A cross-language study on citation practice in PhD theses

CARMEN SOLER-MONREAL
LUZ GIL-SALOM*
Universidad Politécnica de Valencia, Spain

ABSTRACT

Citation practice provides justification for arguments and allows a writer to indicate a rhetorical gap for her/his research and adopt a tone of authority. Claims must be supported with evidence, and writers must demonstrate an understanding of approaches and knowledge in their fields of specialisation, in order to persuade the examiners that the thesis is worthy of the award of a doctorate (Thompson, 2005b). Candidates need to keep the adequate interpersonal relationship with the immediate audience (the examiners). They have to evaluate the previous research in an area of study and to be respectful with previous claims from authorities in the disciplines. They also need to position themselves in relation to other disciplinary members and highlight their individual claims. This paper investigates contrastively how interactional resources of citation and, in particular reporting verbs, are deployed in the literature review chapters of PhD theses written in English and in Spanish.

KEYWORDS: Citation practice, reporting verbs, academic writing, PhD theses, evaluation, stance, variation, cross-language study

RESUMEN

El uso de citas y referencias proporciona una justificación para la argumentación y permite al académico identificar un hueco o espacio retórico donde situar su investigación y adoptar cierto tono de autoridad. Las hipótesis o afirmaciones científicas deben apoyarse en evidencias, y los investigadores deben demostrar un buen conocimiento de los enfoques de su campo de especialización, para así persuadir a los miembros del tribunal examinador de la tesis de que ésta es digna de merecer el doctorado (Thompson, 2005b). Los doctorandos necesitan mantener una relación interpersonal apropiada con su audiencia (el tribunal examinador). Asimismo, necesitan evaluar la literatura de su área de estudio y ser respetuosos con las afirmaciones sobre investigaciones previas de autoridades de la disciplina. Necesitan adoptar una postura ante los miembros de la comunidad científica a la vez que enfatizan sus propias afirmaciones. En este trabajo se lleva a cabo un estudio contrastivo sobre el uso de las citas como recursos interactivos y, en particular, el uso de los verbos introductorios en los capítulos que tratan la revisión de la literatura en tesis doctorales en inglés y en español.

PALABRAS CLAVE: Citas y referencias, verbos introductorios, escritura académica, tesis doctorales, evaluación, postura, variación, estudio contrastivo.

*Address for correspondence: Carmen Soler-Monreal/Luz Gil-Salom. Universidad Politécnica de Valencia. Departamento de Lingüística Aplicada. Camino de Vera 14, 46022. Valencia. SPAIN. E-mail: csoler@upv.es

1. INTRODUCTION

Research theses constitute a key genre used by scientific communities for the dissemination of knowledge. They are widely considered to be the first step to access academic discourse and obtain membership in disciplinary communities. For new claims made in theses to become accepted and for titles to be awarded, examiners need to be persuaded of the validity of these claims. Writers are conscious of what that position implies. Thus, the process of thesis writing is done with the examiners' expectations and requirements in mind. Examiners are people of authority who have the power to accept and award the PhD title (Koutsantoni, 2006). Thesis writers are of lower status, which prevents them from claiming expertise and authority of knowledge. However, they still have to obtain acceptance for their claims. In order to justify the value of the research, claims must be supported with evidence, and writers must demonstrate appropriate understanding of approaches and previous knowledge in their fields of specialization.

Reference to previous studies is typically found in the literature review sections (LRs) of the theses. LRs are important constitutive sections of theses. They provide the background for the research described in the theses and create a context within which the writer's study is situated. They serve to define the conceptual and theoretical framework of a thesis and establish the niche (Swales, 1990) for the writer's research study. The findings from the related work are considered in terms of how they could be used to inform the research proposed. Citation practice provides justification for arguments and allows the writer to show why her/his research is different from what is being documented in the literature or fills in a gap in research. Writers need to evaluate the previous research in an area of study and to be respectful with previous claims from authorities in the disciplines. They also need to position themselves in relation to other disciplinary members and highlight their individual claims. In this context of social interaction, it is necessary for writers to maintain appropriate relations with the immediate audience (the examiners) and the disciplinary community. Brown and Levinson (1987) introduce the concept of *face* to refer to one's self image. They argue that in interactions it is important to protect the participants' faces from potential face-threatening acts. The LR of a doctoral thesis entails critical evaluations which may involve face threatening acts. In order to mitigate the threat, the writers' choices of content, language and style of theses aim at protecting three faces: the writer's, the examiners' and the reviewed authors'.

Research on the Literature Review (LR) chapter of doctoral theses has been carried out on theses produced by native English speaking writers. Rhetorical structures have been examined (Kwan, 2006; Ridley, 2000; Thompson, 2009) together with specific aspects of citation practice and reporting verbs (Charles, 2006a, 2006b; Hunston, 1993; Hyland, 1999; Shaw, 1992; Swales, 1990; Thomas & Hawes, 1994; Thompson 2005a, 2005b; Thompson & Ye, 1991). However, to our best knowledge there have been no contrastive studies based on

LR chapters in theses written in two different languages. This paper investigates contrastively how interactional resources of citation and, in particular reporting verbs, are deployed in the LR chapters of PhD theses written in English and in Spanish.

2. THE CORPUS

The corpus consists of 20 theses -10 in English and 10 in Spanish- written during the last decade by native speakers, within a single applied discipline: computing. The theses had been successfully defended at the University of Glasgow, UK, and the Universidad Politécnica de Valencia (UPV), Spain, respectively, and were downloaded from the theses repositories of these universities.

Both corpora belong to the sub-fields of computation and computer engineering. The topics of the Spanish theses are related to allocation of resources and management techniques, web engineering, control systems, automation, problem-solving architectures and biological processing. The topics of the English theses are related to control systems, management techniques, automation, problem-solving architectures and biological processing. These topics represent a broad spectrum of the research fields of interest in the area of computing.

