b) sociolinguistic competence, which is related to appropriateness, c) discourse competence, that is based on textual cohesion and coherence, and d) strategic competence, which means the appropriate use of communication strategies. Similarly, Savignon (2002: 9-11) specifies two different items in the communicative curriculum, which the L2 learner needs to acquire in his or her learning process: on the one hand, grammatical competence (sentence-level grammatical forms), in his own words "[…] the ability to recognize the lexical, morphological, syntactical and phonological features of a language", and on the other, discourse competence, related to the interconnection of series of utterances to make a whole.

Language learning is thus considered nowadays to be a process of development of language-learning systems which must be learner-centred. This shows that SLA has shifted its focus from the teacher or the method to the learner, which means that the entire teaching system must be needs-oriented. Furthermore, extra-linguistic factors need to be taken into account in the analysis of how a second language is acquired, as Bardovi-Harlig (1996) points out when the author talks about shared knowledge, and about how much learners know of their own language before they start learning a new one, since a significant amount of knowledge in the L1 can be transferred to the L2 for communicating.

An approach that has been widely employed in SLA is Error Analysis. This discipline has proven to be a useful tool for language improvement by authors such as Larsen-Freeman and Long (1992), James (1998), Levinson, Lessard and Walker (2000), Hunston and Francis (2000), Lee (2004), Carrió (2004) and Salem (2007). Indeed, the idea that errors should be considered not merely as marks of failure but as indicators of the point reached in the learning process has been taken up again by some researchers, for example, Kirkgöz (2010). It can also reveal how the practice of teaching is being accomplished, and the steps taken in interlanguage of the learning and teaching processes.

We will in this study focus on errors produced by students in an L2 classroom in order to provide opportunities for progress; in particular, for the improvement of the discourse practice that may take place within the classroom. Since our interest lies in errors from the point of view of language in context, we will focus on those produced at a discourse level. Within discourse errors, James (1998) distinguishes different production errors: coherence and pragmatic errors, and receptive errors. This author considers discourse errors to be what he calls pragmalinguistic deviations. These are errors with regard not only to language but also to knowledge of externally-related factors. A second group included in this concerns sociopragmatic failures, involving errors due to cultural conflicts. These errors frequently result from intercultural influences, i.e. from cultural differences of view related to what is appropriate social behaviour in certain settings (James, 1998).

Finally, the legal framework for language learning in higher education should be taken into account to observe the importance of discourse errors in SLA. The 1999 Bologna Declaration established the foundations on which language teaching at university level throughout Europe is currently based. In that document, all the member states agreed on the need to calibrate and share language proficiency levels and competences in order to facilitate mobility. For this purpose, a reference framework was produced in 2001 by the Council of Europe, entitled the Common European Framework of Reference for Languages (CEFR). It establishes a set of reference

levels based on the Communicative Approach and a set of skills and abilities related to the different competences a learner must obtain in order to master an L2. A plain description of discourse competence is offered; we can read that (Council of Europe, 2001: 123-130):

> Discourse competence is the ability of a user/learner to arrange sentences in sequence so as to produce coherent stretches of language. It includes knowledge of and ability to control the ordering of sentences. […] Departure from these criteria for straightforward and efficient communication should be for a specific purpose rather than because of inability to meet them.

Furthermore, discourse competence is listed as part of the pragmatic competence a speaker needs to master in order to be proficient in a language in the CEFR. This implies the knowledge of certain principles according to which communication is organised, structured and arranged in discourse competence. It is assessed in terms of: (1) topic vs. focus, (2) given vs. new, (3) natural (temporal) sequencing, (4) cause and effect, and (5) structure and management of discourse. This is done according to some specific principles, such as rhetorical effectiveness, logical ordering, thematic organization, style and register, and coherence and cohesion. It is expected to develop with language proficiency, starting from simple, short sentences, and moving on to longer, more complex utterances at higher levels.

