

# A Lexical Analysis of Modal Expressions of *Certainty* in Scientific Discourse<sup>1</sup>

Teresa Marqués Aguado

*Universidad de Málaga*

## Abstract

Scientific discourse has been traditionally described as a register deprived of subjectivity, where the presence of the author is rarely felt. However, recent studies (Swales, 1990; Hyland, 1998) have shown that authors take a stance on what is being argued by using modality (among other devices), which is conveyed not only by means of modal verbs, but also by way of other lexical items (namely nouns, adjectives, adverbs and lexical verbs). In the light of this, the aim of this paper is to explore the use of these lexical expressions in a corpus of scientific research articles amounting to 170, 339 tokens. For the purpose, the devices used to convey *certainty* will be grouped into word classes, paying attention to any outstanding differences between the abstracts and the articles themselves.

## 1. Introduction

Traditionally, scientific discourse has been regarded as a register characterized by its objectivity. However, several scholars, such as Swales (1990) or Hyland (1998), have researched how authors' visibility is felt in scientific discourse. Accordingly, authors try to mitigate statements in order to gain the acceptance of the scientific community they belong to, thus avoiding categorical assertions of claims. Therefore, although they must devise means of persuading readers (Hyland, 1996: 435), they cannot forget either to keep the

---

<sup>1</sup> The present research has been supported by the Spanish Ministry of Science and Technology (grant number HUM2004-01075/FILO), the Autonomous Government of Andalusia (grant number HUM-2609), as well as the Ministry of Education (FPU grant AP-2004-0244). These grants are hereby gratefully acknowledged.

## PROCEEDINGS 31<sup>ST</sup> AEDEAN CONFERENCE

“neutral and unmarked tone (as this is a basic feature of specialized discourse)” (Gotti and Dossena, 2001: 14). In this social and communicative environment, modality is used for a variety of purposes, such as furthering an argumentation, indicating politeness, or marking the movement from observed data to generalizations, all of which relate to the so-called “interpersonal aspects of language” (Vihla, 2000: 600).

Modality has been defined by Bybee and Fleischman as “a semantic domain [...] whose common denominator is the addition of a supplement or overlay of meaning to the most neutral semantic value of the proposition of an utterance, namely factual and declarative” (1995: 2). Likewise, Quirk *et al.* define it as “the manner in which the meaning of a clause is qualified so as to reflect the speaker’s judgement of the likelihood of the proposition it expresses being true” ([1985] 2003: 219).

In this same line, and as opposed to the traditional view that modality is restricted to modal verbs, recent studies have focused on other expressions of modality (Holmes, 1988; Kourilová, 1994; Hyland, 1998; Koutsantoni, 2004, etc.). Along with central modals (*can, could, may, might, shall, should, will, would* and *must*), we find marginal modals (*ought to, dare, need* and *used to*), modal idioms (*had better*), semi-auxiliaries (*be going to, be bound to*), etc. (Quirk *et al.*, [1985] 2003: 135-146). However, the expression of modality extends beyond the scope of the verb phrase to other lexical items, such as nouns, adjectives and adverbs, grouped as “lexical modals” (Huddleston and Pullum, 2002: 173). In the same way, Gotti and Dossena refer to “quite diverse items, the most common being verbal, adverbial, adjectival and nominal expressions, together with particles, clitics” (2001: 10-11). The same view is also found in Perkins—who takes modal verbs as the starting point for his analysis of modal expressions in English (1983: 2)—, *English Grammar: an Outline* (Huddleston, 1988: 79), or the *Oxford English Grammar* (Greenbaum, 1996: 81).

Three kinds of modality have been established, following the approach set up by von Wright in the field of modal logic (1951: 1-2 in Hoye, 2005: 1486):

- Deontic modality, which involves obligation, permission and prohibition.
- Epistemic modality, which comprises possibility and necessity.
- Dynamic modality, which involves ability and willingness.