The sections with headings explicitly related to previous work were considered to be literature reviews (LRs) and selected for analysis. In most theses, these sections constitute a chapter and follow the introduction. In the English corpus the second chapter following the introduction is devoted to the LR in nine theses. In one thesis four reviews of existing literature occupy the first section in four different chapters. In the Spanish corpus, the LR is found in the second section of eight theses, after the introductory section. It is found in the third section of one thesis, after the section devoted to the introduction and the technological context, and it is found as a subsection of section 3 in one thesis. Although some LR texts are realized in one chapter, extensive sections are divided into thematic units that are marked by topic-based headings. This sophistication in the schematic structure of LR sections can be attributed to the complex nature of the writer's research topics and the varied objects that are studied.

Tables 1 and 2 describe the length, format and headings of the LRs in the two sets of theses. Spanish theses are longer than English theses. The average length of Spanish theses is 227.7 pages, while the average length of English theses is 204.2 pages. The average length of the LR sections in the English corpus is 35.5 pages, which represents 17.3% of the whole text. The average length of the LR sections in the Spanish corpus is 30.3 pages (14.6% of the whole text is dedicated to this section). The data show that there is greater variability in the length of these sections in Spanish theses.

The format of the theses falls into one main category: the traditional thesis with introduction, literature review, method, results/discussion and conclusions. Traditional

patterns in turn, can be simple or complex. A thesis with a simple traditional pattern is one which reports on a single study and has a typical macro-structure of introduction, review of the literature, materials and methods, results, discussion and conclusion. On the other hand, a complex traditional thesis is one which reports on more than one study. It typically starts with introduction, review of literature, can also include a general methods section and concludes with a general conclusions section. The intermediate sections reproduce the simple traditional structure of introduction, methods, results and discussion for each of the individual studies reported (Paltridge, 2002). In Spanish theses the simple format is the most frequent, while in English theses the complex format is the most repeated one.

In both corpora, the LR sections have generic headings. The most usual Spanish heading refers to the state-of-the-art. Most of the English headings include *background* or *literature review*.

Theses in English	Thesis length/ LR length	Thesis format	LR labels
TE1	163 pp./51 pp.	Topic-based	Background
TE2	185 pp./52 pp.	Traditional: Simple	Literature Review
TE3	189 pp./17 pp.	Traditional: Complex	Literature Review
TE4	192 pp./26 pp.	Traditional: Complex	Visualisation Background
TE5	290 pp./42 pp.	Traditional: Complex	Background
TE6	223 pp./37 pp.	Traditional: Complex	Literature Review
TE7	265 pp./39 pp.	Traditional: Complex	Background and Related Work
TE8	181 pp./36 pp.	Traditional: Complex	Interactive Information Retrieval: An Overview
TE9	170 pp./42 pp.	Problem/solution: Complex	Literature Review
TE10	184 pp./13 pp.	Traditional: Simple	Background and Motivation

Table 1. English theses: length, format and headings

Theses in Spanish	Thesis length/ LR length	Thesis format	LR labels
TS1	264 pp./57 + 47 pp.	Traditional: Simple	Estado del arte
TS2	195 pp./5 pp.	Topic-based	Estado del arte
TS3	245 pp./90 pp.	Traditional: Simple	Estado del arte
TS4	306 pp./5 pp.	Compilation of RAs	Estado del arte
TS5	206 pp./5 pp.	Problem/solution: Complex	Estado del arte
TS6	293 pp./16 pp.	Traditional: Simple	Estado del arte
TS7	272 pp./54 pp.	Traditional: Simple	Estado del arte
TS8	103 pp./29 pp.	Traditional: Complex	Antecedentes
TS9	239 pp./62 pp.	Traditional: Simple	Estado del arte
TS10	155 pp./33 pp.	Traditional: Complex	Revisión general del estado del arte

Table 2. Spanish theses: length, format and headings

3. THE METHOD

The research design was based on Swales' (1990) classification of citations and on the taxonomy of reporting verbs proposed by Thompson and Ye (1991) and Hyland (1999, 2002). Thompson and Ye (1991) distinguish three categories of reporting verbs according to the process they perform: textual verbs, in which there is an obligatory component of verbal expression (e.g. *state, write, point out),* mental verbs, which refer to mental processes expressed in the author's text (e.g. *believe, think, focus on*), and research verbs, which refer to processes that are part of the research activity and indicate results or experimental procedures (e.g. *find, demonstrate, calculate*). Thompson & Ye also analyse the evaluative potential of reporting verbs. A number of verbs show the author's stance towards the report, which may be positive, negative or neutral. Other reporting verbs construct the writer's stance of acceptance, neutrality or rejection towards the cited research through factive, non-factive and counter-factive options. Further, they allow the writer's interpretation of the author's behaviour or discourse, and the functional status within her/his own framework of the reported information.

Hyland (1999, 2002) also classifies reporting verbs according to the type of activity they refer to, although he uses the terms *discourse* and *cognition* for Thompson and Ye's (1991) textual and mental verb categories. He elaborates a modified scheme of supportive, tentative, critical or neutral options of stance towards the reported claims.

In order to carry out the present study, we first searched each set of LRs for the sentences in which reference was made to the author/s cited, and focus was placed on reporting verbs. The approach was both qualitative and quantitative. The counting of items was conducted manually and particular attention was paid to the context. Second, we gathered all the occurrences of a reporting verb, with its voice and associated syntactic structure. We classified samples from the corpus using the models discussed above. Finally, we compared the data obtained for each set of theses so as to determine variation in the way English and Spanish thesis writers choose to present arguments and report the work of others.

Following Thompson and Ye's (1991) convention, we refer to the person citing as the *writer* and the cited person as the *author*. Examples from the English corpus are coded as TE1: Thesis in English 1, TE2: Thesis in English 2…, and examples from the Spanish corpus as TS1: Thesis in Spanish 1, TS2: Thesis in Spanish 2…

4. RESULTS OF ANALYSIS

4.1. Types of citations

There are several tendencies in the choice of citation forms, which reflect how source material is used in the writer's argument (Dubois, 1988; Thompson, 1996, cited in Hyland, 1999).

