Teaching Modality: A corpus analysis of the use of must by university Spanish learners of English

Hanane Benali Taouis[1]; Sidoní López Pérez[2]

[1]Department of Lingüística Aplicada, Universidad Politécnica de Madrid (Spain)

[2]Department of English Didactics of the English Language, Universidad Internacional de La Rioja (Spain)

Abstract: Corpus-based studies have brought interesting empirical results to the field of foreign language teaching at its different levels, including higher education. The main objective of these studies is to improve the teaching methodology by identifying students' errors and presenting methods that can adapt to their needs. In addition, learner corpora are generally considered as a rich source of information when dealing with the use of expressions and errors produced by the learners in their different productions (Castillo Rodríguez, Díaz Lage and Rubio Martínez, 2020). This research flows in the same stream through detecting the errors that our Spanish learners make in their use of the modal verb *must*. It studies a corpus that includes 246 writing samples, 4,816-word types and 107,042-word tokens produced by 155 students in an online forum of a subject within the framework of the Degree in Early Childhood Education. The free concordance software program AntConc was used to detect errors in the use of the modal verb must by our students. This monolingual corpus is compiled from the participation of students in the compulsory subjects that use English as a Medium of Instruction (EMI). The results of this study show 4 types of errors: the use of infinitive with to after *must*, putting two modal verbs consecutively, the incorrect use of *must* in interrogative forms and the incorrect use of *must* in passive voice. The analysis of these errors in context draws from error analysis in second language (Ellis 1990) to distinguish between native language interference, overgeneralization and target language rule ignorance.

Keywords: Teaching method, linguistics, adult students, learning method, second language.

1. Introduction

In the literature we can still observe a clear agreement on the concept Downing and Locke (2006) introduce about the strong relation between modality and the semantic notions of '… possibility, probability, necessity, volition, obligation and permission' (308). This shows the subjectivity of modality in English and explains the difficulties that non-native speakers might encounter to encode these structures; especially if the mother tongue does not include a wide range of modal verbs and uses other structures like subjunctive mood as the most frequent form to express modality (Palmer 2001, Hoye 1997 & Kerl 1861). In addition, some English modal verbs like must are characterised by having two different deontic meanings; one being necessary for and the other expressing obligation (Quirk, Greenbaum, Leech & Svartvik 1985). Due to these particular qualities that Palmer (1979) summarises in 'there is, perhaps, no area of English grammar that is both more important and more difficult that the system of the modals' (x), and to the difficulties that modal verbs entail for Spanish learners of English, we considered

studying the use of the modal verb must to detect the reasons of this phenomenon and to improve its teaching.

Modal verbs generally express a variety of semantic meanings the same as permission, possibility, intention etc., (Oxford Learner's Dictionaries 2022). To recognise the complexity of these language elements, we should be aware of the possible variations encompassed in each modal verb and the relation that these meanings have with the attitude of the producer (Palmer 1979). In other words, modality depends not only on grammatical features but also on lexical and prosodic ones.

Basically, modal verbs are divided into three different categories including epistemic, dynamic and deontic modals. According to Palmer (1990), epistemic modality includes the speaker's judgement about a proposition based on both evidence and knowledge. Deontic modality involves expressing obligation, permission, or forbiddance, whereas dynamic modality is concerned with the abilities and properties of the subject of the sentence. It is particularly important to mention that the modal verb we are analysing in the current investigation has been defined by many researchers (Palmer 1990, Portner 2009 & Sweetser 1990) as carrying the three categories of modality. Thus, we can deduce its possible ambiguity for non-native speakers of English and expect it to be one of the most complicated modals in English.

In line with the previous investigations conducted on modal verbs, the current study provides an analysis of the use of the modal verb *must* by Spanish intermediate learners of English. The field of modal verbs has been generally investigated in contrastive studies between English and many other languages (Bonilla 2017, Qian 2017 & Orta 2010), but it does not count on an overflowing research when it comes to Spanish learners of English. Having this limited number of studies based on learners' corpus to compare to the use of modal verbs by English non-native speakers, makes it an interesting field of study to scrutinise the Spanish learners' difficulties and provide a better understanding of this language element. The main aim is to explain students' errors using authentic production of Spanish students and suggest methodologies to help the students either avoid or overcome these errors. Hitherto, our research covers both contrastive and errors analysis.

