Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance

JoAnne Neff, Emma Dafouz, Honesto Herrera, Francisco Martínez & Juan Pedro Rica

Universidad Complutense de Madrid

Mercedes Díez

Universidad de Alcalá

Rosa Prieto

Escuela Oficial de Idiomas de Valdezarza

Carmen Sancho

Universidad Politécnica de Madrid

Abstract

This article presents part of the results from research carried out by the SPICLE[1] *team on argumentative texts written in English by student writers, both native and non-native speakers from several L1 backgrounds. The aim of the study was to compare how these writers construct stance by examining their use of devices of evidentiality, specifically, modal verbs (*can, could, may, might *and* must*) and nine reporting verbs* (suggest, wonder, argue, explain, express, recognise, say, show, *and* state). *The texts of American university writers were contrasted with those produced by five EFL groups (speakers of Spanish, Dutch, Italian, French and German). The results showed that the EFL writers either overuse or underuse modal verbs in comparison with the American writers. Regarding the use of reporting verbs, native writers use a wider range of verbs, many of which carry a higher pragmatic import for stance taking. This research is significant not only for the comparison of typological and pragmatic differences but also for the study of interlanguage features and the teaching and learning of writing conventions.*

1. Introduction

It is well-known that one of the most difficult aspects of constructing argumentative texts for both native and non-native university writers is learning how to modalise both the propositions which they themselves put forth or those which they attribute to someone else. Holmes (1983: 100) has noted that the sociolinguistic competence of English as a Foreign Language students (EFL) in both speaking and writing involves learning the social and cultural values of the target community. She divides the grammatical classes of modal verbs, lexical verbs, adverbial constructions, nouns and adjectives into three categories of

devices used in expressing propositions: *personalised* (*subjective modality* in Halliday 1985: 333); *impersonalised* (*objective modality* in Halliday 1985: 333) and *depersonalised* (*abstract rhetors* in Hyland 2000). She points out that these devices serve at least two simultaneous functions: the expression of certainty/doubt concerning the proposition but also the speaker/writer's attitude towards the audience.

In line with this finding, Hinkel (1995) has shown that non-native speakers' use of modal verbs reflects the pragmatic frameworks and norms specific to first language (L1) environments, which may differ considerably from those expected in second language (L2) conceptual structures. In the scrutiny of more than seven hundred essays written by Asian non-native speakers of English, Hinkel found that, besides being topic-bound, the use of a number of modal verbs (mostly those of obligation and necessity) heavily depended on authority or moral values intrinsic to the concrete Oriental philosophy of the students.

As well, Thompson and Yiyun (1991: 366) point out that it is a common experience for teachers of Academic English to have difficulties in identifying clearly the kind of stance that EFL student writers intend to construct when citing other authors, that is, whether the student writers intend to accept the author's viewpoint or reject it. Their conclusion is that students should be trained in identifying the different layers of reporting structures and the evaluative meanings attached to them.

It appears, then, that EFL students may experience difficulties in academic writing for a number of reasons. There may be typological differences between the L1 and L2, such as the existence of a subjunctive mood, which can be used to express modality, instead of a modal verb. There may also be differences in the politeness strategies used by different discourse communities, either in the types or frequencies of the hedges and boosters used as well as in the types and number of devices used for personalised, impersonalised or depersonalised constructions, as noted by Holmes (1983). Finally, EFL writers may not realise that different evaluative meanings are attached to different reporting verbs or they may not have sufficient command of a variety of devices (hedges and boosters) with which to express the appropriate sociocultural competency.

In this study, part of a larger project[2] funded by the Spanish Ministry of Education, we compare the different constructions of stance by native and non-native writers with regard to their modalisation of propositions (Halliday 1985: 335) through the use of modal verbs (*can, could, may, might* and *must*) and reporting verbs (*argue*, *explain*, *express*, *recognise*, *say*, *show*, *state*, *suggest*, and *wonder*). These nine verbs were found to be among the twelve most frequently used in a preliminary analysis of thirty reporting verbs used by both Spanish and American university writers.[3]

By *stance*, we mean "the dialogically enacted positioning of a social agent with respect to alignment, power, knowledge, belief, evidence, affect and other socially salient categories" (Du Bois 2001). We define *modality* in the broad sense, meaning the source and the reliability of speaker/writer's knowledge, with the co-hyponyms *evidentiality*, referring to source and *modality*, referring to certainty of knowledge (Dendale and Tasmowski 2001). Thus, we accept

Palmer's (1986) proposal that evidentiality be included within modality, the former being a means for signalling epistemic attitude towards the information. The focus on modality here follows from a wider comparative study of the construction of stance by native writers (American university students and professional editorialists) and Spanish EFL university students.

Three research questions were posed:

1) Are there quantitative and/or qualitative differences between the native (American university writers) and non-native use (Dutch, French, German, Italian and Spanish) of modal and reporting verbs as evidentiality devices?
2) Are there quantitative and/or qualitative differences in the use of these devices among the various non-native groups?
3) Do the results point to typological and/or sociocultural differences in the construction of writer stance with modal and reporting verbs?