Citations can be classified into integral or non-integral (Swales, 1990). Besides they can either reproduce the author's original words through direct quotations or summarize them. Integral citations include the name of the cited author in the reporting sentence. They make the author prominent and tend to be associated with an extensive comment on an individual study. In non-integral sentences emphasis is given to the product and not to the original author. The name of the cited author has no syntactic function but is referred to in brackets or by numbers, which reduces the prominence of the cited author considerably. The distinction is somehow blurred in our corpus because numbers in brackets, and not only names, also have syntactic functions in integral citations. For our analysis, we assumed that this is a usual way to refer to an author in technological fields and thus decided that both numbers and names plus date of publication in brackets would be considered alike.

Table 3 shows the distribution of citation types and active and passive forms in the two sets of LRs. In the English LRs writers use the original wording of the source material extensively, either in indented blocks or integrated within the text. In contrast, the overwhelming preference in the Spanish corpus is for summary: to summarize the cited author's work. The high ratio of integral citations (48.85% of all the citations) in the English LRs shows the tendency to make authors prominent. This can also explain the high number of active verb-controlling forms (84.93%). Spanish writers, on their part, prefer non-integral forms that downplay the role of the author and reduce her/his visibility (around 62% non-integral citations vs. 38 % integral citations). The quantitative analysis shows that passives and *se pasivo-reflejo* constructions are more frequently used than active forms in Spanish (56% vs. 44%, respectively).

	English LRs	Spanish LRs
Direct quotations	7.70%	0.02%
Integral citations	48.85%	38%
Non-integral citations	43.44%	61.98%
Active forms	84.93%	44%
Passive forms	15.07%	56%

Table 3. The ratio of citation types and active and passive forms in the two sets of LRs

4.1.1. Quotations

Unlike Spanish theses, where only two direct quotations were found to give the author's original words for the definition of a concept, nine English theses present 128 cited messages. The different choices to present the cited work concern the extent of the cited message: short and direct quotations (whose length ranges from 1 up to 4-5 lines) and block direct quotes (whose length ranges from 5 up to 10 lines), where writers use the original wording extensively. Either in indented blocks or integrated within the text, writers develop their arguments based on authorities in their subject fields. Direct quotes are used either to give definitions or to reproduce the author's statements:

Sound, as defined by Moore **[100]**, *"originates from the motion or vibration of an object. This motion is impressed upon the surrounding medium (usually air) as a pattern of changes in pressure"*. TE6

In 2008, Marissa Mayer, the Vice President of Search and User Experience of Google Inc. predicted in an interview held at the LeWeb conference that *"in the future personalized search will be one of the traits of leading search engines"* [Mayer, 2008]. TE7

Citations of citations are also frequent:

Robertson and Hancock-Beaulieu (1992), as cited by Borlund (2000), states that
The conflict between laboratory and operational experiments is essentially a conflict between, on the one hand, control over experimental variables, observability, and repeatability, and on the other hand, realism. TE8

4.1.2. Integral citations

Within integral citations, greater emphasis is given to authors by choice of syntactic position: the author is the subject in active clauses and the by-agent in passive clauses and shortened passive constructions:

In [127] *Taylor expresses* concern over this current 'virtual' versus 'real' divide in the current research community. TE5

Similar results *are reported by Porkaew and Chakrabarti [1999] and Zhai et al. [2006],* who both expand search queries using content-based visual features. TE7

[…] una de las primeras contribuciones que aparecen en la literatura se remonta al año 1999, y *la firman dos de los investigadores más influyentes en el tema Grid: Ian Foster y Karl Kesselman.* En este trabajo, *Foster y colaboradores presentan GARA* [23], una arquitectura para la gestión de recursos distribuidos construida sobre Globus Toolkit. TS2

El campo de la ingeniería de sistemas ha adoptado la teoría de la lógica borrosa *enunciada por LA. Zadeh* en 1965 [Zadeh 1965], como base para desarrollar todo un conjunto de técnicas de modelado, identificación y control. TS1

Explicit mention of the author is also found in both corpora in constructions where the author is named in noun clauses or prepositional phrases:

> *In [40, 44, 67, 70] centralised constraint models were proposed* that could be used within standard constraint solvers. TE9

> *A study on newsgroup articles by Morita and Shinoda (1994) provided evidence* that suggest a correlation between the time spent on reading an article and user preference. TE8

> *Ha sido comprobado* que bajo ciertas condiciones las redes recurrentes pueden utilizarse para aproximar a una precisión arbitraria una descripción discreta en variable de estado, *según [Nikiforuk & Gupta 1995], [Sontag 1993] y [Pham & Xing 1995*]. TS1

> La técnica de identificación de modelos fuzzy mediante la sectorización de no linealidades (sector nonlinearity), aparece en la literatura por primera vez en [Kawamoto et al. 1992], pero es *en [Tanaka & Wang 2001] donde es ampliamente desarrollada y analizada*. TS1

4.1.3.   *Non-integral citations*

In non-integral citations the name of the researcher does not perform an explicit syntactic role in the sentence but appears elsewhere in brackets either in numerical references or in references in which the author name/s and date/s of publication are given. Thompson's (2002) categorization framework is useful to classify the examples. He identifies *Ident, Origin, Source* and *Refer* non-integral citations.

