UNIVERSIDAD NACIONAL DE RÍO CUARTO
ESCUELA DE POSTGRADUACIÓN
FACULTAD DE CIENCIAS HUMANAS

MAESTRÍA EN INGLÉS

TESIS DE MAESTRÍA

STANCE ADVERBS IN NSE AND NNSE WRITERS' RESEARCH ARTICLES

de

Renata Fabiana Cardinali

Directora: Iliana A. Martínez, MA
Co-Directora: Silvia C. Beck, MA

Año 2008

ABSTRACT

The research article is the most outstanding form of scientific interaction. It is a complex construct due to the fact that it is highly interpersonal in spite of its apparent objectivity in presenting information. It has been proposed that the aim of the research article is to persuade so as to gain acceptance from the scientific community for the knowledge claims it presents. The effective writing of this type of texts depends as much on the content encoded as on the way writers communicate it and project themselves into the text by manipulating different linguistic resources such as stance devices. This is a demanding task for native speakers of English as well as for non-native speakers of English. To better understand the problems NNSE writers face, this thesis describes, interprets and explains the distribution and use of –ly adverbs in the different sections of research articles written and published in English and produced by NSE and NNSE writers. The findings show that there are differences in the use and distribution of stance adverbs in the RA in English in the field of biology in two corpora, a learner corpus and a reference corpus. Pedagogical applications and suggestions for further research are provided.

Dedico este trabajo a mi familia y a mis amigos.

Pero especialmente lo dedico a Gabriel, Anita y Juanba porque lo *padecieron*.

Acknowledgements

I began writing this section long before I set out to actually do it. But it's time, so here it goes.

Writing this thesis dissertation has been a very long journey during which I have accumulated many debts. I would like to thank my mentors, Iliana A. Martínez and Silvia C. Beck, who supported and guided my work. The task was not easy and, at times, even overwhelming. But I truly believe that in the end we made a good team.

Along the way, there were people who made a significant difference. I am deeply grateful to Graciela and María Inés who, as heads of the language department, made it their responsibility to see that every staff member had the chance to finish the time-consuming task of writing her dissertation. Also, I am thankful to Luis Parodi for his valuable assistance with the statistics in the present research work.

There were other people who made the journey a bit easier. This is why I would like to thank my colleagues at the UNRC: Elisa, Silvia, Mabel, María Laura, María Inés, Laura, Andrea, Verónica, Carolina, and Eugenia for all their support and encouraging words. I thank especially María Elena and Inés for their comments on earlier drafts of this work. I am also thankful to my colleagues at New English Institute, Celi and Mer, for being there for me.

And there were other people who kept me going like my family and friends.

I know I owe them all for putting up with me after many sleepless nights.

Thanks!

CONTENTS

1. INTRODUCTION 1

2. REVIEW OF THE LITERATURE 6

2.1. Studies characterizing the research article as a rhetorical construction 6

2.2. Studies on evaluation 9

2.3. Studies that propose categories for the analysis of evaluation 10

2.4. Studies comparing NSE – NNSE writing 12

3. THEORETICAL FRAMEWORK 15

3.1 GENRE ANALYSIS 15

3.1.1. The New Rhetoric 16

3.1.2. Systemic Functional Linguistics 16

3.1.3. English for Specific Purposes 17

3.1.4. The research article as a genre 18

3.2. SECOND LANGUAGE ACQUISITION 21

3.2.1. Characteristics of the field 21

3.2.2. SLA and Corpus Linguistics 23

3.3. CORPUS LINGUISTICS 25

3.3.1. Evolution of the field 25

3.3.2. Corpus linguistics 27

3.3.3. Corpus-driven research and corpus-based research 29

3.3.4. Corpora 31

3.3.4.1. Definition 31

3.3.4.2. Corpus design and construction – principles and specific criteria 32

3.3.4.3. Corpus typology 36

3.3.4.4. Reference Corpus 37

3.3.4.5. Computer learner corpora 38

3.3.4.5.1. Computer Learner Typology 39

3.3.4.6. Methods and tools in corpus linguistics 40
3.4. EVALUATION 42
3.4.1. Epistemic stance markers 48
3.4.2. Attitudinal stance markers 49
3.4.3. Style of speaking markers 50
3.4.4. Stance Adverbs 50
3.4.4.1. Classification 51
3.4.4.2. Position 51
3.4.4.2.1. Adverbs as modifiers 51
3.4.4.2.2. Adverbs as adverbials 52
3.4.4.3. Syntactic roles 55
3.4.4.4. Semantic functions 56
3.4.4.4.1. Epistemic Stance Adverbs 57
3.4.4.4.2. Attitude Stance Adverbs 58
3.4.4.4.3. Style Stance Adverbs 59

4. METHODOLOGY **61**
4.1. The corpora 61
4.1.1. The reference corpus 61
4.1.2. The corpus of learner English 64
4.1.3. General considerations for both corpora 66
4.2. Tools for the analysis 67
4.3. Method 69

5. RESULTS **74**
5.1. General frequency and distribution of derivational adverbs in the RC and the CLEMPP 74
5.1.1. Distribution of stance adverb categories across the RC and the CLEMPP 80
5.1.1.1. Distribution of stance categories across the sections of the RA 83
5.1.2. Analysis of over and underused stance markers 85

5.2. Analysis across the sections of the research article 94

5.2.1. Overused items in the CLEMPP Introduction section 95

5.2.2. Overused items in the CLEMPP Methods section 96

5.2.3. Overused items in the CLEMPP Results section 98

5.2.4. Overused items in the CLEMPP Discussion section 99

6. DISCUSSION **102**

6.1. Variety of stance markers 102

6.2. Over and underrepresentation of stance adverbs 105

6.3. Over and underrepresentation of stance categories 106

7. PEDAGOGICAL IMPLICATIONS **109**

7.1. Conclusions and Pedagogical Implications 109

7.2. Strengths of The Study 112

7.3. Limitations of The Study 113

7.4. Suggestions for Further Research 113

7.5. Personal Comments 114

APPENDIX A: Reference Corpus references 115

APPENDIX B: CLEMPP references 131

APPENDIX C: Table C.1. Stance adverbs used according to categories 134

REFERENCES 135

TABLES AND FIGURES

Table 3.1: Stance categories, syntactic role and semantic function 59

Table 3.2: Position of adverbs and adverbials 60

Table 4.1: Distribution of running words in the Reference Corpus 62

Table 4.2: Journals in the RC with impact factor 64

Table 4.3: Distribution of running words in the CLEMPP 65

Table 4.4: Journals and impact factor of the published articles corresponding to the CLEMPP 66

Table 4.5: List of –ly adverbs as stance markers 71

Table 5.1: Frequency of stance adverbs in the RC and the CLEMPP. Raw and normalized numbers 75

Table 5.2: Frequency of stance adverbs types in the RC and the CLEMPP 75

Table 5.3: Derivational stance adverbs used in RC and CLEMPP 77

Table 5.4: Frequency of stance adverbs types in the sections of the RC and the CLEMPP 78

Table 5.5: -ly stance adverbs in the sections of each corpus covering around 50% of the tokens used 78

Table 5.6: Frequency of -ly stance adverbs in the sections of the RC and the CLEMPP ($\alpha < .05$) 79

Figure 5.1: Distribution of stance markers in the sections of the RC and the CLEMPP 80

Table 5.7: Distribution of stance adverb categories across the RC and the CLEMPP - Raw numbers and normalized numbers (at 10,000) ($\alpha \leq .05$) 81

Figure 5.2: Distribution of stance adverb types across the RC and the CLEMPP - Raw numbers and normalized numbers (at 10,000). 83

Table 5.8: Distribution of Stance adverb categories across the sections of RC and CLEMPP. ($\alpha \leq .05$) 84

Table 5.9: Items over and underused in the CLEMPP ($\alpha \leq 0.05$) 85

Figure 5.3: Items over and underused in the CLEMPP 86

Table 5.10: Most frequent patterns associated with the adverb interestingly in the RC 90

Table 5.11: Most frequent patterns associated with the adverb presumably in the RC 92

Table 5.12: Most frequent patterns associated with the adverb likely in the RC 93

Table 5.13: Distribution of stance adverbs across the sections of the RC (Normalized at 10,000) 94

Table 5.14: Overuse and underuse in the Introduction Section of the CLEMPP 95

Table 5.15: Lists of stance adverbs used in the Introduction section of the RC and CLEMPP 96

Table 5.16: Overuse and underuse in the Methods Section of the CLEMPP. 96

Table 5.17: Most frequent patterns associated with the adverb approximately in the RC 97

Table 5.18: Lists of stance adverbs used in the Methods sections of the RC and CLEMPP 98

Table 5.19: Overuse and underuse in the Results Section of the CLEMPP 98

Table 5.20: Lists of stance adverbs used in the Results Sections of the RC and CLEMPP 99

Table 5.21: Overuse and underuse in the Discussion Section of the CLEMPP 100

Table 5.22: Lists of stance adverbs used in the Discussion Sections of the RC and CLEMPP 101

Table C.1: Stance adverbs used according to categories 135

CHAPTER I

INTRODUCTION

Language plays a central role in the process of production and dissemination of scientific activity. The global scale of scientific interactions has resulted in the preference of one language, English, as the vehicle for communication. This has resulted in the emergence of English as an international language, a status that is worldwide recognized. Therefore, effective participation in academic interaction requires that scientists be able not only to understand but also to produce texts in English. Although the preference of one language as the language for international communication may seem to be positive for scientific production, interaction, communication and growth, it places researchers who are native speakers of English (NSE) in a privileged position, and researchers who are non-native speakers of English (NNSE), but nonetheless experts in their fields of study, in an unfavorable position. This is a fact that has not passed unnoticed to scholars in the field of English for Specific Purposes (Flowerdew, 2002; Swales, 1990), and has been criticized on the basis of the hegemony of one language at the expense of other languages (Hyland, 2006; Swales, 1990). However, it has also been claimed that "to deny people access to the linguistic, social, and educational capital that English represents is irresponsible" (Flowerdew, 2002, p. 7), as it would imply denying access to participation in the scientific community.

For many years, researchers at the National University of Rio Cuarto, Argentina, have undertaken research and attempted to publish their production in international journals which publish in English. For most of them, this interaction has been difficult, as their papers often face rejection on the basis of language oddities that they find difficult to improve. Unfortunately, in this geographical area it is very difficult to contact native speakers of English who are also specialists in specific academic fields as well as expert writers (Martínez, 2000). Although researchers may turn to language consultants and seek help to improve their writing of research papers, it is not enough. Writing research articles is a complex process for it encompasses the use of many different types of knowledge that may not be readily available even to proficient speakers of English. Published academic writing is seen "as the outcome of social

interaction" where writers use language in a variety of ways by selecting "one form or structure over another" (Hyland, 2000, p. 132). Expert members of the scientific community are able to engage in such type of disciplinary interaction because they can resort to their knowledge of the disciplinary cultures they belong to and are able to manipulate the accepted rules and conventions that allow them to achieve their aims. This type of knowledge is accessible not only to expert members, but also to newcomers and NNSE provided these are trained to use and exploit the devices effectively (Hyland, 1994; Salager-Meyer, 1994). Academic writing is packed with routine interactions that "represent, through their regularity and repetition, institutionally recognized communicative preferences through which writers construct and engage in disciplinary realities" (Hyland, 2000, p. 133). These preferences may not be obvious to NNSE, unless they are made aware of them.

The complexity of research articles has attracted the attention of numerous researchers interested in helping NNSE achieve native-like production (Dudley-Evans, 1994; Hyland, 1994; Swales, 1990). These authors have highlighted the value of corpus-based methods for the description of academic genres, for they can help reveal characteristic features of the genre and may contribute to identify problematic features, as those that make NNSE writing sound unidiomatic or non-native like (Granger, 2002; Hyland, 2006).

Of the many forms of scientific interaction, the most outstanding is the research article (RA), which has "become one of the main channels for distributing and advancing scientific knowledge among scholars world-wide" (Kanoksilapatham, 2005, p. 270). The RA is an established genre with socially recognized purposes realized through a characteristic structure and particular constellations of formal features. Even when the RA may be perceived as an objective report of information, it has long been accepted that scientific discourse is not purely informational and that there is more to it than merely representing the world as it is. In fact, the aim of the RA is to persuade, rather than just inform, the scientific community to accept the knowledge claims in it (Hunston, 1994). Therefore, the effectiveness of writing this type of text depends not only on the content to be encoded but also on the way writers communicate it and on how they project themselves into the text by manipulating different linguistic devices. In other words, effective writing is the result of the application of different kinds of knowledge. As Hyland points out:

> The means by which academics present knowledge claims and account for their actions thus involves not only cognitive factors, but also social and affective elements, and to study these necessarily moves us beyond the ideational dimension of texts to the ways they function at the interpersonal level.
>
> (Hyland, 2000, p. 12)

When writing research articles, writers engage in a social process and construct their texts including propositional content, their view of reality, along with their own projection by means of stance devices. Stance, defined as the speaker or writer's way of conveying his/her "personal feelings, attitudes, value judgments, or assessments" (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 966), refers to the way writers define their *persona* by communicating "their integrity, credibility, involvement, and a relationship to their subject matter and readers" (Hyland, 1999, p. 101).

Different authors have pointed out that the use of interpersonal devices is problematic for writers, non-experts and/or NNSEs, and have argued for the validity of approaches that aim at raising awareness on the uses and functions of interpersonal features (Hyland, 1994, 2000; Martinez, 2005; Salager-Meyer, 1994; Swales, 1990). Some authors have concentrated on the expression of evaluation in the RA (Hunston, 1994) and abstracts (Hyland & Tse, 2005; Stotesbury, 2003). Others have concentrated on hedging (Hyland, 1994; Salager-Meyer, 1994), commitment and detachment (Vassileva, 2000), and even on the use of one specific adverbial as is the case of *evidently* in Silver (2003). Although the existing literature provides information on adverbs as adverbials to express stance (Biber & Finegan, 1988), they are studied in texts of different genres produced by native speakers. To the best of my knowledge, however, I do not know of studies on the use of stance adverbs in one particular genre, the RA, across native and non-native English speaking corpora. The aim of this thesis is to fill this gap by focusing on derivational –ly adverbs in their role as stance markers. They are linguistic devices that may serve to comment on views, to project attitudes, and to establish relationships with the readers, and are known to be the most widely used in academic prose (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 540).

To better understand the problems that NNSE face when writing RAs in English, this thesis intends to describe, interpret and explain the distribution and use of –ly adverbs in the different sections of RAs in English produced by NSE and NNSE writers.

The present study was carried out using two corpora of RAs in English: a learner corpus and a reference corpus. The specific questions to be answered focus on the frequency of stance adverbs and stance types across the corpora, and across the sections of the RAs in each corpus. The present study also intends to identify items that are overused or underused by NNSE and the patterns associated with them.

The thesis aims to answer the following general question:

How are derivational stance adverbs used in a learner corpus and a reference corpus (RC) of RAs in English? [1]

In addition, the study aims to answer the following specific research questions:

- How does the frequency of derivational stance adverbs compare in a learner corpus and a RC?
- What is the distribution of –ly adverbs used as stance markers across the sections of the RAs in the two corpora?
- How are overused and underused stance –ly adverbs distributed in the sections of the RA of both corpora?
- How are overused and underused stance –ly adverbs used in the two corpora?
- What patterns are associated with the overused and underused -ly stance markers used in the two corpora?

In the present chapter, I have introduced the study, its aims and research questions. The rest of the thesis is organized as follows:

Chapter 2 presents the literature reviewed to identify a knowledge gap in the field.
Chapter 3 presents the theoretical framework underlying this study.
Chapter 4 describes the corpora, the software for analysis, and the methodological procedures followed to carry out the study.
Chapter 5 presents the results obtained from the analysis of the data of the two corpora of RAs used in this work.

[1] The learner corpus in this study is the CLEMPP, a collection of NNSE manuscripts of research articles in English. The Reference Corpus is a collection of published research articles.

Chapter 6 presents a discussion of the results obtained and the main conclusions of the study.

Chapter 7 describes the pedagogical implications of the findings.

CHAPTER II

REVIEW OF THE LITERATURE

In this section I describe the literature reviewed for this study, taking into consideration different aspects that have contributed to the delimitation of this research work. I have organized the section as follows: studies which focus on the characterization of the research article in English as a rhetorical construction, studies which focus on the expression of evaluation in research articles, and studies which propose different categories for the analysis of evaluation in scientific discourse. I have also reviewed studies which compared NSE and NNSE written production and identified features that may be troublesome for NNSE.

2.1. Studies characterizing the research article as a rhetorical construction

The existing literature characterizes the research article (RA hereafter) taking into account its rhetorical organization and its objectives. This has been the focus of attention of different studies (Adams Smith, 1984; Gledhill, 2001; Hunston, 1994; Swales, 1990; Thompson, 1993). These studies are concerned with assessing rhetorical purposes, describing information structure, and accounting for syntactic and lexical choices (Swales, 1990).

Swales describes the RA in English as a "remarkable phenomenon" that is "engineered by rhetorical machining" (1990, p. 125). In his textual overview of this genre, he proposes that the macrostructure of the research article, the hourglass structure proposed by Hill et al. in 1982 (in Swales, 1990), typically consists of the four sections: Introduction, Methods, Results, Discussion and Conclusion. Swales points out that these sections perform different rhetorical functions realized by different linguistic devices. He also reviews previous studies carried out by West and Heslot (in Swales, 1990) who studied that-nominal clauses and tenses, respectively. These studies show that their

distribution in the experimental research article is related to the macrostructure of the genre.

Kennedy (1998) reports a study carried out by Biber and Finegan (1994) who analyzed the linguistic characteristics of 19 medical research reports from 1985. They found significant differences on the use of certain linguistic features across the four sections, Introduction, Method, Result, and Discussion (IMRD), of the research article, showing that linguistic differences characterize the different sections. Luzon Marco (2000) analyzed collocational frameworks of patterns of two function words which frame lexical items in a corpus of 100 medical research papers. She described the "intermediate words or collocates" and concluded that these are distinct and respond to the conventions that characterize the genre (p. 63). Also Gledhill (2001), in an analysis of a corpus of 150 research articles, observed that phraseology, defined as the preferred ways of expressing semantic and communicative roles, varied systematically in the different sections in relation to their rhetorical functions.

Another important contribution to the rhetorical interpretation of the RA is that proposed by Tarone et al. (1998), who studied the occurrence of active and passive verb forms in astrophysics articles and found that the use of active verb forms is a strategic choice made by writers in response to their communicative intentions and the rhetorical goals of the genre. It may be pointed out that in spite of the fact that the articles reviewed in this study do not fit the macrostructure of experimental research articles, the findings are regarded as important because they support the assumption that linguistic choices are shaped by rhetorical purposes.

The studies reviewed above are concerned with the RA and its sections. However, there are studies which have focused on one section or another rather than on the analysis across the sections of the research paper. For example, the Introduction section was the central contribution of the book *Genre Analysis*, published in 1990, in which Swales proposed the revised *Create a Research Space* (CARS) model (p. 140), which captures the characteristic moves of RA introductions. Another study concerned with a specific section was Thompson's (1993), who studied the Results sections of experimental RAs in biochemistry and identified six rhetorical moves that rendered the Results persuasive. Her study shows that in this section writers argue for the validity of the data they present. Martínez (2003) studied the thematic structure in the Methods and Discussion sections of the experimental RA in the field of biology. Her findings show that the two sections feature differences in thematic structure which may be attributed to

the distinctive rhetorical purposes of each of these sections. According to her results, the texts in the Methods sections are descriptive while the ones in the Discussion are argumentative.

Current literature also offers some studies that are concerned with the characterization of academic prose through the analyses across registers. This is the case of the studies carried out by Conrad (1999), Biber et al. (1999), and Conrad and Biber (2000), who compare corpora of academic prose, newspaper reportage, fiction, and conversation. Although their findings do not inform about scientific genres in particular, they do provide valuable information about the academic register.

Conrad (1999) studied linking adverbials – their frequency, semantic category, grammatical structure, placement within the clause, the specific item used, and variation across the registers of the above mentioned corpus of academic prose, newspaper reportage, fiction, and conversation (Biber et al., 1999). Her findings show that linking adverbials are most common in academic prose, where result/inference adverbials account for the largest proportion, followed by apposition and contrast/concession. These findings contribute to the characterization of academic prose as a register which favors the explicit marking of development and support of arguments and logical connection in response to its rhetorical purposes.

Biber et al. (1999) present corpus findings for the use of stance devices across different registers and observe that these are distributed differently. They point out that adverbial stance markers are less frequent than other markers but they are still relatively common in academic prose. In addition to this, they report that single adverbs are the most common category for the realization of stance in all the registers. Particularly, as regards academic prose, single adverbs are the most commonly used to indicate epistemic stance. Interestingly, and especially relevant to the present study on –ly adverbs, their findings show that 55% of the common adverbs in academic prose are –ly adverbs.

Also applying corpus-based methods, Conrad and Biber (2000) studied different linguistic resources used by native speakers and writers to mark their personal stance. The study was conducted on three collections of texts: conversation, academic writing, and news reportage. The authors characterize the adverbial marking of stance along three parameters, namely, semantic class, grammatical realization, and placement in the clause. Their findings indicate that there are distinct differences in the use of stance adverbials across the different registers due to the different communicative purposes.

They conclude that generalizations of findings from one register to another are misleading.

To conclude, some of the studies reviewed here provide support for the characterization of the RA in English as a rhetorical construct exhibiting regularities of linguistic realizations that relate to the communicative purpose/s of the genre and/or genre sections. Using corpus-based methods, other studies have contributed to the description of academic prose as characterized by distinct features such as collocational frameworks, phraseology, active and passive verb forms, moves, thematic structure, linking adverbials, and stance devices. However, it is necessary to carry out further research into the use of specific –ly stance adverbs that may show similarities and differences within the RA and across its sections. This is one of the aims of this study.

2.2. Studies on evaluation

There is a growing interest in the way evaluation, as the overt expression of an author's or speaker's attitudes, feelings, judgments, or commitment concerning the message, is realized in the writing of research articles. This has resulted in studies that focus on the interpersonal aspects of scientific discourse. Several authors have studied the interpersonal role of language in the RA from different perspectives. Hunston (1994) views evaluation in the experimental RA as having three different functions, namely, conveying status, value, and relevance. Stotesbury (2003) studied the expression of evaluation in RA abstracts and found that there are differences in the forms of evaluation used in abstracts across disciplines (humanities, social and natural sciences). The study considered quantification of attitudinal lexis such as attributes, modifiers, adjuncts in the pattern of "x is y", and evaluative nominal and verbal groups. These were observed taking into account the different rhetorical organizational structures of the abstracts in the three disciplines and the findings revealed that evaluation occurred in different discourse moves, for example, those indicating a gap in knowledge, counter claims, findings in relation to previous claims and claims of new knowledge. Stotesbury (2003) argues that evaluation in abstracts is expressed implicitly and explicitly, and that the humanities and social sciences prefer evaluative attributes, while the natural sciences prefer to resort to modality as a way of expressing authorial stance.

Other researchers have been concerned with specific aspects of evaluation and have studied author's comments (Adams Smith, 1984), the use of first person pronouns (Adams Smith, 1984, Hyland, 2004, Martinez, 2005, Salager-Meyer, 1994), hedging (Hyland, 1994, 2004; Salager-Meyer, 1994), and stance (Biber & Finegan, 1988; Biber, et al. 1999; Conrad & Biber 2000). Adams Smith's study (1984) focused on the different ways that author's comments are realized in three genres of medical discourse. She analyzed cases of verbal modality, non-verbal modality and non-modal forms used to mark comments. In the final group, the following forms are included: adverbs, reporting verbs or verbal phrases, qualification of the reporting verbs, reporting nouns, qualification or modification of reporting nouns, emotive nouns, use of first person, use of metaphor, antithesis, and analogy, and selection of discourse structure. Martínez (2005) focused on the use of first person pronouns across the sections in two corpora of biological RAs (NSE and NNSE) and concluded that this feature has a strategic function that relates to idiomaticity and effectiveness. Salager-Meyer (1994) concluded that the choice of expression of tentativeness and flexibility is dictated by the general structure of the discourse, by its communicative purpose, by the level of the claim the writers wish to make, and by the author's pretension to universality and generalization. These studies have contributed to the understanding of how evaluation works in scientific discourse and have also highlighted its role in effective academic writing. It has been noted that the expression of evaluation is an area of difficulty that affects student writing at all levels and that the role of the writer's personal voice is worthy of attention (Johns & Swales as cited in Charles, 2003). There are still unexplored areas in the field of evaluation, such as the use of stance adverbs in RAs and across their sections in native and non-native corpora.