2. The modalisation of propositions

Modal verbs are used to express the speaker/writer's attitude toward the non-factual and non-temporal elements of the situation under consideration. Thus, the use of a modal verb always entails a speaker/writer's judgement or opinion. The function of non-epistemic *can*,[4] as Hoey (1994: 42) notes, is evaluative. It is less concerned with the hypothetical (signalled by *may* and *might*) than with assessing the possibility, which may be physical, of permission, or non-restriction (Lewis 1986: 104). With *can*, the situation is judged as a truly existing possibility. It signals that, in the writer's judgement, a definite possibility exists. For this reason, and particularly when denoting non-restriction, *can* is often involved in presenting a change from problem to solution, as in this text written by an American college student:

(1) What do Cindy Crawford, Kate Moss, Naomi Campbell, and Kristy Turlington have in common? All of these women are extremely beautiful, all are top notch models, but most importantly all are understood to be the "perfect and ideal" women. The 1990's base these standards or expectations on every woman, and every woman who doesn't resemble Barbie is perceived as imperfect.
… every woman **may** not look at the famous models and want to resemble them, but I admit that I am the opposite -- I only wish that I **could** be as flawless as Cindy Crawford.
… Plastic surgery is on an increase these days, it seems that everywhere you look, girls as young as fourteen are having breast implants, or facial operations, what they don't realize though, is that implants or tucks don't last forever, and are a leading cause of Cancer. Liposuction is used to remove fat, but is a very dangerous procedure that **could** lead to bleeding to death, at the slightest mistake made by a surgeon.

> … Any wealthy woman **could** go to a plastic surgeon, and explain what she wants to look like, but we all **can** understand and be truly happy with ourselves… (LOCNESS, AUW)

The function of *could* is to signal the existence of remote possibility. In example (1), stressed *could* deductively evaluates the hypotheses of being beautiful, of bleeding to death, and of wealthy women going to plastic surgeons. These instances of *could* occur in clauses for which it is not possible to either affirm or deny the hypothesis, given that the events have not taken place. The only other possibility for the reader, then, is to agree or disagree with it (Winter 1982: 198). What differentiates *could* from *might* is that the former signals *unilateral* possibility, while *might* maintains open possibility, i.e. various options may be considered (Gresset 2001).

Writers must advance their own viewpoints but should counterbalance them by introducing the arguments of others, a rhetorical strategy that student writers may not always be aware of. It is in the critical examination of evidence that the epistemic modals *may* and *might* are often used. The same relationship is true of this pair of modal verbs as between *can* and *could*, in that *might* signals a more remote possibility than *may*. But what differentiates them from *can* and *could* is that with the former the *volitionality* of the speaker/writer is involved in creating the possibility (Lewis 1986: 113). Propositions including *may* and *might* are speculations or hypotheses.

These modals and other signals of suspension of fact, such as irrealis constructions, for instance, concessive constructions with *if*, establish a speculative context in which the writer can put forward a hypothesised situation.

3. Stance-taking with reporting verbs

In their study on evaluation in reporting verbs, Thompson and Yiyun (1991: 372-373) first make a classification of these verbs into those that denote the *author's stance* (the person whose information is being reported) and the *writer's stance* (the stance adopted by the student writers themselves), both of which are concerned with the truth/correctness of the reported proposition. However, they also note that if a (student) writer reports information from another author using, for example, the verb *recognise*, the writer is giving, at the same time, an interpretation of the author's stance (in their framework, the *author's behaviour interpretation*). Thus, verbs like *suggest* and *wonder* signal the non-factive when the (student) writer puts forward a proposition; but when the proposition being presented is that of another author, the writer is also signalling his/her interpretation of the author's stance. This latter type also includes *argue* and *recognise*, as compared to factive but neutral (non-interpretative) verbs such as *say*, *state*, and *express*. *Explain* and *show*, on the other hand, signal presentation of facts, either by the writer or another author.

4. Analysis

In order to compare the expression of modality and evidentiality in a variety of non-native writing, we analysed argumentative texts from the *International Corpus of Learner English* (*ICLE*), written by non-native speakers of the following languages: Dutch, French, German, Italian and Spanish.[5] The performance of these EFL groups, representative of Romance languages (French-speaking university writers, FUW; Italian university writers, IUW; and Spanish university writers, SUW) and Germanic languages (Dutch university writers, DUW; and German university writers, GUW), was contrasted with that of a reference group of American university writers (AUW), using material from the *LOCNESS* corpus (Louvain Corpus of Native English Essays).

The analysis of data in all corpora was conducted with OUP WordSmith Tools 3.0, specifically the Wordlist and Keywords tools, and was divided into five steps. After a search for the five modal verbs and the nine reporting verbs mentioned above, the frequencies for each item in every corpus were counted. These data were then run through the Keywords tool in order to calculate the chi-square test of significance with Yates correction for a 2 X 2 table. Then, a comparison between each non-native corpus and the native reference corpus was carried out. All items showing a p value lower than 0.05 were considered statistically significant. To compensate for the uneven number of words among corpora, we normalised the statistically significant figures by 10,000 words. Finally, a graph of every item showing the statistical differences was produced. The chi-square values for all the quantitative results appear in the Appendix.

5. Findings for the modal verbs

In answer to the first two research questions, the results concerning the use of modal verbs by the non-native writers show patterns of over- and underuse of certain modals. The case of *can* is especially interesting since it is overused by all non-native writers, the highest frequencies being found in the Italian and Spanish learner writing (Figure 1).