4.1.3.a. Ident. One type of non-integral citation serves to identify the actor in the sentence, but her/his name has been de-emphasised:

> Similarly, *other researchers (Harman 1992b, Ruthven 2003)* experimented with query expansion through probabilistic term weighting. TE10

> In musical composition studies, *it has been suggested* that waveform can be correlated to the "texture" of tactile stimuli *[54]*. TE6

> *Otros autores la han estudiado* desde un punto de vista empírico para la corrección de errores sistemáticos *[Borenstein et al. 94]* o estimación de distintos parámetros del error *[Martinelli 02] [Kleeman 95]*. TS4

Por otra parte la transmisión de errores en los modelos cinemáticas directos o inversos, […], *ha sido estudiada por distintos autores* a través de la caracterización de matrices isotrópicas *[Saha et al. 95] [Low et al. 05] [Kim2 et al. 04] [Kim et al. 05]*. TS4

The subject of a sentence in a non-integral citation can be the human activity behind the research:

Various research has been done to optimise parameters and algorithms [Zhou and Huang, 2001; Doulamis and Doulamis, 2003; Huang and Zhou, 2001; Porkaew and Chakrabarti, 1999]. TE7

Although the accuracy of implicit approaches has been questioned (Nichols 1997), *recent studies have shown* that they can be an effective substitute for explicit relevance feedback (White et al. 2002b). TE10

Análisis empíricos han demostrado que el rendimiento de APTEEN está entre LEACH y TEEN en términos de energía disipada y tiempo de vida de la red. TS3

Para mejorar sus prestaciones *se investiga* en técnicas para identificar los modelos dinámicos eficientemente y en métodos de control de articulaciones que compensan no-linearidades y acoplamientos [An et al. 88], así como en optimización dinámica y control adaptativo para distintas condiciones de trabajo [Craig 88] [Ortega et al. 89]. TS4

4.1.3.b. Origin. Writers prefer to focus attention on results and it is usual to find a non-human entity related to an author's proposal or contribution. The non-integral citation then indicates the originator of the model, research or technique:

Past scientific literature has focused on classifying implicit feedback sources based on the underlying user behaviour *(Kelly & Teevan 2003, Stevens 1993, Nichols 1997)*. TE10

Path consistency [28, 63, 90] assumes that there is a constraint linking each pair of variables, meaning that the constraint graph would be a clique. TE9

Una versión posterior de GARA [26] *exploró* la posibilidad de utilizar información del estado del sistema en tiempo de ejecución para mejorar la selección de recursos. TS2

En SOP [SUBRAMA01], se diseña una arquitectura donde cada nodo posee sus propias capacidades y funcionalidades. TS3

4.1.3.c. Source. Non-integral citations also serve to express generalization when reference is made to groups of authors and no specific author is mentioned. The citation tells where the information comes from. The function is that of attribution. In this way, emphasis is given to the information contained in the proposition:

Ontologies are "content specific agreements" on vocabulary usage and sharing of knowledge *[Gruber, 1995]*. TE7

Bridging the Semantic Gap is considered one of the most challenging research issues in multimedia information retrieval today [*Jaimes et al., 2005]*. TE7

Debido a que los *clubs* no son más energéticamente eficientes que los árboles de expansión para conectar nodos en una red de gran extensión, *DMSTRP* es una solución elegante para redes amplias, *según los autores*. TS3

En primer lugar, a pesar de resultar computacionalmente simples, son enormemente eficaces y eficientes *[Goldberg, 1989]*. TS9

4.1.3.d. Refer. Non-integral citations can be used to refer the reader to a text to find further details:

For a more detailed review of vibrotactile devices *see Summers [136]*. T6

There are many different complete search strategies *[38, 39, 46, 50, 51, 77, 88, 97]*. TE9

Por otro lado, para la robótica móvil existe una reciente bibliografía*, ver por ejemplo [Inoue et al. 97] [Lyshevski et al. 00] [O'Connor et al. 96] [Canudas et al. 97] [Samson 95]*, que aborda su modelado cinemático, dinámico y/o control. TS4

> Seguidamente se citarán los esquemas de selección más comunes*, pudiendo consultar para más información los siguientes trabajos [Goldberg and Deb, 1991] [Bäck et al., 1991] [Blickle and Thiele, 1995] [Herrera et al., 1998].*TS9

4.2. Categories of reporting verbs

In the corpus, reporting verbs are used in relation to the review of existing computer systems, applications and techniques. Past proposals for solving problems and initial findings are also reported. As stated previously, Thompson and Ye (1991: 369) associate reporting verbs to three groups of processes: textual, mental and research. We found that each reporting verb in the texts analysed belonged to one of these three groups. However, it was not always easy to include a reporting verb into one specific category. Certain English verbs, such as *note or find*, and certain Spanish verbs, such as *concluir,* can be classified as referring to either mental or textual processes. Likewise the English *observe* and the Spanish *representar* can be either research or textual verbs. When ambiguity was found, we included the verbs in the category we considered most adequately reflected their use in context.

There is greater variation in the choice of reporting verbs in the English corpus. However, in both sets of theses, nearly half these forms occurred only once. Looking at the choice of reporting verbs used in the corpus of English theses, we find that writers employ 143 different verbs in 932 occurrences to refer to their literature in their LR chapters. The most common verb is *state*, followed by *suggest*. *Propose*, *find* and *show* occur with similar frequency, followed by *present*, *note* and *argue*. Our data show a high presence of verbs related to textual processes, which report on existing techniques, highlight existing problems and provide solutions for these problems. Mental verbs (*assume, view, believe*) are much less frequently used. As for the distribution of the verbs along the English theses, *suggest* is found in all theses, *state* and *find* in eight theses, and *propose* and *present* in seven and five theses, respectively. Some of these verbs occur in a high number of instances in one thesis, for example: *suggest* occurs in 22 instances in TE8, *propose* in 21 instances in TE9 and *state* in 20 instances in TE5. Other reporting verbs, although not so frequent, are present in many theses, this is the case of: *develop, note, define, argue* and *highlight*, among others.

In the Spanish corpus a total of 110 different reporting verbs and 464 occurrences of these verbs have been found. Some verbs are used only once and in only one thesis (e.g. *investigar, detallar, citar, cubrir, dedicarse, enunciar, comprobar, señalar, afirmar, firmar, explorar, establecer*), but others are found in several theses and in a great number of examples (*proponer* is found in all the theses; *presentar* and *emplear* are found in eight theses*; realizar* and *demostrar* are found in seven theses; *desarrollar* and *utilizar* appear in six theses; *analizar, mostrar, centrarse, plantear* and *abordar* are found in five theses). As happens in the English texts, the Spanish verbs belong mainly to textual and research processes. In fact, few mental verbs have been found (*considerar, asumir*).

Table 4 shows the most frequent forms found in each corpus and the proportion of total reporting verbs which they comprise.