Nevertheless, studies on the use of modal verbs by Spanish learners of English are very few and most of them are dissertations or theses. They all agreed on the difficulties faced by non-native students in using modal verbs. This makes our research relevant to the field of language teaching, but it also limits the scope of our theoretical background. In addition, and taking into consideration that most research on the use of modal verbs focuses on the writing of specific and standardised assignments by non-native students of English, another outstanding aspect of this research is the use of spontaneous learners' written production extracted from activities in a non-linguistic subject (López Pérez & Benali Taouis 2019, López Pérez 2021).

2. Research method

2.1. Research objectives

This research draws on a corpus-driven approach (Biber 2010) and concept-oriented approach (COA) to analyse the subject's errors with the aim of providing insight into understanding the learning process (Bardovi-Harlig 1995, 2000; Dietrich, Klein & Noyan 1995) and explaining the reasons behind the difficulties that our intermediate Spanish learners of English face with modal verbs. The choice of modal verbs is not arbitrary, but based on a deep analysis of the types of errors and on the results of previous studies in which modal verbs were included in the first ten most frequent errors of Spanish learners of English (MacDonald 2016). The purpose of this research is to bring to light the errors that students make while using the modal verb must. A total of 166 hits were assigned with the token must and 8 errors were analysed. Although the number of these errors is not very high, it demands our attention because it shows that students continue to make a wrong use of the structure that is required with modal verbs. These errors showed four categories of grammar errors. The analysis goes a step further

to compare the structures of these errors with the students' native language (NL) to the English grammatical rules to find the possible reasons.

2.2. Participants

Our monolingual corpus, ENTECOR, (see Table 1 below) includes 786 participants, 527,099 tokens and 13,148 types. This comparable corpus encompasses two sub-corpora (TICOR & SECOR). The latter comprises 120 subjects' written production in a forum from a subject within the framework of the Master's Degree in Secondary Education taught totally online at Universidad Internacional de La Rioja (UNIR). The TICOR sub-corpus embraces 246 writing samples, 4,816-word types and 107,042-word tokens written by 155 students. Again, this sub-corpus comprehends 2 components (ICT composed of spontaneous written texts of 155 students in an online forum of a subject within the framework of the Degree in Early Childhood Education, and TIC, including written samples of 511 participants in a similar online forum as part of another subject within the Degree in Primary School Education).

Table 1: Components of corpus ENTECOR

Sub-corpora	TICOR		SECOR
Components	ICT	TIC	TRAINCOR
Number of students	155	511	120
Tokens	107,072	317,759	102,268
Types	4,821	9,320	6,380

The actual study uses the second sub-corpus of TICOR (ICT). Similar to all our previously stated corpora, this one is composed of multiple contributions of the subjects in a forum of a compulsory non-linguistic subject (ICT Tools Applied to the Learning of English), in which English is the medium of instruction and assessment. The instructors' role in the forums is limited to monitoring and observation without any type of intervention. The students were given a topic for discussion based on the content of the subject and they were required to use English as their only medium of communication.

2.3. Research Methodology

To analyse the corpus the free concordance program Antconc (3.3.4) was used by searching for the strings *must, must not and mustn't* to be able to find the different forms of this modal verb in the corpus. We assigned three levels of search (1R, 2R and 3R) to make sure lines in Key Word in Context (KWIC) could be generated in order to allocate this word in context and see it in its affirmative, interrogative and negative forms.