The principal difference between the use of *can* in the native texts (AUW), on the one hand, and in the EFL student texts, on the other, is that the former use *can* less in the dynamic sense of ability and more in the sense of possibility of non-restriction. Also native uses of *can* denote a **definite possibility**, as in:

(2) Money of itself is neither good nor bad. One's attitude about money is the deciding factor that determines what influence money has. It can be an influence for evil if the need for whatever it buys is possessing the buyer … (LOCNESS, AUW)

In two L1 Romance language EFL groups, the Spanish and the Italian, *can* (which signals a **definite possibility of non-restriction**) is sometimes incorrectly used, when *could* (the remote possibility) would be more appropriate. The use of

can in this example from the Italian corpus can be compared to the use of *could* in example (1) above, or to the use of the more hypothetical *might*:

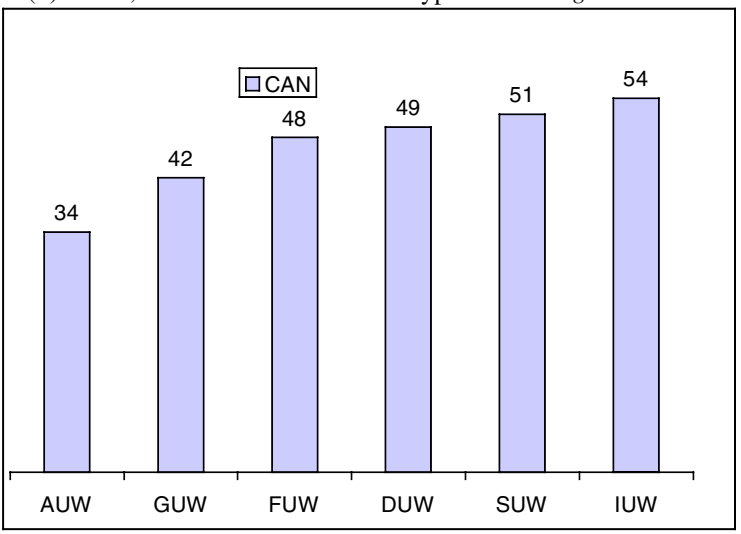

Figure 1. The use of *can* by native and non-native writers

(3) This is the first of a hundred questions that I ask to myself every time that I read or I hear about a capital execution. Sometimes I think that this is the right solution and that it can be the only way to improve our society but then when I understand that it is very difficult to lower the criminality I ask to myself how the death of someone can be considered a right thing. (IUW)

We assume that this erroneous use is caused, at least in part, by the use of the modal *can* in both the Spanish and Italian texts in an epistemic sense, which reflects the more hypothetical senses of the Spanish modal *poder* and the Italian modal *potere.* This over-extension in meaning may account for part of the drastic overuse of *can* by both the Spanish and Italian EFL writers, the two groups with the highest frequency for this modal. The overuse may also be due to a teaching effect, in that this is the first modal EFL students learn and they may thus believe that it can be used in exactly the same contexts as their own modals *poder* and *potere.*

This mistaken use of *can* contrasts sharply with the correct use of this modal by the Dutch and German EFL writers, who have more tokens of *can* + *be*, all indicating a definite possibility, as in:

(4) … is the automobile man's worst friend? It is a fact that it can be very helpful, it makes us much more independent and mobile.' (GUW)

Most of the use of can by the Dutch writers is in passive voice clauses, as in:

(5) … the same rule can be applied to artists (DUW)

In the French corpus, on the other hand, there are only fourteen instances of *can + be*, most indicating a definite possibility, although there are also three erroneous uses, as in:

(6) … now they favour a world of power and economy. This can be a reason for the disappearance of dreaming …' (FUW)

Here *could* would have been more appropriate to signal a unilateral possibility, rather than *can*, which indicates a definite possibility.

While the Dutch, French and German EFL writers also showed a significant overuse of *can*, this is not caused by choosing this modal instead of the more remote and unilateral *could*, as with the Spanish and Italian writers.

One other similarity between the Spanish and the Italian EFL writers should be noticed. That is that these two groups of writers manifest the same type of error in constructions with anticipated *it*. This results in the construction of sentences with two subjects, like the following example from the Italian corpus:

(7) Television influences people beyond the limit of danger; it can be noticed a relationship between the unjustified death of a student and the previous telecast of a film in which ... (IUW)

Or this one from the Spanish corpus:

(8) It can be appreciated a contrast between the archetypal couple that will finally get married and the secondary characters in which the author ... (SUW)

We assume that this type of error is due to typological differences of Italian and Spanish with English. That is, these constructions in the EFL student texts are probably due to transfer from the native languages, both of which are null-subject, allow for a flexible word order and permit a 'reflexive passive' construction. In this construction, a verb carries a reflexive preposed particle to construct an impersonal subject and the topic of the clause is introduced in the direct object position, as in the Spanish translation of example (8):

(9) Se puede apreciar un contraste entre la pareja arquetípica que finalmente se casa y los personajes secundarios en los que el autor ...

The opposite tendency can be observed for *could*, which was underused by all EFL student writers except for the Italian group (Figure 2).

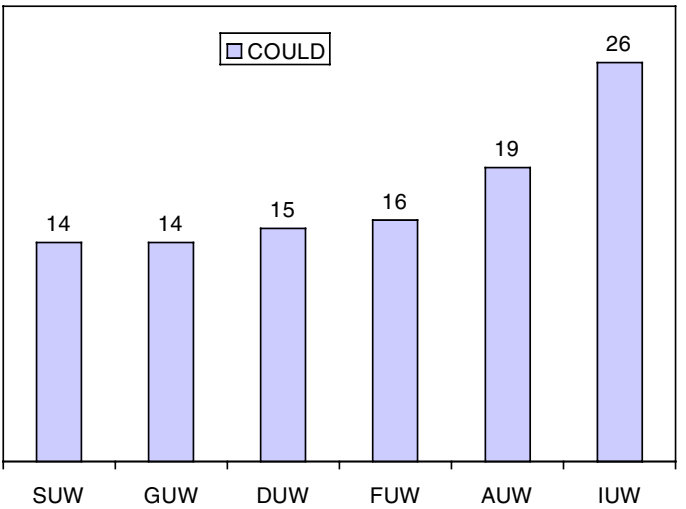

Figure 2. The use of *could* by native and non-native writers

The Spanish writers show the lowest frequency for *could*, a result which may be related to their great overuse of *can*. On the other hand, the Italian overuse of *could* may be related to having misjudged this modal (instead of *may*, or *might*) as a signal for hypothetical reasoning, as in these examples from the IUW corpus:

(10) Following these considerations for every single woman will be very difficult to obtain the permission to have an artificial insemination but according to me it would be the right thing to do: for instance, when the child will go to school or will establish contacts with other children, he could ask to himself why he has not a family as the others have, with a father and a mother and not only with a mother and this could be a problem for him. A further reason why I am against the artificial insemination for single women is that if some of them have never had sexual intercourse, the decision to gave birth to a son is against God and against human nature. (IUW)

(11) … These people are unfortunately not used to arms and so they make some fatal mistakes by exercising. But, not only may arms be deadly for people who use them, they could also injure a member of the gun owner's family as, for instance, a curious child who identifies his father's gun with a toy. (IUW)

In example (10), apart from the null subject of the first clause (also frequent in the Spanish texts), this Italian student writer uses the auxiliary *will* to indicate future in the temporal clause beginning with 'for instance,..', as it would be in Italian, and then constructs the hypothetical clause with *could*. If this is a hypothetical proposition – which it seems to be since it begins with 'for instance' – the best

solution in English would be to use the present tense in the temporal clause and the most hypothetical modal (*might*) in the main clause.

In example (11), the second clause with *could* is as speculative as the previous clause with *may*, yet the Italian student writer has used *could* as if this modal signalled a hypothesis. The usual combination of clauses with a 'not only/but also' construction involves two equal verbs, often in the present tense, but when modal auxiliaries are used, the first clause frequently has a verb in the present tense and a modal auxiliary – usually a speculative one -- in the second, as in: 'Not only is it dangerous personally, but it might also provoke an international crisis.' As Quirk *et al.* (1985: 941) note, this construction suggests 'that the content of the first clause is surprising, and that of the second clause, often reinforced by an adverb such as *also* or *even*, is still more surprising.' However, in example (11), the student writer uses a modal auxiliary in the first clause to indicate a hypothesis, but then regresses to a less speculative modal auxiliary in the second clause (*could*). This seems to suggest that the Italian writer believes that *may* and *could* have the same epistemic value. Mistaken uses of *could*, such as the ones in examples (10) and (11) on the part of the Italian student writers account for approximately 15 per cent of their 596 tokens of *could.*

As Biber *et al.* (1999) have noted, *may* is extremely common in academic prose. Thus, it is not surprising that this modal has a high frequency in the American writers' texts. In the EFL texts, *may* was underused by all the EFL groups except the French (Figure 3), who show no significant difference from the AUW, and, are, therefore, not represented on the chart. Here, the Spanish sub-corpus exhibits once again the lowest frequency, a trend which is sustained by a similar low occurrence of *might* (Figure 4). As previously noted, *may* is frequently used to put forward a hypothetical proposition. This is later negated, sometimes in a clause marked by an adversative conjunction, such as *yet* or *but*, as in example (12) or the combination of a concessive clause and a negative adjective which indicates the opposite of the expected, as in example (13), both (12) and (13) being taken from the AUW texts:

(12) She compares the lack of coverage of female athletes to the overwhelming responses to the swimsuit issue. Some may argue that the women in the magazines are better to look at than those on the court. But players are not asking for a comparison of looks, but for a sense of respect as a woman. (LOCNESS, AUW)

(13) They want the audience to see this as an uncontrollable problem, but as any college student can tell you, while all these facts may be true, they are not uncontrollable. (LOCNESS, AUW)

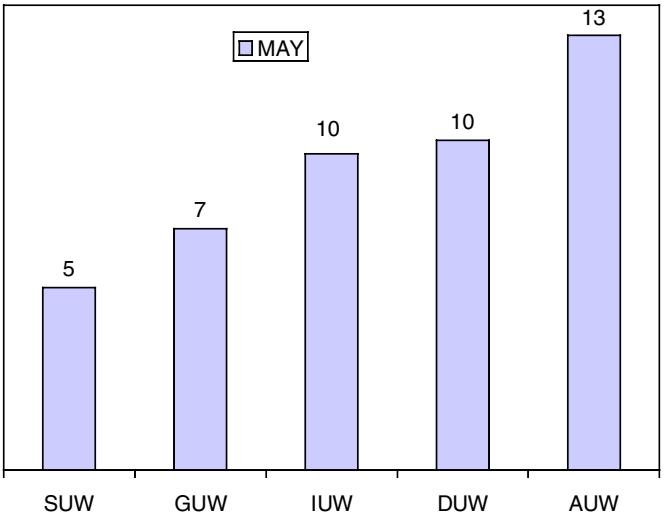

Figure 3. The use of *may* by native and non-native writers

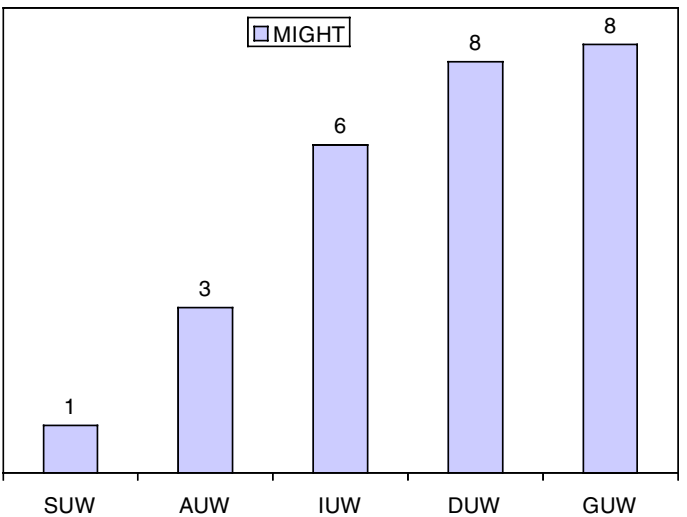

Figure 4. The use of *might* by native and non-native writers

These rhetorical strategies are infrequently used by the Spanish writers, which would account, at least in part, for the low numbers of both *may* and *might*.

