This paper investigates the use of epistemic expressions in scientific English. The main aim of this research is to analyse if native speakers of English use epistemic modality in the same way than non-native speakers of English and to detect the most outstanding cognitive implications of this fact. The corpus used in this research contains 50 research papers written by native English speakers and 50 scientific papers written by Spanish researchers who use English to communicate internationally. As epistemic modals are used to indicate the possibility of some piece of knowledge, this paper focuses on epistemic modal verbs in order to detect if native speakers of English and non-native speakers of English communicate modality in the same way, or if there are differences in frequency and use. The results obtained in this analysis indicated that there are differences in the frequency of use of epistemic expressions, even if the intention of the writers is the same.

Keywords: epistemic modality, non-native English speakers, native English speakers, contrastive study.
1. Introduction

This paper deals mainly with the role of modality in scientific discourse. If we agree that humans have a natural tendency to make modal judgments, we will also consider that this fact involves the use of concepts such as possibility, necessity and probability. These concepts may be used in different ways by speakers, as language is used depending on the linguistic background of the writer, the academic genre and the expected reader of the text. In this way, the main purpose of this paper is to analyse if native speakers of a language use epistemic modality, more precisely, modal verbs.

One of the standpoints of this paper is to consider language as an individual act, used in a unique way by different speakers. The way we use language creates our own style. We could express the same ideas in such different ways that our audience feel sympathy for the authors’s position depending on the rhetorical strategies deployed. Language changes and linguistic variation can be observed in the same language (Samraj, 2002, 2004; Freddi, 2005; Charles, 2007; Ozturk, 2007), or even more if we contrast speakers with different linguistic background (Salager-Meyer et al., 2003; Martín Martín, 2003; de Haan and van Esch, 2005; Moreno and Suárez, 2008; Hinkel, 2009; Schleef, 2009). In this way, the main aim of this paper is to detect if there is variation in the use of epistemic modality in scientific English. As White (2003) citing Volosinov (1995, p. 139) explains: “The actual reality of language-speech is not the abstract system of linguistic forms, not the isolated monologic utterance, and not the psychological act of its implementation, but the social event of verbal interaction implemented in an
utterance or utterances”. The implementation of utterances expressing modality in scientific discourse is one of the aims of this study.

Another standpoint of this analysis is to determine the expressions of epistemic modality used in scientific English that mark the necessity/possibility of an underlying proposition relative to some body of evidence. There are several studies that suggest the expressions to be included in the studies of epistemic modality, such as Palmer (1986), Huddleston (1988), Lock (1996), Thue Vold (2006), Alonso Almeida and González Cruz (in press) and Alonso Almeida (in press-a). The classification used for the purposes of this research was the one proposed by Fintel and Gillies (2007), who include in this category the modal verbs must, might, may, should, can, could, need.

This study also considers essential to know the specific text in which the expression of modality is placed. The study of modality in different text types has been carried out by scholars such as Chafe (1986), Carretero (2002, 2004), Hidalgo-Downing (2004), Marín-Arrese (2004), Marín-Arrese et al. (2004), Ferrari (2009), Ortega-Barrera and Torres-Ramírez (2010) and Alonso-Almeida and Cruz-García (2011). They have shown that, depending on the intention of the writer expressions of modality are used with different purposes. This research focuses on the study of modality in a specific genre, although our main aim is the detection of the possible causes of variation in the use of epistemic modality in scientific research papers.

In the epistemic use, modals can be interpreted as indicating inference or some other process of reasoning involved in coming to the conclusion stated in the research being described. However, epistemic modals do not necessarily require inference, reasoning, or evidence. We should take into account that one effect of using an epistemic modal is a general weakening of the speaker’s commitment to the truth of the sentence containing the modal. However, it is disputed whether the function of modals is to indicate this weakening of commitment, or whether the weakening is a by-product of some other aspect of the modal’s meaning. In this study these considerations are taken into account in my description of epistemic modality.

The choice of a contrastive approach is motivated by the reason that English scientific language changes in the way it is transmitted; it has similarities and differences. In this way, with a contrastive analysis, students and researchers
from non-English backgrounds can easily compare and contrast the scientific language used by the research community. The objectives of this paper are; in the first place, to identify modal verbs that express epistemic modality in scientific research taking into account the mother tongue of the writer; in the second place, to contrast the frequencies of the use of modality and, in the third place, to reveal the causes of variation in the use of epistemic modality.