English theses	n	%	Spanish theses	n	%
state	69	7,40	proponer	64	13,79
suggest	60	6,43	presentar	36	7,75
propose	45	4,83	utilizar	27	5,81
find	42	4,51	desarrollar	26	5,60
show	41	4,40	emplear	25	5,39
present	34	3,64	describir	14	3,02
note	22	2,36	plantear	12	2,59
argue	20	2,15	realizar	11	2,37
discuss/report	13/13	1,39	introducir	9	1,94
highlight	12	1,29	abordar/aplicar	8/8	1,72
Total	371	39,80	Total	240	51,72

Table 4. The ten most commonly used reporting verbs used in the English/Spanish theses. The first figure indicates the number of instances of the verb. The next figure expresses this as a percentage of all the reporting verbs used in the two sets of texts

Two verbs have their corresponding form in both sets of LRs, although they are not given the same prominence: *propose* occupies the third position in English and *proponer* is the most often used verb in Spanish. And *present* is in place 6 in English and *presentar* is the second most frequently used verb in Spanish. The ten most commonly used verbs in the Spanish corpus concentrate more than half of all the occurrences of reporting verbs in the corpus. The ten most frequently used verbs in the English corpus represent more than 39% of all the occurrences of reporting verbs in the corpus. This leads us to infer that English writers recur to a greater variety of choices while Spanish writers seem to choose among a more limited range of lexical verbs. In both sets of theses, the most common verbs belong to textual and research process categories. There is predilection in the English corpus for textual verbs (*state, suggest, propose, present, note, argue, discuss, highlight*) which convey an argument scheme, which regards explicit presentation, interpretation and speculation as accepted aspects of knowledge and focus on the cited author's propositions. The Spanish writers' choices of textual verbs facilitate speculative (*proponer*) and expository writing (*presentar, describir, plantear, introducir*). Besides, the data for the Spanish LRs show that textual verbs are used in the same proportion as research verbs referring to procedures (*utilizar, desarrollar, emplear, realizar, aplicar*), which reveals that Spanish thesis writers in computing tend to alternate expository and speculative writing with experimental explanatory schema, which view research activity as inductive, impersonal and empirically based.

A number of these verbs are used denotatively, but others perform evaluative functions that reflect either the author's or the thesis writer's interpretation and position with respect to the reported information. In the following sections, the uses of denotative and evaluative reporting verbs in the corpus will be commented upon.

4.2.1. Denotative reporting verbs

The detailed discussion of specific previous studies enables the thesis writer to display mastery of the literature before the examiners, which is one of the objectives the doctoral candidate bears in mind when she/he writes her/his text. About 50% of all the reporting verbs in both corpora provide information objectively, without interpretation and contribute to the impartial reporting style of academic writing. This strategy may be due to caution: it allows the writer both to be faithful to and respectful with the reviewed author's findings, while protecting her/him from refutation and conforming to politeness conventions. The power asymmetries between the writer and the examiners, seen as the disciplinary gatekeepers, lead the writer not to rely on her/his own voice and position (Koutsantoni, 2006).

In the English theses, the most common denotative textual verbs are: *present, note, argue, report, discuss,* and *define*. The research verbs: *do, find, show, develop, examine, investigate, measure, observe, discover, prove* and *perform* are mainly used denotatively. The examples given below illustrate the uses of some denotative verbs:

> In 2001, Bessiere et al. [21], and Zhang et al. [98], *presented* AC2001 and AC3.1 respectively. TE9

> Fukumoto et al. [44] proposed using voice coils to provide tactile feedback of button pushes and *found* an increase in dialling speed when compared to using audible beeps for button push feedback. TE6

In the Spanish corpus reporting verbs are mainly denotative, thus contributing to the objective and impersonal reporting style of scientific discourse. The textual verb *presentar* is the most widely used verb in the corpus. However, analysis of all the reporting verbs revealed a preference for reporting information as research acts categories expressing both procedures and findings (35 different verbs, such as . *utilizar, desarrollar, emplear, realizar* and *aplicar*):

> Zheng y colaboradores [34] *presentan* tres algoritmos para encontrar la distribución de la carga que minimice el tiempo de respuesta o el coste computacional en sistemas Grid. TS2

> En estos trabajos *se describen* y *analizan* los problemas temporales (retrasos aleatorios y falta de sincronismo) que aparecen en los SCBR, desde un punto de vista fundamentalmente estocástico. TS5

4.2.2. Stance of reporting verbs

The ultimate goal of the writer of the thesis is to obtain acceptance and credibility so as to be awarded the doctorate. This means that the information and the manner in which claims are

presented are selected according to this purpose. The writer's choices are intended to persuade the audience. The writer needs to show she/he has an original contribution to make in order to be accepted and be worthy of the award of the doctorate (Thompson, 2005). Once the writer has established the territory of the research (Swales, 1990) and explained the background for her/his work by situating cited authors within the field, the writer must place the work of the thesis in relation to the cited research and indicate a gap or need so as to justify her/his claims. This move demands the writer's involvement and positioning towards the information she/he is reporting. The means of reporting on previous research and presenting her/his own also convey an assessment of the activity reported. The choice of a reporting verb allows the writer to exploit its evaluative potential.

An examination of the structure of LRs reveals that at some point in the literature review the writer needs to convey her/his own purposes and prepare for the contrast between others' views and her/his own for her/his occupation of the niche. Reporting verbs that introduce the work of others are used together with other linguistic elements which open an evaluative space (Thompson & Ye, 1991: 369) for the information in the reported proposition.

Either alone or in combination with other elements in the context, different reporting verbs carry an evaluative stance that is related to the author or the writer. In other words, they express the author's attitude towards the validity of the reported information (author acts) or the writer's view on the reviewed author's information/opinion (writer acts).

4.2.2.a. Author acts

Certain reporting verbs attribute a position to the original author. Following Thompson & Ye (1991), authors can be reported as having a positive, negative or neutral attitude towards the reported information.