As for the use of the modal verb *might*, all the EFL groups, except the SUW, have at least double the frequency shown in the texts of the American writers (Figure 4). Since *may* is almost five times as frequent in academic texts (Biber *et al.* 1999) as *might,* which signals a more remote possibility than *may*,

the EFL writers, except for the Spanish, may be interpreting *might* as more mitigating. Thus, its use (instead of *may*) may constitute a politeness strategy, as it has been found to do in unsigned editorial texts (Neff *et al.* 2001).

Regarding the use of *must*, only the writing of native speakers of Germanic languages present statistically significant differences with the reference group (Figure 5) in their underuse of this modal. This may be caused partially by a preference for verbs like the German *sollen* (should) as a politeness mitigator.

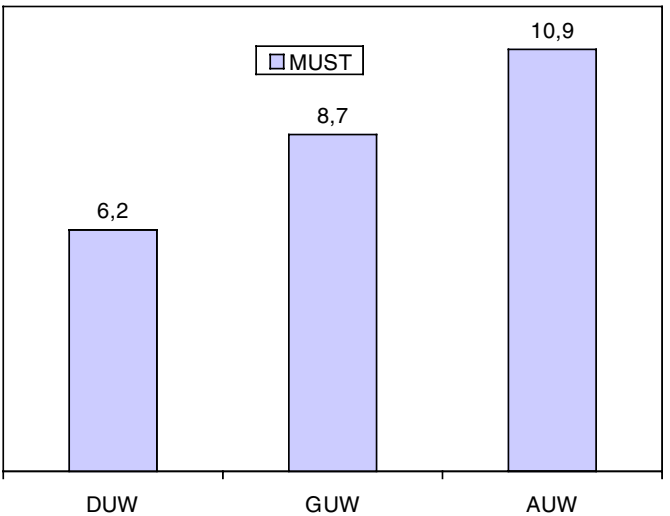

Figure 5. The use of *must* by native and non-native writers

In a previous analysis by Neff *et al.* (2000), the construction *we can* proved to have the highest frequency for two-word clusters within the SUW subcorpus. This result motivated an investigation into the uses of *we can* and several other clusters consisting of *we* + modal verb among the DUW, FUW, GUW, IUW and SUW in comparison to AUW. As can be seen in Table 1, which shows the frequencies per ten thousand words, all of the EFL texts had a statistically significant overuse of *we can*, except for the German writers.

Table 1. Cluster *we* + modal verb

	we can	*we could*	*we may*	*we might*	*we must*
AUW	1.5	0.5	0.4	0.1	1.5
DUW	3.9	—	—	—	—
FUW	11.6	2.7	2.4	0.6	2.6
GUW	—	—	—	—	0.6
IUW	7.0	1.3	—	—	—
SUW	17.7	—	—	—	2.8

The groups with an L1 Romance language — the Italians, the French, and most of all, the Spanish — had the highest frequencies for this cluster, which most frequently involved a lexical verb denoting mental or verbal processes, such as *see*, *observe*, *find*, and *say*, as in the following three examples from the Spanish texts:

(14) Regarding love we can say that it is also attainable if you have enough money to afford or if your parents agree to it (SUW)
(15) We can see the double personality of young Marlow, who is not able to maintain his position in love affairs. On the contrary, Hastings can do it very well (SUW)
(16) The problems that we can find are two: on the one hand, the overcrowding of the prisons that makes difficult, to a great extent, the application of these plans of rehabilitation (SUW)

The French texts show an overuse of *we* + all of the modal verbs: *we can* (11.6), *we could* (2.7), *we may* (2.4), *we might* (0.6) and *we must* (2.6), in comparison to the AUW frequencies. Particularly notable was their use of the clusters involving *say*: we *can say*, *we could say*, and *we may say*. Both the Spanish and the Italian texts showed the highest frequencies for *we can see*, *we can find*, and *we can say*, the SUW being the group that most overused these clusters.

Finally, the overuse of *we must* deserves further comment as it is twice as frequent in the SUW texts as in those of the AUW. A full 53% of the tokens of *we must* are used with reporting verbs such as *point out*, *add*, *appreciate*, *say*, *underline*, and *state*, as shown in the following examples:

(17) To finish, we must indicate that in 19th century people had not liberty to choose between to be Christian, atheistic or something like that;… (SUW)
(18) But the world of women is still a dropout in the political world. At last, we must state that it is due to the feminist struggle that the sexuality of women is put on the same level as the masculine one, and it is not so hidden. (SUW)

Some of the other uses of *we must* by the Spanish writers, as in example (19) appear to be quite similar to the way in which the American writers use this expression, as in (20):

(19) In my opinion we have to forget the feminism of the seventies because this term identifies us with inequality, oppression and victimization. And instead we must focus our energies on the realization of our personal and political power. However, we must take into account that we have some qualities counting against us … (SUW)
(20) Therefore, we must take responsibility for the water... (AUW)

However, the profusion of their use in the SUW texts, one coming right after another as in example (19), tends to maintain the emphasis on the **speaker's**

necessity and not on a more abstract necessity. Even if the two uses of *must* in (19) are quite different rhetorically, when *must* is used so frequently with *we*, it may appear to border on face-threatening acts which make strong impositions on the reader for common action, a tendency which Sancho (2001) found in her analysis of deontic modals in aeronautical research articles written in English by Spanish engineers.