2.3. Studies that propose categories for the analysis of evaluation

In the literature reviewed, there are studies that propose categories and highlight different aspects of evaluation in academic discourse and across genres. Salager-Meyer (1994) studied hedging in medical discourse using a three dimensional concept of hedges, involving purposive fuzziness and vagueness, the expression of modesty and avoidance of personal involvement, and the impossibility or unwillingness of reaching absolute accuracy and qualification. She proposed a taxonomy of hedges based on

formal and functional criteria that included the categories of Shields, Approximators, Expressions such as "I believe", Emotionally-charged intensifiers, and Compound hedges. Taking a broader perspective, Hyland (1999) studied stance as "the ways writers project themselves into their texts to communicate their integrity, credibility, involvement, and a relationship to their subject matter and their readers" (p. 101). He developed a taxonomy of stance features that comprises categories with five main functions: hedges, emphatics, attitude markers, relational markers, and person markers. In his study, he proposed a category labeled "hedges" as in Salager-Meyer (1994). Further relations can be established between Hyland's *emphatics* and *attitude markers* with Salager-Meyers' *emotionally-charged intensifiers* since these include the same type of markers with similar functions. In addition, some of the items in Hyland's *attitude markers* and *person markers* can be related to the categories of *shields* and *expressions such as I believe* proposed by Salager-Meyer. These studies are relevant to the present thesis because they contributed information to the analysis of stance in the sections of RAs in English and across NSE and NNSE corpora carried out in the present thesis.

Another study that contributed to the delimitation of the categories used for this thesis was Biber and Finegan's study (1988) on the occurrence of adverbials as one of the primary lexical markers of stance in English in the Lancaster-Oslo-Bergen and London-Lund corpora of written and spoken British English, respectively. In this study, their main goal was to describe the "ways in which texts represent stance, and to identify various speech styles that are marked by stance" (p. 3). They observed these items across fifteen genres and categorized the stance adverbials on the basis of their semantic fields assigning them the following labels: *honestly, generally, surely, actually, maybe, and amazingly* adverbials. However, the taxonomy and examples provided by Biber et al.'s (1999) *Longman Grammar of Spoken and Written English* (LGSWE) are, in my opinion, the most comprehensive and useful for the current study. The authors provide a syntactic and semantic classification for stance markers together with lists of specific items that exemplify the categories. I have used these, together with the listings provided in Quirk, Greenbaum, Leech, and Svartvik (1985), to build the list of derivational adverbs that I observed in the present study. These adverbs have not received much attention in the literature so far.

2.4. Studies comparing NSE – NNSE writing

The existing literature shows that there is a growing interest in studies that compare NSE and NNSE productions with the aims of providing interlanguage descriptions and identifying features which seem to be troublesome for specific groups of NNSE (Granger, 2002; Meunier, 2002). Studies on NNSE corpora are inherently contrastive and valuable for they make contributions which are focused on NNSE. On the one hand, by virtue of their NNSE learner specificity, they contribute to Second Language Acquisition theory by providing descriptions of interlanguage and of the factors that influence it. On the other hand, they also contribute to the development of learner oriented pedagogical tools and classroom practices (Granger, 2004).

Coates, Sturgeon, Bohannan and Passini (2002) carried out a study on 120 medical articles submitted for publication in 1999 and 2000 produced by NSE and NNSE from the United States, the United Kingdom, Sweden, Japan, Germany, France, Spain and Italy. Their aim was to establish the extent to which submitted manuscripts were rejected for publication on the basis of language problems. Their findings indicated that written articles which exhibit serious language problems correlated with a high rejection rate.

Connor, Precht, and Upton (2002) studied business English using learner corpora from Belgium and Finland and a native corpus from the United States of America which contain letters of job application written by NSE and NNSE. They sought to demonstrate the value of combining the traditional text linguistics tool of genre analysis with a genre-specific corpus to make broader statements about how different writers approach writing for a specific purpose. Their findings suggest the writing of letters of application does not differ significantly across cultures, Western cultures, though there is variability in the use of optional moves within the genre. The researchers argue in favor of the textlinguistic approach for the analysis of genre specific learner language combined with a computerized analysis.

Granger and Tyson (1996) studied the use of connectors in English in essay writing using the French section of the International Corpus of Learner English (ICLE) corpus and the control corpus of English in the ICLE. The findings revealed that NNSE tend to overuse connectors which add to, exemplify, or emphasize a point, and underuse those that change the direction of the argument or develop it.

Aijmer (2002) reports on a study on the use of modality in argumentative writing produced by NNSE (Swedish) advanced university students. She compares these findings with essays from the LOCNESS (Louvain Corpus of Native English Essays)[1], the British LOB Corpus, and the corpus findings in the Longman Grammar of Spoken and Written English (LGSWE). The aim of her study was to investigate the use of modality using the following categories: modal auxiliaries, modal adverbials, and harmonic and disharmonic modal combinations. Her findings reveal a global overuse of modal auxiliaries by all the L2 writers, a tendency which may be partly developmental and partly interlingual (Aijmer, 2002). The results also show that there was evidence of underuse at a functional level where learner writers failed "to use *may* at all in its root meaning" (Aijmer, 2002, p. 72). Aijmer (2002) affirms that the distribution of expressions of modality affect the style and tone of a piece of writing. She considers that *style* is affected if learner writers use expressions which are more frequently associated with speech and that *tone* is affected because the significant use of modal expressions of tentativeness or certainty and their mixed use in contradictory strategies give the impression of non-nativeness.

Vassileva (2000) studied the degree of commitment and detachment in research RAs published in leading international and Bulgarian journals. Her corpus consisted of texts in English, Bulgarian and Bulgarian English produced in the field of linguistics. She studied the Introduction, Discussion and Conclusion sections with the aim of identifying differences in the overall distribution of hedges and boosters, as interpersonal devices, which may reveal cross-cultural variations leading to misunderstandings in communication. The study revealed similarities and differences. On the one hand, she concluded that the three different groups of writers employ hedging and that they use approximately the same linguistic means of expressing hedging and boosting. On the other hand, her results showed that hedging was more frequently used in the texts in English, which indicates high degree of detachment on the writers' part, whereas boosting was much more frequent in the Bulgarian and Bulgarian English texts, which indicates that the writers show a higher degree of commitment, or in other words, a lower degree of deference towards the scientific community. The evidence points out that there are differences in the means of expressing boosting in the three varieties, and also, that hedges are distributed

[1] For a full description of the LOCNESS see Granger, S. (1998). The computer learner corpus. a versatile new source of data for SLA research. In S. Granger (Ed.), p. 3-18

differently throughout the sections studied. In her conclusion, she advises Bulgarian writers to master the conventions of English if their aim is to participate successfully in the English-Speaking academic community. The findings of this study are similar to the ones reported by Hyland and Milton (1997), who observed that the writing of Hong Kong students displays firmer assertions, which endow their writing with an authoritative tone and give the impression of stronger commitments to the statements presented when compared to similar writing by NSE. It still remains to pay attention to the use of specific resources for the expression of stance, in particular adverbs, in NNSE writers' productions.

To conclude, the studies reviewed in this section have contributed to the delimitation and characterization of the object of study of the present thesis – the RA in English. Some of these studies provide evidence for its characterization as a complex rhetorical construction. Others describe its interpersonal nature by depicting the different linguistic forms in which writers express evaluation in them. Finally, some other studies contribute to the identification of features that seem to be troublesome for NNSE writers.

In Chapter III, I present the theoretical framework underlying this study.

CHAPTER III

THEORETICAL FRAMEWORK

In this section I describe the theoretical framework for this thesis, which is based on the contributions of Genre theories, Second Language Acquisition, Corpus Linguistics and approaches to Evaluation. I have chosen to develop this section beginning with the description of the frameworks that I consider to be more general in scope in relation to the present thesis. Genre theories provide the framework for characterizing the object of study of this thesis, the research article, and Second Language Acquisition provides the framework for characterizing the language user and learner language that is used as data for the study. Following this, I describe the frameworks that I consider to be more specifically related to the aims of the present study, namely Corpus Linguistics and the approaches for the study of Evaluation. Corpus Linguistics provides a linguistic theory and the methodology for the study of collections of texts to uncover linguistic patterns, and studies on Evaluation provide the categories used in the analysis of the data.

3.1 GENRE ANALYSIS

Genre theory, as reviewed in Hyon (1996), has mainly developed from the contributions of three different scholarly traditions: North American New Rhetoric (Bazerman, 1988, Myers, 1990), Australian Systemic Functional Linguistics (Bloor & Bloor, 1995; Halliday, 1994; Martin, 1992), and English for Specific Purposes (Bhatia, 1993, 2001; Gosden 1992; Hyland, 1994, 1999, 2002; Swales, 1990).

The theorists and practitioners in these schools can be grouped into two categories depending on the aspect of discourse they foreground: language and text structure or context. Flowerdew (in Johns, 2002, p. 4) labels the first of these approaches *linguistic*, where he places the Australian Systemic Functional Linguistics (hereafter SFL) and English for Specific Purposes schools (hereafter ESP), and the other *contextually grounded*, where he places the New Rhetoric school (hereafter NR).

3.1.1. The New Rhetoric

New Rhetoric research draws on North American research into L1 rhetoric, composition studies and professional writing (Hyon, 1996). The NR defines genre as a form of social action that is "flexible, plastic, and free" (Bakhtin as cited in Hyland 2004, p. 35). Its focus is mainly placed on the contexts where genres are used by expert members and on the way that genres incorporate the beliefs and purposes of particular social groups. Thus, genres are forms of social action, and studies within this tradition are "centred not on the substance or the form of the discourse but on the action it is used to accomplish" (Miller as cited in Hyland 2004, p. 35). Therefore, they are concerned with investigating the social, cultural and institutional contexts where genres take place rather than with describing the lexico-grammatical forms and rhetorical patterns of the language used. The studies in this framework are characterized by their use of ethnographic research tools, such as participant observation, interviews and descriptions, as well as analyses of texts which are "regarded as evidence of how people respond to routine situations in ways that differ by culture and by community" (Hyland 2004, p. 36).

3.1.2. Systemic Functional Linguistics

The SFL model of genre evolved from the theoretical work of Michael Halliday (1994). In his framework, language is seen as a system of choices by means of which users communicate certain functions and which allows them to express their experience of the world, interact with others, and create coherent messages (Halliday, 1994). Theorists in this framework are concerned with the relationship between language and its functions in social settings which are connected to particular contexts at the levels of register and genre. Register shapes the forms of language for it is determined by the activity taking place (*field*), the relationship between participants (*tenor*), and the role of language (*mode*). At the level of genre, this theory observes that the linguistic choices of the writer are influenced by the "writer's social purpose in using language" and therefore the notion of genre is a concrete expression of the register variables for it involves conventions for organizing messages so that readers can recognize its purpose/s (Hyland, 2004, p. 27). SFL defines genre as "a staged, goal-oriented social

process" (Martin, 1992, p. 505) and regards genres as having a linguistic structure which is purposeful, interactive, sequential and linked to the context. Unlike the NR school, this tradition emphasizes the study of genres in terms of their functions as well as the lexico-grammatical elements used to realize them, which is considered to be valuable for identifying genre types (Hyland, 2004). In this respect, this tradition is closer to the ESP tradition, described in the following section.

3.1.3. English for Specific Purposes

Dudley-Evans (1994) points out that the term *genre* was first used in an ESP context by Tarone et al. in 1981 in an article about the use of active and passive forms in astrophysics journal articles to show that it is the writer's communicative purpose that determines choices at the levels of grammar and lexis (Dudley-Evans, 1994, p. 219). In the ESP tradition, genre is defined as a class of structured communicative events sharing some set of communicative purposes which are recognized by the expert members of a discourse community. These factors shape the schematic structure of the discourse and influence and constrain the choice of content and style (Swales, 1990). Genres, thus, are not only recognized by their communicative purposes but also by their structure, style, content, and intended audience (Swales, 1990). Bhatia (1993), following Swales (1990), defines genres as:

> ...a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs. Most often it is highly structured and conventionalized with constraints on allowable contributions in terms of their intent, positioning, form and functional value. These constraints, however, are often exploited by the expert members of the discourse community to achieve private intentions within the framework of socially recognized purpose(s).
>
> Bhatia (1993, p. 13)

Hence, genres are shaped by social context and communicative purposes, as well as defined by formal properties. Taking this into consideration, Hyland (2004) highlights the pragmatic diversity of the ESP field, which draws on an eclectic theoretical foundation since it uses notions of dialogism and contextual situatedness, like the NR, and it also focuses on text structure, like the SFL. However, it differs from

the other traditions in that it addresses cross-cultural and L2 instruction (Hyland, 2004). Theorists and practitioners in the ESP school are concerned with genre "as a tool" (Hyland, 2004; Hyon, 1996) for understanding how experts use genres in specific academic and professional contexts and this is why their studies focus on the structures and meanings of texts and the way these are related to the context of use.

According to Swales (1990) "genres are communicative vehicles to achieve goals" (p. 46). This is so much so that it is the "communicative purpose" that shapes the genre as well as its internal structure since it is what brings text-internal factors and text-external factors together to assign generic values to different forms (Bhatia, 1993, 2001). Although communicative purpose may be difficult to specify in a clear cut way, it is essential in identifying different types of texts as belonging to particular genres because purpose helps to link the relatively regular structures in texts to the function/s of particular genres or genre stages in the "real world of discourse" (Bhatia, 2001, p. 82). This is what Bhatia (2001) calls *generic integrity* and which

> ...may be understood in terms of a socially constructed typical constellation of form-function correlations representing a specific professional, academic, or institutional communicative construct realizing a specific communicative purpose of the genre in question.
>
> Bhatia (2001, p. 88)

The notion of generic integrity is useful in helping to see how it is that different community members recognize, construct, interpret and use different generic artifacts to achieve their private intentions and goals. Different authors seem to agree on the assumption that understanding how this is done by expert members may prove to be essential to inform practitioners and novice writers (native and non-native) and guide them into successful participation (Hyland, 2004; Salager-Meyer, 1994).

3.1.4. The research article as a genre

Language is an integral element in scientific activity since it is by means of written texts that scientists interact with the scientific community (Gledhill, 2001). The role of language in this interaction is so important that "you cannot separate science from how it is written" (Halliday, 1993, p. 70). The nature of this interaction is complex

due to the fact that research writing is influenced by social, historical, philosophical, and psychological factors (Bazerman, 1988; Bhatia, 1993; Thompson, 1993) which are reflected in the written product. Therefore, the written product, the research article, is part of an ongoing interaction that requires participants to share and use different kinds of knowledge related to the genres of science.

It has long been accepted that scientific writing is not purely informational and impersonal and that there is more to it than merely complying with conventions (Bazerman, 1988; Hyland and Tse, 2005; Myers, 1989, 1990; Salager-Meyer, 1994; Thompson, 1993). From this perspective, the effective writing of research articles entails participation in a complex process that involves the use of the different kinds of genre knowledge, namely, communicative purpose, participant roles, context, formal text features (conventions), text content, register, cultural values, and awareness of intertextuality (Johns, 1997).

The research article (RA), as an established genre for scientific communication, is the means by which scientists present their claims and findings to the scientific community, with the expectation of their being accepted as their contribution to knowledge. This requires rhetorical effort on the part of writers, as they need to persuade other scientists of the validity of their claims. This effort usually involves the expression of evaluation, that is, the expression of personal judgment and an appeal to shared norms and values, rendering the RA interpersonal rather than objective (Hunston, 1994).

The study of the RA in English has been informed by the insights provided by genre theories and, in this study in particular, by the view of genre proposed by the ESP tradition that emphasizes the means by which the text realizes its communicative purpose (Duddley-Evans, 1994). In addition to this, the study of the RA can also be informed by the insights gained by evaluation theories since purposes, ideologies and values are built up in texts and usually transmitted overtly (Thompson & Hunston, 2000).

Bathia's definition of genre (1993), cited in section 3.1.3, provides a clear framework for identifying the RA as a socially recognized scientific genre for it has a specific communicative purpose: presenting claims, results or findings to a scientific community. In addition, it is structured into sections – Introduction, Methods, Results, Discussion, and Conclusions-, which also have specific communicative functions that constrain the content as well as the style in which the content is presented. This

definition considers the role that private intentions may play in determining the strategic use of conventions and the linguistics choices the writer makes at any given point in the text.

According to Swales (1990), the RA comprises a class of communicative events with shared communicative purposes recognized by the discourse community and realized through a schematic structure which constrains the content and style choices. Each section, Introduction, Methods, Results and Discussion, exhibits particular linguistic features determined by its specific rhetorical function. The Introduction has the function of contextualizing the research being presented in the relevant literature, claiming its novelty, and presenting the main features of the study. The Methods section describes the procedures used in the study. The Results section presents the findings in a way that "seems to be designed to deny on the author's part any associative contamination with commentary or observation" (Swales, 1990, p. 171). However, this section has also been shown to be rhetorically geared (Thompson, 1993). The Discussion section contextualizes the study by relating it to previous work in the field, reflecting the writer's sense of membership in the scientific community. In the light of this, it may be expected that this genre will favor some linguistic structures and elements (Luzon Marco, 2000) or word patterns (Hunston, Francis & Manning, 1997) over others, whose identification may be useful for the characterization of this academic genre. Similarly, it is expected that each of the sections of the RA (Swales, 1990) will favor the use of some linguistic elements over others and that this will be directly related to each section's rhetorical function.

Hunston (1994) argues that the main goal of this genre is to persuade a particular community, the academic community, to accept the new knowledge claims, and that this goal is potentially face-threatening. In relation to this, she highlights the central role of evaluation in the RA, because it is by means of evaluation that persuasion is partly achieved. Hunston observes that, essentially, each section has a persuasive goal: the Introduction section attempts to persuade the reader that the research carried out is necessary and worthwhile; the Methods section attempts to persuade the audience that the research was well done and avoided distortion; the Results section has the goal of persuading the readers that the statistical packages used were useful and informative. Finally, the Discussion section attempts to persuade the readers that the results presented make sense and move forward to building a consistent body of knowledge.

In this section I have described the object of study of this thesis and I have characterized it as an evaluative and interpersonal artifact shaped by communicative purposes. In Section 3.2, I will describe some of the contributions to this theoretical framework that have been derived from theories on Second Language Acquisition (SLA).

3.2. SECOND LANGUAGE ACQUISITION

One of the objectives of this study is the description of the language produced by NNSE writers. Therefore, the theoretical framework can gain of insights from Second Language Acquisition (SLA) theories which can contribute to the description of the nature of the data collected as well as the characterization of NNSE writers.

3.2.1. Characteristics of the field

Second Language Acquisition research established itself as a field of enquiry by the end of the 1960s and since then, it has developed in different ways. The scope of its field of enquiry shifted attention from the linguistic characteristics of learner language to pragmatic issues. In addition to this, SLA research changed from being only a consumer of linguistic theories to becoming a contributor as well, and it increased its theory-led research (Ellis, 1994). As a result of these developments, SLA research has become a complex field with interrelated sub-fields, in Ellis' words (1994) "a rather amorphous field of study with elastic boundaries" (p. 2) which draws on and contributes to a number of different disciplines such as linguistics, cognitive psychology, psycholinguistics, sociolinguistics, and education.

SLA as a field of enquiry covers studies on different types of "non-native Englishes" which Granger (1996, p. 13), following Kachru and McArthur, proposes to subcategorize into the following: English as an Official Language (EOL), English as a Second Language (ESL), and English as a Foreign Language (EFL) to embrace and identify the different groups that use English as a means of communication. This subcategorization is important because it embraces the different types of English produced by language users in different settings. This is central in the present study, as it focuses on the description of a specific kind of non-native English, that is, the one

produced by Spanish native speakers working at the National University of Río Cuarto (UNRC), in Río Cuarto, Argentina.

For many years the central goal of SLA has been to describe and explain the learner's linguistic communicative competence and its development by analyzing learner language. Linguistic competence consists of implicit knowledge which is unconscious and proceduralized, and explicit knowledge which is conscious and declarative. Linguistic competence is reflected in the learner's linguistic performance, that is to say "learner language". From this perspective, learner language, defined as the "oral or written language produced by learners", and characterized as being "highly variable" (Ellis & Barkhuizen, 2005, p. 4), is capable of providing information on both of these types of knowledge. Learner language reflects what learners have acquired and what they are capable of producing and thus can be regarded as a good source of data for linguistic enquiry. Ellis and Barkhuizen (2005) identify three different ways of gathering data: data from clinical elicitation, data from experimental elicitation, and data from naturally-occurring language use. This is language that is produced in a real life situation to achieve a communicative or aesthetic goal.

Since the present study is concerned with the description of the language in RAs in English produced by non-native language users, the data used is naturally-occurring language collected in a corpus (see section 3.3.4.1). This data is considered to be authentic due to the fact that it is the product of real communicative events. Although the writers are not students, the data collected is regarded as *learner language*, as defined above, because it has been produced by non-native writers of English. This is so because these writers may be expert writers in their own language, but they are not expert writers in English. Investigating learner language usually has an applied aim, and it involves the comparison of non-native productions with those of native writers' data to highlight similarities and differences that may contribute information to SLA theory and may further the development of pedagogical tools and improved forms of NNSE classroom practices. In the case of this thesis, the data obtained from the learner language corpus (see section 3.3.4.5) are compared with the data from the corpus of published RAs that is regarded as the reference corpus (see section 3.3.4.4.).

The study of learner language has evolved over the years from Contrastive Analysis in the 1960s –which sought to identify the linguistic differences between L1 and L2, to Error Analysis in the 1970s –which provided the methodology for investigating learner language. Since recently, SLA research has begun to gain insights

into L2 from current corpus linguistic methods with an innovative form of contrastive analysis as is Granger's Contrastive Interlanguage Analysis (CIA) (1996, p. 17). CIA is the linguistic approach to the study of learner language that consists in carrying out quantitative and qualitative comparisons between native and non-native data or between different varieties of non-native data with the aim of obtaining evidence for the nature of interlanguage and its development. The term *interlanguage* was coined by Selinker "to refer to the mental grammar that a learner constructs at a specific stage in the learning process" (in Ellis & Barkhuizen, 2005, p. 54). Although there is no complete agreement on this issue, *interlanguage* is considered to consist of implicit linguistic knowledge, and knowledge that is systematic, permeable, transitional, and variable; it is also considered to be the product of the application of learning and communication strategies. In addition, *iterlanguage* is explained as a process that can stop developing, i.e. something that may fossilize (Ellis & Barkhuizen, 2005). Therefore, and as conceived from the CIA perspective, comparisons of learner language can provide information on learner knowledge. Specifically, comparisons of NSE and NNSE data can help identify non-native features of learners' productions and highlight features of interlanguage which are developmental (Granger, 2002).

3.2.2. SLA and Corpus Linguistics

In spite of the fact that the collection of learner language samples is a long standing practice in SLA research, the analysis of learner language has been greatly improved by means of corpus-based methodologies which take into account aspects such as the type of language collected for data, the size of the collection and other variables that render the samples comparable and enriching for the investigation of issues which are central in SLA research (Ellis, 1994; Granger, 1996). For example, one of the best known projects which has been carefully designed to provide comparable and useful samples is the International Corpus of Learner English (ICLE). The ICLE is a computerized corpus of argumentative essays written by advanced learners of English of different L1 backgrounds. It contains over 2 million words of EFL writing from 11 categories of learners: Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Polish, Russian, Spanish and Swedish. Granger considers this corpus a "large learner corpus" (Granger, 2002). It is the first of its kind and it has been collected following

strict principles that ensure that the eleven sub-corpora are the same size and cover the same type of learner so that they are fully comparable (Granger, 1996).

The use of Corpus Linguistics methodologies to investigate learner data compiled in digitalized learner language corpora is a relatively new development that has gained recognition as such for it can help uncover regularities at the lexicogrammatical level of texts. The insights gained from the field of Corpus Linguistics have not only informed the conceptualization of learner language samples and provided a definition of learner corpora, but they have also provided adequate methodological tools for their analysis.