2. Methodology

The corpus used in this study comprises 50 research papers written by native English speakers (NES) and 50 research papers written by non-native English speakers (NNES) who are Spanish. The linguistic background of the English writers was identified considering their affiliations and names, as well as other research papers written by the same authors. In the case of the Spanish researchers, the papers were collected from researchers who work at the Polytechnic University of Valencia and who publish in international journals. The papers were collected during 2010 and 2011 from international journals specialised in the specific field of engineering.

Once collected, the 100 papers were processed to be analysed with the help of the software WordSmith Tools 5.0. The modal verbs were identified and their epistemic use checked in order to observe their use in this specific context. The criteria followed to identify epistemic modality markers in the texts were based on Thue Vold (2006):

a. The marker had to explicitly qualify the truth value of a propositional content.

b. The marker has to be a lexical unit (not phrases or paragraphs). From those I selected modal verbs for this study. Modal verbs have been argued to be in the borderline between grammatical and lexical (Aikhenvald, 2004), and this is the reason I include them here.

After the analysis, the results were presented and a statistical analysis was carried out in to determine their applicability to other similar texts. The relative risk and the chi-square test were calculated and the results contrasted. Finally, conclusions were discussed and commented on.
3. Results and discussion

The total amount of running words of the corpus analysed are shown in Table 1:

<table>
<thead>
<tr>
<th>STATISTICAL DATA</th>
<th>OCCURRENCE NNES (%)</th>
<th>OCCURRENCE NES (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running words</td>
<td>184357 (47.11%)</td>
<td>206907 (52.89%)</td>
</tr>
<tr>
<td>Word list</td>
<td>10590 (45.43%)</td>
<td>12716 (54.57%)</td>
</tr>
<tr>
<td>Sentence No.</td>
<td>9017 (50.00%)</td>
<td>9017 (50.00%)</td>
</tr>
<tr>
<td>Average word per sentence</td>
<td>20.44 (46.11%)</td>
<td>22.94 (53.89%)</td>
</tr>
<tr>
<td>Paragraph No.</td>
<td>1145 (55.51%)</td>
<td>916 (44.49%)</td>
</tr>
<tr>
<td>Word No. per paragraph</td>
<td>161.29 (41.58%)</td>
<td>225.88 (58.12%)</td>
</tr>
</tbody>
</table>

The sentence number was the same in both groups of texts, but we can observe that NES used more than 5% words in the texts. The average word per sentence of NES is also higher than the words found in the texts written by NNES. Once analysed the corpus under study, the total amount of modal verbs selected in this research were counted. Table 2 shows the total occurrences found after the analysis of the corpus:

<table>
<thead>
<tr>
<th>MODAL VERBS</th>
<th>OCCURRENCES NNES (%)</th>
<th>OCCURRENCES NES (%)</th>
<th>RELATIVE RISK</th>
<th>CHI-SQUARED (χ²)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN/</td>
<td>877 (59.82%)</td>
<td>589 (40.18%)</td>
<td>1.26 (1.16-1.36)</td>
<td>P = 0.00</td>
<td></td>
</tr>
<tr>
<td>BE ABLE</td>
<td>78 (76.47%)</td>
<td>24 (23.53%)</td>
<td>2.75 (1.75-4.32)</td>
<td>P = 0.00</td>
<td></td>
</tr>
<tr>
<td>COULD</td>
<td>166 (48.82%)</td>
<td>174 (51.18%)</td>
<td>0.81 (0.66-0.99)</td>
<td>P = 0.03</td>
<td></td>
</tr>
<tr>
<td>MAY</td>
<td>181 (39.69%)</td>
<td>275 (60.31%)</td>
<td>0.56 (0.47-0.66)</td>
<td>P = 0.00</td>
<td></td>
</tr>
<tr>
<td>MIGHT</td>
<td>13 (24.07%)</td>
<td>41 (75.93%)</td>
<td>0.27 (0.14-0.50)</td>
<td>P = 0.00</td>
<td></td>
</tr>
</tbody>
</table>
The modal verbs were analysed manually and the epistemic modal verbs were detected in the corpus. These epistemic modals selected considering the occurrences found are *can*, *could*, *may*, *might* and *must*. Of these, the form *can* is certainly controversial. While Biber et al. (1999) consider this form within the domain of epistemic modality, other scholars (cf. Carretero, 1992; Collins, 2009) disagree. I will comment on this in the light of the examples selected.