In the theses, the author's reported attitude is mainly neutral, with verbs such as *focus on, approach, propose, find, show // estar enfocado a, centrarse en, analizar, presentar, estudiar, desarrollar, describir, abordar, plantear, orientarse a:*

> Recent research *has focused* on the technology used to provide the feedback from touchscreens on mobile devices.TE6

> Van Zwol et al. [2008] *approach* this problem by transferring video annotation into an online gaming scenario. TE7

> Los autores *presentan* EEDUC como solución al problema del *hotspot* (punto caliente) en WSNs. T3

Sus últimos trabajos presentados ([Bau03], [Lor03a], [Lor03b]) *se centran*, fundamentalmente, en la influencia de las pequeñas derivas en la frecuencia de los relojes, físicamente separados, en los SCBR. TS5

Tentativeness is expressed by hedges like *propose, suggest // proponer, plantear, sugerir, pretender, intentar*. The writer withholds full commitment and this allows him to present a contrast with a new view:

In one of the earlier efforts for supporting video retrieval, Arman et al. [1994] *proposed* to use the concept of key frames (denoted as Rframes in their paper), which are representative frames of shots, for chronological browsing the content of a video sequence. TE7

Through digital presentations it is not as easy to '*forget'* and move on as Grudin *suggests* in [76]. TE5

Algunas de estas técnicas *intentan* construir de manera simultánea las reglas del sistema, las funciones de pertenencia y el mapeo entre el conjunto de datos de entrada y salida. TS1

Con este algoritmo *se pretenden* superar algunos de los problemas del FCV. TS7

Positive attitude is conveyed in a number of cases by *point out, highlight, claim, support, emphasise, state, stress // señalar, destacar, apoyar, dar soporte, defender, afirmar*:

Glass et al. state that "There is a severe decoupling between research in the computing field and the state of the practise of the field". *Tichy et al. is even more firm, stating that* there is active apathy in producing empirical work. TE2

Chalmers also goes on to stress that while it is inevitable that designers have influence over meaning, through the finite nature of computational models of context, it is often good to leave as much as possible of this interpretation open to the users. TE5

He suggests that this can be achieved by revealing what the underlying system is doing, *which is also supported by Dourish [39]* to provide users with a means of understanding and predicting how their actions will be reflected by the system. TE5

GARA también sentó las bases para la discusión en torno a la efectividad de las estrategias de asignación de recursos en el Grid. A partir de la publicación de este trabajo, *muchos investigadores apoyaron* la tesis de que la reserva anticipada de recursos era la única forma de alcanzar unos niveles de fiabilidad y de calidad de servicio razonables en el contexto de la computación Grid. En cambio, *otros investigadores seguían apoyando* el aprovisionamiento bajo demanda, a pesar de los problemas que presentaba en situaciones de sobrecarga. Esta discusión se ve reflejada en *varios trabajos que apoyan* la superioridad de la reserva anticipada sobre la reserva bajo demanda [24] y [25]. TS2

Los autores *dictaminan* que *HECTOR* es el primer protocolo de enrutamiento geográfico basado en coordenadas virtuales que es tanto eficientemente energético como fiable en cuanto a garantía de entrega. TS3

Diferentes autores [50, 51, 52] *han dado soporte* a esta nueva ingeniería para cubrir las necesidades introducidas por los sistemas basados en la Web. TS6

The writer's interpretation of the author's attitude as negative may be felt as challenging and inappropriate by examiners due to the differences in relations of power among the participants (Koutsantoni, 2006). This explains why generally writers impute positive and neutral positions to authors and avoid direct criticism and actual refutation. In the Spanish theses, only one instance of a negative reporting verb has been found, with *criticar,* and in that case the criticism is softened by means of a *se pasivo-reflejo* construction. The strategy aims at shielding author acts. However, a number of examples have been found in the English LRs, where authors are attributed negative attitudes towards the work of their colleagues. Criticism helps writers to indicate a need or a gap that the work presented in the thesis intends to fill in:

Kelly [2004] criticises the study approaches that focus on display time as relevance indicator, as she assumes that information-seeking behaviour is not influenced by contextual factors such as topic, task and collection. TE7

Despite their intuitive and straightforward character, *researchers have begun questioning* the level of support these techniques offer (Bates, 1990). *Buckley, Salton and Allan (1994) argue that the design of existing RF systems does not provide* adequate information to support the effective operation of the underlying query re-formulation heuristics and algorithms, discouraging users from applying relevance assessments to the viewed items. TE8

Debido a que *se critica* el uso de receptores GPS, asumimos que HECTOR no los incorpora [...] TS3

4.2.2.b. Writer acts

The writer's evaluative judgements towards the work of others and her/his own work are also made manifest with reporting verbs. The writer commits herself/himself and assumes responsibility for claims by employing verbs which imply a personal stance.

Non-factive stance shows no clear signal as to the writer's attitude towards the reliability of the author's findings. This is expressed through verbs such as *propose, claim, indicate, distinguish, introduce, cite, denote, examine, note, pose // utilizar, ofrecer, realizar, diseñar, encontrar, desarrollar*. The writer's stance towards the acceptance of the author's results and conclusions is factive with *ensure, enforce, demonstrate, show, prove, agree, support, exhibit, back up // demostrar, dar una solución, aportar, dejar constancia, posibilitar, proporcionar, quedar claro, confirmar, permitir, mejorar, optimizar, destacar*, which reveal the writer's agreement with a prior statement:

However, *as the work of Goffman has shown,* while we may not perceive ourselves as engaging explicitly in characterisation we do implicitly change our character befitting any given occasion or situation. TE5

It has been proven that the set of unmatched participants is the same for all stable matchings for a given instance of SMI [37]. TE9

El protocolo DMSTRP [HUANG06] *mejora a* BCDCP mediante la construcción de MSTs (Minimum Spanning Trees) en vez de los clubs que conectan los nodos en los clústeres. TS3

Esta estrategia ha demostrado su eficiencia para resolver algunos tipos de problemas. T2