Corpus Linguistics methodology offers different possibilities for the analysis of learner corpora depending on the type of questions guiding the searches to be answered by corpora comparison. Different authors agree on the point that learner corpus research is valuable since it makes it possible to investigate aspects of learner language which have been previously difficult or impossible to explore (Aijmer, 2002; Granger, 2002). In addition, this type of research is considered to be of great value because it can shed light on the difficulties that learners have and thus contribute to organizing the teaching agenda in relation to the needs of specific learner populations (Granger, 2002; Meunier, 2002). Computer learner corpus research is inherently contrastive and makes use of frequency data. This data can provide information on the properties of interlanguage, particularly highlighting features of overuse, underuse, and misuse (Aijmer, 2002; Barlow, 2005), which are considered to be characteristic of non-nativeness (Granger, 2002). Over and underuse are identified in terms of comparison of frequencies. If a certain item in the learner corpus has a frequency higher than expected when compared to the reference corpus, it will be considered to be overused. Conversely, if an item in the learner corpus has a frequency lower than expected in relation to the reference corpus, the item will be considered to be underused. Misuse may be identified as the result of grammatical and/or lexical errors. The identification of phenomena such as over and underuse can "help explain why a text which may contain no overt grammatical/lexical errors nevertheless creates the impression that it has not been written by a native speaker" (Aijmer, 2002, p. 57). In the context of the present study, the identification of these phenomena in the corpus of learner writing may provide hints that indicate where it is necessary to perform in-depth analysis of salient items which may be identified as problematic for a specific NNSE group.

3.3. CORPUS LINGUISTICS

The research carried out for this thesis develops within the theoretical framework and methodology provided by theorists and researchers in the field of Corpus Linguistics. This section describes the evolution of the field, its theoretical assumptions and methodology.

3.3.1. Evolution of the field

Corpus Linguistics deals with the description of language by resorting to naturally occurring language data compiled in collections of texts built specifically for linguistic research. Studies in the field of Corpus Linguistics have a longstanding tradition that, according to Kennedy (1998) began long before the introduction of computers and software packages as tools for the analysis of language. The use of some kind of corpora in linguistic studies has always had a role in the description of language. For example, historical linguistics has always used corpus-based methods to study language change and evolution since it is in the collection of data from different time periods that the changes can be observed (Flowerdew, 2002; Kennedy, 1998). Other fields of scholarship that traditionally used corpus-based analysis are biblical and literary studies, lexicography, dialect studies, language education studies, and grammatical studies (Kennedy, 1998).

The field of Corpus Linguistics has evolved over the years. Although many studies in Corpus Linguistics have been carried out without the use of computers, different authors consider that the use of computer technology is currently a defining feature of the field (Ooi, 2001; Tognini-Bonelli, 2001; Widowson, 1996). In spite of the fact that the beginning of Corpus Linguistics as a field of language study is located around the late 1950s and early 1960s (Tognini-Bonelli, 2001), it was in 1991 that the use of computer technology for language study resulted in the emergence of a new view of language. In the publication of *Corpus, Concordance, Collocation* (1991), Sinclair introduces a new principle, the idiom principle, for the interpretation of how meaning arises from a text. The principle postulates that the words in a text occur in "semi-preconstructed phrases that constitute single choices" even when they may seem to be analyzable into segments. What is outstanding about this is that the meaning of these semi-preconstructed phrases is attached to the whole sequence rather than to the

individual items in it (Sinclair, 1991, p. 110). This principle contrasts with the open-choice principle, which is the one on which most traditional and structural grammars rest. In the light of the open-choice principle, a text is seen as a series of slots that are filled from the lexicon following only grammatical restraints. Sinclair suggests that meaning is constructed and understood by means of one of these principles because meaning arises from the phrase in accordance with the idiom principle, or it arises from the individual words complying with the grammatical rules of the language.

This new view of language results in the redefinition of units of meaning which are extended to include the node word as a core as well as the words in the co-text which are co-selected with it and form a regular pattern (Tognini-Bonelli, 2001). This is illustrated in collocations and colligations. Collocation is defined as "the occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991, p. 170) and colligation is "the collocation between a lexical word and a grammatical one" (Hunston, 2002, p. 12n1). These linguistic combinations are constructed from the lexical and grammatical choices that the system offers to language users, and they give discourse its fluency, naturalness and idiomacity (Beaugrande, 2001).

The availability of computers since the 1960s has had a great impact on the development of the field in relation to the way corpora are stored and analyzed. Corpus studies evolved from manual analysis in rather small corpora to computerized analysis in huge collections. The literature recognizes Quirk 1968's Survey of English Usage (SEU) Corpus as the first major pre-electronic corpus and the Brown University Corpus of Present-Day American English, built by Francis and Kučera in 1964 commonly known as the Brown Corpus, as the first computer readable corpus (Kennedy, 1998). The Brown Corpus was used as a model for the building of different corpora. For example the Lancaster-Oslo/Bergen (LOB) Corpus was built by Johansson et al. in 1978 (Kennedy, 1998) to be a collection representative of written British English. The aim of the designers was to make the LOB the counterpart of the Brown Corpus. Similarly, other first generation corpora were built following the model established by the Brown Corpus to represent written varieties of Indian, New Zealand and Australian English (Kennedy, 1998). Regarding the collection of spoken data, the first machine-readable spoken corpus was the London-Lund Corpus (LLC), which compiled 100 texts, 87 of which were part of Quirk 1968's Survey of Spoken English (SSE) Corpus, and it contained about half a million words of transcribed speech (Kennedy, 1998).

Since the 1980s, a new type of corpora began to be built. These are referred to as the second generation mega-corpora due to the fact that they are capable of containing several millions of words. The most important of these are the Cobuild Corpus, which later became The Bank of English Corpus, the Longman Corpus Network, the British National Corpus (BNC), and the International Corpus of English (ICE) (Kennedy, 1998).

With the evolution of the field, different views on what Corpus Linguistics implies have emerged. Many authors (Biber, Conrad & Reppen, 1998; Granger 2002; Kennedy, 1998; Sinclair, 1991; Tognini-Bonelli, 2001) agree that Corpus Linguistics is a new approach to linguistic description which is capable of changing our perspective on language due to the fact that it is an approach that brings together "data gathering and theorizing" (Halliday as cited in Tognini-Bonelli, 2001, p. 1), and that this approach leads to a qualitative change in the understanding of language. However, there are different views on whether Corpus Linguistics is a theory, the corpus-driven approach, or a methodology, the corpus-based approach. Tognini-Bonelli (2001) argues that although Corpus Linguistics is within the scope of applied linguistics, it stands out as a discipline due to the fact that it is a "pre-application methodology" that establishes "its own set of rules and pieces of knowledge before they are applied" (p. 1). This approach to the study of language is referred to as the corpus-driven approach. Other authors (Biber, Conrad & Reppen, 1998; Granger 2002; Kennedy, 1998) characterize Corpus Linguistics as a methodology for the study of language that is corpus-based and uses evidence from naturally occurring language collected to test and exemplify already existing categories. Although the latter is the approach used in this thesis, both of them will be described in detail in section 3.3.3.

From either perspective, different authors (Biber, Conrad & Reppen, 1998; Tognini-Bonelli, 2001) agree on the fact that Corpus Linguistics can provide input from which different areas such as lexicography, language teaching, translation, stylistics, grammar, and gender studies among others, can benefit.

3.3.2. Corpus linguistics

Different definitions of Corpus Linguistics have been proposed by different authors. However, the one proposed by Tognini-Bonelli (2001) stands out for its clear

definition of the field along three lines. She characterizes Corpus Linguistics as an empirical approach to the description of language use that operates within the framework of contextual and functional theory of meaning by making use of new technologies. The approach is empirical in the sense that it is a scientific enquiry that involves different steps and uses authentic data consisting of actual instances of language used by people in real communicative events. An important characteristic of Corpus Linguistics is that it has the aim of describing language use by using texts collected in a corpus, as its source of evidence - *corpus evidence*. Corpus evidence is assumed to be "the main vehicle for the creation of meaning" and it consists of different text fragments that, when brought together, can show salient features in patterns of repetition and co-selection. Therefore, the information corpus evidence provides may be "evaluated as meaningful in that it can be generalized to the language as a whole" (Tognini-Bonelli, 2001, p. 2).

According to Tognini-Bonelli (2001), Corpus Linguistics works within the Firthian framework of contextual and functional theories of meaning where "a text is a single, unified, meaningful event" and the items it consists of are determined by the context of situation that plays a determining role in the creation of meaning (Sinclair, 2001, p. xi). Although corpus evidence does not qualify as a single text, for it consists of the different texts, the framework is useful for it is assumed that clauses, phrases, words, and morphemes occur in a context, the verbal co-text, which is considered to be as important as the wider cultural context. This co-text is highly relevant for determining the meaning of any linguistic unit. It is from this perspective that a corpus is considered to be a multiple set of events - texts that can be studied for the similarities and differences among them (Sinclair, 2001).

Finally, another defining feature of Corpus Linguistics is the use of the new technologies. The use of computers and analysis software has had a great impact on the analysis of linguistic phenomena. First of all, computers have made it possible to observe linguistic phenomena in large amounts of language data at the click of a button, and they have also provided the tools for the description of corpora in terms of quantitative data such as frequencies, frequencies higher or lower than expected by chance, and frequency of co-occurrence.

In addition, the use of computers has also allowed for the performance of systematic analyses of the data by giving the linguist access to "a quality of evidence" that was impossible before (Sinclair, 1991, p. 4). This has influenced the field

considerably because it has given Corpus Linguistics methodology the potential of claiming the validity and reliability of the evidence it provides. Corpus Linguistics is concerned not only with what is possible in language "but also with what is probable- what is likely to occur in language use" (Kennedy, 1998, p. 8). Therefore, the concept of frequency, which is tied to the use of computers, plays an important role both for linguistic description and for addressing language-related issues and problems, since frequencies provide quantitative information that can become the basis for more qualitative studies.

It is also necessary to highlight the fact that studies in the field of Corpus Linguistics can provide input from which different areas such as lexicography, language teaching, translation, stylistics, grammar, and gender studies among others, can benefit (Biber, Conrad & Reppen, 1998; Tognini-Bonelli, 2001). In the case of the study in this thesis, Corpus Linguistics has provided a theory and also a methodology to study a particular academic genre – the research article in English.

3.3.3. Corpus-driven research and corpus-based research

The Corpus Linguistics approach to the study of language can be observed from two different research perspectives: corpus-driven research and corpus-based research. Work along the lines of corpus-driven research observes language in context and aims to derive linguistic categories from the patterns that are seen to emerge in the analysis (Tognini-Bonelli, 2001). It is a holistic approach to the study of language where the repeated instances in a text are seen to reflect the semiotic system, since it is assumed that the contextual elements, participants and objects, will somehow be realized in the linguistic co-text. It is important to make clear that in this approach, the co-text is the verbal environment that is to be formalized and the context refers to the situational and cultural parameters where the interaction takes place (Tognini-Bonelli, 2001). The corpus is seen as evidence of language use and the theoretical statements derived from it are fully consistent with this evidence. Instead of using categories determined a priori, this approach derives the linguistic categories from the recurrent patterns and frequency distributions. Hunston and Francis (2000) point out that corpus-driven research is based upon the assumption that the behavior of individual words and their patterning are important in the description of language use. Therefore, they highlight the importance

of the study of concordance lines which are lists of lines shown in the computer screen where the investigated word, node word, is centralized in each case and shown with the words that come before and after it to the left and to the right. These authors point out that they are the raw material for linguistic investigations and the study of language patterns in large amounts of data.

A number of authors have applied the corpus-driven approach to the study of language. Hunston and Francis' *Pattern Grammar: A Corpus-driven approach to the lexical grammar of English* (2000) is the result of this approach applied to the study of the English language. This methodology challenges traditional ideas and concepts about grammar such as word classes, constituent units, functional categories, distinctions between lexis and grammar, and grammar and semantics, as well as the role of grammar in indicating what can or cannot be said. These are replaced by the redefinition of grammar as including lexis and syntax, the association of pattern and meaning, and the role of grammar in describing what is typically and usually said or used. Tognini-Bonelli (2001) argues in favor of the setting up of corpus-driven linguistics as a discipline on its own due to the fact that it has its own set of goals, specific philosophical perspective, methodology, and theoretical framework as well as an accumulating body of knowledge which is particular to this specific domain of study.

Work in line with the corpus-based approach aims to explain, test or exemplify theories and descriptions that have already been formulated (Tognini-Bonelli, 2001). Linguists working within this approach analyze corpus data using categories established a priori and regard the corpus as a source of quantitative evidence that can validate these categories up to a certain extent or that can provide probabilistic information for the underlying theory. Biber, Conrad and Reppen (1998) characterize corpus-based analysis as an empirical analysis of language use collected in corpora using computers and techniques which are both quantitative and qualitative. These authors argue that these characteristics result in a scope and reliability of analysis that is not possible to achieve in other ways. Another defining feature of corpus-based studies is that they require both quantitative and qualitative analyses, since numbers alone do not say much about the functions of linguistic items. Quantitative analyses are useful in determining major patterns in language, and they are necessary to determine the extent to which a certain pattern is common or otherwise in relation to others (Conrad, 1999). Qualitative analyses are necessary to describe the communicative functions of the patterns studied.

Biber, Conrad and Reppen (1998) compare the corpus-based approach to other approaches in the field of linguistics to emphasize their complementary relationship and to point out to the fact that many research questions for corpus-based studies often grow out of other more traditional approaches. These authors emphasize the fact that research questions of a corpus-based study may grow out of different sources such as prior structural analyses, a hypothesis or theoretical framework, or intuition or anecdotal evidence and consider that different areas in Corpus Linguistics can benefit from the use of corpus-based studies.

A number of authors have applied the corpus-based approach to the study of language. Kennedy (1998) reviews corpus-based descriptions of English that focus on word classes and lexical items, syntactic processes, and language variation. Similarly, Biber, Conrad and Reppen (1998) describe studies using the corpus-based approach which focus on lexicography, grammar, lexico-grammar, and discourse characteristics.

Of the two approaches described in this section, this thesis follows the corpus-based approach to provide a description of a particular aspect of language use in the research article in English in the field of biological sciences using categories established a priori.

3.3.4. Corpora

3.3.4.1. Definition

Sinclair defines a corpus "as a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or a language variety as a source of data for linguistic research" (Sinclair, 2005). He observes that a corpus is a collection of samples of language texts or complete documents that is different from other collections, such as text archives or databases, in that it is compiled in a systematic, planned and structured way.

Computer readability has become a defining feature in corpus studies, although there were corpora before the availability of computers and there are many investigations that can be carried out without using computer technology.

External criteria are given priority in selecting the texts to include in a corpus because they relate to the communicative function texts have and because they work towards making a corpus representative. A corpus is considered to be representative of a

language or a language variety if it is a collection of language use that is authentic in the sense that it contains instances of language taken from genuine communications of people interacting in real life situations. In addition to this, the purpose for building a corpus is linguistic research. This is a defining feature of this type of collection because it sets it apart from other collections such as the World Wide Web, databases, and archives which have different purposes.

Since the scope and validity of any corpus study is directly related to the corpus it uses for the enquiry, issues of corpus design and construction are central in Corpus Linguistics and deserve special attention. In the following section I will discuss the principles and criteria of corpus design and construction.

3.3.4.2. Corpus design and construction – principles and specific criteria

A corpus is a valuable source of evidence for the description of language use provided it is designed and built following certain principles. Different authors coincide in that there are central notions to be considered in corpus design and these are representativeness, balance and sampling (Hunston, 2002; Sinclair, 1991, 2005; Tognini-Bonelli, 2001).

Representativeness refers to whether the findings derived from the analysis can be generalized to the population the corpus aims to represent and also to other corpora (Leech as cited in Kennedy, 1998; Tognini-Bonelli, 2001). Sinclair (2005) proposes that there are certain steps to be followed to build a corpus as representative as possible. Among these are the decisions the corpus designer makes on the structural criteria used to build the corpus, the inventory of text types that are to be included based on external criteria, the priority these criteria will have at the point of deciding which texts are more or less important as part of the sample, the size of the sample and its components, and the step by step planning and documentation of the design of the corpus. In Tognini-Bonelli's terms, representativeness is "assessed by comparing the range of variability of a sample with the characteristics of a target population" (2001, p. 79). In other words, the collection should resemble the whole with its characteristic features. Corpora can be said to represent the total statistical population or to be a representative sample of the total population of discourse. An example of a corpus representative of the total statistical population would be one which contains all the works of Charles Dickens and

by this it aims to represent the total population of works produced by that writer (Kennedy, 1998). In contrast, a corpus can be said to be a representative sample of the total population of discourse when it is not attainable to compile the total population so a sample, a representative one, is built. For example, since compiling all the production of a language variety is not possible, a corpus of one-million or more words is built according to specific design criteria to represent the target language variety. This kind of corpus is also called sample-text corpus.

Representativeness is related to authenticity. According to Tognini-Bonelli (2001), all the material included in a corpus is assumed to have been taken from genuine communications because Corpus Linguistics deals with language in use. Hunston (2002) suggests that to build a representative corpus the designer has to break down the whole into its component parts and s/he has to include equal amounts of texts for each part. Put this way, it can be observed that representativeness is tied to balance, which will be discussed later in this section. It seems worthy to exemplify this notion with a brief description of a well known corpus, the British National Corpus (BNC). The BNC, built between 1991 and 1995, is a corpus that contains 100 million words of contemporary written and spoken language that aimed to represent the British English language as a whole. In order to attain representativeness the designers included written and spoken texts from different sources. The written texts consisted of informative and imaginative (literary) prose from different categories, whereas, the spoken texts were recorded in a variety of educational and informative events, tutorials and classrooms, news reports, and business events. The recordings were collected in 12 regional sampling areas throughout the United Kingdom (Kennedy, 1998).

Balance is another central notion in corpus design. According to Sinclair (2005) a corpus is considered to be balanced when the proportions of the different texts it contains correspond with informed and intuitive judgments. Balance does not always have to do with the compilation of an equal number of texts from different genres or modes; it has to do with the proportions of the texts included in relation to the use that language users make of these different texts (Kennedy, 1998; Sinclair, 2005;). Balance is affected by use and specialization of the texts included. A general corpus would be considered to be balanced if, for example, the proportions of spoken and written texts as used by the population are reflected by the proportions of spoken and written texts collected in the corpus. However, this is difficult to attain because even though it is known that spoken language is far commoner than written language, the collection of

spoken corpora are more difficult, time-consuming and expensive than that of written language. For example, in spite of the fact that the BNC was a major project designed to obtain a balanced corpus, the proportions of spoken and written texts in it (10% and 90%, respectively) are not balanced in relation to language use. Sampling is a central notion in corpus design because the reliability and validity of the generalization depends on adequate sampling (Kennedy, 1998).

Sampling relates to the type of linguistic study that the corpus will give evidence for. The type of linguistic query that will be carried out with the corpus together with the methodology used in the query will determine the minimum size needed to obtain valuable evidence from the corpus (Sinclair, 2005). For example, a query on the occurrence of two or three word combinations will require a minimum corpus size larger than the one required for the study of a single item because the occurrence of two or more words together "is inherently far rarer than either on its own" (Sinclair, 2005). Methodological decisions will affect corpus size because further data analysis using computers requires larger amounts of data "to penetrate below the surface variation and isolate the generalities" (Sinclair, 2005). Another important issue in sampling is related to the selection and inclusion of whole texts or text samples of the same size. On the one hand, complete texts that vary in size in relation to each other may exert an influence on the results of the study; on the other hand, text samples of the same size may seem more balanced but there is nothing in the part chosen that ensures its representativeness of the whole. Thus, Sinclair (2005) affirms that "the integrity and representativeness of complete artifacts is far more important than the difficulty of reconciling texts of different dimensions".

The selection of the texts to be included in a corpus is based on two types of criteria which are reflected in the following principles:

- The contents of a corpus should be selected without regard for the language they contain, but according to their communicative function in the community in which they arise.
- Any control of subject matter in a corpus should be imposed by the use of external, and not internal, criteria (Sinclair, 2005)

External criteria are derived from the communicative event where the texts are used, while internal criteria reflect the details of the language in the texts. According to Sinclair (2005), external criteria should exclusively determine corpus design and

construction, while the selection of each text should be based on common specific criteria. Sinclair outlines the following specific criteria (1991, 2005): text mode, type, domain, language specifications, location and time. A common specific criterion is the text mode which refers to collections of transcribed spoken texts to represent speech and collections of written texts. Whether spoken or written, the texts compiled can be grouped into text types, or genres, such as books, letters, and research articles. The register or text domain is also considered to be a specific criterion by means of which texts are grouped into, for example, academic, popular, news reportage and fiction. Another criterion is related to language specifications, which involves the representation of a language or a language variety. Finally, the last two common specific criteria relate to the context of production of the texts, in other words, they take into account where and when the texts were produced: country of origin and time of production. Specific design criteria aim to ensure that the structure of the corpus works towards representativeness and comparability with other similarly structured corpora.

Some of the principles established by Sinclair (1991, 2005) focus on the format and documentation of the text in the corpus to ensure, on the one hand, that only the relevant texts in plain text format are included and, on the other hand, to make sure that everything has been appropriately documented as regards the contents and arguments on which the corpus rests. This is very important not only to keep the corpus growing in size, if this is an aim, but also to have a source of information that can be useful to explain unexpected results (Sinclair, 2005).

The existing literature on corpus design discusses notions and principles that mainly focus on the building of corpora that are general (Hunston, 2002; Kennedy, 1998; Sinclair, 1991, 2005). Thus general corpora are taken as the default in corpus design and related definitions. The central notions (representativeness, balance, and sampling) as well as the principles of corpus building, some of which have been mentioned above, are discussed in relation to general corpora instead of other type of corpora. Therefore, they deserve special attention and should be discussed in relation to other types of corpora specifically. This will be done in sections 3.3.4.4 and 3.3.4.5.

3.3.4.3. Corpus typology

There are various types of corpora that can be classified according to the purpose, representativeness (described above in section 3.3.4.2), organization and format. Taking purpose into account, Kennedy (1998) defines general corpora as collections of texts which are used to describe the vocabulary, grammar or discourse structure of the language. These corpora are also called balanced or core corpora because they typically contain texts from different genres and domains (spoken and written, private and public). These corpora contrast with specialized corpora, as the ones used in this study, which are corpora delimited by a specific register and discourse domain (Beaugrande, 2001) and they are built to answer questions for particular projects such as the description of child language development or the English used in petroleum geology exploration, just to provide a few of examples. Other specialized corpora are compiled to answer questions about regional or sociolinguistic variation and examples of these are dialect, regional, non-standard and learners' corpora (Kennedy, 1998, p. 20). Hunston (2002) adds to this typology comparable, parallel and pedagogic corpora. Comparable corpora are collections of different languages or language varieties that can be compared to highlight their characteristic features. Parallel corpora are collections of the same text/s, for example novels, that have been produced in different languages and can be used to identify equivalents and differences between those languages. Pedagogic corpora are the collections of the language a learner has been exposed to.

Taking into account the mode of language, corpora can be classified into spoken and written, whether they are collections of transcribed oral interactions or collections of written texts, respectively. When time period is taken into account, corpora are also referred to as synchronic, when they attempt to represent language at a particular time, and diachronic, when they attempt to represent language over a period of time.

As regards format, corpora can be tagged, parsed or concordanced. A tagged corpus is one where each word is given a grammatical label; a parsed corpus is one that is labeled to show sentence structure and the function of different word classes in the sentence; a corpus is concordanced when all the words in the corpus are listed together in their context (Kennedy, 1998, p. 21). These are also referred to as annotated corpora for they contain additional information, that is, data other than the plain text, on the different parts of the collection that can be used to retrieve and analyze the data they

provide. The counterparts of annotated corpora are referred to as raw corpora which are collections of plain text with no extra information included (Granger, 2002).

Taking into account the size of a corpus and its possibility for growing, corpora are called dynamic or monitor when they are open-ended compilations since the design allows new texts to be added. In addition, corpora can be regarded as large or small depending on the number of running words they contain. However, this assessment is relative to the date of compilation of the corpora. For example, the Brown Corpus, built by Francis and Kučera in 1964, contained a million words and it was regarded the largest at the time, it must be noted that it was the only one then. Nowadays, a million-word corpus is regarded as small due to the fact that there are corpora which contain hundreds of millions. Size is so relative that in a few years any collection with less than a few billion words will be considered small (Sinclair, 2001).

Since in the present study two corpora will be compared - a Reference Corpus and a Learner Corpus -, it is reasonable to devote some space to the specific description of each of these types.