For the most part, these *we* clusters are used by writers with L1 Romance languages to present new topics, as in examples (14) through (16) above. However, these clusters also have the pragmatic function of including the reader in the writer's discourse community and assuming that the information presented is common knowledge, instead of constructing a more impersonal reader-in-the-text stance, such as *it might be argued* (Neff, *et al.* 2001), which does not oblige the reader to take on board the proposition.

6. Findings for the reporting verbs

Table 2 displays the frequencies of reporting verbs per ten thousand words. The Dutch writers' behaviour is the closest to that of the reference group since they do not show statistically significant differences for any of the reporting verbs except for *argue*, which is underused by all the EFL writers. This reporting verb constitutes an important rhetorical device, since it allows the writer to put forward another author's argument without presupposing its acceptance, either by the writer or by the reader. It does imply, though, that the author has produced some sort of *evidence* in favour of his/her proposition.

Table 2. Reporting verb use

	suggest	*wonder*	*argue*	*explain*	*express*	*recognize*	*say*	*show*	*state*
AUW	1.2	1.2	5.3	1.4	1.5	1.6	13.2	7.4	8.6
DUW	—	—	1.4	—	—	—	—	—	—
FUW	—	—	1.1	3	2.9	0.3	19.2	9.5	0.8
GUW	0.5	2.7	1.3	—	—	0.4	—	4.8	1
IUW	2.6	—	1.8	2.6	2.6	—	15.9	—	1.4
SUW	—	—	0.9	—	—	0.6	21.4	—	0.6

Another verb predominant in the native texts but infrequent in most non-native ones is *state*, which serves approximately the same rhetorical function as that of *argue*, except that it more neutrally signals *affirm* or *express belief.*

The reporting verb *par excellence* in all sub-corpora is *say*, which, indicating *utter* or *speak*, carries less pragmatic import than either *argue* or *state.* The speakers of Romance languages use this verb much more often than the Dutch or German speakers, who show no significant difference in comparison to the AUW. Of the three L1 Romance language groups, the Spanish writers use *say* most often. The Italian and the French groups, on the other hand, show a preference for two verbs, *explain* and *express*, which are rare both in the American texts and in the rest of the EFL texts.

The general underuse of the verbs *argue* and *state* is most marked in the SUW corpus. This fact may be attributable to insufficient training or to the excessive formality and/or contextual specificity surrounding some of these verbs in Spanish: *argüir* (contend), *aducir* (adduce), *argumentar* (argue), *declarar* (state), *manifestar* (show), *sostener* (claim), *defender* (advocate). The main Spanish translation for *state* in dictionary entries, for example, is *declarar*, a verb mainly used in administrative and legal settings.

Furthermore, the overuse of *say* by all Romance language groups might bear a relationship to the contextual meanings of some verbs in the native languages with regard to the channel through which the reporting act is realised. For example, the Spanish *decir* (say) is commonly used to report both oral and written messages, which in English might be rendered differently, as in *The man says*, *The law states.*

One last case worth mentioning is the exclusive overuse of *wonder* by the German writers. Most occurrences of this verb are accompanied by a first person singular pronoun, but there are also examples with *you* and *we*, all of which denote an informal stance, as can be seen in the following examples from the GUW corpus:

(21) This means that I will have to break this habit. I wonder what the conclusion of my next essay is going to be like (GUW)
(22) You should not wonder if a man says: "The woman has to do the household, she has to cook and to stay at home with the baby and give him/her the love (GUW)
(23) The same is found in much smaller units of life: How often do we wish that the working day should be over? How often do we wonder about the ending of a book, the outcome of a story? (GUW)

In summary, the AUW show a rather balanced use of the reporting verbs *say, state, show* and *argue*, whereas in some of the non-native groups there is a heavier reliance on one reporting verb (*say*) and a much reduced repertoire of other verbs which might allow the EFL writers to report authors' propositions with greater or lesser grades of certainty or doubt. These examples from the AUW corpus, all of which allow the writers to present counterbalanced argumentation, are not common in the EFL corpora:

(24) Statistics recently state that out of all the "hate crimes committed in the United States, robbery and murder are amongst the highest" with robbery being ... (Locness, AUW)
(25) …The pro-gun activists, however, argue that firearms actually prevent murders, rapes and burglaries (AUW)
(26) Some people argue that professional soldiers would make up a group of mercenaries, way too distant from the rest of the society (AUW)

7. Conclusion

This study confirms previous research (Cook 1978; Hinkel 1995) which has shown that, owing to typological, instructional or sociocultural factors, non-native speakers encounter difficulties in the use of English modals. A case in point is the overuse of *can* by Spanish and Italian writers, in part due to transfer from the L1 epistemic meanings of the Spanish modal *poder* and the Italian modal *potere*. A large part of the EFL students' problems must surely be caused by not distinguishing between *can* as a definite possibility, and *could* as a remote possibility. On the other hand, they may also have difficulties in discriminating between *could* as a unilateral possibility, and *might* as signalling various options and therefore facilitating the construction of a reader-in-the-text. The occurrence of clusters of *we* + modal verb + verb of mental/verbal process in the texts of writers with an L1 Romance language may also reveal a transfer of positive politeness strategies (i.e. towards affiliation and commitment). The Spanish results, for example, suggest a transfer from Spanish to English of the sociolinguistic norms for formal writing.