Las condiciones de estabilidad de los sistemas borrosos han sido estudiadas de manera intensiva a lo largo de la década de los 90, *destacan* trabajos como [Tanaka & Sugeno 1992], [Wang et al. 1995], [Kang et al. 1998] o [Tanaka & Wang 2001]. TS1

When the writer's purpose is to establish the niche for her/his own alternative claim that justifies her/his work, counter-factive stance portrays the author's judgements as false, incorrect or incomplete. These verbs are used in order to justify the value of the writer's contribution. They signal the absence of an act that might have been expected from the author

and are found with negations and expressions with negative meaning such as *discuss, fail, ignore, lack, oversimplify // no concretar, poseer puntos débiles, no terminar de encontrar, no comentar, no optimizar*. English LRs show instances of negation reporting verbs (*ignore*) and writer's personal commitment to propositions (*I feel, to the best of our knowledge*) whereas Spanish writers hedge their criticisms with negative forms:

> These textbooks all focus on one particular aspect of the results: that fixing bugs is a small proportion of what maintenance programmers do, while changes to functionality (Adaptive and Perfective maintenance) are more important. However, *I feel that this ignores* the original tone of the Swanson categorisations, which make it clear that Corrective and Adaptive maintenance should be considered together as unavoidable sources of maintenance whilst Perfective maintenance represents voluntary reasons to make changes. TE2

> Further, *to the best of our knowledge*, *hardly anything has been done* to incorporate implicit relevance feedback in the video retrieval and recommendation domain. TE7

> Ambas aproximaciones poseen deficiencias en cuanto a eficiencia en el mantenimiento de su topología y la dispersión de sus mecanismos de actualización, que si bien incorporan ideas buenas, como la topología jerárquica en dos niveles, *no terminan de encontrar una solución óptima ni interesante*, aunque implantan estrategias interesantes a seguir para alcanzar una solución de compromiso aceptable. TS3

> La propuesta presentada por Hera *no aborda* cómo se implementan estos servicios Web *ni propone* un método para derivar servicios Web a partir de los modelos Hera. TS6

In the Spanish corpus, the option of challenging and criticising is avoided, as the writer assumes she/he must show modesty and prudence. He is aware that she/he is placed in an unequal situation to that of the other participants and criticism of the work of others can constitute a face-threatening activity (Myers, 1989). Textual and mental verbs are used to mitigate this threat. In some cases, there is a non-integral citation in which a generic noun refers to the criticised author and her/his name appears elsewhere in the text. Another way of softening the negative evaluation is to criticise the product, thus making the author invisible and avoiding personal confrontation:

> *Los autores no dejan constancia* de los detalles del algoritmo, *únicamente comparan* sus resultados con lo que ellos consideran buenas estrategias, pero no con ningún protocolo conocido hasta la fecha ni con un enfoque específico. TS3

> Una crítica general a todas las propuestas es que *ninguna de ellas propone* una guía para derivar automáticamente los servicios Web a partir de los modelos conceptuales que proponen. Además, *tampoco proponen* servicios Web que den soporte a la navegación definida en sus métodos. TS6

A typical strategy that allows the writer to occupy the niche for her/his work but reduces the threat to the author's acts and protects the writer's face is to combine factive and counterfactive stance. Making omissions explicit or assessing that particular acts were not performed but, at the same time, recognising successful achievements can be seen as a mitigating politeness strategy:

> This is not a criticism of the approach but it shows how hard it is to provide a flexible yet formal method of teaching, and shows the high value of system experts, mentors, in the learning process. *They point out* that there is a high upfront cost with producing the materials but *do not discuss* the problems of maintaining the materials to keep them current. TE2

> The authors *do not give* a complexity argument for their solutions. However, they *do give* empirical results of some experiments comparing their model against a distributed version of the EGS algorithm.TE9

> WSDM *permite* el uso de servicios Web externos [77, 76] *pero no soporta* el diseño ni la implementación de los servicios Web propios. Esta propuesta *no presenta* un método para derivar los servicios Web a partir de sus modelos conceptuales. TS6

> También se plantea la posibilidad de emplear diferentes frecuencias de muestreo en controlador y planta, lo cual lleva a la consideración de una estructura de control multifrecuencia que*, si bien en estos trabajos se presenta de forma poco desarrollada, demuestra su potencial* para la resolución de problemas como el propuesto en el desarrollo del proyecto. TS5

5. CONCLUSION

In the academic environment of the doctoral thesis, the review of existing studies creates the context for both situating one's work within an area of research and interacting with one's community effectively. The options offered by citations and the network of reporting verbs help writers to appropriately link their local contributions to a wider disciplinary framework and so persuade readers of their individual claims. In this paper we have compared the uses and functions of citations and reporting verbs in the LR chapters of PhD theses written in English and in Spanish in the area of computing. We have based our study on Swales' (1990) classification of citations and on Thompson and Ye's (1991) categorisation of reporting verbs.

The English LRs show a tendency to duplicate the author's original wordings and make authors prominent through integral active verb-controlling forms. On the contrary, the overwhelming preference of Spanish writers is to use their own words rather than those of others in non-integral citations that downplay the role of the author, thus reducing her/his visibility. Evidence is provided by the high number of passive and impersonal constructions. The analysis of reporting verbs has shown that there is greater variation in the choice of reporting verbs and higher occurrence of these forms in the English corpus. In both sets of theses, the most common verbs belong to textual and research process categories. There is predilection in the English corpus for textual verbs which convey an argument scheme. The Spanish writers' choices alternate arguments with experimental explanatory schema.

The choices made by English writers reveal personal commitment, whereas in the Spanish corpus, the implication of human intervention is reduced. Evaluation is mainly positive and factive in both corpora. Negative and counter-factive stances are expressed in order to justify the validity of the claims made in the theses. In these cases, and although individual styles must be taken into account, English writers highlight weaknesses so as to justify the validity of their contribution but Spanish writers tend to avoid personal confrontation and mitigate the strength of their arguments. These behaviours reflect cultural differences and show that English writers are more assertive than Spanish ones, who are prudent and seem to be conscious that they are placed in an unequal situation to that of the gatekeepers of the discipline and the examiners.