### 3.1. Epistemic modals

In this section, I describe the use of epistemic modals in the sub-corpora of NES and NNES. As already pointed out, modals are said to be both lexical and grammatical. The difference is not clear, but what is obvious is that they have a scope over the proposition they hedge. One of the most declared functions of modals is the mitigation of claims (Kranich, 2009; Mur Dueñas, 2011), and so authors prevent future events threatening their public self-image. In what follows, I describe the use of the epistemic modals *can*, *could*, *may*, *might* and *must* in the corpora in this order.

#### 3.1.1. Can/Could

The form *can* is certainly one of the most frequent ones in both corpora. Cases of this form are given in the following instances:

**NES:**

(1) The detailed construction and operation *can* be found. The structure of the muscles gives the actuator a number of desirable characteristics. (NE3)
(2) To enhance the safety of the system, the workers can now be removed from the immediate area of the pond, reducing the radiation risk and the ‘manipulator’ will be tele-operated from a remote and shielded site. (NE9)

(3) A small dead band can be observed for low pressures. This is due to the pressure required to overcome the radial rubber elasticity, i.e. to partially inflate the rubber liner to make contact with the outer braid. (NE17)

NNES:

(4) Skinner bases the trade-off concept on the argument that no technologically based system can perform equally well on every performance capability, and that managers should design the system considering a few strategically significant capabilities. (NNES2)

(5) The power loss in PGTs however can be surprisingly high, and consequently the efficiency will be very low. Therefore, an estimation of the efficiency is one of the most important steps in the design of a PGT. It could not only give a numerical estimate of the efficiency, but also its analytical expression. (NNW17)

(6) We appreciate in this new figure that the energy can be redistributed more easily within the rubber, thus temperature profiles are flatter, which illustrates the dependence of temperature profiles with the thermal conductivity of rubber. (NNW23)

The modal can is frequently included in the domain of dynamic modality (Palmer, 1991, pp. 7-10), and less frequently within the domain of epistemic modality (cf. Huddleston and Pullum, 2002, p. 180). Dynamic modality is related to the expression of volition and ability (Palmer, 2001, p. 10). This modality is also called intrinsic modality: “intrinsic modalities typically involve a person’s or thing’s intrinsic disposition, which has the potential of being actualised” (Radden and Dirven, 2007, p. 246). This idea of potentiality is extremely useful in scientific writing to describe the functions and uses of gadgets and substances that are able to, and has the potential of being actualised. Examples (1)-(6) above can be labelled as examples of dynamic modality, since they seem to indicate the disposition of the entities to perform the actions described in the proposition rather than to
indicate the chances of the propositions to be true. In this sense, the dynamic meaning of *can* implies factuality rather than possibility and probability (cf. Alonso-Almeida forthcoming-a).

There is a slight difference between the uses of *can* in (2) and (4), for instance. While the modal verb in (4) clearly appears to show a fact concerning “technologically based systems”, this same modal in (2) shows the disposition of the workers to be outside the area of the pond, but it also seems to indicate a weak sense of possibility of carrying out the described situation in $P$. A rephrasing of the modal would be clarifying, and so *can* can be replaced by the matrix *it is now possible that...* However, this epistemic meaning would contradict the rest of the message in running discourse. If safety is pursued, and safety certainly concerns workers, then there is not an evaluation of chances but an actualisation of the event. The use of *can* not only allows this but, in some ways, it also seems to suggest some mitigation of the claims. Alonso-Almeida (forthcoming-b) states that “dynamic modality, like epistemic modality, functions as a device to downtone assertiveness”. This idea is also present in Perkin’s definition of dynamic modality where the expression “empirical circumstances” indicates some degree of factuality: “[dynamic modality] is concerned with the disposition of certain empirical circumstances with regard to the occurrence of the same event” (1982, p. 252).