Various modifications of this basic taxonomy have been proposed, such as Palmer's grouping into propositional and event modality, the former including epistemic and evidential modality, and the latter comprising deontic and dynamic (2001: 8). Other approaches have to do with the classification into epistemic and root modality (Coates, 1983: 18-21; Gotti *et al.*, 2002: 21), the distinction between intrinsic and extrinsic (Greenbaum, 1996: 80; Biber *et al.*, 2000: 485); and Bybee's classification (1985) into epistemic, speaker-oriented and agent-oriented modality (in Bybee and Fleischman, 1995: 6). Notwithstanding these terminological differences, the analysis of *certainty* presented in this paper is to be located within the framework of the classical three-fold distinction, and more specifically within the scope of epistemic modality, as it forms part of the wider notion of possibility.

In the light of this, the aim of this paper is to carry out a study of modality conveyed by lexical expressions in a corpus of scientific research articles (RA), which amounts to 170, 339 tokens. More specifically, we will taxonomize those lexical items (nouns, adjectives, adverbs and lexical verbs) that exclusively convey *certainty*, which is one of the values under the scope of epistemic modality. In general terms, "with certainty markers authors emphasise their certainty and conviction to claims and ideas" (Koutsantoni, 2004: 172), and they also address their readers as peers who are familiar with the reasoning reflected in the RA.

The classification of these expressions into word classes will be based on the qualitative analysis of the statistical data retrieved, and major differences arising between the texts and the abstracts will be noted. The results will be eventually compared to those presented by Holmes (1988), Hyland (1998) and Koutsantoni (2004), who have also examined the value of *certainty* in ESP. Alternative expressions of *certainty* not found in these works have been supplied by the *Longman Synonym Dictionary* (1986). Accordingly, the paper is divided into three sections: first, the methodology; second, the analysis of the data; third, the conclusions.

## 2. Methodology

The corpus used for the present study was compiled from the journal entitled *Thin-Walled Structures*, published by ELSEVIER<sup>2</sup>. It has been chosen on account of two reasons: first, it publishes articles by authors from different countries and backgrounds on a world-wide scale; and second, it is ranked among the leading publications in the field.

Volume 42, which comprises 12 issues and which was published in 2004, was selected and downloaded, but some RA were excluded, as they focused almost entirely on mathematical demonstrations, and so have been those sections within articles dealing with formulae, thus rendering a total of 44 RA. The reason is that *Wordsmith Tools* (Scott, 1998), the software tool used for extracting the data, counts formulae and scientific notations as words, thus distorting the results obtained.

For the purpose of analysis, the articles were downloaded into Word files and then, each of them was split into 3 smaller files: one including the abstract, another with the main text, and the third one with the references, which were discarded. This procedure will allow us to detect variations in the use of modal expressions between the abstracts and the body of the text.

When downloading texts from webpages, some distortions and errors may occur, and they had to be amended manually. This process of revision comprised changes such as deleting the remaining formulae and hyphenation, given that *Wordsmith Tools* understands hyphenated words as two independent items.

The frequency lists were retrieved by means of the programme *Wordlist*. Taking as input the .txt files generated from the Word files mentioned above, two wordlists were generated, one relating to the abstracts and one relating to the texts. Once the relevant items were selected from the wordlists, their concordances were retrieved by way of the programme *Concord* in order to analyse the context where the items occur so as to discern whether they convey modal meanings, as in

---

<sup>2</sup> I would like to thank Melchor Serrano Blanca for allowing me to consult and use the data of his end-of-the-year project (Mechanical Engineering), entitled *Creación de una base de datos léxica de términos utilizados en ingeniería relacionados con las estructuras metálicas y de hormigón*.

the case of *certain*, which can be also used as a restrictive adjective. In this fashion, the quantitative analysis will be supplemented by a qualitative analysis of the data (Koutsantoni, 2004: 165).

### 3. Analysis of the lexical expressions conveying *certainty*

This section presents the analysis of the expressions conveying *certainty* and their classification into word classes, along with a short general statistical analysis. Those cases in which clauses, the tone of the sentence or “discourse-based expressions of confidence” (Koutsantoni, 2004: 172) convey *certainty* will be left aside, as they fall outside the scope of our investigation.