3.3.4.4. Reference Corpus

One of the two corpora used in this study is called Reference Corpus (RC). The RC is characterized as being a specialized corpus, for it compiles texts of a particular type, a genre, with a particular communicative function. This collection differs from a general language collection in that it aims to be representative of a particular text type, instead of holding different text types, and it aims to provide evidence for a particular language use (Hunston, 2002). A corpus is considered to be a reference corpus because its purpose is to serve as a sort of model of specialized language use in a particular scientific field. It is built to answer specific questions about language use and it aims at being representative of a specific population. The purpose for the creation of a reference corpus is to render it comparable to other corpora following the same design to highlight characteristic similarities and differences.

Granger (2002) calls this type of corpus a control corpus because in her studies this corpus is compared and contrasted with a computer learner English corpus to spot similarities and differences.

3.3.4.5. Computer learner corpora

One of the corpora built for this study is a learner corpus because it comprises research article manuscripts produced by non-native speakers of English (NNSE). Granger (2002) refers to learner corpora as *Computer Learner Corpora* and provides the following definition based on Sinclair's definition (in Granger, 2002). Corpora are

> ...electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance.
>
> (Granger, 2002, p.7)

Although some of the notions in the definition above have already been described, it seems important to expand on them specifically in relation to computer learner corpora since they may acquire a different shade of meaning. As regards authenticity, a collection is considered to be so if the textual data it contains is the outcome of "genuine communications" (Sinclair as cited in Granger 2002, p. 8). This notion, when applied to computer learner corpora, needs to be expanded in a way that allows for the inclusion of different degrees of authenticity since non-native learners' productions usually originate in classrooms which essentially imply a certain level of artificiality. Therefore, according to Granger (2002), authenticity should range from "gathered from genuine communications of people going about their normal business" to "resulting from authentic classroom activity" (p. 8).

As regards the variety of the textual data mentioned in the definition as foreign and second language, non-native varieties of English are grouped into different types of English around the world and they are labeled in different ways, namely, English as Official Language (EOL), English as a Second Language (ESL) and English as Foreign Language (EFL). Granger (2002) considers that EOL includes indigenized or nativised varieties such as Nigerian or Indian English. ESL refers to the language acquired in English speaking countries such as Australia, the UK and the USA. Finally, EFL covers English acquired in classrooms in non-English-speaking countries, such as Belgium, Germany and Argentina, just to name a few.

As regards textual data, learner corpora include continuous stretches of discourse which include correct as well as erroneous use of language. Although the design criteria taken into consideration for the building of corpora described above hold for the criteria to build learner corpora, Granger (2002) adds some which are specific for the design of Computer Learner Corpora. Some of the criteria relate to the learner and others to the task settings. As regards the learner, Granger points out that it is important to consider and document criteria such as the learning context where the learner is immersed, the learner's mother tongue and other languages that s/he may speak, and the level of proficiency. As regards task settings where the learner produced the text compiled in the corpus, Granger considers that it is important to note the time limit posed on the task, whether the learner used reference tools or not, if the text was produced in an exam situation or otherwise, and also the type of audience or interlocutor the task demanded. These are just some of the considerations regarded as important in the design of computer learner corpora. These should be taken into account to ensure that the collection is as valid a source of data as it is meant to be.

According to Granger, purpose is another important notion in the learner corpus definition provided above because different corpora are compiled to answer different questions that can aim to improve some aspect of language theory or contribute to the production of better foreign language teaching methods and tools (2002). Learner corpora are especially interesting for language teaching and learning because they offer the possibility of identifying errors and areas of language that pose difficulties for learners (Tognini-Bonelli, 2001).

In order to ensure comparability, computer learner corpora must be standardized and documented appropriately. As explained above in relation to other corpora, a learner corpus can be raw or annotated based on standardized annotation software. In addition, it is important that the information about the learners that produced the texts included and the information about the tasks are recorded.

3.3.4.5.1. Computer Learner Typology

Granger (2002) describes the different types of computer learner corpora in terms of dichotomies. The following are the ones considered by the author to be

particularly relevant: *monolingual – bilingual, general- technical, synchronic – diachronic, and written – spoken* (p. 10).

Current research is mainly focused on monolingual corpora and aims at providing evidence on what Granger refers to as "non-nativeness, translationese, and un-Englishness" (2002, p. 10). In addition, existing corpora tend to be general rather than focusing on specialist or technical language use, and they also tend to be synchronic since they attempt to represent learner use at a particular point in time. Finally, research mostly focuses on the study and compilation of written language as opposed to spoken.

3.3.4.6. Methods and tools in Corpus Linguistics

There are different methods that can be used to investigate a corpus. Different authors agree on the fact that the choice of method depends on the type of research questions that are to be answered using the evidence provided by the corpus (Biber, Conrad & Reppen, 1998; Hunston, 2002). Corpora are analyzed using tools of different kinds, some of which are computer programs developed for specific analyses like the ones described in Biber, Conrad and Reppen (1998), and others are commercially available software packages like WordSmith Tools (Scott, 2004), which is the one used in this study.

The most basic method, and more widely used in early linguistic studies, is the counting of occurrences of certain linguistic forms: lexical and grammatical items. These frequency counts can be observed in wordlists that the software produces from the search in the corpus. The behavior of each of the items in the wordlist can be studied obtaining concordance lines which display each item in its context of occurrence. Frequency lists are lists of all the types in a corpus with their corresponding frequency of occurrence that can be displayed in frequency and alphabetical order. This method is particularly useful for providing information that can be used in reference materials for language teaching, translation and basic searches for stylistic indicators. Different authors have referred to specific uses of frequencies. According to Nation (2001), frequency or word counts can be used to identify if specific items are frequent or rare and thus determine their importance in a particular context. Similarly, Flowerdew (2001) highlights the use of frequency data as a powerful tool in helping course designers select the vocabulary items to be incorporated into a course and also in

providing information to use in grading these items. In addition to this, frequency data can be used to evaluate and revise existing courses. Hyland (2000) has argued in favor of frequencies as a point of departure for more qualitative studies "using the quantitative findings as a basis for characterizing broad similarities and differences in the genres produced by particular communities" (p. 141).

There are other methods of studying corpora that can be used to obtain additional kinds of information useful to identify and analyze complex association patterns such as "the systematic ways in which linguistic features are used in association with other linguistic features and non-linguistic features" (Biber, Conrad & Reppen, 1998, p. 5). Association patterns can help to answer two main kinds of research questions: questions that focus on the use of a linguistic feature, lexical or grammatical, and questions that focus on the characteristics of texts or varieties. The use of a linguistic feature can have linguistic and non-linguistic associations. Linguistic associations can be lexical or grammatical, and non-linguistic ones can be related to the distribution of the features across registers, dialects and time periods (Biber, Conrad & Reppen, 1998). Association patterns are considered to have an important role in the description of texts and varieties such as registers, dialects, styles or individual literary works (Biber, Conrad & Reppen, 1998). Collocation and colligation are examples of association patterns.

A common method used in many corpus studies, in fact the main investigative technique, is the comparison of results obtained from one corpus with the results obtained from another corpus or other corpora. For example, the comparison of frequency lists obtained from different corpora can give interesting information about the texts collected in the two different corpora. Kennedy (1998) mentions a study that compares a corpus of general English and another of economics texts in English and the results obtained from the comparison help identify the lexical items which are more common in the field of economy (p. 101-2). Comparison of frequency wordlists can provide information on keywords, words which are significantly more or less frequent than expected in the smaller of two corpora when compared. The suite of programs WordSmith Tools (Scott, 2004) has a tool that does this automatically and uses two different statistical tests to calculate keywords: the classic chi-square test of significance with Yates correction for a 2 X 2 table, and Ted Dunning's Log Likelihood test (Dunning, 1993), which gives a better estimate of keyness, especially when contrasting long texts or a whole genre against a reference corpus (Scott, 2004). Gledhill (in

Hunston, 2002) uses the identification of keywords as a point of departure in investigating the linguistic characteristics of the different sections of the research article. Comparison is also very important in register variations. According to Biber, Conrad and Reppen (1998), frequencies in a register can be judged as common or rare only when compared to other registers. In addition, comparison is central in learner corpora studies because comparisons between native and non-native productions can highlight a range of features of non-nativeness such as instances of under- and overrepresentation of words, phrases and structures (Granger, 2002). Different types of studies comparing a learner corpus with a reference corpus have been carried out on a variety of lexical and grammatical topics such as complement clauses, direct questions, tenses, modals, hedges/certainty markers, adjective intensifiers, and connectors (Barlow, 2005).

In this study two corpora will be used to observe and compare the behavior of a list of derivational –ly adverbs across the sections of RAs in English: a reference corpus and a learner corpus.

3.4. EVALUATION

The expression of writer's and speaker's opinions is an important feature of language that has been studied by different authors using a wide range of terms: *connotation, affect, attitude, evaluation, appraisal,* and *stance* (Thompson & Hunston, 2000). Thompson and Hunston define evaluation as "the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about" (2000, p. 5). The opinions can be expressed along two parameters: good/bad and certain/uncertain.

The study of the realizations of evaluation has been the focus of attention of different authors following two different approaches: the *separating approach* (Halliday, 1994; Martin & Rose, 2003) and the *combining approach* (Biber & Finegan, 1988; Conrad & Biber, 2000; Thompson & Hunston, 2000). The distinction between the two approaches is based on whether the opinions along the good/bad parameters and the certain/uncertain parameters are treated separately or in combination. In other words, separating approaches deal with modality separated from attitudinal meaning and combining approaches deal with modality and attitudinal meanings together. In Halliday's (1994) model, both modality and attitudinal meanings belong in the category

of interpersonal meanings, with modality having two subcategories: modalization, for propositions related to probability and usuality, and modulation, for proposals related to obligation and inclination. They are ways of constructing the semantic space between the positive and negative poles. The following examples for the categories are provided in Halliday (1994, p. 357):

3.1) *There can't be many candle-stick makers left.* (modality: probability)
3.2) *It'll change right there in front of your eyes.* (modality: usuality)
3.3) *The roads should pay for themselves, like the railways.* (modulation: obligation)
3.4) *Voters won't pay taxes anymore.* (modulation: inclination)

Attitudinal meanings can be expressed by the epithet, one of the functional elements of the nominal group, which indicates some quality. When the quality is objective the epithet is considered to be defining and experiential, but when it is subjective the epithet is considered to be attitudinal and interpersonal. The following examples are provided in Halliday (1994, p. 184):

3.5) *the **long** train* (experiential epithet)
3.6) *there came a **mighty** train* (interpersonal epithet)

Martin and Rose (2003), within the systemic framework, expand the treatment of attitudinal meanings and use the term "appraisal" for the system of interpersonal meanings that are negotiated in texts. They establish three categories of resources for expressing attitudes: affect, judgment, and appreciation. *Affect* is used to refer to the way "people express their feelings", *judgment* is used to refer to resources used for "judging character", and *appreciation* is used to refer to "the resources for valuing the worth of things" (p. 24). These are resources for expressing attitudes, amplifying them and attributing them to sources. The authors provide the following examples of these resources:

Affect:

3.7) *We were **ecstatic**.* (p. 25)

3.8) *I was **torn to pieces**.* (p. 25)

Judgment:

3.9) *Our leaders are **too holy** and **innocent**. And **faceless**.* (p. 29)

3.10) ***Respectable** members of their communities.* (p. 31)

Appreciation:

3.11)*a **beautiful** relationship* (p. 34)

3.12)*my **unsuccessful** marriage* (p. 34)

Among the authors who adhere to the combining approach we find Thompson and Hunston (2000) who use the term *evaluation* to refer not only to the manifestation of speaker/writer's attitudes – which may be related to certainty or obligation or desirability or any number of other sets of values-, but also to feelings and viewpoint. Their approach to evaluation aims at accounting for "why, when, how, and what speakers and writers evaluate" (p. 6). According to Thompson and Hunston (2000), evaluation is used to perform three functions, which are not exclusive since one instance of evaluation may perform more than one function at the same time. Hence, evaluation is used for the expression of the writer's or speaker's opinion and thus reflects the value system of that person and the community s/he belongs to. It is also used for the construction and maintenance of relations between the producers and audience, and for the organization of discourse.

Biber et al. (1999) and Conrad and Biber (2000) use the term *stance* to refer to the communication of speakers and writers' personal feelings, attitudes, value judgments, and assessment, alongside with the communication of propositional content. For this reason they are also considered to be representative of the combining approach. Other researchers mentioned in Biber and Finegan (1988) discussed similar communicative functions in terms of *posture and stance, hedges, footing*, and *evidentials*.

Biber et al. (1999) propose three categories for stance: epistemic, attitudinal and style. Stance can be expressed in different ways which include paralinguistic devices, word choice, and grammatical devices. Non-linguistic, that is, paralinguistic, devices are associated to speaking. In this study, the focus is on the linguistic devices that are used in writing.

Word choice, meaning value-laden word choice, includes affective or evaluative words which depend on the context and shared background for their interpretation, as

the grammatical structure of these expressions does not show that they mark stance. This type of stance is inferred from the use of an evaluative lexical item such as an adjective, main verb, or noun (Biber et al. 1999). Value-laden word choice is a form of expressing stance that is not grammatically marked and that is effective provided that the addressee is able to recognize its use. In other words, the grammatical structure of these expressions does not show that they mark stance and that is why it is considered that stance is "embedded in these structures" (Biber et al. 1999, p. 969). The following is an example of this kind of stance where the attributive adjective is a purely lexical expression of stance which can only be interpreted in relation to the context and shared background knowledge. The use of the adjective *appropriate* is not grammatically marked as an expression of stance, so its effective use in this sentence will depend on the addressee's ability to identify it as an evaluative item with positive connotation. However, the evaluative force this adjective describing "moral regulation" has needs to be interpreted in the light of the shared values of the cultural group where the text is produced and its specific context of use (Channel, 2000).

3.13)*The division of economic functions has temporarily outstripped the development of* ***appropriate*** *moral regulation.* (Acad) (p. 969)

Grammatical stance devices include two linguistic components: one that presents the stance and the other that presents a proposition that is framed by that stance. As in the following examples from Biber et al. (1999):

3.14)***Typically****, the Urgonian limestones are thought of as rudist reef deposits. (Acad)* single adverb (p. 972)

3.15)*It's* ***amazing*** *that judges can get away with outrageous statements (News)* stance complement clause: extraposed structure (p. 969)

The structural relation between the constituent presenting the stance and the constituent presenting the qualified proposition can vary since the latter can be part of the main clause or an embedded structure. In the first case, the proposition is given in the main clause and the stance marker occurs in some peripheral or embedded structure as in the case of adverbials and constructions with modal verbs. Thus, in the sentence in Example 3.16 from Biber et al (1999), the propositional information is given in the core

of the main clause, shown here between square brackets, while the device marking stance, in this case a single derivational adverb, is shown in bold.

3.16)***Sadly**, [the troubles ended all that].* (News) (p. 971)

In other cases, the proposition is embedded as in the following sentence:

3.17)**I was getting a bit upset** [*that my voice was going a bit*]. (Conv) (p. 971)

Here the propositional information is expressed by a that-complement clause, shown between square brackets.

As regards position in the clause, in most cases the stance marker precedes the informational proposition, although adverbials can also occur in other positions. According to Biber et al. (1999), this ordering reflects the "primary function of stance markers as a frame for the interpretation of the propositional information" (p. 971).

Grammatical stance can be marked by devices from different structural levels: clauses (such as adverbial clauses – *as one might expect*), phrases (for example, prepositional phrases functioning as adverbials, such as *in actual fact*), and words from a wide range of classes (such as nouns – *importance*, predicative adjectives – *essential*, modal verbs - *might*, lexical verbs - *love*, and adverbs - *unfortunately, surprisingly*).

According to Biber et al. (1999, p. 969) the different grammatical devices that express stance form a cline in the extent to which they represent the grammatical marking of stance. In decreasing order of grammatical integration, the devices proposed by Biber et al. (1999) are the following: premodifying adverbs, stance noun + prepositional phrase constructions, modals and semi-modals, complement clauses and stance adverbials. Premodifying adverbs are incorporated into a phrase and frame the phrase locally, rather than reporting a stance towards an entire proposition. The following examples from Biber et al. (1999) show this:

3.18)*I'm **so** happy for you. Honestly, I'm **really** happy for you. (Conv)* stance adverb + adjective (p. 970)

3.19)*Orogenies and accompanying metamorphism of* ***about*** *this age (that is,* ***about*** *478 million year B.P.) have been recognized. (Acad*)* stance adverb + noun phrase (p. 970)

Stance noun + prepositional phrase constructions have two structural components but they are more integrated than stance adverbials and stance complement clauses because it is not always clear that the prepositional phrase actually represents a proposition. This is an example from Biber et al. (1999):

3.20)*They deny* ***the possibility of*** *a death wish lurking amidst the gardens of lust. (Acad)* (p. 970)

Modals and semi-modals are incorporated into the main clause, which is also the one containing the framed proposition, as shown in the following examples from Biber et al. (1999):

3.21)*I* ***might*** *be up before you go. (Conv)* (p. 970)
3.22)*She* ***has to*** *go to a special school. (Conv)* (p. 970)

Complement clauses and stance adverbials express stance by means of two structural components: one clause expressing the stance and the other presenting the framed proposition.

The following are examples from Biber et al. (1999, p. 969) show the framed proposition between square brackets:

A. Stance complement clauses

3.23)*I just* ***hope*** *[that I've plugged it in properly].* (Conv) controlled by a verb

3.24)*I'm* ***very happy*** *[that we're going to Sarah's].* (Conv) controlled by an adjective

B. Stance adverbials

3.25)***Unfortunately****, [we cannot do anything about it].* (News) Single adverbs and adverb phrases

3.26)***In actual fact*** *[only a fraction of this number actually occurs].* (Acad) Prepositional phrase

3.27)***As one might expect,*** *[Gauss didn't collaborate much with others].* (Acad) Adverbial clause

3.28)*[You just have to try and accept it],* ***I guess.*** (Conv) Comment clause

Biber et al. (1999, p. 972) distinguish three different types of meanings realized by stance markers: Epistemic, Attitudinal, and Style of speaking. These are semantic categories that can be realized by means of a variety of grammatical devices.

This study focuses on the marking of stance as realized by derivational –ly adverbs, the most frequent in academic discourse according to Biber et al. (1999, p. 540). Thus, the following description of the different semantic categories of stance only includes information on the realization of stance by derivational -ly adverbs.

The following sections (3.4.1/2/3) deal with the description of the semantic categories of Epistemic, Attitudinal and Style of Speaking stance, proposed by Biber et al. (1999). They are defined and exemplified.

3.4.1. Epistemic stance markers

Epistemic stance markers are used to present speaker/writer's comments on the status of information in the proposition. They can mark certainty (doubt), actuality, precision, or limitation; or they can indicate the source of knowledge or the perspective from which the information is given. Although these can be realized through different grammatical structures, adverbials are common in the realization of every semantic category of epistemic stance. Adverbials as realized by comment clauses, hedges (a subcategory of adverbs), prepositional phrases and single adverbs can be seen as denoting *certainty (or doubt), actuality, precision, or limitation.* Adverbials as clauses can also be seen *to mark source or perspective of knowledge.*

Marking certainty (or doubt), actuality, precision, or limitation

3.29)*It was **definitely** a case of exploiting child labour.* (News) - Single adverb (p. 972)

3.30)*In the tropics, hsemonchosis must be considered, **possibly** originating from hypobiotic larvae.* (Acad) (p. 972)

3.31)***Typically**, the Urgonian limestones are thought of as rudist reef deposits.* (Acad) - Single adverb (p. 972)

Marking the source or perspective of knowledge

Adverbial -

3.32)*Durkheim's emphasis upon the importance of constraint is **evidently** directed primarily against utilitarianism.* (Acad) - Marking source (Conrad and Biber 2000, p. 59)

3.4.2. Attitudinal stance markers

The attitudinal stance markers in this category report personal attitudes or feelings. According to Biber et al. (1999, p. 974), some of the forms seem clearly attitudinal (e.g. *fortunately, interestingly*), while some others clearly mark personal feelings or emotions (verbs such as *fear* and *love*, or adjectives, such as *happy* and *angry*, and controlling complement clauses). As regards grammatical realization, this stance category can be realized by the following structures: verb/adjective/noun + complement clause, stance noun + prepositional phrase, modal verbs, and adverbials, which may be realized by single derivational adverbs. The following examples from Biber et al. (1999, p. 974) are considered to be markers of attitude:

3.33)***Fortunately** this did not stop the women from trying.* (News)

3.34)***Interestingly** sudden electrical death is more likely following right coronary artery occlusion.* (Acad)

3.35)***Amazingly**, the ghost disappeared after the exorcism.* (News)

3.36)***Sadly** it is still not known if there are infinitely many regular primes.* (Acad*)

3.4.3. Style of speaking stance markers

The stance markers in this category present the speaker/writer's comments on the communication itself, in other words, on the manner of conveying the message. According to Biber et al. (1999, p. 857), these can be glossed as "I am being X when I say..." and in this way they are more focused on the speaker/writer than epistemic and attitude adverbials. The grammatical realizations of this category are more limited than those for the other two categories since only constructions with a verb + complement clause and adverbials (single adverbs, prepositional phrases, and adverbial clauses) can realize this type of stance. What is more, stance "adverbials are the primary grammatical device" used to present this stance (p. 975). The following examples were chosen from Biber et al. (p. 975) to show single adverbs presenting style of speaking stance:

3.37)***Honestly**, I've got no patience whatsoever.* (Conv) – single adverb
3.38)***Quite frankly**, we are having a bad year.* (News) – adverb phrase
3.39)"***Strictly speaking**", he says, "this isn't my beat".* (Fict) - adverbial clause

From the previous description, it can be observed that the different stance categories - Epistemic, Attitude and Style of Speaking stance - can be realized by adverbials which are realized by different grammatical devices. One of such devices is the adverb as a word class. Since the focus of this study is on derivational –ly adverbs, it is necessary to devote some space to the description of the adverb as a word class.

3.4.4. Stance Adverbs

Adverbs are used by speakers and writers to add information and a variety of meanings that are not indicated by the other parts of speech. Halliday (1994, p. 25) points out that the term *adverb* is a *class label* as opposed to *functional labels* used to name linguistic units such as subject, object, complement, to name but a few. The adverb class is so heterogeneous that it is often described as "an amorphous collection of linguistic items" (Bloor & Bloor, 1995, p. 22) and a "ragbag category where words

get put if they do not fit anywhere else" (Thompson, 1993, p. 14). Such descriptions of the class may partly explain the versatility of its members.

3.4.4.1. Classification

Adverbs are classified in a variety of ways depending on the parameter taken into account, namely *form, position, function* and *meaning*. As regards form, Quirk et al. (1985) offer a morphological classification, by grouping them into three classes: *simple* (just, well, down, near), *compound* (somehow, somewhere, therefore, hereby) and *derivational* (oddly, interestingly, economically, primarily). The first two are also considered to be closed class adverbs, and the last group is considered to be open class adverbs. Biber et al. (1999) add to the morphological classification of adverbs the category of *fixed phrases*, phrases which are invariant in form such as *of course*, *kind of*, and *at last,* whose components rarely retain their independent meaning.

The adverbs studied here are -ly derivational adverbs functioning as stance markers in the roles of modifiers and adverbials. Section 3.4.4.2 describes the positions of -ly derivational adverbs, together with their functions and meanings.

3.4.4.2. Position

3.4.4.2.1. Adverbs as modifiers

Adverbs modifying adjectives usually precede the adjective that they modify, but postmodification also occurs. The following are examples provided in Biber et al. (1999).

Premodification:

3.40) *This is **slightly** larger than the calculated value.* (Aca*) (p. 544)

3.41) *The hospital mortality at 21 days for those who received streptokinase was not **significantly** different from the control group. (Aca*)* (p. 546)

Postmodification:

3.42)*It is rich* ***nutritionally*** *with high calcium content.* (Aca) (p. 545)

3.43)*Several preparations are available* ***commercially****.* (Aca*) (p. 545)

Adverbs modifying other adverbs: in this case adverbs usually precede the adverb they modify.

3.44)*The do-it-yourself builder* ***almost*** *always uses water-repellent plywood, oil tempered hardboard or fibre cement sheet.* (Aca*) (p. 546)

Adverbs modifying other elements: most of these adverbs occur as premodifiers (3.45/6/7), however, postmodification also occurs, especially with locative information about a noun phrase (3.48) (Biber et al., 1999).