The following are instances of *could* in both corpora:

NES:

(7) Since the operator can move the input joystick many times faster than the bandwidth of the pole mechanism this *could* easily make the system unstable. (NES45)

(8) This simple test was required to verify that the designed pMA system *could* move the pole and a load (combined mass 70 kg) through a work volume measuring $3\text{ m} \times 3\text{ m}$ at the pole tip. The tests revealed that this *could* easily be achieved satisfying the power and motion range requirements. (NES32)

NNES:

(9) A method that *could* provide the same results, regardless of the local conditions of the undertaking (such as size, geographic location and population characteristics) would be extremely convenient. (NNW22)
Collins (2009) describes the epistemic use of the modal *could* in the following terms: “*could* appears to be undergoing a similar semantic development, with a weak epistemic use evolving from the unreal use via bleaching of irrealis meaning, and subsequent shedding of its tentativeness marking” (2009, p. 107). This is clearly the case of example (8), the second case of *could* therein, and example (9) that corresponds to NNES. The examples in (7) and (8), the first example therein, show also a dynamic meaning. Thus, *could* in (7) indicates both the disposition and the potentiality of the entity described, that is “Since the operator can move the input joystick many times faster than the bandwidth of the pole mechanism”, and also the possibility of $P$ to occur. In this case, *could* shows the chances the entity has to “make the system unstable”.

As in (2) above, the use of *could* in (8), that is “This simple test was required to verify that the designed pMA system *could* move the pole and a load...”, indicates a dynamic meaning but, at the same time, it also evaluates the chances of $P$ to occur. Actually, *could* also indicates a hypothetical situation providing the conditions stated later, i.e. “through a work volume measuring $3 \times 3$ m at the pole tip”, are fulfilled.

3.1.2. May/might

The modals *may* and *might* present similar meanings in the two sub-corpora. The form *may* is used more frequently in the NES sub-corpus than in the NNES. Examples are given, below:

NES:

(10) Other debris on the floor of the pond *may* require that the retrieved material must be lifted vertically up to 2 m to be clear of entanglement. To prevent jamming of the waste in the removal skip, sections must not be larger than 1 m. If the size is greater than 1 m the waste must be manipulated into the hydraulic clamp where it is reduced, before being located in the skip. During the movement of the fuel rods silt within the pond is disturbed reducing the visibility. Reductions in the visibility *may* eventually mean that work must stop until the silt has resettled. When lifting very heavy loads that need to be size-reduced, material is
often dragged to the cutter, a process that is particularly prone to displacing silt and causing loss of visibility. (NES1)

(11) These results show that there is promise for developing an integrated process that **might** produce syngas, methanol, and power from a methane source. (NES14)

NNES:

(12) Likewise, the sum of the torques applied to a link has to be zero. Taking into account that these are the torques applied to the elements of the links, the equation for link _n_ **may** be written as 0. (NNW18)

(13) It is hence necessary to provide the Administration with tools that allow monitoring water distribution systems conditions and the way they are technically managed. The problem is to find tools that are not subject to interpretation, specially regarding the regulative role that they **may** play. (NNW22)

(14) Nevertheless, this solution **may** carry several shortcomings. First, according to results plotted in Figures 7 and 8, a mold temperature of about 100 ºC seems not to be sufficient for these low-thermal rubber samples, because the surface temperature is not high enough. (NNW23)

(15) The need to differentiate the speed ratio expression **might** appear to be an obstacle against translating the above procedure into a programming language lacking symbolic manipulation capabilities. (NNW17)

All the examples of **may** above are epistemic, except for the case of NNES **may** in (12) that is dynamic. This dynamic meaning indicates the potentiality of the equation to be written as 0 rather than an evaluation of the chances of the equation to be written as 0. In this particular instance of **may**, the proposition flanked by the modal is seen as factual rather than as possible. This **may** in (12) can be nicely replaced by **can** without really affecting the meaning of the proposition. The other examples above are certainly epistemic, since the writers seek to mitigate the propositions hedged by epistemic **may**.

In (10), for instance, the authors indicate their hesitancy towards their conclusion concerning what “the reductions in the visibility **may** eventually mean”. This **may** refers to some hypothetical situation in the future that cannot be effectively
asserted. The same is true of the example in (14) where the authors convey lack of precision and a low degree of commitment towards the information given. The authors protect their image by using may since this modal helps reducing imposition on the readers/listeners, and the fact that the solution carry some problems is perceived as possible rather than as factual.