The main objectives of the study presented in this paper are, firstly, to propose some guidelines to analyse discourse competence in a corpus of written texts produced by students of English as a second language at higher education level. Then, the second objective is to detect discourse competence errors. Finally, the third objective is to identify the causes of these errors.

2. Methodology

A corpus of 206 texts produced in English by ninety Spanish university students between 2008 and 2011 was analysed. All the students involved in this study had a B1 level of proficiency in English. They were asked to undertake a series of written homework assignments, with all of the texts being of types proposed by the CEFR as part of the ability to identify and reproduce specific text types, such as personal letters, descriptions, giving reasons or opinions (Council of Europe, 2001). For the analysis, three text types were collected: summaries (S), opinions (O) and formal letters (L), given that the students were involved in a Tourism degree at Universitat Politècnica de València and these assignments could be incorporated in their weekly classes.

An analysis grid based on the one proposed by Mestre (2011: 145) was used to analyse all the discourse competence errors, putting them into context in order to identify the exact causes of the error. The items analysed were based on the recommendations of the CEFR. All error types were tagged to simplify the analysis of the data by raters (three colleagues from the university). The following table shows the items analysed in this particular study as well as the type of errors and the tags used to identify them.

Table 1. Grid used for the analysis of discourse errors based on Mestre (2011).

Item analysed	Competence: skill	Error: Inability to	Tag
Topic/focus, given/new, Structure	Identify unfamiliar words	Extrapolate meaning	DVM
	Natural sequencing: temporal	Sequencing	DSS
	Cause/effect	Verb tense	DSV
	Thematic organisation	Word order	DSWO
	Cause /effect. Linking	Connectors / transitions	DSC
	Thematic organisation: Summarise	Summarising (text types)	DSTH
	Link simple elements	Main points	DSFocus
Coherence and cohesion	Repetition	Coherence and cohesion	DCC
	Synonymy		
	Antonymy		
	Pro-forms		
	Collocation		
	Enumeration		
	Parallelism		
	Transitions		

The texts were identified by the year of production (08, 09, 10), the text type chosen in this study (O, S, L) and correlative numbers. They were marked by three different raters and tagged using the above grid. In this study, a quantitative and qualitative analysis of the data was applied to obtain a broader, in-depth analysis of the results. Below, Table 2 offers some examples of the tagging of the errors found in the texts produced by the students. It shows the type of errors found, the tags used to classify them and the sentences extracted from the texts.

Table 2. An example of discourse competence errors identified.

Error	Text	Example
VERB TENSE (DSV)	08O1T1O	Will they can interact
WORD ORDER (DSWO)	09L2T3	We send you a relation of our hotels around the world that you can choose either.
		You hope you enjoy of your stay.
CONNECTORS/ TRANSITIONS (DSC)	10S3T3	The best of this trip was to met Italy and her cities especially the town called Luca, in this town the people go to the places with bike and all the town it's whole of little shops very interesting.
SUMMARISING (DSTH)	08S1T12	He is a participant of Indian game "Who wants to be a millionaire?" and when he going to win 20 million rupees, the police arrest him because they thought that he cheat. Desperate to prove his innocence
MAIN POINTS (DSFOCUS)	08S1T8	The last movie I saw was the Twilight. Is the story of a young human and a vampire who fall in love. Bella Swan (Kristen Stewart) has always been different to others. When his mother married a second time, she decides to go live in a remote and rainy town called Forks

3. Results

The analysis carried out by raters using the grid identified 2,874 discourse errors, including all the categories and texts analysed. They were classified depending on the type of error recognized and the text type in which they had been produced. The distribution shows clear differences in the amount of errors for each category and text type, as can be seen in the figures below. Figure 1 shows all discourse errors found in the text grouped by categories.
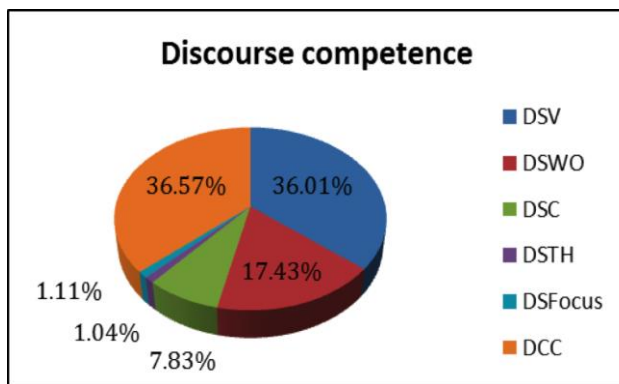

Fig. 1. Discourse competence errors.