It remains to be seen how developmental factors (i.e. interlanguage stages) shape the informants' modal choices in each subcorpus. It will also be necessary to identify the causes of large quantitative differences in the use of certain reporting verbs among EFL writers with L1 languages pertaining to the same family, for example, regarding the overuse of the reporting verbs *explain* and *express* in both the FUW and IUW subcorpora, but not in the Spanish subcorpus. However, in other cases, writers with a Romance L1 do behave similarly. For instance, although frequently used by all EFL groups, *say* is – of all reporting verbs with a low illocutionary force – most overused by the EFL writers with an L1 Romance language. This suggests that these non-native students, even at quite advanced stages, have not acquired a broad enough range of reporting verbs from which to select the most appropriate for the context. In this way, the study has implications for the teaching of EFL academic writing. Further contrastive studies should be carried out in the three Romance languages – French, Italian and Spanish – to determine the extent of possible L1 influence on the use of reporting verbs, that is, whether the L1 also uses a narrow range of reporting verbs in argumentative writing.

Two further teaching implications can be drawn. Firstly, it would be useful to raise awareness of the inherent difficulties presented by certain modal verbs because of their lack of exact correspondence with those of the native language. For this purpose, modal verbs in native texts should be analysed *in context*, so that EFL students can better grasp the difference between *sets of modals*. Secondly, it would be advisable to provide non-natives with a detailed contrastive view of epistemic and pragmatic modal contexts in L1 and L2, placing emphasis on the stylistic and communicative effects produced by different hedges and boosters.

Notes

1 *SPICLE* is the Spanish team which contributed texts of Spanish university writers to the *International Corpus of Learner English* (*ICLE*), Centre for English Corpus Linguistics, Université catholique de Louvain, Belgium.

2 The larger study is titled 'A Contrastive Analysis of Evidentiality in English and Spanish: A Corpus Study', BFF2000-0699-C02-01, Universidad Complutense, Madrid.

3 One reporting verb frequently used by the American university writers, *claim*, was used only seven times by the Spanish EFL writers, and, thus, was not considered in this analysis.

4 *Can* is only epistemic in interrogative and negative clauses, as in '*Can* it be that ...?'

5 The number of words contained in each EFL corpus were as follows: Dutch, 237,631 words; French, 287,683 words; German, 203,647 words; Italian, 226,988 words and Spanish, 194,845 words. The American university texts from the *LOCNESS* corpus totalled 149,790 words.

References

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999), *Longman Grammar of Spoken and Written English*. London: Longman.

Cook, W. (1978), 'Semantic structure of English modals', *TESOL Quarterly*, 12: 5-16.

Dendale, P. and L. Tasmowski (2001), 'Introduction: evidentiality and related notions', *Journal of Pragmatics*, 33: 339-348.

Du Bois, J. (2001), 'Taking a stance: constituting the stance differential in dialogic interaction', paper, American Anthropological Association, San Francisco (November).

Gresset, S. (2001), 'Towards a contextual micro-analysis of the non-equivalence of MIGHT and COULD', *I International Congress on Modality*, University of Verona.

Halliday, M. A. K. (1985), *An Introduction to Functional Grammar*. London: Edward Arnold.

Hinkel, E. (1995), 'The use of modal verbs as a reflection of cultural values', *TESOL Quarterly*, 29: 325-341.

Hoey, M. (1994), 'Signalling in discourse', in: M. Coulthard (ed.), *Advances in Written Text Analysis*. London: Routledge. 26-45.

Holmes, J. (1983), 'Speaking English with the appropriate degree of conviction', in: C. Brumfit (ed.), *Learning and Teaching Languages for Communication: Applied Linguistic Perspectives*. London: Center for Information on Language Teaching and Research. 100-121.

Hyland, K. (2000), *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.

Lewis, M. (1986), *The English Verb: An Exploration of Structure and Meaning*. London: Language Teaching Publications.

Neff, J., E. Dafouz, M. Díez, H. Herrera, F. Martínez, R. Prieto, J. P. Rica and C. Sancho (2001), 'Subjective and objective modalization of evidentiality in native and non-native argumentative texts', *I International Congress on Modality*, University of Verona.

Neff, J., E. Dafouz, M. Díez, F. Martínez, R. Prieto and J .P. Rica (2000), 'The construction of writer stance in native and non-native texts', *XVIII Congreso Nacional AESLA*, Universidad de Barcelona.

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

Palmer, F.R. (1986), *Mood and Modality*. Cambridge: Cambridge University Press.

Sancho, C. (2001), 'Epistemic tides and personal strongholds: embedded discursive spaces in the research article', *XIII European Symposium on Language for Specific Purposes*, University of Vaasa, Finland.

Thompson, G. and Y. Yiyun (1991), 'Evaluation in the reporting verbs used in academic papers', *Applied Linguistics*, 12(1): 365-382.

Winter, E. (1982), *Towards a Contextual Grammar of English: The Clause and its Place in the Definition of Sentence*. London: Allen & Unwin.

Appendix

Modal verb use per EFL group as compared to the control group, the American university writers (WordSmith Output)
[D = Dutch; A = American; S = Spanish; G = German; F = French; I = Italian]

	Word	Freq. DUW	% DUW	Freq.AUW	% AUW	Keyness	P (Chi square)
D vs. A	can	1.173	0,49	514	0,34	47,6	0,000000
D vs. A	might	189	0,08	48	0,03	33,1	0,000000
D vs. A	may	236	0,1	196	0,13	7,9	0,004881
D vs. A	could	365	0,15	290	0,19	8,5	0,003598
D vs. A	must	148	0,06	163	0,11	24,2	0,000001
	Word	**Freq. SUW**	**% SUW**	**Freq.AUW**	**% AUW**	**Keyness**	**P (Chi square)**
S vs. A	can	997	0,51	514	0,34	54,7	0,000000
S vs. A	could	281	0,14	290	0,19	12,2	0,00048
S vs. A	might	18	0,01	48	0,03	21,8	0,000003
S vs. A	may	107	0,05	196	0,13	54,7	0,000000