#### 3.1. Statistical data

In table 1, the ten most frequent words in the texts and the abstracts are compared in order to check whether any differences emerge. The table also includes the data relating to the COBUILD corpus (Sinclair, 1991) to check the reliability of our corpus, as well as to test whether there are any salient differences which might be attributed to the scientific character of the corpus under scrutiny:

Table 1. Ten most frequent words

POSITION	TEXTS	ABSTRACTS	COBUILD
1	The	The	The
2	Of	Of	Of
3	And	And	And
4	In	In	To
5	A	To	A
6	Is	Is	In
7	To	A	That
8	For	Are	I
9	With	For	It
10	Are	With	Was

## PROCEEDINGS 31<sup>ST</sup> AEDEAN CONFERENCE

Most of the words in the three lists are functional words, but there are minor differences regarding the position of some of them (for instance, *to*, *a* or *in*). Moreover, when comparing the texts and abstracts wordlists to the COBUILD one (Sinclair, 1991: 143), the main difference lies in the presence of personal pronouns, which are not included in the top positions in our corpus, as they reveal a certain degree of subjectivity.

Although the focus of our analysis lies in expressions other than modal verbs, those which can potentially express *certainty*, among other values —i.e. *can*, *could*, *must*, *will*, *would* (Holmes, 1988: 28)—, are frequent: ×881 in the case of texts and ×27 in the case of abstracts. To avoid any possible corpus-length dependency (the set of texts amounts to 162,830 tokens, as opposed to the 7,509 tokens of abstracts), frequencies have been normalised to a text of 104 words (see Biber, 1988: 13–14), thus rendering 54.1 and 35.95 in the case of modal verbs, respectively.

### 3.2. *Nouns*

Nominal expressions conveying *certainty* have received little attention in previous research (for instance, they are not mentioned by Koutsantoni, 2004) and, as a matter of fact, they are not very frequent in the corpus under study either. The list of nouns below (see table 2) is supplemented by a couple of examples:

- (1) The VALIDITY of the proposed technique has been established (abstracts)
- (2) Numerical simulations have provided clear EVIDENCE that, besides “rigid-body” motions, the compressed range may also experience non-negligible flexural deformations (texts)

Table 2. Nouns conveying *certainty*

TEXTS		ABSTRACTS	
<i>Item</i>	<i>Frequency</i>	<i>Item</i>	<i>Frequency</i>
Validity	14	Reliability	3
Verification	9	Validation	2
Evidence	6	Validity	2
Reliability	5	Effectiveness	1
Assertion	1		
Certainty	1		
Clarity	1		
Reality	1		
TOTAL	38	TOTAL	8

Only two of the nouns listed by Holmes (1988: 36-37) are found in our corpus (i.e. *certainty* and *evidence*), which can be attributed to the fact that she deals with *doubt* rather than with *certainty*.

### 3.3. Adjectives

Some of the occurrences of the adjectives listed below (see table 3) express *certainty*, whereas others do not, as evinced by the following pair of sentences: in (3) *certain* functions as a restrictive adjective, whereas it expresses *certainty* in (4):

- (3) The frequency increases initially with the increase of the tensile load, but at CERTAIN loading level it starts to decrease (texts)
- (4) From the above discussion, IT IS CERTAIN THAT the new method presented for building the dynamic equation of sandwich plate with VEM core has good accuracy and reliability (texts)

Table 3. Adjectives conveying *certainty*

TEXTS		ABSTRACTS	
<i>Item</i>	<i>Frequency</i>	<i>Item</i>	<i>Frequency</i>
Significant	34	Good	5
Effective	13	Effective	3

**PROCEEDINGS 31<sup>ST</sup> AEDEAN CONFERENCE**

Clear	12	Correct	2
Evident	12	Excellent	1
Reliable	6	(To make) sure	1
Obvious	4	Universal	1
True	4	Valid	1
(To be) sure	2		
Certain	1		
Inevitable	1		
TOTAL	89	TOTAL	14

As shown in (4), some of these adjectives (*clear, obvious*) tend to appear in the structure *it is clear/obvious (from X) that*, where X stands for the (optional) source from which the conclusion arises, whereas others (*inevitable, reliable or universal*) are found in either attribute or predicative positions:

- (5) The <GBT>-based formulae are both accurate and UNIVERSAL (abstracts)
- (6) Therefore RELIABLE results can be expected (texts)

*3.4. Adverbs*

Most of the modal adverbs occurring in our corpus (see table 4) are mobile insofar as they can occupy different positions in the sentence, but they usually modify the whole clause. The group of so-called *downtoners* has been excluded as they typically have “a generally lowering effect” (Quirk et al., [1985] 2003: 445), which clashes with our evaluation of certainty markers:

- (7) ACTUALLY, the failure of thin-walled cold-formed members in compression always occurs with a local plastic mechanism (abstracts)
- (8) The test results summarised [...] CLEARLY show that, generally, good concrete compaction results in higher beam-column member capacities (texts)



Table 4. Adverbs conveying *certainty*

TEXTS		ABSTRACTS	
<i>Item</i>	<i>Frequency</i>	<i>Item</i>	<i>Frequency</i>
Clearly	19	Effectively	2
Indeed	12	Actually	1
Effectively	9	Clearly	1
Essentially	8	Essentially	1
Obviously	7		
Correctly	5		
Actually	3		
Invariably	2		
Definitely	1		
Evidently	1		
Inevitably	1		
Markedly	1		
Truly	1		
TOTAL	70	TOTAL	5

3.5. *Lexical verbs*

Most of the lexical verbs listed below (see table 5) indicate “that any reasonable and informed reader would draw the same conclusion” (Hyland, 1998: 123), while others (called *epistemic evidential verbs* in Hyland’s terms [1998: 124]) simply echo others’ opinions or suggestions (*assume, conclude, claim*):

Table 5. Lexical verbs conveying *certainty*

TEXTS		ABSTRACTS	
<i>Item</i>	<i>Frequency</i>	<i>Item</i>	<i>Frequency</i>
Assume	92	Show	15
Know	23	Find	7
Conclude	21	Assume	5
Demonstrate	20	Conclude	4
Confirm	16	Validate	4
Reveal	10	Demonstrate	3

## PROCEEDINGS 31<sup>ST</sup> AEDEAN CONFERENCE

Verify	8	Report	2
Prove	6	Verify	2
Believe	5	Confirm	1
Hold	2		
Approve	1		
Claim	1		
TOTAL	205	TOTAL	43

Although Hyland states that the active prevails in epistemic modality (1998: 122), the most frequent lexical verb (*assume*), among others, is in the passive most of the times, which may serve to emphasise *certainty*:

- (9) The study further CONCLUDES that the behavior of perforated rectangular plates (abstracts)
- (10) It must be remembered that axisymmetry IS ASSUMED in the current study (texts)

### 4. Conclusions

In this corpus-driven analysis of modality, lexical items that convey *certainty* have been classified, whereby authors try to publicise and ratify their claims, as well as to persuade readers. Although authors often adopt cautious positions, they may also employ certainty markers to emphasise a particular point, addressing their readers as knowledgeable peers.

If compared with previous research, our taxonomy is somewhat more reduced than those provided by Kourilová, Holmes or Hyland, but this might be attributed to the fact that the focus of our research lies exclusively on *certainty*. Some items which were expected *a priori* (such as *faithful* or *truth*, or negative items recorded by Holmes e.g. *unquestionably*, *undoubtedly*, *(no) doubt*, etc.) are not, however, found in our corpus, although the set of certainty marker varies from corpus to corpus. Similarly, the number of items varies in each word class, although only some of them are used regularly (such as *clearly*, *significant*, *assume* or *validity*), and this may evince certain preferences on the part of

the authors of these RA, although the scope of the present research does not allow for such generalizations.