It can be observed that the greatest number of errors belong to the categories of coherence and cohesion (DCC, 36.57%) and verb tense (DSV, 36.01%). These two error types account for more than 70% of all errors found in the texts. Errors due to word order (DSWO, 17.43%) were also reasonably prominent. However, the other three items —transitions and connectors (DSC, 7.83%), summarizing an idea (DSTH 1.04%) and identification and reproduction of main points (DSFocus, 1.11%)— display a low number of occurrences, and can be considered to be of little importance:
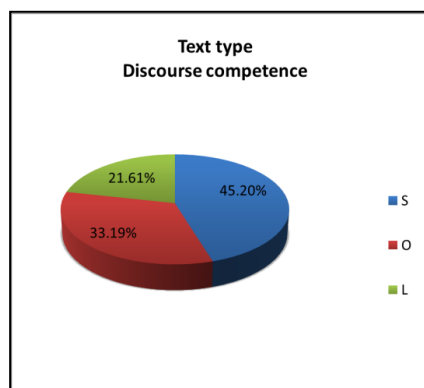

Figure 2. Discourse competence distributed per types of texts.

It can clearly be seen that more than 45% of errors, that is, a total of 1,299 errors, were found in summaries (S). This is followed by errors found in opinions (O), of which there were 954, 33.19% of the total. The lowest amount of errors was encountered in letters (L), in which 621 errors were detected, that is, a little below 22%. We could observe clear differences in the amount of errors produced by students. The standarisation of the patterns of business letters seem to be affecting the production of errors in students with a pre-intermediate level of English. We could detect fewer errors than in summaries, a kind of text that has not fixed patterns and

students explain emotions and feelings. It seems that providing patterns to students avoids the errors in second language acquisition, as we could observe in our analysis.

Finally, Figure 3 shows the combination of these two analyses, of the types of errors identified and where they were found, clearly showing the texts in which students produce a greater amount of errors. As in the previous figures, considerable differences can be found in the results.
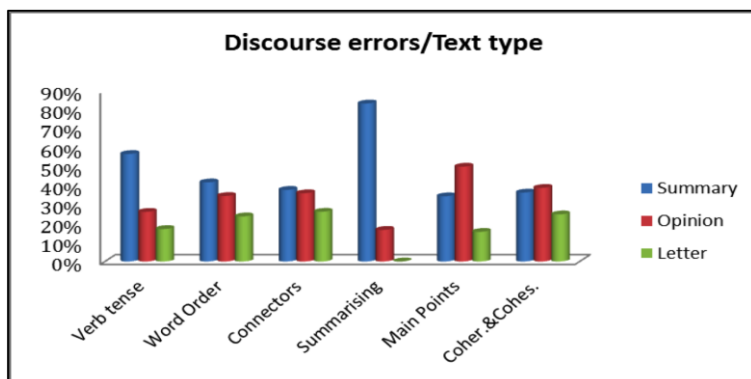

Figure 3. Discourse error categories distributed per text types.

Not surprisingly, the ability to summarize has produced the greatest amount of errors in summaries, since many students proved unable to identify the focus in a summary and sequence it. Apart from this, the results show that the least amount of errors in all categories is found in letters, and the greatest amount of errors in all categories was found in summaries, with two exceptions.

4. Discussion

The analysis of the results found in the present study shows that students with a B1 level of proficiency produce a significant amount of errors when they write. It can also be deduced from the analysis that not all text types produce the same amount of errors, thus indicating that students do not encounter the same level of difficulty when they learn the different skills involved in the acquisition of discourse competence in different kinds of texts.