3.45)***Nearly*** *everybody came to our party.* Pronouns – (Quirk et al., 1985, p. 449)

3.46)***Virtually*** *all the students participated in the discussion.* Predeterminers – (Quirk et al. ,1985, p. 449)

3.47)***Nearly*** *a thousand demonstrators attended the meeting.* NP with indefinite article - (Quirk et al., 1985, p. 450)

3.48)*Thus, in [the ammonia example]* ***above****, if ammonia, NH3, is allowed to escape from the reaction system, the reaction cannot achieve equilibrium.* (Aca). Postmodification of a noun phrase - (Biber et al., 1999, p. 548)

3.4.4.2.2. Adverbs as adverbials

The single adverb functioning as an adverbial has a distinctive difference with other elements of the clause due to the relative freedom with which it can be placed in different positions in a sentence. The range of possible positions varies according to the type of grammatical realization of the adverbial. However, for the sake of clarity, I will use the examples provided in Quirk et al. (1985, p. 491), which show the range of possible positions by moving the same adverbial to the different positions:

Initial:	**By then** the book must have been placed on the shelf.
Initial Medial:	The book **by then** must have been placed on the shelf.
Medial:	The book must **by then** have been placed on the shelf.
Medial Medial:	The book must have **by then** been placed on the shelf.
End Medial:	The book must have been **by then** placed on the shelf.
Initial End:	The book must have been placed **by then** on the shelf.
End:	The book must have been placed on the shelf **by then**.

Initial position is the position that precedes any other clause elements, that is, the position before the subject. Although Quirk et al. (1985) consider that in subordinate or coordinate clauses initial position is the position following the conjunction, in this particular study adverbs were classified as being in initial position only when they were sentence initial.

3.49)***Suddenly**, the driver started the engine.*

3.50)*I had scarcely got into the taxi when **suddenly** the driver started the engine.*

According to Biber et al. (1999), stance adverbials of all sorts can occur in initial position. This is shown in the following example:

3.51)*But **personally** I find it quite a relief.* (Conv) (p. 773)

Medial position is the position immediately after the Subject and the operator, when there is one, as shown in the following examples from Quirk et al. (1985):

3.52)*The driver **suddenly** started the engine.* (Subject – verb / initial Medial)

3.53)*They are **seriously** considering him for the post.* (Subject – operator/ initial Medial)

3.54)*He was* **probably** *unhappy* (Medial)

3.55)*They will have **seriously** considered him for the post.* (Before the main verb / end Medial)

In the examples above, it can be observed that medial position has several variants, namely, initial medial, medial, medial medial, and end medial. It is important to mention that although stance adverbials can be freely placed in all positions, corpus findings discussed in Biber et al. (1999, p. 772) show that these preferably occur in medial positions:

3.56)*They are **probably** there now.* (Conv)

3.57)*There is **actually** a very sound reason why Ray chose this amount.* (News)

Another common position for stance adverbials is the position immediately after the subject in clauses with and without an operator as shown in the following examples:

Clause with an operator:

3.58)*Caring workers **generally** are "entrusted" with a burden of responsibility which many people would find impossible.* (Aca)

Clause without an operator:

3.59)*The word gossip **actually** means "God's kin".* (News) (p. 773)

End position is the position in the clause following all obligatory elements; it is also the position of the obligatory adverbial when this follows the other obligatory elements. The following examples show the range of possibilities of adverbials realized by different grammatical structures (Quirk et al. 1985, p. 498).

After SV:	The light was fading **rapidly.**
Adverbial in SVA:	Dr Blackett is **in Tokyo.**
After SVO:	She was digging a trench **in the garden.**
After SVOO:	He gives his car a wash **every week.**
Adverbial in SVOA:	He put the vase in the cabinet **without a word.**
After SVC:	They became teachers **in the end.**

According to Biber et al. (1999), stance adverbials of all sorts can occur in end position, as shown in the following example:

3.60)*It was already out* ***actually***. (Conv) (p. 774)

3.4.4.3. Syntactic roles

As regards syntactic roles, different authors observe that adverbs can function as modifiers of adjectives, adverbs, and other elements such as noun phrases, pronouns, predeterminers, prepositional phrases, particles of phrasal verbs, numerals, and numerical expressions, and they can also function as complements of prepositions and clause element adverbials (Biber et al., 1999; Quirk et al., 1985). The following exemplify some of these functions:

3.61)*This is* ***slightly*** *larger than the calculated value.* (Acad*) - Modifying an adjective (Biber et al, 1999, p. 544)

3.62)*They'll figure it out* ***really*** *fast.* (Conv) - Modifying an adverb (Biber et al, 1999, p. 546)

3.63)*It is still not clear whether the* ***approximately*** *250 people still listed as missing include those whom ex-detainees say were still alive in May.* (News) - Modifier of numeral (Biber et al, 1999, p. 548)

3.64)*Tosi (1984:27-34, Figure 3) estimates that* ***roughly*** *one-quarter to one-third of the total surface of the area of four sites in "prehistoric Turan" was devoted to different craft activities.* (Acad) Modifier of a numerical expression (Biber et al, 1999, p. 548)

3.65)*He took it in* ***slowly*** *but* ***uncomprehendingly***. (Fict) Circumstance adverbial (Biber et al, 1999, p. 549)

However, the classifications these authors provide of adverbs as clause element adverbials are based on different principles since Quirk et al. (1985) provide structurally-based classes while Biber et al. (1999) base their classification on semantic criteria. Thus, Quirk et al. (1985) classify adverbs as clause element adverbials into

Adjuncts, *Subjuncts*, *Conjuncts, and Disjuncts*. Adjuncts and subjuncts are relatively integrated within the structure of the clause, while Disjuncts and Conjuncts have a more peripheral relation in the sentence.

Biber et al. (1999) group adverbs as clause elements following semantic criteria and provide the following categories: *Circumstance adverbials* (A_c), *Linking adverbials* (A_l), and *Stance adverbials* (A_s) depending on the function of each class of adverbials and the extent to which they are integrated into the clause structure. Circumstance adverbials are the most varied class of the three and they are more integrated into the clause structure. Linking and Stance adverbials are more peripheral and Stance adverbials are optional.

3.4.4.4. Semantic functions

Different authors seem to agree on the semantic functions of adverbs although they refer to these functions using different terms: modal adjuncts (Halliday, 1994; Bloor & Bloor, 1995), adjuncts of modality and disjuncts (Quirk et al., 1985), and Epistemic, Attitudinal and Style of Speaking stance adverbs (Biber et al., 1999). Halliday (1994) refers to them as *modal adjuncts* and sub-classifies them into two main types: one includes probability, usuality, typicality and obviousness; while the other includes opinion, admission, persuasion, entreaty, presumption, desirability, reservation, validation, evaluation, and prediction (p. 49). According to Bloor and Bloor (1995), these adjuncts have the function of "hedging or modulating the proposition" (p. 55).

Quirk et al. (1985) use *adjuncts of modality* and *content and style disjuncts* to discuss the role of adverbs. Modality has three aspects: emphasis, approximation and restriction. Adjuncts of modality can be used to enhance or diminish the truth value or force of a sentence as shown in the following examples from Quirk et al. (1985, p. 485):

Positive Emphasis:

3.66)*She has **certainly** been enthusiastic about her work.*

Negative Emphasis:

3.67)*She has **hardly** been enthusiastic about her work.*

In between these two there is what Quirk et al. (1985, p. 485) call *approximation,* which can also include a comment on the form of or the authority for the statement. The following examples show this:

3.68)*She has **probably** been enthusiastic about her work.*
3.69)*She has **allegedly** been enthusiastic about her work.*

The third aspect is called *restriction,* where adverbs are used to focus upon a particular part of the statement as in the case of *only* in the following:

3.70)*She has been enthusiastic **only** about her work.*

In relation to the function of adverbs as *disjuncts*, Quirk et al. (1985) group them into two classes, *style* and *content*. Style disjuncts "convey the speaker's comment on the style and form of what he is saying" providing in this way clues as to the conditions under which the utterance is produced. Content disjuncts "make observations on the actual content of the utterance and its truth conditions" and they are sub-classified into two groups depending on whether they make an observation in relation to "degree of or conditions for truth of content" or the "value judgment of content" (p. 615).

According to Biber et al. (1999), adverbs can be used to express the three types of stance: Epistemic, Attitudinal and Style of Speaking and they propose the following subcategories. Epistemic stance adverbs are similar to *Type a content disjuncts* which, according to Quirk et al. (1985), express a degree of truth and doubt, or make reference to the "reality or lack of reality in what is said", or express a contrast with reality (p. 621).

The following are the sub-categories for Epistemic Stance provided and exemplified by Biber et al. (1999, p. 557).

3.4.4.4.1. Epistemic Stance Adverbs:

a. they can show different levels of **certainty** or **doubt**
e.g. probably / definitely

3.71)*My ideas about food are **definitely** passé. (News)*

b. they can comment on the **reality** or **actuality** of a proposition
e.g. actually / really

3.72)***Actually** I'm not very fussy at all. (Conv.)*

c. they can be used to show that a proposition is based on some **evidence,** without specifying the exact source
e.g. apparently/reportedly

3.73)*The supernumerary instar is **reportedly** dependent on the density of the parental population. (Acad)*

d. they can be used to show the **limitations** on a proposition
e.g. mainly/typically

3.74)***Typically**, the front top six teeth will decay because of the way the child has sucked on its bottle. (News)*

e. they can be used to convey **imprecision** (also called hedges which are in turn called approximators when they modify numerical of other quantifying expressions)
e.g. kind of/ sort of/roughly

3.75)*The ratio of clerks to total employees in the same manufacturing industries **roughly** averages 9% (Acad*).*

3.76)*So now I needed a job that I could do for **approximately** [four months]. (Fict).*

3.4.4.4.2. Attitude Stance Adverbs

These tell a speaker's or writer's attitude toward a proposition. They do this by typically conveying an evaluation, value judgment, or assessment of expectations (Biber et al., 1999). These are similar to *Type b content disjuncts* which, according to Quirk et al. (1985), "convey some evaluation of or attitude towards what is said" (p. 621).

3.77)***Curiously**, an (at first sight) almost diametrically opposed argument may be advanced without contradicting the above. (Acad).* (Biber et al, 1999, p. 558)

3.4.4.4.3. Style Stance Adverbs

These adverbs comment on the manner of speaking which the speaker/writer is adopting. They manifest whether the speaker or writer is using the language sincerely, frankly, or simply. These are similar to *style disjuncts* which, according to Quirk et al. (1985), can be sub-classified into two groups: modality, and manner and respect. Style disjuncts draw attention not only to what is said but also to how it is being said.

3.78)***Briefly**, the aim was to encourage particular schools to develop and implement learning resources plans...* (Acad) (Conrad and Biber, 2000, p. 60)

In this section I have described the characteristic features of adverbs and adverbs functioning as adverbials in relation to semantic class, position, and syntactic role. Tables 3.1 and 3.2 include the categories that will be used in the analysis of frequency and distribution of stance adverbs in the two corpora used in this study.

Table 3.1.: Stance categories, syntactic role and semantic function

Stance Categories	Syntactic Role and Semantic Function of Stance Markers
Epistemic Stance	Adverbs marking • certainty (or doubt) • reality or actuality • evidence • limitations • imprecision
	Adverbials marking • the source or perspective of knowledge
Attitudinal Stance	Reporting personal attitudes or feelings
Style of speaking Stance	Presenting the speaker/writer's comments on the communication itself

Table 3.2.: Position of adverbs and adverbials

Position	
Adverbs as modifiers	· Premodification · Postmodification
Adverbs as adverbials	· Initial · Medial · End

In the following chapter, Chapter IV, I will describe the methodology and analytical tools used in this study.

CHAPTER IV

METHODOLOGY

This chapter describes the corpora used in this study by making reference to the theoretical principles and concepts for corpus building described in the theoretical framework in Chapter III. It also describes the methodology and analytical tools used to collect, classify, and process the data provided by the corpora used in this work.

4.1. The corpora

The data for this study was gathered from two corpora of research articles (RA) in English built taking into consideration the notions and principles described and proposed by Sinclair (1991, 2005). One of the corpora can be described as a Reference Corpus (RC) and the other one is a corpus of learner English. The two corpora used in this study were built by the research team directed by Iliana A. Martínez and Silvia C. Beck, of which I am a member.

In sections 4.1.1 and 4.1.2, I will describe in detail each of the corpora used in the present study.

4.1.1. The Reference Corpus

The Reference Corpus is a one-million word corpus of experimental research articles produced in the field of biological sciences and published in countries where English is the native language. This corpus is considered to be a reference corpus because of the purpose for which it was compiled, since it is used as a baseline in comparison with other more specialized corpora (Hunston, 2002). The evidence it provides shows characteristic features of RAs published in countries where English is the first language. The language used by the authors of the texts in the corpus may be regarded as the target to attain by NNSE writing RAs.

The RC is also representative of the genre of published research writing in English in the specific field of biology. It comprises 213 research articles published in international journals. Thus, it is also a specialized corpus because it is delimited by a specific genre and discourse domain and it is expected to provide evidence for questions of language used in that specific genre and domain (Beaugrande, 2001). Although the

notion of balance (See chapter III, section 3.3.4.2) was taken into account in the design of this corpus, it was not problematic in this particular case due to the fact that it is a specialized corpus that seeks to represent a very specific population of writers and writing. In addition, the notion of sampling (See chapter III, section 3.3.4.2) was also taken into consideration for the inclusion of texts. The ones selected were those that were available in our institution - the National University of Río Cuarto (UNRC) - and on the Internet at the moment the corpus was built, and had the four sections: Introduction, Methods, Results, and Discussion (Swales, 1990). As regards size, the decision to include whole texts prevailed over the possibility of inclusion of samples of identical size. Thus, the compilation aimed at maintaining the integrity of each text as a whole document (Sinclair, 2005). In addition to this, the texts have also been separated into four different sub-corpora that contain each of the four sections of the RAs to allow for comparisons across sections. Table 4.1 shows the number of tokens (running words) in the corpus and in each of its sections. The size of the corpus is in accordance with what is expected for a reference corpus, which must be larger than the corpus of study. For certain analyses, Berber Sardinha suggests that the reference corpus should be about 5 times the size of the corpus being studied (in Scott & Tribble, 2006), although, as Scott and Tribble say, "This does not invalidate choosing one much bigger, of course!" (p. 65).

Table 4.1: Distribution of running words in the Reference Corpus

Introduction	**Methods**	**Results**	**Discussion**	**Reference Corpus**
152,048	329,301	424,972	293,444	1,199,765

In addition to the notions of representativeness, balance, and sampling described above, the selection of the texts included in the RC was done considering the external criteria proposed by Sinclair (2005), namely, research field, genre, sections, country and date of publication, journal impact factor, and availability.

As already mentioned, the texts of the RC were selected according to the research field where they were produced, the field of biological sciences, and the genre of published RAs. In addition, each RA was selected considering its internal structure including only those articles which presented the IMRD format: Introduction, Methods or Methods and Materials, Results and Discussion sections (Swales, 1990). The articles

in the corpus were drawn from journals published after 1992 in English-speaking countries. This allowed us to control for representativeness, as it was the closest form of ensuring that the authors were either native speakers or that the text had been revised by a native speaker. The impact factor of journals, the frequency with which a published article is likely to be cited in a particular year, was chosen as another criterion for it is well known that high impact factor journals place heavy demands on the language of the articles they publish. From the journals available, either in the library of the National University of Río Cuarto or from online journals, the ones selected were those which have an impact factor of 1 or higher as reported in the S.C.I. Journal Citation Report of 1995, which was the version available to us at the moment of the corpus collection. All the journals complied with this criterion, except for the *Microbiology Journal*, which was included because it was pointed out as a journal widely used by local researchers and because of its availability at the time of compilation. Table 4.2 shows the list of journals and their impact factors at the time of construction.

Table 4.2: Journals in the RC with impact factor

Journals	Impact Factor
Applied and Environmental Microbiology	3.211
Biochemistry	4.159
Blood	3.569
Cell	40.481
Cellular Signalling	2.110
Ecology	3.131
Faseb Journal	13.404
Febs Letters	3.842
Infection and Immunity	3.721
Journal of Applied Ecology	1.382
Journal of Bacteriology	3.903
Journal of Biological Chemistry	7.385
Journal of Clinical Investigation	8.788
Journal of Clinical Microbiology	3.913
Journal of General Virology	3.410
Journal of Virology	6.033
Microbiology	0.287
Molecular And Cellular Biology	10.498
Mollecular Biology of The Cell	9.376
Proceedings National Academy of Sciences USA	10.521
The Embo Journal	13.505
The Journal of Cell Biology	12.480
The Journal of Immunology	7.412

4.1.2. The Corpus of Learner English

The learner corpus, called *Corpus of Learner English Manuscripts for Publication Purposes* (CLEMPP), was built with manuscripts of experimental RAs in English produced by native speakers of Spanish. The CLEMPP comprises 33 RA manuscripts from the field of biology published in English in international journals after 1992. The articles were produced by researchers who are native speakers of Spanish working at the National University of Rio Cuarto. The manuscripts were collected by the director of the project of which I form part. They were drawn from two sources: a.) final exams of the courses she teaches for doctoral candidates or b.) manuscripts submitted to her for language editing. In all cases the manuscripts included in the corpus were the version that the writers considered to be ready for publication. This version was selected because it is the one that exhibits the characteristics of scientific writing as produced by NNSE which is precisely what renders the comparison of the CLEMPP to the RC interesting. Of the manuscripts received, only those that were published in international journals were included in the CLEMPP. This is a form of validation, since these published manuscripts have successfully undergone the revision process and have been accepted, which implies legitimization (Hyland, 2000; Martín-Martín & Burgess, 2004).

As regards the structure of the research papers, only those articles which conform to the Introduction, Methods or Methods and Materials, Results and Discussion format (Swales, 1990), were included in the corpus. Table 4.3 shows the running words in the corpus and the sub-corpora. Table 4.4 shows the list of journals where the articles were published and their impact factor.

Table 4.3: Distribution of running words in the CLEMPP

Introduction	Methods	Results	Discussion	CLEMPP
16,501	29,125	29,194	28,676	103,496

Table 4.4: Journals and impact factor of the published articles corresponding to the CLEMPP

JOURNALS	Impact factor
Physiologia Plantarum	1.700
Polymer Degradation and Stability	0.880
Plant Growth Regulation	1.286
Seed Science and Technology	0.246
Journal of Agricultural Science	0.581
Photochemistry and Photobiology	2.215
FEBS Letters	3.842
Photochemistry and Photobiology	2.215
Current Microbiology	0.962
Annals of Botany	not indexed
FEMS Microbiology Letters	1.488
Nematologia Brasileira	not indexed
Dyes and Pigments	0.338
Biocell	not indexed
Cuad. Herpetol	not indexed
Dyes and Pigments	0.338
Genetics and Molecular Biology	not indexed
Current Microbiology	0.962
Annals of Botany	not indexed
Biocell	not indexed
Plant Growth Regulation	1.286
Canadian Journal of Botany	0.917
Acta Limnologica Brasiliensia	not indexed
Plant Growth Regulation	1.286
Crop Protection	0.401
Letters in Applied Microbiology	0.764
Pest Management Science	not indexed
Hidrobiología	0.666
Seed Science Research	not indexed
Plant Science	1.179
Canadian Journal of Chemistry	1.074
European Journal of Plant Pathology	0.844
FEBS Letters	3.842

As some of the journals where the CLEMPP articles were published were not indexed in the available version of the S.C.I. Journal Citation Report (1995), their impact factor is not included in the Table.

4.1.3. General considerations for both corpora

To build these corpora, the RC and the CLEMPP, the following concepts (see Chapter III, section 3.4.2.) were taken into consideration: authenticity, components, text

format, and documentation. First, authenticity was a notion considered in the building of the corpora. This was attained by choosing texts published in international journals, which is one of the steps in the process of scientific interaction. These texts are thus part of "genuine communications", hence authentic material (Sinclair as cited in Granger, 2002, p. 8).

Secondly, the corpora have 5 components each, the whole text and the four sections of each text. This organization allows the corpus to be compared with other corpora designed in a similar way and it also allows for the comparison of the sub-corpora containing the different sections in the collection. As regards text format, the texts have been stored as plain text. The texts have been edited to exclude the captions for tables and figures, section titles, and formulae which were tagged, that is labeled, so as to allow the program to read or ignore the elements selected.

Finally, the origin of the texts in the corpora has been carefully documented so that each file can be traced back to its source. In the RC this has been done by giving each file a label composed of part of the *journal title* and a number that refers to the number of texts taken from the specified journal. In the CLEMPP this has been done by giving each file a label composed of the word *text* and a number that refers to a specific RC manuscript. In addition to this, and for ease of identification and accessing, the selected texts were digitalized and saved in files to which different extensions were assigned: ".int" for introduction, ".met" for method, ".res" for result and ".dis" for discussion.

4.2. Tools for the analysis

The corpora were analyzed using the software WordSmith Tools (Scott, 2004), which is a suite of computer programs with three tools: Wordlist, KeyWords and Concord. The first tool used in this study was the Wordlist. This tool creates lists of words from the source texts chosen by the program user and organizes them according to different criteria such as frequency, alphabetical order, and file order. Wordlists are created on the basis of the different *types* in the corpus, that is, each word in the text counted only once, as different from the running words – or *tokens*. In this study, wordlists were the first step required in the process of analysis because wordlists form the basis for the performance of different actions with other tools. One of the functions of the Wordlist tool is the Wordlist Match List. This function helps to filter the lists

built with the Wordlist tool using a Match list with the items to be matched. A Matchlist is a file to which the software resorts to find the items to match. Once the items in the Matchlist are identified, the items can be marked to be isolated or deleted. In the present study the items in the RC and CLEMPP that matched the items in the Matchlist were marked and isolated since they were the focus of the study.

Another tool in the suite of programs is the KeyWords Tool. The first step for using this tool is to build separate wordlists from the texts/corpora to be used. This tool can locate and identify keywords in a text or corpus by comparing the two wordlists, one of which, the larger one, is referred to as the *reference file* (Scott, 2004). A word is defined as a *keyword* if its frequency is outstanding, that is unusually higher or lower than expected, in relation to the tokens in the text or corpus which is being studied, and in relation to the frequency it has in the reference corpus. The program provides information on the frequency of each keyword in raw numbers and percentages related to the tokens in the text or corpus being studied together with the frequency in raw numbers and percentages of the item in the reference file. Depending on whether their frequency is outstandingly higher or lower than expected, keywords may be identified as having *positive* or *negative keyness*. This is computed by the program which crosstabulates the frequencies and tokens in each of the texts/corpora compared and calculates keyness using statistical tests such as the chi-square and Dunning's Log function (Dunning, 1993), which are provided with the suite of programs. In this study the identification of keywords in the CLEMPP when compared to the RC provided the significant items for the study, that is, those that were found to be overused and underused in NNSE writing (See Chapter III, section 3.2.2). The statistical test used was the Log Likelihood test.

The Concord tool provided with WordSmith Tools (Scott, 2004) is a tool that searches the item indicated by the user, the *node* (Scott & Tribble, 2006), and it displays all its occurrences in a corpus showing its immediate context. The concordance of an item can provide rich contextual information on the meaning and characteristics of use of the node word within the text/corpus studied (Scott & Tribble, 2006). This tool has other functions such as identifying *collocates*, and patterns that highlight the words found in the neighboring context of the node, and may help identify characteristic *association patterns* in the corpus studied. In this study this tool was used to observe the immediate context of the words that had been identified as key words in the corpora and their sections. This was obtained using the *Pattern* function that comes with WordSmith

Tools, which allows one to see the items that are most frequently found to the left (L) and to the right (R) of a search-word or node. These items are organized in terms of frequency and in the different positions to the R or L of the node. The results obtained show the lexical patterns associated with the search-word and help make sense of great numbers of individual concordances. For the identification of patterns, the horizons were set at 5 places to the R and L, with a minimum frequency of 5 occurrences. In other words, this means that the computer searched concordance lines for the node for items that are located in positions R5 to R1 and L1 and L5 which occur at least 5 times in the corpus. The strength of collocational relationship of some of the items in the pattern was calculated by means of the *Mutual Information* (MI) score provided with the WordSmith software (Scott, 2004). MI compares the frequency of co-occurrence of two words in a given scope with their predicted frequency of co-occurrence. The higher the MI score, the more genuine the association between two words (Wang, 2005).