	Word	**Freq. GUW**	**% GUW**	**Freq.AUW**	**% AUW**	**Keyness**	**P (Chi square)**
G vs. A	might	169	0,08	48	0,03	35,7	0,000000
G vs. A	can	858	0,42	514	0,34	13,4	0,000247
G vs. A	could	294	0,14	290	0,19	12,4	0,000432
G vs. A	may	148	0,07	196	0,13	29,4	0,000000
G vs. A	must	177	0,09	163	0,11	4,1	0,043297
	Word	**Freq. FUW**	**% FUW**	**Freq.AUW**	**% AUW**	**Keyness**	**P (Chi square)**
F vs. A	can	1.377	0,48	514	0,34	41,7	0,000000
F vs. A	could	458	0,16	290	0,19	6,6	0,010031
	Word	**Freq. IUW**	**% IUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
I vs. A	can	1.215	0,54	514	0,34	72,5	0,000000
I vs. A	could	596	0,26	290	0,19	18	0,000022
I vs. A	might	144	0,06	48	0,03	16,9	0,00004
I vs. A	may	216	0,1	196	0,13	10,2	0,001405

Use of *we* + modal verb per EFL group as compared to the control group, the American university writers (WordSmith Output)
[D = Dutch; A = American; F = French; G = German; I = Italian; S = Spanish]

	Word	**Freq. DUW**	**% DUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
D vs. A	we can	92	0,04	23	0,02	16,1	0,00006
	Word	**Freq. FUW**	**% FUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
F vs. A	we can	335	0,12	23	0,02	121,9	0,000000
F vs. A	we could	77	0,03	7	>0,01	23,9	0,000001
F vs. A	we may	69	0,02	6	>0,01	21,8	0,000003
F vs. A	we must	74	0,03	22	0,01	5	0,025697
F vs. A	we might	16	>0,01	1	>0,01	4,9	0,027209
	Word	**Freq. GUW**	**% GUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
G vs. A	we must	12	>0,01	22	0,01	6,1	0,01386

	Word	**Freq. IUW**	**% IUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
I vs. A	we can	159	0,07	23	0,02	54,8	0,000000
I vs. A	we could	29	0,01	7	>0,01	5,4	0,020342
	Word	**Freq. SUW**	**% SUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
S vs. A	we can	344	0,18	23	0,02	205,3	0,000000
S vs. A	we must	54	0,03	22	0,01	5,9	0,014792

Use of reporting verbs per EFL group as compared to the control group, the American university writers (Wordsmith Output)
[D = Dutch; A = American; F = French; G = German; I = Italian; S = Spanish]

	Word	**Freq. DUW**	**% DUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
D vs. A	admit	39	0,02	10	>0,01	6,1	0,013229
D vs. A	indicate	15	>0,01	2	>0,01	4,1	0,042514
D vs. A	maintain	4	>0,01	14	>0,01	10	0,001547
D vs. A	argue	33	0,01	80	0,05	47,9	0,000000
D vs. A	state	51	0,02	129	0,09	81,3	0,000000
	Word	**Freq. FUW**	**% FUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
F vs. A	imply	72	0,03	5	25,1	0,3	0,000001
F vs. A	say	552	0,19	198	0,13	20,2	0,000007
F vs. A	explain	87	0,03	21	0,01	9,9	0,001693
F vs. A	express	83	0,03	22	0,01	7,7	0,005659
F vs. A	show	272	0,09	111	0,07	4,5	0,034371
F vs. A	recognize	9	>0,01	0,3	0,02	0,0	0,000008
F vs. A	argue	33	0,01	80	0,05	65,5	0,000000
F vs. A	state	22	>0,01		0,09	0,0	0,000000
	Word	**Freq. GUW**	**% GUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
G vs. A	wonder	54	0,03	18	0,01	8,2	0,004163
G vs. A	admit	33	0,02	10	>0,01	5,7	0,017139
G vs. A	suggest	10	>0,01	18	0,01	4,6	0,031205

G vs. A	conclude	7	>0,01	16	0,01	5,9	0,01521
G vs. A	note	3	>0,01	11	>0,01	6,1	0,013516
G vs. A	show	98	0,05	111	0,07	9,4	0,002142
G vs. A	recognize	9	>0,01	24	0,02	11,2	0,000803
G vs. A	argue	27	0,01	80	0,05	44,7	0,000000
G vs. A	state	20	>0,01	129	0,1	117,4	0,000000
	Word	**Freq. IUW**	**% IUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
I vs. A	suggest	60	0,03	18	0,01	8,4	0,003798
I vs. A	explain	60	0,03	21	0,01	5,9	0,015098
I vs. A	express	60	0,03	22	0,01	5,2	0,022655
I vs. A	say	360	0,16	198	0,13	4,1	0,043375
I vs. A	argue	40	0,02	80	0,05	35,2	0,000000
I vs. A	state	31	0,01	129	0,09	109,9	0,000000
	Word	**Freq. SUW**	**% SUW**	**Freq. AUW**	**% AUW**	**Keyness**	**P (Chi square)**
S vs. A	say	416	0,21	198	0,13	31	0,000000
S vs. A	disagree	4	>0,01	12	>0,01	5,3	0,021868
S vs. A	recognize	11	>0,01	24	0,02	8	0,004711
S vs. A	claim	16	>0,01	32	0,02	9,6	0,001951
S vs. A	argue	17	>0,01	80	0,05	58,5	0,000000
S vs. A	state	11	>0,01	129	0,09	133,1	0,000000