In this way we accessed all the forms of *must* and we could verify its different collocations in each of the forms previously mentioned. There was no negative form in the corpus and the results were 0 hits when we used the strings *mustn't*, *must not and must**. Since students were required to use only formal language, no contraction was found. After allocating the sentences with the targeted modal verb, we analysed all the hits. With a deep and separate analysis of each

sentence in context, the researchers determined which hits contained errors in the use of *must*. All the detected errors were categorised and labelled in tables with the hit number and the student ID. The student ID is a hypothetical number assigned to each student during the coding of the corpus, following the international coding standards, to keep the corpus both anonymous and controlled (Castillo Rodríguez & Díaz Lage 2015). At the same time, all the examples are provided verbatim and the correct form is also supplied. In addition, in this actual study, the allocated errors had a specific category that was repeated many times, and obviously, we had to demonstrate that it was not an error made by the same student for it to be significant and representative of our participants.

2. Research method

We can observe that there are a total of 7 errors with *must* in the corpus some of which are more frequent than others. As previously mentioned only the errors that were repeated are identified by students' ID for the credibility of this study. However, as explained above the actual investigation did not rely only on this simple search; since we also conducted a search with *mustn't*, *must not* and *must** to make sure we could detect the different forms of this modal verb together with any symbol or character that was not properly saved during the coding process. Each of the sentences was carefully scrutinised to find any error related to the use of *must* in clusters. In table 2 below we delineate the errors and provide the possible correction of each.

Table 2: Errors in using *must* with hit numbers and student ID

Hit number	Error	Correct form
23	*must be** answer	
70	*do** teachers *must** focus…?	must teachers focus
125	They *will must* same characteristics	
141	*must to** create (student Id 67)	must create
142	*must to** do (student Id 43)	must do
143	*must to** evolve (student Id 118)	must evolve
144	*must to** know (student Id 08)	must know
145	*must to** learn (student Id 08)	must learn

The errors were then classified, calculated and represented in table 3 below. As shown in the percentages and error frequency columns, the use of *must* followed by *to infinitive* verbs is the most frequent error with a percentage of 71.42% errors. The rest of the errors comprising the misuse of *must* in interrogative form and the future both demonstrate a lower percentage with not more than 14.28%.

Basically, in Error analysis (EA) learners' language reflects learning difficulties in terms of errors. Brown (1980) cites four main sources of errors by foreign/second language learners: Interlanguage Transfer, Intralanguage Transfer, context of learning and communication strategies. This classification of causes has its roots in the belief that learners first use their native language (NL) as the first linguistic system of reference resulting in an inevitable interference. Then, as the learners advance in the target language system more and more errors originating from the misuse of the target language (TL) rules appear. The other cause is the context of learning, also named ¨false concepts¨, includes all the misleading information or confusing explanations that students receive from the instructor or other sources of knowledge. The last category is the ability of learners to use certain techniques to convey the message when their language knowledge is limited.

In addition to causality, the theory of Error Analysis also classified second language acquisition errors into four basic categories as detailed in the surface structure taxonomy by Dulay, Burt, and Krashen (1982). These types being omission, addition, misformation and misordering errors which reflect students' misuse of the target language system. From the

description of the types of errors in table 3 below we can conclude that the main focus of this analysis is the second category of errors (addition) as they all share the characteristic of including an added word that alters the meaning or the grammatical structure of the sentence.

The first error in table 3 below described as the use of *must* with the auxiliary *do* in interrogative form is a clear inaccurate use of the rule of simple present interrogation in English through which students learn that to make a question the sentence should start with an auxiliary verb in simple present inverted with the subject, followed by the main verb in infinitive. This seems to be the exact construction that our subject applied in the sentence 'do* teachers must* focus'. For a comparative analysis to attribute this error, we can conclude that it is an intralingual one. This is an error originating from the misuse of the target language (TL) rules and has no relation to the target language of our students (Richards & Schmidt 2002, Benali Tauois & López Pérez 2019). The Spanish linguistic system has no inversion and no auxiliary addition in its interrogative structures. In this case students are not aware of the exception of the rule stating that if the construction of the interrogative sentence is different when the main verb is an auxiliary, in this case a modal verb.