4.3. Method

The aim of this study was to compare the use of derivational stance adverbs in the CLEMPP and the RC. The steps prior to the analysis were to build the corpora described in this chapter in Section 4.1. and compile a list of derivational adverbs to be studied. In corpus linguistics research there are different ways of delimiting the focus of study. Researchers may choose to delimit their work by selecting a specific combination of linguistic resources or by using selected lists. For example, Charles (2003) investigates the construction of stance through nouns when they function as retrospective labels that occur with the deictic *This* in sentence initial position. The use of selected lists is a common practice in corpus linguistics research. In some cases the lists are compiled from different sources and combined with items found in the corpus studied. For example, Hyland (2000) used a list of 180 lexical expressions to study hedging and boosting which was compiled from the research literature, reference grammars, and the author's earlier research findings. Similarly, Hyland (1999) studied stance markers using a list compiled from different sources and he combined it with the most frequently occurring items found in the corpus of research articles that he analyzed. In other cases, researchers use lists that are compiled from other sources and do not add other items. For example, Hiltunen (2006) used the list 58 coming-to-know verbs found in Meyer 1997's book for his study on the use of these verbs in three

academic disciplines. Biber and Finegan (1988) worked from the subcategories described in Quirk et al. (1985) and provided lists of keywords used in the identification of stance markers proposing six categories of stance adverbials.

For the present study the list of 141 –ly stance adverbs was compiled taking into consideration criteria that would render the results comparable with other research as suggested by Granger and Tyson (1996). In their study, these authors argue in favor of selected lists "if results are to be meaningful and open to comparison" (p. 20). A selected list makes it possible to count on a set of items previously identified by other researchers as relevant because of their frequency in other corpora. The results obtained from the analysis of this set of items may be used to compare the way the items are used in corpora from similar or different genres or discourse domains. It must be noted that lists do not imply an exhaustive compilation because they are usually built on corpora which provide information about what is present in them and not what is absent (Hunston, 2002). Many authors have provided lists that have been used by others for research for example a well known list such as Coxhead's *Academic Word List* (Coxhead, 2000) has been widely used for research and materials development although recent research has demonstrated that it is not exhaustive (Chen & Ge, 2007; Hyland & Tse, 2007; Martínez, Beck, & Panza, manuscript submitted for publication).

As mentioned above the first step in the research consisted in building wordlists from the two corpora, the RC and the CLEMPP, which provided information to describe them in relation to the number of *tokens* or running words in the corpora and sub-corpora. The function Matchlist, within the tool Wordlist, was used to build a matchlist with the –ly stance adverbs that were the focus of this study. Table 4.5 shows the list of 141 stance markers used for the Matchlist. This list is the result of a compilation of the –ly adverbs identified as stance markers in Biber et al. (1999) and Quirk et al. (1985).

Table 4.5: List of –ly adverbs as stance markers

accordingly	actually	admittedly	allegedly	amazingly	amusingly
annoyingly	apparently	appropriately	approximately	arguably	artfully
Assuredly	astonishingly	avowedly	basically	bluntly	briefly
Broadly	candidly	certainly	clearly	cleverly	conceivably
confidentially	conveniently	correctly	crudely	cunningly	curiously
Decidedly	definitely	delightfully	disappointingly	disturbingly	essentially
Evidently	factually	figuratively	flatly	foolishly	formally
fortunately	frankly	fundamentally	funnily	generally	happily
Honestly	hopefully	hypothetically	ideally	importantly	incontestably
incontrovertibly	incorrectly	incredibly	indisputably	indubitably	inevitably
interestingly	ironically	justly	largely	likely	literally
Luckily	mainly	manifestly	mercifully	metaphorically	mildly
Naturally	nominally	obviously	oddly	officially	ostensibly
Outwardly	patently	personally	plainly	pleasingly	possibly
predictably	preferably	presumably	primarily	privately	probably
Prudently	purportedly	really	reasonably	refreshingly	regretfully
regrettably	remarkably	reportedly	reputedly	rightly	roughly
Sadly	seemingly	sensibly	seriously	Shrewdly	significantly
Simply	strangely	strictly	superficially	Supposedly	surely
surprisingly	suspiciously	technically	thankfully	Theoretically	tragically
Truly	truthfully	typically	unaccountably	Unarguably	unbelievably
undeniably	understandably	undoubtedly	unexpectedly	unfortunately	unhappily
Unjustly	unluckily	unnaturally	unquestionably	unreasonably	unwisely
Usually	wisely	wrongly			

The Matchlist tool is capable of performing different actions after the initial matching of the items specified in the Matchlist built for that purpose. In this case, the action specified was to delete all the unmatched items so as to be left with the items that were the focus of this study, that is to say, the *types* of each distinct –ly stance adverb that was attested in the corpora. Although this action may have resulted in the exclusion of a number of -ly adverbs that may be present in RAs in English and prove interesting to study, this may be expected as it is one of the limitations of corpus-based studies (Hunston, 2002). The study of all the –ly stance adverbs used in the corpus would require a different study.

The next step involved the use of the Concord Tool to make concordances of each of the items. Each concordance output was studied to delete the cases where adverbs were not functioning as stance markers. The following step involved the analysis of each concordance output to assign each stance adverb to one of the categories of stance markers considered in this study: Epistemic, Attitude or Style of Speaking stance markers.

The steps described above were followed to obtain information on the following:

- General frequency and distribution of stance adverbs in the RC and CLEMPP.

- Semantic frequency and distribution of stance marker types in the RC and the CLEMPP.

Next, the two corpora were compared as complete texts and across the sections of the RAs. As the collections are different in size, a procedure called *normalization* was needed to adjust the raw frequencies (tokens) and make the results obtained comparable (Biber, Conrad, & Reppen, 1998). Normalization is achieved by dividing the frequency count of the item being studied by the total number of running words (tokens) in the corpus or sub-corpora, and then by multiplying this result by the basis chosen for normalization. In the present study, I calculated this using a basis of 10,000. To compare differences of normalized proportions, the statistical significance test of differences of proportions for unrelated samples was used, with a p value of .05.

Since one of the aims of this study was to describe the use of a specific set of –ly adverbs (141 items contained in the Matchlist) in the CLEMPP and RC, the number of stance adverb *types* used by each group of writers and their *coverage* in relation to the total of instances in each corpus was considered to indicate variety of use of certain items (Hyland and Milton, 1997). In this study, the focus was on those types whose frequency covered 50% of the total number of occurrences in each corpus. When the number of types used to cover 50% of all occurrences was higher, the writing exhibited the use of a broader range of resources, while fewer adverb types to cover 50% of the uses attested in each corpus indicated a more limited range of resources.

After this, the KeyWords Tool was used to identify the –ly stance adverbs that are key in the corpora using Dunning's Log function (Dunning, 1993). The minimum frequency was set to 1 and the p value was set to 0.05 to identify the items that were over and underused in the CLEMPP (see Chapter III, section 3.2.2). At this point, it is important to mention that a single occurrence of a certain item in the CLEMPP and its absence in the RC will result in the identification of such item as a negative keyword and, therefore, it will be regarded as an instance of overuse in the CLEMPP. This single occurrence in the corpus under study is worth noting if one considers that its absence in the RC, which is considerably larger than the CLEMPP, may indicate that it is a dispreferred choice for NSE writers. In spite of the fact that single occurrences do not lend themselves to further analysis, their identification in the present study may provide information on features of non-nativeness which are characteristic of the NNSE writers of the CLEMPP.

Once the keywords had been identified, the Concord tool was used again to perform an in-depth analysis of these specific items. When frequencies of occurrence allowed, the *pattern function* (within the Concord tool) was used to identify lexical association patterns between the keyword and its context. To measure the strength of collocational relationships, the MI score provided with the WordSmith software (Scott, 2004) was used. This calculates the probability of the relationship between words based on the total corpus size in tokens, that is, it provides statistical information.

The steps described above were followed to obtain information on the following:

- Underuse and overuse of stance adverbs in the RC and CLEMPP.
- Frequency of overused and underused stance adverbs across the sections of the RC and CLEMPP.
- Association patterns of key stance markers in the RC and CLEMPP.

In the present chapter, I have described the corpora used to carry out this study, their building, the tools used for the analysis and the methodological steps followed. In the following chapter I will present the findings of this work.

CHAPTER V

RESULTS

This chapter shows the results obtained from the analysis of –ly derivational adverbs used as stance markers in the two corpora studied, the Reference Corpus (RC) and the Corpus of Learner English Manuscripts for Publication Purposes (CLEMPP).

The results are presented in sections 5.1 and 5.2. The following aspects are analyzed:

- General frequency and distribution of stance adverbs in the RC and CLEMPP
- Semantic frequency and distribution of stance marker types in the RC and the CLEMPP
- Underuse and overuse of stance adverbs in the RC and CLEMPP
- Frequency of overused and underused stance adverbs across the sections of the RC and CLEMPP
- Association patterns of key stance markers in the RC and CLEMPP

5.1. General frequency and distribution of derivational adverbs in the RC and the CLEMPP

Using the Matchlist of adverbs ending in –ly compiled using Biber et al.'s (1999) and Quirk et al's. (1985) lists, which is described in Chapter IV, containing 141 –ly derivational adverbs, I identified the target adverbs present in the RC, the CLEMPP, and their respective sections. This rendered two lists which were edited to exclude all the -ly adverbs that were not functioning as adverbs marking stance

This section presents the results in two stages: a) frequency of stance adverbs in the corpora as complete documents and b) frequency and distribution of stance adverbs across the IMRD sub-corpora. In addition to presenting results in relation to frequency, during the first stage (stage a) there is also a description of the use of stance adverb *types*. In this case the results are not normalized because the focus is on lexical variety

as indicated by number of distinct occurrences of each adverb type and their coverage. (See Methodology, section 4.3.)

a) Table 5.1 shows stance adverb tokens, the running words in a text or corpus, and normalized frequency counts. Although the raw numbers differed greatly due to the different sizes of the corpora, the normalized figures showed a surprisingly similar frequency of occurrence of stance adverbs for native speakers and non-native speakers of English.

Table 5.1: Frequency of stance adverbs in the RC and the CLEMPP. Raw and normalized numbers

	RC (1,199,765)	CLEMPP (103,496)
Tokens	**2187**	**195**
Tokens normalized at 10,000	**18,22**	**18.84**

Table 5.2 shows the number of stance adverb types used in the two corpora. The *types* are the adverbs in the Matchlist counted only once[1]. For example, the adverb *interestingly* occurs 153 times in the RC but it counts as only one type. Relating the number of types used in each corpus to the total number of types (141) available in the list, it is possible to see that NSE writers use a wider variety of stance adverbs, the specific set of resources that are the focus of study in this research work. The difference in the use of types may be due to the different size of the corpora. Still, the RC provides a more numerous list of stance adverbs used by native speakers, which can be used as a pedagogical tool.

Table 5.2: Frequency of stance adverbs types in the RC and the CLEMPP

	RC (1,199,765)	CLEMPP (103,496)
Types	**57**	**32**

[1]The use of the term *type* here is not related to categories of adverbs.

Table 5.3 shows the list of adverbs that were used by both groups of writers - a total of 30 items, the list of adverbs used by only NSE (27 items) and those used by only NNSE (2 items). The learning of vocabulary items seems to place different demands on the part of NNS learners depending on the linguistic distance between the L1 and target L2 (Granger, 1993; Melka, 1997; Schmitt, 1997). One way of looking at the items in the lists is to observe the similarities and differences they share when comparing English (L2) and Spanish (L1). *Cognate* is the term used to describe words in different languages which are historically related or borrowings which are similar or identical in form and meaning (Granger, 1993; Schmitt, 1997). Granger (1993) makes a distinction between *good cognates*, which are words with the same meaning in both languages (e.g. *similarly*), and *deceptive cognates*, which have partially or totally different meanings (e.g. *actually*). These are also called *false friends* (Schmitt, & M. McCarthy, 1997). A third group, *non-cognates* (Granger & Swallow, 1988), may also be identified. These are the words in L2 that are not formally related to their equivalents in Spanish (e.g. *remarkably*). They are shown in Table 5.3. I have identified the adverbs that may be described as good cognates in normal print. Those perceived as deceptive cognates or false friends in bold, and the ones perceived as non-cognates in italics. Of the total number of stance adverbs used in both corpora, 66 % were transparent, 32 % were non-cognates and 2% were deceptive cognates. The Table shows that most of the adverbs used by the NNSE were good cognates (75 %), with only 22 % non-cognate adverbs used. This is evidence of the effect of L1 influence. In addition, it may also be argued that deceptive cognates are used less frequently because they place greater demands on the part of the NNSE writer, who may feel reluctant to use them and choose the ones that are good cognates instead. Of the 27 stance adverbs that only NSE used, 45 % were not good cognates. Two adverbs occurred only in the NNSE corpus, *basically* and *definitely*. These may be associated to a negative influence of the mother tongue, as the words "*básicamente*" and "*definitivamente*" are very common in the Spanish spoken in our country. In section 5.1.2 I will analyze the use of *basically* and *definitely* in the NNSE corpus. The adverb *actually* derives form the adjective *actual* that has been identified in the literature as a "confusing false cognate" for Spanish L1 users (Laufer, 1997, p. 162), usually wrongly used by NNSE meaning "presently". Coates et al. (2002) present *actually* as an example of inaccurate word use in the category they labeled "jargon". In their study the researcher/editors replaced *actually* with "now".

Table 5.3: Derivational stance adverbs used in RC and CLEMPP

Common to RC and CLEMPP	**RC only**	**CLEMPP only**
1. Actually	*1. Accordingly*	1. basically
2. apparently	*2. admittedly*	2. definitely
3. appropriately	*3. arguably*	
4. approximately	*4. broadly*	
5. briefly	5. *conceivably*	
6. certainly	6. correctly	
7. clearly	7. decidedly	
8. curiously	8. formally	
9. essentially	9. fundamentally	
10. evidently	10. hypothetically	
11. fortunately	11. ideally	
12. generally	12. incorrectly	
13. importantly	13. inevitably	
14. interestingly	14. metaphorically	
15. largely	*15. mildly*	
16. likely	16. nominally	
17. mainly	*17. ostensibly*	
18. obviously	18. reasonably	
19. possibly	*19. reportedly*	
20. presumably	*20. roughly*	
21. primarily	*21. seemingly*	
22. probably	22. superficially	
23. really	*23. surely*	
24. remarkably	24. surprisingly	
25. simply	25. theoretically	
26. strictly	*26. truly*	
27. typically	27. unfortunately	
28. undoubtedly		
29. unexpectedly		
30. usually		

Note. Good cognates in normal print, deceptive or false cognates in bold, non-cognates in italics

The distribution of types across the sections in the two corpora showed wider variability of -ly stance adverbs in all the sections in the RC with respect to the CLEMPP. The results are shown in Table 5.4. The findings in the RC and the CLEMPP show that the Discussion is the section where more variety of –ly stance adverbs is observed, followed by the Results, Introduction, and the Methods. The results indicate that the writers of both corpora use more stance markers in the section where there is more argumentation, the Discussion section.

Table 5.4: Frequency of stance adverb types in the sections of the RC and the CLEMPP

	Introduction		Methods		Results		Discussion		Total	
	RC (152,048)	CLEMPP (16,501)	RC (329,301)	CLEMPP (29,125)	RC (424.972)	CLEMPP (29,194)	RC (293,444)	CLEMPP (28,676)	RC (1,199,765)	CLEMPP (103,496)
Types	37	15	21	7	44	17	51	21	57	32

Table 5.5 shows the coverage of adverb types used in the two corpora. This was calculated taking into consideration the top adverb types in the wordlists and their cumulative frequency. The purpose was to identify the types that represented approximately 50% of the -ly stance adverbs tokens in each corpus to account for their coverage. Variety was measured taking into consideration the number of types that covered 50% of the stance adverbs occurring in the corpora. The results show that NNSE used a more limited range of items and that they repeated these more often than NSE writers. The data showed that NSEs' use of adverb types was more varied in most sections except for the Methods sections where the use of only two types accounted for over 50% of the adverbs used in each corpora. Since the difference might be attributed to the difference in the sizes of the corpora used, a sample of 20 full texts (100,000 tokens) of the RC was compiled to observe if the coverage differences were due to size. The results of this search showed that the difference was not due to the size of the corpora, as the coverage remained constant.

Table 5.5: -ly stance adverbs in the sections of each corpus covering around 50% of the tokens used

INTRODUCTION		METHODS		RESULTS		DISCUSSION		FULL TEXT	
RC	CLEMPP	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP
primarily probably presumably usually likely interestingly typically generally	Mainly approximately generally primarily	briefly approximately	briefly usually	approximately likely clearly interestingly presumably	approximately mainly clearly	likely probably clearly interestingly approximately possibly presumably	mainly probably clearly approximately	likely approximately probably interestingly clearly briefly primarily presumably	mainly approximately clearly probably

b) In a second step, the frequencies of stance adverbs in the corpora were compared across the sections of the RA. The resulting frequencies were normalized for comparison (Table 5.6). The differences were not statistically significant. Table 5.6 shows that the distribution of stance adverbs in NNSE and NSE corpora was similar across the sections of the RA. The Discussion sections in both corpora were the ones where stance adverbs occurred most frequently. This was followed by the Results sections. The Introduction and Methods sections were the third and fourth, respectively. This similarity in the distribution suggests that NNSE writers are aware of the rhetorical functions of each of the sections in the RA.

Table 5.6: Frequency of -ly stance adverbs in the sections of the RC and the CLEMPP ($\alpha < .05$)

Test of differences of proportions for unrelated samples										
	SECTION								TOTAL	
Section	Introduction		Methods		Results		Discussion		Total	Total
	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP
N	152048	16501	329301	29125	424972	29194	293444	28676	1199765	103496
Tokens	274	28	369	24	768	60	776	83	2187	195
Tokens normalized at 10,000	18.02	16.97	11.21	8.24	18.07	20.55	26.44	28.94	18.22	18.84
$z \geq 1.96$	0.3035		1.4656		-0.961		-0.7833		-0.4428	
Significance $\alpha \leq .05$	0.3808		0.0714		0.1683		0.2167		0.3290	

Figure 5.1 presents the findings described above. It shows the distribution of stance markers across the sections of the research article in English in the two corpora analyzed in the present study. Although the distribution across the sections was similar in both corpora, Figure 5.1 clearly shows that the normalized frequencies of stance adverbs, although not significant, are surprisingly higher in the Discussion and Results sections of the CLEMPP than in the same sections of the RC. These findings are broadly consistent with those of Salager-Meyer (1994) with respect to the occurrence of hedges, particularly *shields,* which are included in the Epistemic Stance category of the present study. This may be attributed to the fact that *mainly* is overused in both sections (see Tables 5.21 and 5.19) and may also be related to the fact that it is an Epistemic stance adverb which may be safely used in a variety of linguistic contexts. This item is further analyzed in section 5.1.2.

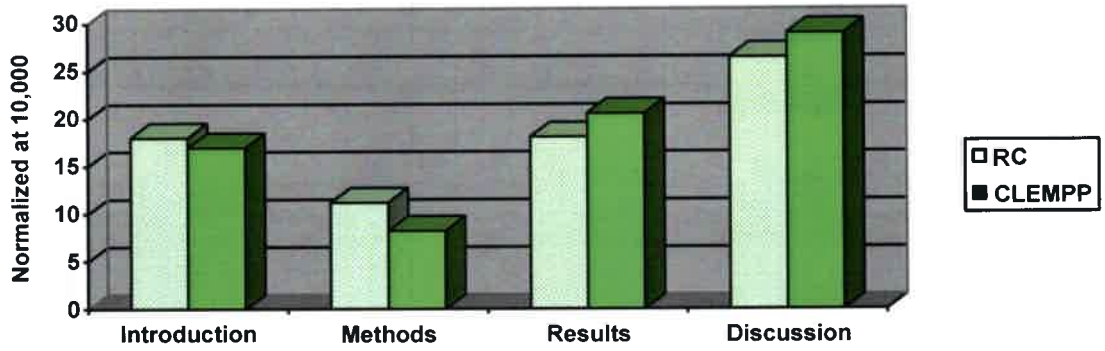

Figure 5.1: Distribution of stance markers in the sections of the RC and the CLEMPP.

To conclude, the results indicate that NNSE writers make use of a smaller range of the resources identified as available to express their stance in the RA, both in the whole texts and in the sections. Still, it may be observed that the use of –ly adverbs across the sections in the CLEMPP and the RC are similar. This indicates that NNSE writers are aware of the need to use stance adverbs in the different sections in spite of the fact that their resources may be more limited than those of the writers of the texts in the RC. Some authors have investigated into the features that make words hard or easy to learn. Laufer (1997) suggests that there may be some grammatical categories more difficult to learn and this may be the case of adverbs, however, the evidence is far from conclusive.

5.1.1. Distribution of stance adverb categories across the RC and the CLEMPP

The stance adverbs identified were classified into the three sub-categories of stance marking proposed by Biber et al. (1999), namely Epistemic, Attitudinal, and Style stance adverbs taking into consideration their use in the corpus. This was done by studying the concordance lines where they occurred. The procedure was performed in both corpora to exclude those adverbs that were not functioning as stance markers. For instance, in these corpora the adverb *naturally* which was used as a manner adverb to describe the way in which something is done or happens.

The results revealed that adverbs used as Epistemic Stance markers were the most frequently used in research articles in English produced by both NSE and NNSE. The

second most frequent category was the category of Attitude stance markers and the least frequently used were the Style markers of stance in both corpora. Table 5.7 shows the distribution of the categories of stance markers in the RC and the CLEMPP. There were differences in the normalized frequency counts of the adverbs marking the three types of stance. NNSE writers used more Epistemic markers than NSE writers. However, NNSE used fewer Attitudinal and Style markers than NSE writers. The differences are statistically significant for the normalized frequencies of Epistemic and Attitude markers but not for Style markers. These results are in agreement with the corpus findings reported in Biber et al. (1999) that indicate that epistemic adverbials are much more common than the other two classes across the registers they studied. Similarly, Salager-Meyer (1994) found that shields, approximators and compound hedges accounted for over 90% of the total number of hedges in the genres she studied (Case Reports and Research Papers). The findings in relation to the occurrence of Epistemic and Attitude Stance adverbs indicate that they are linguistic resources used in compliance with the rhetorical conventions and communicative purposes of the genre of the RA in English.

Table 5.7: Distribution of stance adverb categories across the RC and the CLEMPP - Raw numbers and normalized numbers (at 10,000) **($\alpha \leq .05$)**

	REFERENCE CORPUS	**CLEMPP**
N	1199765	103496
Epistemic	1701	175
Tokens normalized at 10,000	14.17	16.9
z / Sig.	**-2.2235 / 0.013***	
Attitude	338	12
Tokens normalized at 10,000	2.81	1.16
z / Sig.	**3.123 / 0.001***	
Style	148	8
Tokens normalized at 10,000	1.23	0.77
z / Sig.	**1.2996 / 0.097**	

*$p < .05$

Since the use of adverbs in the categories Epistemic and Attitude stance markers were found to be statistically significant, I will present some of the adverbs found in the

corpora[2]. The five Epistemic adverbs most frequently used in the RC were *approximately, probably, clearly, likely,* and *primarily*, which represent 52 % of the total number of -ly stance adverbs used in the RC. The Epistemic adverbs in the CLEMPP were *mainly, approximately, clearly, probably,* and *usually,* which represent 66 % of the total number of -ly stance adverbs used in the CLEMPP.

The five Attitude stance adverbs most frequently used in the RC were *appropriately, correctly, curiously, fortunately*, and *importantly*, which represent 25 % of the total number of ly-stance adverbs used in the RC. The five most frequent Attitude stance adverbs in the CLEMPP were *interestingly*, *remarkably*, *importantly*, *appropriately*, and *curiously*, which represent 84 % of the total number of ly-stance adverbs used in the CLEMPP. The results indicate that NNSE are well aware of the use of Epistemic Stance markers as distinctive resources of research articles in English.

The data in Figure 5.2 show that there was a markedly lower presence of Style and Attitude markers in both corpora. This provides further evidence to the corpus findings reported in Biber et al. (1999), which indicate that epistemic adverbials are much more common than the other two classes across the registers they studied. In relation to academic prose, these authors consider that this may be attributed to the fact that writers are concerned with expressing their stance in terms of the certainty of information, with the indication to the sources of information, limitations and perspectives as well (Biber et al., 1999).