Likewise, might in (11) and (15) is used in an epistemic sense, and it shows a lower degree of certainty than may concerning the propositions hedged in both cases. There, this modal indicates possibility (cf. Biber, 1999, pp. 489-491). In (11), might seems to indicate the expected logical result of the research carried out, but this remains highly hypothetical. The use of might in (15) clearly indicates the authors’s stance concerning the information given. They are not really sure that “the need to differentiate the speed ratio expression” may result in the described situation. The form may could have also been used in this context, but might expresses “a little less certainty about the possibility” (Palmer, 2001, p. 58).

3.1.3. Must

The form must is deployed with similar functions in NES and NNES to indicate both epistemic and evidential meanings. The use of must in NNES outnumbers the use of this modal in NES. The following examples illustrate the uses of must in the two corpora:

NES:

(16) Since the streams enter at a temperature less than the desired temperature of 673 K, either heat must be added to the reactor or sufficient heat from the oxidation must be supplied to allow water to attain its supercritical state. According to the flow sheet in Fig. 1, we examined reactions and the heat required by the reactor. (NES28)

NNES:

(17) The genesis of a typology is theoretical, trying to define ideal types that might exist; meanwhile, taxonomies attempt to classify real organizations. But at the end, some kind of similitude exist between both configuration types, since a typology must prove its usefulness in real organizations, and one of the main applications of taxonomies is to identify the best
types. They also coincide in considering the choice of the construct on which they are based as one of the most important decisions on their development. The variables that sustain them must be carefully selected and based on existing theory. (NNES31)

(16) and (17) show the use of *must* with an epistemic sense, since, in both cases, it shows the possibility of *P* to occur. Another reading of the modal verb in these examples indicates an evidential meaning. Evidentiality is “concerned with indicating the information source the speaker is relying on to make a claim. This places this category next to epistemic modality without, however, merging them into one” (Diewald, Kresic and Smirnova, 2009, p. 190). Cornillie (2009) clearly shows a more disjunctive model concerning evidentiality and its relation to epistemic modality. He argues that the modes of knowing and degrees of speakers’s commitment towards the proposition do not necessarily go hand by hand. Cornillie (2009, p. 47) claims that evidentiality “refers to the reasoning processes that lead to the proposition”.

In (16) and (17), *must* reflects what Van der Auwera and Plungian (1998, p. 86) describes as inferential evidentiality, which is a case of epistemic necessity: “The claim is only that inferential reading amounts to epistemic modality and more particularly epistemic necessity”. The modal *must* in (16) shows the necessity of adding the heat or the heat from the oxidation to “attain a superficial state”, while it also shows the likelihood of this event to happen. In (17), *must* indicates a logical, and thus expected, requirement of the typology so that its usefulness can be upheld.

4. Conclusions

Epistemic modals are interesting in part because their semantics is bound up both with our information about the world and with how that information changes as we share what we know. This study shows the way in which epistemic modals are used and how they are distributed in two corpora of scientific English texts written by NES and NNES. Concerning frequency, there is certainly a distinction between the use of *can, may* and *must* in both compilations. While
native English writers seem to deploy *may* more often than Spanish writers, the latter tends to use *must* and *can* more frequently, even when *can* can be safely replaced by *may*. This situation may arise from the direct translation of the modal verb *puede* in Spanish, which is used to indicate both potentiality and possibility/probability.

In the case of *must*, this form in the NNES sub-corpus patently outnumbers the occurrences in NES. The meanings of this form in both sub-corpora happen to be identical. There is not an apparent reason to describe the massive use of this form in NNES, and only a transposition of thoughts from Spanish to English can clearly justify this in the same fashion of *can*.

Concerning their meanings, *can* is unambiguously used to indicate actualisation and potentiality. It has also been argued that *can* also expresses downtoning of the claim in modal hedges. The form *could* also indicates a dynamic meaning, but it also shows evaluation of chances as to \( P \). The forms *may* and *might* are barely epistemic, although dynamic cases are also attested in our corpus. Finally, *must* indicates an inferential process which results in the propositions it accompanies in the two sub-corpora under survey. This article represents work in progress, and more thorough analyses must be carried out in order to verify initial findings.

**References**


Ferrari, L. (2009). Marcadores de modalidad epistémico y evidencial en el análisis de las conclusiones de artículos de investigación de disciplinas distintas. ALED.