Our second type of errors is obviously another addition error in which the subject is using the auxiliary *will* before *must* trying to form the future tense without taking into consideration the exception that modal verbs have in carrying the tense. The student also seems to confuse the use of *must* with that of *have to* and use *must* in context where *have to* would be the accurate choice. We can only attribute this to the overgeneralization (future with will) of a target language rule resulting in a relatively meaningless linguistic construction. Accordingly, we can deduce the teaching of *must* and *have to* together in the same lesson as most of the grammar textbooks suggest might need some modifications to avoid these types of errors. This is an error resulting from what Richards (1971) defines as 'context of learning' including misleading explanations or improperly contextualised items. In this context, students are being introduced to a linguistic rule by contracting two modal verbs with raising their awareness about the possible errors and the same happens with the teaching of future tense rule when the students are not introduced to exceptions. Definitely, we cannot claim that this is the only teaching scenario, but it is a possible one. This can also be a result of not paying attention to the exceptions explained by the instructor or forgetfulness in the long term. Together with the previously specified causes we can also think of other sources of information that students might have access to without any supervision from the instructor as one of the reasons generating these types of errors.

Finally, the most frequent error is the use of *must* with *to infinitive* verbs. This is another additional error and a consequence of an interlanguage transfer. In Spanish the modal verb *must* is either translated as *tener que* or *deber*. When the Spanish English learner thinks of it as *tener*, this structure is transferred into the English structure and appears as the one we have in this case with the modal verb *must* followed by a *to infinitive* form, making the sentence grammatically incorrect.

Table 3: Types and numbers of errors

Types of Errors	Error Frequency	Percentage of Errors
Must with the auxiliary do in interrogative form	1	14.28%
Must in future preceded by will	1	14.28%
Must following by to infinitive	5	71.42%

4. Conclusions and discussion

The pedagogical implications of this corpus-based empirical study are determined by our belief that these errors are a source of information about the learning process of the learners, an orientation for instructors in setting their teaching objectives, and a hint of what grammar rules should be explained in contrast with others. The present research aims at raising the instructors'

awareness about the importance of adapting the lesson and the grammatical rules we share with our students to their knowledge and their unconscious background, both from their NL and from previously introduced TL rules.

The actual study is set to detect Spanish university learners' errors in English aiming to help instructors predict these errors in their teaching context and better structure its content. Focusing on must as one of the most confusing modal verbs for Spanish learners of English, we analysed the errors to observe that the difficulties come from both first language transfer and target language rule overgeneralization. Consequently, we can deduce that it is not accurate to focus the language teaching methodology on one of the causes. It is crucial to be aware of the contrastive errors and the ones resulting from the structure of the target language rules to include both in the teaching content. This means that although Spanish learners of English have some specific common errors, they also share other difficulties with all learners of English independently of their first language background. Obviously, the teaching design should take care of these problematic areas for the learners, and include them in the teaching goals. This probably accounts for the importance of compact corpus analysis studies as the actual one in orienting instructors, content designers, textbook writers and educational websites developers. In this way, the results of this study are expected to help instructors to first consider the difficulties that the Spanish learners of English have with the use of must in the future, in interrogative form and choosing the correct form of the verb that comes after this modal verb. Then, it can also be a way to encourage them to analyse the errors of their students as a tool to assess the learning process and evaluate the teaching content. This might also help the students with a self-evaluation resource or peer learning material if used appropriately by the instructor. Another way to benefit from the results presented in this study is to focus on the presented errors in follow up and practice activities or customise them to suit the needs of each learners group. It might be useful to build up a repertoire of errors and deduce possible patterns to have a solid reference for both learners and instructors through the development of classified samples of error-based teaching material. With the access to the variant learning material that the internet provides, language instructors are no longer the unique source of information for language learners; thus, they became responsible for the competence of unknown teaching volunteers who might intervene in the learning process of their students. This cannot be considered as a fully negative interference, but we should not neglect its ugly part. Hence, instructors need to plan ahead and prevent the errors that can be caused by the different resources students are exposed to. This implies having to control language problems before they are reflected in the learner's interlanguage, which can be possible only through error analysis either based on recent studies or on the analysis of the production of each student.