Regarding the text types in which the errors were found, since most discourse errors were detected in summaries, this may mean that learners find it more difficult to produce texts which require clear structuring, the identification of main points and sequencing. We must not forget that the ability to summarize, at least for adults (here, students at university level), is acquired in the L1, and is then transferred to the L2. This could be due to a lack of knowledge in the student's own L1, as well as an inability to reproduce structures and standards in L2. The fact that the least amount of errors is found in letters can lead us to the supposition that the more structured and the more formulaic the text type, the easier it is for students to follow established patterns and to write correct structures.

Regarding the error types found in the texts, it could be deduced from the analysis that a majority of the errors detected are related to sequencing, more specifically, verb tense and word order. In other words, students find it difficult to identify the correct verb tense and use it throughout their texts. It is a common error for students to use several verb tenses in the same paragraph, or use impossible verb combinations (example 08O1T0). Furthermore, students often use incorrect word order in their writings, possibly because they do not know the correct English language structure and use L1 structures with English words. This is in line with the fact that many errors concern coherence and cohesion issues (collocations, enumerations, and transitions), with students not reproducing the standard structures, but their own. With regard to transitions, there is a considerable lack of connectors throughout the texts (example 10S3T3), consistent with their lack of sequencing.

However, students find fewer problems in using correct English structures when writing formal letters. In all the categories of error, it can be seen that letters seem to present fewer problems to students. The letters written by students were formal, and followed a previously explained and repeatedly practiced pattern. It seems

that the less formulaic the structure of the text type, then a greater the number of errors was produced, and the more constrained the format, the greater level of proficiency that was displayed. It can thus be said that text type is an issue of concern for students with a B1 level of proficiency.

Finally, we believe that further studies should be undertaken to detect the errors of students. Errors could be associated to the different levels of language learning or interlanguage to link certain aspects of language acquisition to learning strategies. This is a preliminary study that focused on discourse but our purpose is to include further aspects such as pragmatic or grammatical competences.

References

Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bardovi-Harlig, K. (1996). Pragmatics and Language Teaching: Bringing Pragmatics and Pedagogy together. In L. F. Bouton (Ed.) *Pragmatics and Language Learning*. Urbana IL: University of Illinois at Urbana-Champaign.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards and R.W. Schmidt (Eds.). *Language and Communication*. London: Longman.

Canale, M. and Swain M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-7.

Carrió, M. L. (2004). Las implicaciones de los errores léxicos en los artículos en ingles científico-técnico. *RAEL: Revista Electrónica de Lingüística Aplicada*, 3, 21-40.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press

Hunston, S. and Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

James, C. (1998). *Errors in Language Learning and Use*. London: Longman.

Kirkgöz Y. (2010). An analysis of written errors of Turkish adult learners of English. *Procedia Social and Behavioral Sciences,* 2: 4352–4358.

Larsen-Freeman, D. and Long, M. H. (1992). *An Introduction to Second Language Acquisition Research*. London: Longman.

Lee, I. (2004). Error correction in L2 secondary writing classrooms: the case of Hong Kong. *Journal of Second Language Writing*, 13, 285-312.

Levison, M., Lessard, G. and Walker, D. (2000). A multi-level approach to the detection of second language learner errors. *Literary and Linguistic Computing*, 15- 3, 313- 322.

Mestre, E. M. (2011). *CEFR & error analysis in Second Language Teaching at University level*. Saarbrücken: Lambert Academic Publishing.

Salem. I. (2007). The lexical-grammatical continuum viewed through student error. *ELT Journal*, 61-3, 211-219.

Savignon, S. J. (Ed.). (2002). *Interpreting Communicative Language Teaching: Contexts and Concerns in Teacher Education*. New Haven: Yale University Press.

Sinclair, J. and Coulthard, M. (1975). *Toward an Analysis of Discourse: the English Used by Teachers and Pupils*. Oxford: Oxford University Press.