An interesting aspect for further research will be to study how each instance of report contributes to the overall tone of the text in every LR. The results can be used as a basis for helping students understand and interpret linguistic choices in authentic literature reviews.

REFERENCES

Brown, P. & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press.

Charles, M. (2006a). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, *25*, 310-331.

Charles, M. (2006b). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics*, *27*(3), 492-518.

Dubois, B. (1988). Citation in biomedical journal articles. *English for Specific Purposes*, 7, 181-194.

Hunston, S. (1993). Professional conflict: Disagreement in academic discourse. In M. Baker, G. Francis and E. Tognini-Bognelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 115-134). Amsterdam: Benjamins.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, *20*(3), 341-367.

Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 115-130). London: Longman.

Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes, 5,* 19-36.

Kwan, B. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, *25*, 30–55.

Paltridge, B. (2002). Thesis and dissertation writing: An examination of published advice and actual practice. *English for Specific Purposes*, *21,* 125-143.

Ridley, D. (2000). The different guises of a PhD thesis and the role of a literature review. In P. Thompson (Ed.), *Patterns and Perspectives: Insights into EAP Writing Practice* (pp. 61–76). Reading: University of Reading.

Shaw, P. (1992). Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics*, *13*(3), 302-319.

Swales, J. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.

Thomas, S. & Hawes, T.P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes*, *13*(2), 129-148.

Thompson, G. & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics*, 12(4), 365-382.

Thompson, G. (1996). Voices in the text: discourse perspectives on language reports. *Applied Linguistics, 17,* 501-530.

Thompson, P. (2002). Manifesting intertextuality in the PhD thesis. *Revista Canaria de Estudios Ingleses*, *44/ April*, 97-114.

Thompson, P. (2005a). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, *4*, 307–323.

Thompson, P. (2005b). Aspects of identification and position in intertextual reference in PhD theses. In E. Tognini-Bonelli & G. Del Lungo Camiciotti (Eds.), *Strategies in Academic Discourse* (pp. 31-50). Amsterdam: John Benjamins.

Thompson, P. (2009). Literature reviews in applied PhD theses: Evidence and problems. In K. Hyland. & G. Diani (Eds.), *Academic Evaluation. Review Genres in University Settings* (pp. 50-67). London: Palgrave Macmillan.

APPENDIX. THESES CORPUS

English theses

TE1. Hall, Malcolm. 2008. *Contextual Mobile Adaptation.* Department of Computing Science.

TE2. Hutton, Alistair J. 2008. *An Empirical Investigation of Issues Relating to Software Immigrants*. Department of Computing Science.

TE3. Kildal, Johan. *Developing an Interactive Overview for Non-Visual Exploration of Tabular Numerical Information*. Department of Computing Science

TE4. Jakubowska, Joanna. 2008. *Genome Visualisation and User Studies in Biologist-Computer Interaction*. Department of Computing Science.

TE5 Sherwood, Scott Caldwell. 2008. *Designing to Support Impression Management*. Department of Computing Science.

TE6. Hoggan, Eve Elizabeth. 2010. *Crossmodal Audio and Tactile Interaction with Mobile Touchscreens*. Department of Computing Science.

TE7. Hopfgartner, Frank. 2010. *Personalised Video Retrieval: Application of Implicit Feedback and Semantic User Profiles*. Department of Computing Science.

TE8. Arapakis, Ioannis. 2010. *Affect-Based Information Retrieval*. Department of Computing Science.

TE9. Unsworth, Chris. 2008. *A Specialised Constraint Approach for Stable Matching Problems*. Department of Computing Science.

TE10. Psarras, Ioannis. 2009. *Colombus: Providing Personalized Recommendations for Drifting User Interests*. Department of Computing Science.

Spanish theses

TS1. García-Nieto Rodríguez, Sergio. 2010. *Identificación y Control Predictivo Fuzzy T-S en Espacio de Estados. Una Aproximación al Control No Lineal*. Departamento de Ingeniería de Sistemas y Automática.

TS2. Torres Serrano, Erik. 2010. *Técnicas de Monitorización y Diferenciación de Servicios para la Asignación de Recursos en Entornos de Computación Grid, en base a Indicadores de Nivel de Servicio*. Departamento de Sistemas Informáticos y Computación.

TS3. Capella Hernández, Juan Vicente. 2010. *Redes Inalámbricas de Sensores: una Nueva Arquitectura Eficiente y Robusta basada en Jerarquía Dinámica de Grupos*. Departamento de Sistemas Informáticos y Computación.

TS4. Gracia Calandín, Luis Ignacio. No date. *Modelado Cinemático y Control de Robots Móviles con Ruedas*. Departamento de Ingeniería de Sistemas y Automática.

TS5. Casanova Calvo, Vicente. 2005. *Sistemas de Control basados en Red. Modelado y Diseño de Estructuras de Control*. Departamento de Ingeniería de Sistemas y Automática.

TS6. Ruiz Server, Marta. 2010. *Generación Automática de Servicios Web a partir de Modelos Conceptuales*. Departamento de Sistemas Informáticos y Computación.

TS7. Díez Ruano, José Luis. 2003. *Técnicas de Agrupamiento para Identificación y Control por Modelos Locales*. Departamento de Ingeniería de Sistemas y Automática.

TS8. Chiú Nazaralá, Raúl. 2009. *Análisis y Desarrollo de Observador Empleando LMI Aplicado a Bioprocesos*. Departamento de Ingeniería de Sistemas y Automática.

TS9. Valero Cubas, Soledad. 2010. *Arquitectura de Búsqueda Basada en Técnicas 'Soft Computing' para la Resolución de Problemas Combinatorios en Diferentes Dominios de Aplicación*. Departamento de Sistemas Informáticos y Computación.

TS10. Cervantes Posada, Mariamar. 2010. *Nuevos Métodos Meta Heurísticos para la Asignación Eficiente, Optimizada y Robusta de Recursos Limitados*. Departamento de Sistemas Informáticos y Computación.