5. References

Bardovi-Harlig, K. (1995). 'A Narrative Perspective on the Development of the Tense/Aspect System in Second Language Acquisition '. *Studies in Second Language Acquisition*, 17, 263-91.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: A study of form, meaning, and use*. Language Learning Monograph Series. Blackwell.

Benali, T. H. & López, P.S. (2019). 'Analysing EFL Written Corpus for a Better Teaching and Learning of Gerund and to Infinitive Structures'. *Journal for Educators, Teachers and Trainers*, 10 (1), 85-100.

Biber, D. (2010). 'Corpus-based and Corpus-driven Analyses of Language Variation and Use'. In B. Heine, & H. Narrog, (Eds.), *The Oxford Handbook of Linguistic Analysis*. pp. 160-91. Oxford: Oxford University Press.

Bonilla, J.F. (2017). 'The Use of the Epistemic and Deontic Modality in the Modal Verb "can" Between English Native and Spanish Non-native Students in the Written Discourse'. *Journal of Artistic Creation and Literary Research*. 5(2), 13-24.

Castillo Rodríguez, C., & Díaz Lage, J. M. (2015). Exploitation of a learner corpus: Analysing opening and endings in academic forums. *Revista Opción*, Supl. 31(6), 192-210.

Castillo Rodríguez, Cristina; Díaz Lage, José María y Beatriz Rubio Martínez (2020). Compiling and analyzing a tagged learner corpus: a corpus-based study of adjective uses. *Círculo de Lingüística Aplicada a la Comunicación.* 81, 115-136,

Dietrich, R., Klein, W., & Noyau, C. (1995). *The acquisition of Temporality in a Second Language*. Amsterdam: John Benjamins.

Downing A. & Locke, P. (2006). *English Grammar: A university course*. London: Routledge.

Dulay, H., Burt, K. M., & Krashen, S. (1982). *Language two*. Rowley, MA: Newbury House.

Ellis, R. (1990). *Instructed Second Language Acquisition: Learning in the Class-room*. Oxford: Blackwell.

Hoye, L. (1997). *Adverbs and Modality in English*. London: Longman.

Kerl, S. (1861). *A Comprehensive Grammar of the English Language*. New York: New.

Li, Q. (2017). 'A Corpus-based Study of Modal Verbs Use in English Writing by EFL Learners'. *Canadian Social Science*, 13(11), 31-5. DOI: http://dx.doi.org/10.3968/9975, accessed 10 April 2022.

López Pérez, S. (2021). Analysis of Subject-Verb Agreement Errors in Third Person Singular Verb Forms by Spanish University Students: A Corpus-Based Study. *International Journal of English Linguistics.* 11(2), 23- 35.

López Pérez, S., and Benali Taouis, H. (2019). Analysis of noun (direct object) collocations with the high-frequency verb DO by Spanish students in an online learner corpus. *Complutense Journal of English Studies.* 27, 99-120.

MacDonald, P. (2016). 'We all Make Mistakes! Analysing an Error-coded Corpus of Spanish University Students' Written English'. *Complutense Journal of English Studies*. 24, 103-29.

Orta, V.I. (2010). 'A Contrastive Analysis of the Use of Modal Verbs in the Expression of Epistemic Stance in Business Management Research Articles in English and Spanish'. *Ibérica.* 19, 77–96.

Oxford Learner's Dictionaries. (2022). https://www.oxfordlearnersdictionaries.com/definition/american_english/modal

Palmer, F. R. (2001). *Mood and Modality*. Cambridge: Cambridge University Press.

Palmer, F. R. (1990). *Modality and the English Modals.* 2nd edition. London: Longman.

Palmer, F. R. (1979). Modality and the English Modals. London & New York: Longman.

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

Quirk, R., Greenbaum,S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English language*. London: Longman.

Richards, J. C., & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Longman.

Sweetser, E. E. (1990). *From Etymology to Pragmatics: The Mind-as-Body Metaphor in Semantic Structure and Semantic Change*. Cambridge: Cambridge University Press.