The findings in the present section are described in relation to the frequency of stance in the two corpora together with the –ly adverbs most frequently used in each category. The Epistemic stance markers in the CLEMPP were used in the following order of frequency: showing limitations on a proposition (e.g. *mainly, generally, typically*) and degrees of certainty (e.g. *clearly, probably, usually*), conveying imprecision (e.g. *approximately)*, commenting on the reality of a proposition (e.g. *essentially, actually*) and indicating source of evidence (e.g. *apparently, presumably*). On the other hand, the Epistemic stance markers in the RC had the same functions, but occurred in different frequency order: showing degrees of certainty (e.g. *probably, clearly, likely*), conveying imprecision (e.g. *approximately, roughly*), showing limitations on a proposition (e.g. *generally, typically, largely*), indicating source of

[2] See Appendix C for the complete list of stance adverbs used grouped into categories

evidence (*presumably, apparently, reportedly*), and commenting on the reality of a proposition (*essentially, actually, truly*).

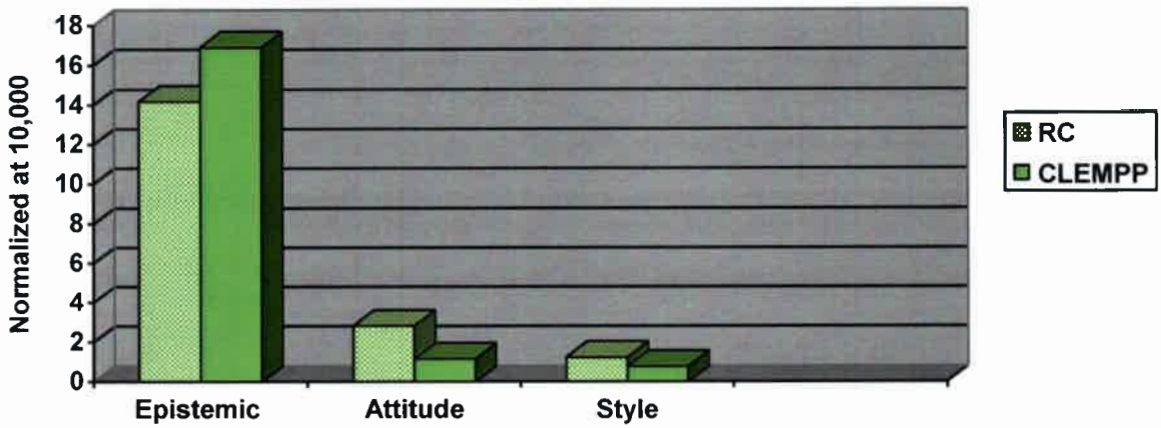

Figure 5.2: Distribution of stance adverb types across the RC and the CLEMPP - Raw numbers and normalized numbers (at 10,000).

The use of Epistemic stance markers by NNSE writers suggests that they are aware of the need to modalize their propositions. NNSE writers use more frequently Epistemic stance adverbs than they use Attitude stance markers. This suggests that NNSE writers intervene with adverbs at the interpersonal level more confidently when they express attitude to truth, but are less prone to express their attitude by using adverbs expressing attitude.

5.1.1.1. Distribution of stance categories across the sections of the RA

As regards the analysis of the sections of the RA, the results revealed that Epistemic stance markers are distributed similarly in the two corpora. Table 5.8 shows that they are most frequently found in the Discussion, followed by the Results, Introduction and Methods. The use differed significantly in the Results section, where NNSE writers overused the items.

The analysis across the sections of the RA in the two corpora revealed that writers in the RC used Attitude stance markers most frequently in the Discussion, followed by the Results, Introduction and Methods. On the other hand, writers in the CLEMPP also used these markers most frequently in the Discussion and less frequently

in the Introduction and Results sections. NNSE writers did not use these markers in the Methods section.

The difference in the distribution of the three types of stance adverbs in the sections of the RA of both corpora was found to be significant in the Results section, where NNSE writers used Epistemic stance adverbs significantly more frequently than NSE, while NNSE used Attitude stance adverbs significantly less frequently than NSE writers. This is shown in Table 5.8. In relation to Style stance markers, the results revealed that they were the least frequently used in the two corpora. The differences were not statistically significant (data shown in Table 5.8).

Table 5.8: Distribution of Stance adverb categories across the sections of RC and CLEMPP. **($\alpha \leq .05$)**

	Introduction		**Methods**		**Results**		**Discussion**		**Total**	
	RC	**CLEMPP**	**RC**	**CLEMPP**	**RC**	**CLEMPP**	**RC**	**CLEMPP**	**RC**	**CLEMPP**
	152048	**16501**	**329301**	**29125**	**424972**	**29194**	**293444**	**28676**	**1199765**	**103496**
Epistemic	228	26	238	17	607	56	628	76	1701	175
Tokens normalized at 10,000	14.99	15.76	7.23	5.84	14.28	19.18	21.40	26.50	14.17	16.90
$z \geq 1.96$	-0.2395		0.8531		-2.1207		-1.7658		-2.2235	
Significance $\alpha \leq .05$	0.405		0.197		0.017*		0.069		0.013*	
Attitude	45	2	7	0	145	3	141	7	338	12
Tokens normalized at 10,000	2.96	1.21	0.21	0	3.41	1.03	4.80	2.44	2.82	1.16
$z \geq 1.96$	1.277		0.7868		2.1834		1.7829		3.1229	
Significance $\alpha \leq .05$	0.101		0.216		0.015*		0.077		0.001*	
Style	1	0	124	7	16	1	7	0	148	8
Tokens normalized at 10,000	0.06	0	3.76	2.40	0.38	0.34	0.24	0	1.23	0.77
$z \geq 1.96$	0.3294		1.1657		0.0917		0.8271		1.2996	
Significance $\alpha \leq .05$	0.371		0.122		0.463		0.204		0.097	
Total	274	28	369	24	768	60	776	83	2187	195
Tokens normalized at 10,000	18.02	16.97	11.21	8.24	18.07	20.55	26.44	28.94	18.22	18.84
$z \geq 1.96$	0.3035		1.4656		-0.961		-0.7833		-0.4428	
Significance $\alpha \leq .05$	0.381		0.071		0.168		0.217		0.329	

*$p < .05$

We found stance adverbs of the three types both in Results and Discussion in both corpora. Their presence in Discussion was expected, as it is an argumentative section. It was surprising to find them in Results, as this section is glossed as being factual and objective. It was also surprising to see that NNSE writers used significantly

more Epistemic stance adverbs in Results than native writers. This suggests that NNSE are conscious of the need to modalize their propositions, but perhaps overdo it.

As expected, NSE writers used significantly more Attitude stance adverbs than NNSE writers. This evidence suggests that NNSE writers are less confident than NSE to express personal feelings.

5.1.2. Analysis of over and underused stance markers

One of the aims of this study was to identify overuse and underuse of –ly stance adverbs in the CLEMPP. The procedure by means of which this was achieved is described in the Methodology. Table 5.9 shows the items over and underused in the CLEMPP when analyzed as complete texts by performing the keyword analysis. In section 4.2, a *Keyword* was defined as a word whose frequency is outstanding, that is, unusually higher or lower than expected, in relation to the tokens in the corpus which is being studied (CLEMPP), and in relation to the frequency it has in the RC. This keyword analysis revealed that *mainly, usually, basically,* and *definitely* were overused in the CLEMPP, since their frequency in this corpus was outstandingly higher than expected in relation to the tokens in the corpus and when compared to the RC. Similarly, the keyword analysis also showed that *interestingly, presumably,* and *likely* were underused in the CLEMPP as their frequency was lower than expected in the CLEMPP when compared to the RC (See Table 5.9).

Table 5.9: Items over and underused in the CLEMPP ($\alpha \leq 0.05$)

N	Key Word	CLEMPP Freq.	%	RC. Freq.	RC. %	Keyness	p value
1	MAINLY	34	0,03	39		77,80	0,0000000000
2	USUALLY	15	0,01	57		11,72	0,0006190842
3	BASICALLY	1		0		5,06	0,0244223345
4	DEFINITELY	1		0		5,06	0,0244223345
5	*INTERESTINGLY*	*3*		*153*	*0,01*	*-10,89*	*0,0009678194*
6	*PRESUMABLY*	*1*		*108*		*-11,58*	*0,0006662849*
7	*LIKELY*	*1*		*110*	*0,01*	*-11,86*	*0,00057353*

The results in Table 5.9 are graphically shown in Figure 5.3.

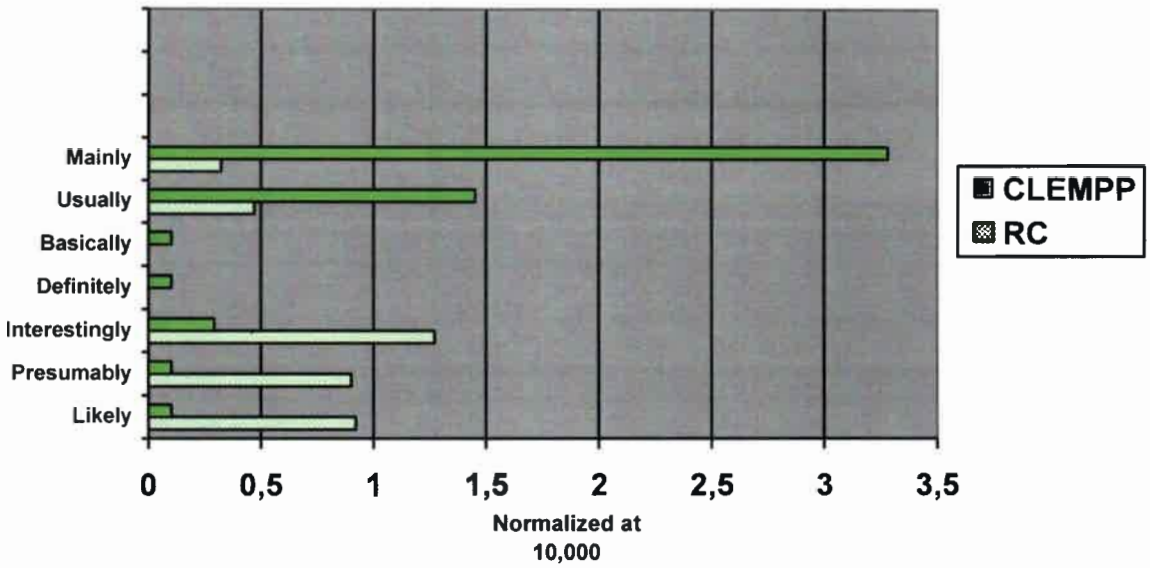

Figure 5.3: Items over and underused in the CLEMPP.

The analysis of the key items, both over and underused, showed that NNSE writers tend to overuse those items which have forms that are transparent to Spanish speakers and that, in general, can be used safely in a variety of contexts. This is the case of the adverbs *usually, basically,* and *definitely*. The item *mainly* is also overused in the CLEMPP, but it is different from the other overused items in that it is not transparent but a frequent item in the RA, and this may be the reason why NNSE writers are familiar with it and tend to use it with excessive frequency in their writing.

The words *likely* and *presumably* are underused in the CLEMPP, although they are frequently used in the RC. This may be attributed to the fact that *likely* is not transparent to Spanish speakers and the cognate word for *presumably* is not common in Spanish. The item *interestingly*, a transparent form for Spanish speakers, was also found to be underused. This may be explained by the fact that it is an Attitude stance adverb, and, as mentioned above, NNSE writers may lack confidence to use this kind of marker.

Some of these items were also found to be over and underused in the sections of the RA. This is the case of the item *mainly* which was overrepresented in each of the sections of the research articles of the CLEMPP, while *usually* was overrepresented only in the Methods and Results sections. The marker *definitely* was overused in the Introduction and *likely* was underused in the Discussion section.

The analysis presented above has concentrated on general features of the over and underused adverbs in the CLEMPP. In what follows, I describe the results for each of these items in particular.

The four derivational stance adverbs found to be overused and the three found to be underused were analyzed in the corpus texts. In the following section I briefly describe the collocational and colligational patterns of each of these items in relation to their syntactic role, position, semantic function, and association patterns. I provide examples with information about the source (corpus), which is given at the end of each example.

The analysis of the data revealed that writers in the CLEMPP and RC use the key stance adverbs studied similarly. This is the case for *mainly*, most of the uses of *usually*, and *interestingly*. The use of each of these will be described in the following sections together with the association patterns in which they occur in the cases that it was possible to obtain these patterns. It is important to point out that some of the concordances extracted from the CLEMPP to exemplify NNSE writers' use of stance markers contain different types of mistakes. These have been inserted in the text of this work as they are in the original manuscripts in the CLEMPP because they are the actual choices made by the NNSE writers.

Mainly

The adverb *mainly* was overused in the CLEMPP (see Table 5.9). As regards syntactic roles, the analysis of the concordances show that *mainly* is similarly used in both corpora, being used in both as a clause element adverbial and a modifier of other elements, but only in the CLEMPP was it used as a modifier of another adverb. The concordance analysis of the CLEMPP showed that *mainly* was used 85% of the times as a clause element adverbial (Example 5.1), 12% as a modifier of other elements (Example 5.2), and 3% as a modifier of another adverb (Example 5.3).

5.1) *It is employed **mainly** as bronchus-dilative and heart stimulant (11). (clempp ft 07\nns008.t07.txt) (clempp text 08)*

*5.2) The emitted radiation (**mainly** 1270nm) was detected at right angles using an amplified Judson J16/8Sp g ....(clempp text 002)*

*5.3) This indicates that communities distribute **mainly** laterally and this would be caused by the asymmetry of transport and deposit processes, which generate a heterogeneous habitat in the shore. (clempp 028)*

The concordance analysis of the RC showed that *mainly* occurred 90% of the times as a clause element adverbial (Example 5.4) and only 10% as a modifier of other elements (Example 5.5). There were no cases of *mainly* as modifier of another adverb, suggesting that this use may not be common in the language of science, although this speculation needs further research.

*5.4) The specificity of this approach depends **mainly** on the number of measured peptide masses and the peptide mass accuracy. (RC 06)*

*5.5) These were mixtures of **mainly** IgG1 and G2a, although the former also contained G2b. (RC 06)*

As regards position, *mainly* was mostly found to be placed in medial position in the clause in both corpora. The following are examples from the CLEMPP and RC, respectively:

*5.6) Soluble activity was **mainly** localized in the embryo plus endosperm tissues with the higher values dis ... (clempp 003)*

*5.7) Chemokines are **mainly** divided into two subfamilies. (*RC 06)

Semantically, Biber et al. (1999) classify this adverb, when functioning as a clause element adverbial, as an Epistemic marker that has the role of showing limitation on a proposition. When the item was found to be a modifier, it modified a numerical expression as in example 5.2. .

No association patterns were found for this item.

Usually

The adverb *usually* was overused in the CLEMPP with a normalized frequency of 1.45. In the RC the normalized frequency of the item was 0.47. Regarding syntactic roles, the data indicate that *usually* is similarly used in both corpora. The findings showed that in all cases it was an adverbial (Examples 5.8 and 5.9).

5.8) *As said in the introductory section the species O2(- and O2(1(g) are* ***usually*** *present in aerobic Rf-phosensitised processes (26). (CLEMPP 031)*

5.9) *Diauxic shift* ***usually*** *occurred at approximately OD600 5.0. (RC 06)*

As regards position, the data showed that it was similar in both corpora. The item *usually* occurred 5 % of the cases in initial position and 95% in medial position in the CLEMPP. In the case of the RC, the data showed that it occurred 73% of the times in medial position and 27% in initial position. These positions are illustrated in examples 5.10 and 5.11 from the CLEMPP and RC, respectively.

5.10) ***Usually****, 16 shots were need for averaging so as to achieve a good signal to noise ratio, from which the decay curve was obtained.(CLEMPP 008)*

5.11) *Despite the availability of PCR amplification of nucleotide sequences of the E. chaffeensis genomic DNA, the diagnosis of HME* ***usually*** *depends on serology. (RC 06)*

According to Biber et al. (1999) this adverb, together with *often*, is considered to contribute "to academic writers' ability to make clear the generalizability of their statements" in this type of register (p. 801).

No association patterns were found for this item.

Basically and definitely

There was only one case of the adverbs *basically* and *definitely* in the CLEMPP and they were both functioning as clause element adverbials (see examples 5.12 and

5.13, respectively). These were identified as instances of overuse because they do not occur in the RC (See Methodology, section 4.3).

5.12) ***Basically*** *the method empoys an specific complexing agent (fluorescamine) which produces a fluorescent complex in the presence of primary amino (using a excess of fluorescamineTHe spectrofluorometeremployed was a Spex Fluoromax, and the exitation and emission wavelengths fro the complex <( )> were 390 and 475 mn respectively). (CLEMPP 006)*

5.13) Hence, the evaluation of such light stabilizers under different types of exposure to luminosity (especially natural day-light) is ***a definitely*** *mandatory test to provide valuable information about the efficiency of a given dopant operating as a light filter. (CLEMPP 002)*

Basically is used as an Epistemic stance marker and *definitely* as an Attitude marker.

The fact that these words were not used in the RC suggests that they may not be appropriate terms for the genre. This, however, needs further research.

Interestingly

The item ***interestingly*** was underused in the CLEMPP, with a normalized frequency of 0.29, functioning in all the cases as a clause element adverbial. The normalized frequency of this item in the RC was 1.27.

5.14) Most ***interestingly****, these authors found that during flower development the different tissues exhibited different ratios of JAs. (CLEMPP)*

5.15) ***Interestingly****, the number of BB1+ basophils was much less than that of EG2+ eosinophils in bronchial mucosa. (RC)*

The analysis lines showed that the adverb is an Attitude stance marker, placed in sentence initial position. This position was the same in both corpora. This is illustrated in examples 5.14 and 5.15 from the CLEMPP and RC, respectively.

Association patterns were identified for *interestingly*. These are shown in Table 5.10.

Table 5.10: Most frequent patterns associated with the adverb *interestingly* in the RC

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	OF	IN	THE	FIG	SHOWN	**INTERESTINGLY**	THE	THE	OF	OF	THAT
2	AND	OF	OF	FIGURE	RESPECTIVELY		**WE**	OF	THAT	THE	TO
3	IN	AND	ET	NOT			IN			IN	IN
4		TO		AND						IS	OF
5		THE		AL							
6				OF							

The pronoun *we* is usually found to occupy the first position to the right and it has the second highest MI score (6,852) of the patterns. The personal pronoun *we* collocates with *interestingly*, 50% of the cases being in the Results and the other 50% in the Discussion. This may be related to the rhetorical goals of the sections of the research article. Examples 5.16 and 5.17 illustrate the pattern.

5.16) ***Interestingly***, *we observed down-regulation of the hSP-C-transgene in these mice in response to allergen when compared with saline-treated controls (Fig. 7). (RC)*

5.17) *Most* ***interestingly***, *we found that under the same conditions of strong activation where it inhibited the above mentioned cytokines, PD98059 instead enhanced the production of IL-4, IL-5, and IL-13. (RC)*

The analysis revealed that the pattern "Interestingly we" is followed by verbs expressing results such as *found*, *recovered*, *showed*, *observed*, and *note*. This was similar for both, the Results and Discussion, sections.

A possible explanation may be the fact that in Spanish the cognate adjective *interesting* is not used as a derivational adverb.

Presumably

The adverb **presumably** was also underused in the CLEMPP, with only one case functioning as clause element adverbial (Example 5.18).

5.18) *Bands 1 and 3 of anionic patterns from the crude extract showed, prior to stain peroxidases, a red brown colour which belong* ***presumably*** *to leghemoglobins. (CLEMPP)*

In the RC the normalized frequency of the item was 0.90. The cases showed that it was used as clause element adverbial both in initial position (example 5.19) and in medial position (example 5.20).

5.19) ***Presumably****, 751R is not an essential part of the epitope. CBL-4, ... (RC)*

5.20) In an attempt to mimic the events ***presumably*** *responsible for docking IN at the nucleopore, recombinant forms of yeast karyopherin- and karyopherin-(18) were successively added to GST-IN (Fig. 4C). (RC)*

The analysis showed that the adverb is an Epistemic marker of stance, expressing some degree of doubt over the proposition it frames.

Association patterns were identified for this item. Table 5.11 shows that the adverb *presumably* usually precedes phrases beginning with the item *by* (Example 5.21), which has a collocational strength with *presumably* of 4.083 as calculated with the MI score.

5.21) Both kinds of organic amendment improved plant establishment, ***presumably*** *by retaining more water. (RC 06)*

Presumably also precedes *due (to)* (MI score: 7.697, example 5.22) and *because* (MI score: 6,985, example 5.23). This adverb is preceded by *would* (MI score:7.297, example 5.24).

5.22) In addition, comparison of cells from acutely infected patients versus asymptomatic patients revealed an increase in apoptotic cell death in cells from the acutely infected patients, ***presumably*** *due to higher viral titers (36). (RC 06)*

5.23) The advantage of the GAL-E1-IN system lies in the ease with which the immunofluorescence analysis can be performed, ***presumably*** *because more IN protein is produced. (RC 06)*

5.24) *Where mammalian FLICE/MACH1 activity is regulated by the action of external ligands (FasL or TNF), Drosophila drICE regulation would* ***presumably*** *be linked to transcriptional regulation of rpr. (RC 06)*

Table 5.11: Most frequent patterns associated with the adverb *presumably* in the RC

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	OF	THE	IN	THE	AND	**PRESUMABLY**	BY	THE	TO	OF	OF
2		OF	THE	AND	IS		DUE	TO	A		THE
3					WOULD		BECAUSE	A	THE		
4								IN	OF		

The analysis of the distribution of the most frequent collocates showed that these patterns tend to occur in the Results, Introduction, and Discussion sections of the research article. The collocations of *presumably* with *due to* and *because of* are more frequently found in the Results sections, where they are used to express a reason. The collocation of *presumably* preceded by *would* is most frequently used in the Discussion sections. The item *would* is considered to be a hedge and found to be one of the most widely used devices to express stance (Hyland, 1999).

Likely

The item *likely* was also underused in the CLEMPP. This item was underused in the CLEMPP where its normalized frequency was 0.10. The analysis showed that *likely* occurred 0.92 times in the RC, where it was used as a clause element adverbial (Examples 5.25 and 5.26). Before describing the results obtained it is worth pointing out that the word *likely* may function both as as an adverb and as an adjective. Since the focus of this thesis is on derivational adverbs, the cases where *likely* was used as an adjective were not included in the analysis.

5.25) *Most **likely**, the very weak SPB pattern shown in Fig. 5 reflects short microtubule arrays at the SPB, which are not able to produce nuclear movement. (RC)*

5.26) *Some of these proteins **likely** interact with viruses to modulate their entry into cells. (RC)*

As regards the position of this adverb in the RC, the data showed that it is used 85% of the times in medial position (5.26), and 15% in initial position in the clause (5.25).

Association patterns were identified for this item. Table 5.12 shows the patterns in which the adverb *likely* most frequently occurs. The item is frequently preceded by *most, more, less, not,* and *also*; and it also shows that it is followed by *due to, the,* and *a.*

Table 5.12: Most frequent patterns associated with the adverb *likely* in the RC

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	OF	OF	IN	IS	MOST	**LIKELY**	DUE	TO	THE	OF	OF
2	IN		AND	WAS	IS		THE	THE	OF		THE
3				OR	MORE		A	A	A		IN

The items that are shown to have high collocational strength with *likely* are *most* (MI score: 7,297) and *due to* (MI score: 6,852).

The analysis revealed that the pattern *likely due to* tends to occur in the Results sections to express reason.

5.2. Analysis across the sections of the research article

Table 5.13 shows the normalized frequency for each of the items found to be overused and underused in the CLEMPP. As pointed out previously, overuse of items such as *usually*, *basically*, and *definitely* may be accounted for by the fact that their meaning is straightforward and clear. In the case of *mainly*, overuse may be related to the fact that it is frequently used in the RA and so NNSE writers are exposed to the item very often as they read. On the other hand, the underuse of items such as *presumably* and *likely* may be explained by the fact that they are not transparent to Spanish speakers and they may not know how to use them. In the case of *interestingly*, which is transparent, the underuse may be attributed to NNSE's lack of confidence in using items that express attitudes.

Table 5.13: Distribution of stance adverbs across the sections of the RC (Normalized at 10,000)

Key Stance Adverbs	Introduction		Methods		Results		Discussion	
	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP	RC	CLEMPP
Mainly	5 0.32	5 3.03	2 0.06	3 1.03	17 0.40	12 4.11	15 0.51	14 4.88
Usually	19 1.29	2 1.21	10 0.30	5 1.71	10 0.23	5 1.71	18 0.61	3 1.04
Basically	0	0	0	1 0.34	0	0	0	0
Definitely	0	1 0.60	0	0	0	0	0	0
Interestingly	16 1.05	0	0	0	69 1.62	0	68 2.31	3 1.04
Presumably	21 1.38	0	1 0,03	0	42 0.98	1 0.34	44 1.49	0
Likely	14 0.92	0	0	0	46 1.08	0	50 1.70	1 0.34

What follows describes over and underused items per section of the RA in the CLEMPP.