At any rate, the distribution of modal expressions into word classes is shown in table 6:

Table 6. Distribution of lexical items into word classes

LEXICAL ITEMS	TEXTS		ABSTRACTS	
<i>Nouns</i>	38	2.33	8	10.65
<i>Adjectives</i>	89	5.46	14	18.64
<i>Adverbs</i>	70	4.29	5	6.65
<i>Lexical verbs</i>	205	12.58	43	57.26
TOTAL	402	24.68	70	93.22

If checked against the whole corpus, the figure for texts and abstracts together ( $\times 472$ ) renders a frequency of 27.7 (normalised to a text of  $10^4$ ); but, if examined independently, abstracts are richer than texts in certainty markers, even though the opposite tendency was observed as regards modal verbs. As for word classes, the data in table 6 show that almost the same order is maintained in both columns, excepting nouns and adverbs, which shift their positions. Irrespective of the normalised figures, lexical verbs are the most common lexical item in both texts and abstracts. Furthermore, the fact that abstracts are typically constrained in terms of length probably accounts for the low frequency of adverbs.

Furthermore, studies of modal expressions used in RA (and not only those expressing *certainty*, but also *possibility*, *obligation*, etc.) can constitute a helpful tool for ESP teachers and textbooks, as they can cater for their students' linguistic needs in depth, rather than by resorting exclusively to modals, because, as the frequencies above evince, modal expressions are by no means subsidiary to modal verbs. In this line, a study of the ratio between the latter and other lexical expressions of modality is necessary in order to gain a global view of the various ways in which modality is expressed, not only in RA as a text type, but also in relation to the particular sections of RA.

References

- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP.
- Biber, D. et al. 2000. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bybee, J. and S. Fleischman. 1995. *Modality in Grammar and Discourse*. Amsterdam & Philadelphia: John Benjamins.
- Coates, J. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
- Gotti, M. and M. Dossena. Eds. 2001. *Modality in Specialized Texts: Selected Papers of the 1<sup>st</sup> CERLIS Conference*. Bern, Berlin, Bruxelles, Frankfurt/M., New York, Oxford, Wien: Peter Lang.
- Gotti, M. et al. Eds. 2002. *Variation in Central Modals: a Repertoire of Forms and Types of Usage in Middle English and Early Modern English*. Bern, Berlin, Bruxelles, Frankfurt/M., New York, Oxford, Wien: Peter Lang.
- Greenbaum, S. 1996. *The Oxford English Grammar*. Oxford: OUP.
- Holmes, J. 2004. "Doubt and Certainty in ESL Textbooks". *Applied Linguistics* 9:1. 21–44.
- Hoye, L.F. 2005. "'You may think that; I couldn't possibly comment!' Modality Studies: Contemporary Research and Future Directions. Part II". *Journal of Pragmatics* 37. 1481–1506.
- Huddleston, R. 1988. *English Grammar: an Outline*. Cambridge: CUP.
- Huddleston, R. and G. K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Hyland, K. 1996. "Writing Without Conviction? Hedging in Science Research Articles". *Applied Linguistics* 17:4. 433–454.
- Hyland, K. 1998. *Hedging in Scientific Research Articles*. Amsterdam, Philadelphia: John Benjamins.
- Kourilová, M. 1994. "Modal Meanings in Medical Science Reporting". *Languages for Specific Purposes*. W. Scott and S. Mülhaus. Eds. 65–67. Kingston upon Thames: CILT and Kingston University.
- Koutsantoni, D. 2004. "Attitude, Certainty and Allusions to Common Knowledge in Scientific Research Articles". *Journal of English for Academic Purposes* 3. 163–182.
- Krug, M. G. 2000. *Emerging English Modals: a Corpus-based Study of Grammaticalization*. Berlin, New York: Mouton de Gruyter.
- Longman Synonym Dictionary*. 1986. Harlow: Longman.
- Palmer, F. R. 2001. *Mood and Modality*. Cambridge, New York: CUP.
- Perkins, M. R. 1983. *Modal Expressions in English*. London: Frances Pinter.
- Scott, M. 1998. *WordSmith Tools Version 3.0*. Oxford: OUP.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Quirk, R. et al. [1985] 2003. *A Comprehensive Grammar of the English Language*. Essex: Longman.

**TERESA MARQUÉS AGUADO**

- Swales, J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.
- Vihla, M. 2000. “Modal Expressions: Modifying Information”. *Neuphilologische Mitteilungen* 101. 600–603.