5.2.1. Overused items in the CLEMPP Introduction section

In the CLEMPP Introductions, four items were overused, *mainly, definitely, fortunately* and *undoubtedly* (see Table 5.14). The adverb *definitely* was found only once in the CLEMPP and did not occur at all in the RC (See Table 5.14). The adverbs *fortunately* and *undoubtedly* were found to be key in this section but not in the research article as a whole document. These items occurred in the RC but only in the Discussion section of the research article as shown in Table 5.22.

Table 5.14: Overuse and underuse in the Introduction Section of the CLEMPP

N	Key Word	CLEMPP (16,501) frequency	%	RC (152,048) Freq.	RC. %	Keyness	P
1	MAINLY	5	0,03	5		10,41	0,0012557217
2	DEFINITELY	1		0		4,65	0,0310958065
3	FORTUNATELY	1		0		4,65	0,0310958065
4	UNDOUBTEDLY	1		0		4,65	0,0310958065

The item *fortunately* is an Attitude stance marker used in sentence initial position in the CLEMPP. The item *undoubtedly* is an Epistemic stance marker used in medial position in the CLEMPP. It is worth pointing out that even though these two items are identified as overused by the computer software, the fact that there is only one occurrence of each does not allow for the description of their behavior.

Table 5.15 shows the list of adverbs that were used in both sub-corpora- a total of 12 items, and the lists of adverbs used by only one group or the other. The list of adverbs used in the introductions of the RC contains 37 items, more than the one used by the NNSE, which contains only 15 items. This shows that writers in the RC use a wider variety of the resources available in the language system, as far as stance adverbs studied are concerned.

Table 5.15: Lists of stance adverbs used in the Introduction section of the RC and CLEMPP

Common to RC and CLEMPP	RC only	CLEMPP only
1. actually	1. accordingly	1. definitely
2. apparently	2. arguably	2. fortunately
3. appropriately	3. broadly	3. undoubtedly
4. approximately	4. certainly	
5. clearly	5. conceivably	
6. essentially	6. correctly	
7. generally	7. evidently	
8. largely	8. fundamentally	
9. mainly	9. ideally	
10. primarily	10. importantly	
11. simply	11. incorrectly	
12. usually	12. inevitably	
	13. interestingly	
	14. likely	
	15. possibly	
	16. presumably	
	17. probably	
	18. remarkably	
	19. strictly	
	20. surprisingly	
	21. theoretically	
	22. truly	
	23. typically	
	24. unexpectedly	
	25. unfortunately	

5.2.2. Overused items in the CLEMPP Methods section

In the Methods section, the following adverbs were found to be overused: *usually, mainly,* and *basically*, whereas the adverb *approximately* was found to be underused (data shown in Table 5.16).

Table 5.16: Overuse and underuse in the Methods Section of the CLEMPP.

N	Key Word	CLEMPP (29,125) Freq.	%	RC. (329,301) Freq.	RC. %	Keyness	P
1	BASICALLY	1		0		5,02	0,0250520818
2	MAINLY	3	0,01	2		8,67	0,0032351103
3	USUALLY	5	0,02	10		7,70	0,0055179140
4	APPROXIMATELY	4	0,01	126	0,04	-5,71	0,0168296639

The analysis of the concordance lines from the methods section of the RC showed that *approximately* is mainly functioning as a pre-modifier of adjectives (Example 5.27), numbers and numerical expressions (Examples 5.28 and 5.29). As such, it is considered to be a stance marker used to convey imprecision, hence this item can be called "hedge" (Biber et al., 1999) and it can be considered to belong to the Epistemic stance type marking imprecision.

5.27) *It was determined that 1 ng of p24 is* ***approximately*** *equivalent to 125 infectious units. (RC)*

5.28) *An ear biopsy culture was performed on each mouse* ***approximately*** *30 days after tick feeding to test for the presence of borreliae in the tissue as previously described (21).(RC)*

5.29) *After 4 h in the presence of galactose, -tubulin levels increased by approximately fivefold and -tubulin increased by* ***approximately*** *twofold in both strains. (RC)*

The association patterns identified for *approximately* (Table 5.17) show that it is preceded by the verb *containing* (4,604), and the prepositions *of* (2.598), *at* (1.960), *with* (1,702), and *for* (1,515) and it is followed by numbers and numerical expressions.

Table 5.17: Most frequent patterns associated with the adverb *approximately* in the RC

n	l5	l4	l3	l2	l1	centre	r1	r2	r3	r4	r5
1	to	a	of		of	*approximately*	2	5	of	of	were
2	of	of	cell		containing		1	to	cells	the	the
3	c		were		with		20	µg		were	of
4	the		was		at		5	h			
5					for		10	µl			
6							50	of			
7							100				

Table 5.18 shows the list of adverbs that were used in the Methods sections in both sub-corpora- a total of 6 items, and the lists of adverbs used by only one group or the other. The list of adverbs used in the Methods Section of the RC shows a wider variety, containing 21 items, than the one used in the CLEMPP, which contains only 7 items.

Table 5.18: Lists of stance adverbs used in the Methods sections of the RC and CLEMPP

Common to RC and CLEMPP	RC only	CLEMPP only
1. approximately	1. accordingly	1. basically
2. briefly	2. actually	
3. essentially	3. apparently	
4. mainly	4. appropriately	
5. typically	5. correctly	
6. usually	6. generally	
	7. importantly	
	8. largely	
	9. likely	
	10. possibly	
	11. presumably	
	12. primarily	
	13. probably	
	14. roughly	
	15. simply	

5.2.3. Overused items in the CLEMPP Results section

In the Results section of the CLEMPP the items overused were *mainly, usually, and obviously* (See Table 5.19).

Table 5.19: Overuse and underuse in the Results Section of the CLEMPP

N	Key word	CLEMPP (29,194) frequency	%	RC (424,972) Frequency.	RC. %	Keyness	P
1	MAINLY	12	0,04	17		28,79	0,0000000776
2	USUALLY	5	0,02	10		9,68	0,0018639012
3	OBVIOUSLY	1		0		5,49	0,0191361401

Interestingly, the item *obviously* was overused in this section and not in the corpora as a whole. All the occurrences of this adverb in the RC occurred in the Discussion section (See Table 5.22). It is an adverbial used in sentence and clause initial position. The item *obviously* is an Epistemic stance marker that indicates degree of certainty. The following examples are from the Results sections of the CLEMPP (Example 5.30) and the Discussions of the RC (Examples 5.31 and 5.32).

5.30) ***Obviously*** *these results indicate the phototransformation of the -NH2 group in the cases of tyr and the dipeptide (initial fraction of the trace) and a generation of the primary amino group in the cases of a positive slope as in the second section of tyr-tyr and the tripeptides. (CLEMPP Results section)*

5.31) ***Obviously***, *however, proteins identified by this technique are only `in vitro' substrates until proved to be a physiological target by in vivo experiments. (RC - Discussion Section)*

5.32) *Access of antibodies is* ***obviously*** *relevant to questions about immunity and its evasion, but what of the role of proteases? (RC - Discussion Section)*

Table 5.20 shows the list of adverbs that were used in both sub-corpora- a total of 16 items, and the lists of adverbs used by only one group or the other. The list of adverbs used in the Methods Section of the RC shows wider variety (44 items altogether) than the one used by the NNSE, which contains only 17 items.

Table 5.20: Lists of stance adverbs used in the Results Sections of the RC and CLEMPP

Common to RC and CLEMPP	RC only	CLEMPP only
1. apparently	1. accordingly	1. obviously
2. approximately	2. actually	
3. briefly	3. appropriately	
4. clearly	4. broadly	
5. curiously	5. certainly	
6. generally	6. conceivably	
7. importantly	7. correctly	
8. largely	8. essentially	
9. mainly	9. evidently	
10. possibly	10. formally	
11. presumably	11. hypothetically	
12. probably	12. incorrectly	
13. strictly	13. interestingly	
14. typically	14. likely	
15. unexpectedly	15. mildly	
16. usually	16. nominally	
	17. ostensibly	
	18. primarily	
	19. really	
	20. reasonably	
	21. remarkably	
	22. reportedly	
	23. roughly	
	24. seemingly	
	25. simply	
	26. surprisingly	
	27. truly	
	28. unfortunately	

5.2.4. Overused items in the CLEMPP Discussion section

In the Discussion section, the items *mainly* and *really* were overused. The adverb *likely* was underused (data shown in Table 5.21).

Table 5.21: Overuse and underuse in the Discussion Section of the CLEMPP

N	Key word	CLEMPP (28,676) Frequency	%	RC (293,444) Frequency	RC. %	Keyness	P
1	MAINLY	14	0,05	15		30,36	0,0000000329
2	REALLY	1		0		4,84	0,0278430525
3	*LIKELY*	*1*		*50*	*0,02*	*-4,32*	*0,03766693*

It is worth pointing out that the item *really* was not found to be a key item when the corpora were compared as full texts. This is probably because it is an item that is not frequently used in the RC. Biber et al. (1999) classify this item as an Epistemic stance adverb that marks actuality and their findings indicate that it is not frequently used in

academic prose. In fact, their corpus findings show that it is most commonly used in conversation to mark actuality (p. 867). In the present study, it occurred only once in the RC.

5.33) *It was reported that peroxidase may be produced in cell suspension cultures to a level greatly in excess of that found in the intact plant (*Parkinson et al., *1990) and although the reasons for the enhanced production of peroxidase in culture are not* ***really*** *known, there are a number of contributing factors to explain this fact. (CLEMPP 014)*

5.34) *To determine whether this* ***really*** *was a direct effect on the regulatory regions of the XFKH1 gene, or whether XSmad2C worked indirectly through its ability to induce other mesoderm-specific genes, we exploited the fact that no transcription occurs in Xenopus embryos before the mid-blastula transition (Newport and Kirschner, 1982), although injected mRNAs are translated from the time of injection. (RC 06)*

Table 5.22 shows the list of adverbs that were used in both sub-corpora, a total of 20 items, and the lists of adverbs used by only one group or the other. The list of adverbs used in the Methods Section of the RC shows a wider variety, containing 52 items, than the one used by the NNSE, which contains only 21 items.

Table 5.22: Lists of stance adverbs used in the Discussion Sections of the RC and CLEMPP

Common to RC and CLEMPP	RC only	CLEMPP only
1. actually	1. accordingly	1. really
2. apparently	2. admittedly	
3. approximately	3. appropriately	
4. certainly	4. arguably	
5. clearly	5. briefly	
6. essentially	6. broadly	
7. evidently	7. conceivably	
8. generally	8. correctly	
9. importantly	9. curiously	
10. interestingly	10. decidedly	
11. largely	11. formally	
12. likely	12. fortunately	
13. mainly	13. ideally	
14. obviously	14. inevitably	
15. possibly	15. metaphorically	
16. primarily	16. mildly	
17. probably	17. ostensibly	
18. remarkably	18. presumably	
19. typically	19. reasonably	
20. usually	20. reportedly	
	21. roughly	
	22. seemingly	
	23. simply	
	24. strictly	
	25. superficially	
	26. surely	
	27. surprisingly	
	28. theoretically	
	29. truly	
	30. undoubtedly	
	31. unexpectedly	
	32. unfortunately	

The data obtained from the analysis across the sections in the two corpora clearly demonstrate that the writers in the RC use a wider variety of stance adverbs in all the sections. There are certain items which are found to be overused in some sections of the research article in the CLEMPP. This is the case of *definitely*, *fortunately*, and *undoubtedly* in CLEMPP Introductions. This is also the case for *basically* in Methods, *obviously* in Results, and *really* in Discussions of the CLEMPP.

The lists of adverbs used by NSE and NNSE in the different sections of the research article in English indicate that NSE writers use a greater variety of the items identified in each class than the NNSE writers.

In the following chapter, Chapter VI, I will discuss the results presented in this chapter in relation to other studies in the field.

CHAPTER VI

DISCUSSION

The aim of the present thesis was to describe the use of derivational stance adverbs as attested in the two corpora used for analysis, the Reference Corpus and the Corpus of Learner English Manuscripts for Publication Purposes. Although differences in the use of stance adverbs were not significant, neither in the complete texts nor in the specific sections (Tables 5.1 and 5.6), the analysis of types and categories provided interesting information. The analysis suggests various points for discussion, of which I will concentrate on the following:

- Variety of stance markers
- Over and underrepresentation of stance adverbs
- Differences in the use of Stance categories across the RA sections

The expression of author's stance can be realized through a variety of linguistic features provided by lexis and grammar. The scope of the present study is limited to the expression of three types of stance, Epistemic, Attitudinal, and Style, as expressed by the specific derivational –ly adverbs focused on in this research.

6.1 Variety of stance markers

The analysis of the data showed that, although interpersonal authorial intervention is considered to be problematic for NNSE (Hyland, 1994; Martínez, 2005; Salager-Meyer, 1994), NNSE writers did use derivational stance adverbs in the text of the RA in English. The normalized frequencies in the CLEMPP and in the RC did not differ significantly. However, the lower number of adverb types in the CLEMPP reveals that NNSE writers used fewer resources than those available to mark stance, both in the texts as complete documents and in the different sections.

There are several likely explanations for the differences in the use of types showing lower variability in NNSE. To begin with, it is possible that NNSE writers

have this vocabulary at the receptive level but have not yet incorporated it for production (Melka, 1997), and this may result in a tendency to overuse the lexical items they are more familiar with. This tendency appears to characterize learner language production, as reported by Granger and Hinkel, who observed that L2 academic essays contain a smaller range of vocabulary than those produced by L1 writers (as cited in Hyland, 2006). The results from the present study indicate that this may be the case for the NNSE writers of the CLEMPP, who used a smaller range of the stance markers identified as available.

Secondly, the NNSE writers in this study tended to use more often derivational adverbs which have forms transparent to Spanish speakers, which are the ones Granger refers to as *good cognates,* provided their similarity in form is also a similarity in meaning (1993). She suggests that cognates can affect the development of L2 vocabulary positively and negatively, and they deserve careful consideration. It is not surprising that, in the present study, the 10 most frequently used adverbs were practically the same in both corpora, although in different order of frequency. Most of these adverbs are Epistemic Stance adverbs and their use in these corpora may be attributed to the fact that a central concern of the discourse of the experimental sciences is the accumulation of knowledge. In the CLEMPP, the 10 top items were *mainly, approximately, clearly, probably, usually, generally, apparently, briefly*, typically, likely**. In the RC the top items were *approximately, probably, interestingly, clearly, briefly*, primarily, likely*, presumably, possibly, essentially*. As can be seen, the most frequent adverbs from both corpora were transparent for Spanish speakers, except for those marked with an *, which are also common to both corpora. A possible explanation for this similarity may be the NNSE writers´ familiarity with these items due to the fact that they are proficient readers of English in their own fields of study. Another point to note is that almost half of the items that were used only by NSE were non-cognate adverbs and their absence in the CLEMPP may be attributed to their intrinsic difficulty for Spanish speakers.

In some cases, the production of NNSE writers seemed to be influenced by L1 transfer (Sharwood Smith, 1994). The data showed that the items *basically* and *definitely* were overused by NNSE writers in sentence initial position probably due to the influence of Spanish. L1 transfer might be considered to negatively affect learners' production, rendering it non-idiomatic. Hoye (in Aijmer, 2002, p. 58) suggests that negative transfer may be promoted by similarities and differences between L1 and L2

which may lead to interference and impede learners' performance in L2. Thus, transfer could be related to language distance. Swan (1997) observes that Spanish and English share a great number of lexical features (the Graeco-Latin vocabulary) and differ in syntax. The similarity in lexical features may be considered to support L2 vocabulary learning. However, it might also promote interference errors which arise when two items are similar but not identical in form or use.

Since there were no instances of the items *basically* and *definitely* in the RC to observe whether NNSE writers' use was correct or otherwise, a search of these items was carried out using the Collins WordbanksOnlineEnglish (http://www.collins.co.uk/corpus/CorpusSearch.aspx). The results of this search revealed that these items are actually used in general English, but they are more frequently used in medial position. It may be argued that the use of these items may be considered a misuse, since they are frequently used in general English but not in the writing of RAs, as attested in the RC. Specificity and register restriction have been identified as sources of errors because some frequent words in one register or field may not be normally used in another (Laufer, 1997).

To conclude, the findings in relation to the use of stance adverbs in the present study suggest that NNSE writers may be aware of the variety of choices in the language system, but prefer to stick to a few items that make them feel confident. According to Hyland (1999) confidence plays an important role in the creation of an authorial *persona*, which emerges from the way writers project themselves into their texts (p.101). In addition, the writers' projection into the text may also be affected by the cultural differences between the NSE and NNSE who produced the RAs in the corpora. Since stance adverbs can contribute to the creation of the writers' voice it seems reasonable to highlight the value of the lists of stance adverbs identified in this study for their pedagogical potential (Tables 5.3, 5.15, 5.18, 5.20, and 5.22). The items in the lists may prove useful to help increase the vocabulary NNSE writers have available for.

Finally, it is worth pointing out that writers may use different interpersonal linguistic devices, of which stance adverbs are just a small part. These adverbs, on their own or in combination with other devices, are used to negotiate writers' claims not only in terms of different degrees of certainty, possibility, and precision but also in terms of authors' evaluation and position, with a purpose in mind: to persuade the reader of the validity of the paper's contribution.

6.2. Over and underrepresentation of stance adverbs

The overuse and underuse of lexical items have been considered to affect the quality of a text and may be regarded as factors contributing to non-nativeness (Granger, Meunier, & Tyson, 1994). The analysis of the data in this study revealed over and underuse of stance adverbs in the RAs produced by NNSE writers. These items were analyzed in relation to their syntactic role, position, stance meaning, and association patterns when it was possible and frequencies allowed for it. In general, and as far as syntactic role is concerned, the items over and underused were used with similar syntactic functions in both corpora. As regards position, the items over and underused did not show significant differences. A similar observation was made in relation to the semantic uses of these items. In particular, the overuse of certain lexical items may have different interpretations. The overuse of *basically* and *definitely*, which may be considered as possible cases of L1 transfer, might also be related to misuse. The overuse of *mainly* and *usually* may be due to the fact that NNSE writers used these adverbs instead of other possible items conveying similar meanings, probably to avoid risks. Considering this, there may be a relationship between NNSE writers' overuse of stance adverbs and their productive knowledge of a small range of such adverbs to express stance. In this sense, the findings in the present study are similar to those in Milton's study, who reports that Hong Kong L2 students overused some prefabricated phrases when compared to L1 production, which rendered the L2 style of writing repetitive (as cited in Hyland, 2006).

This study revealed that certain stance adverbs (*interestingly*, *presumably* and *likely*) are underused in the CLEMPP. This was unexpected due to the fact that these items are some of the most frequently used in the RC, but it may be explained by their already mentioned differences with the L1 (Table 5.9).

The data allowed for the identification of certain associated patterns of the over and underused items. The underlying assumption is that the idiom principle (Sinclair, 1991) is essential for the understanding and production of texts. The language in the scientific RA is considered to exhibit "a large number of semi-preconstructed phrases that constitute single choices" (Sinclair, 1991, p. 110). This study identified patterns that may be regarded as characteristic features of the adverbs studied. The patterns are "*interestingly*/we", "*presumably* / by / due to / because", "would *presumably*", "most *likely*", "*likely* due to", "containing / of / at / with / for *approximately*" and

"*approximately* with numbers and numerical expressions". The identification of collocational and colligational patterns is important for the description of scientific discourse since native-like linguistic selections imply choices not only at the level of syntax but also at the level of frequent and familiar collocations. The underlying assumption is that NNSE writers may benefit from the identification and description of the frequent patterns used by NSE writers of RAs.

The company a word keeps and its possible combinations are an important area of research in second and foreign language study. Hoye studied the combination of adverbs and modality in English and reported that Spanish speakers consistently neglected the potential for modal and adverbs to combine and claimed that this interfered with L2 learners' performance (in Aijmer, 2002). Also, different researchers have highlighted the value of the identification of patterns. Luzon Marco (2000) states that the knowledge of the phraseology of the genre can help students understand and create meanings. Moreover, according to Hyland (2006), the identification of patterns contributes to the making of informed choices, guided by expert practice and disciplinary expectations.

The analysis of over and underrepresentation of stance items in this study has identified a specific set of linguistic items that are worth studying in more detail for they are valuable resources for the expression of different types of stance. The detailed description of these items may prove useful for the development of L2 academic writers.

6.3. Over and underrepresentation of stance categories

The results showed that there is a great difference between the use of Epistemic Stance adverbs and the other two types, Attitude and Style, in both corpora. The differences were significant (Table 5.7). These results are similar to the corpus findings presented in Biber et al. (1999) for NSE users' productions. There are several likely explanations for this use by NNSE writers. Firstly, the high frequency of Epistemic stance adverbs, even greater than the normed frequencies of NSE, may reveal NNSE writers' awareness of the fact that RA article writing is concerned with the degree of certainty of the information presented. On the other hand, the low normed frequency of Attitude stance markers, even lower than that of NSE, may be related to the lack of awareness of the role of the expression of attitudes, opinions and value judgments in the

writing of research articles in English. In addition, it is likely that writers are aware of this, but they do not feel confident enough to use Attitude markers. This has been pointed out as a problematic area by different researchers (Hyland, 1994, 1999, Salager-Meyer, 1994). In their study, Hyland and Milton (1997) conclude that the Hong Kong students that they studied had difficulties manipulating certainty and affect in the writing of academic texts. This may be interpreted as the result of NNSE writers' lack of confidence.

As regards the distribution of the Epistemic, Attitude and Style Stance categories across the sections of the RA, the findings indicate that there were significant differences in the Results section. NNSE writers used significantly more Epistemic markers, while they used significantly fewer Attitude markers in this section. In relation to Style stance markers, the results revealed that they were the least frequently used in the two corpora and, although, there were attested differences, these were not statistically significant.

The only section that showed significant results in the use of Epistemic and Attitude stance adverbs was the Results section, which is thought to be an objective, factual part of the RA. Here there was overuse of Epistemic adverbs and underuse of Attitude adverbs. Again, this overuse may be related to the writers' confidence in relation to the use of adverbs expressing certainty, while affective language does not seem to be part of their productive repertoire. It must be remembered that Thompson (1993), in a study of biochemistry articles, concludes that in the Results section scientists employ a variety of rhetorical moves to argue for the validity of scientific facts and knowledge claims. Similarly, Kanoksilapatham (2005) showed that results in the field of biochemistry are not only presented, but also accompanied by comments. These studies provide evidence that stance is expected in this section.

To conclude, the findings of the present study as regards variety of stance markers, their over and underrepresentation in the corpora and their sections, and the differences in the use of Stance categories across the IMRD sections of the experimental RA, seem to indicate that the realization of stance in the text of the RA in English reflects the rhetorical conventions of the genre and its communicative purposes. The findings also indicate that the use of these devices by NNSE writers seems to be difficult not only because they seem to manipulate only a narrow range of items but also because they lack the confidence necessary to use them strategically. Taking all this into consideration, it seems reasonable to suggest that NNSE writing would benefit from the

findings in studies like the one reported in this thesis which has only scratched the surface of the problem.

In the present chapter I have discussed the findings in relation to the variety of stance markers, their over and underrepresentation in the corpora and their sections, and the differences in the use of Stance categories across the IMRD sections of the experimental RA. In the following chapter, Chapter VII, I will present the pedagogical implications of this study.

CHAPTER VII

Pedagogical Implications

This final chapter presents the conclusions derived from the analysis and interpretation of the results of the study. It also puts forward some considerations for the implications of the present study together with some of its advantages and limitations. Finally, I include a personal comment on the present thesis.

7.1. Pedagogical implications

The role of English as the language for scientific communication is widely acknowledged. According to Hyland (2006), over 90% of the journal literature in some domains is printed in English, which has resulted in the growth of English as the language of research at the expense of other languages and researchers who are non-native speakers of English. In the current "Anglophone research world" (Swales as cited in Hyland, 2006) NNSE researchers face the challenges of meeting the rhetorical standards posed by the community gatekeepers who frequently reject non-standard varieties (Flowerdew & Gosden as cited in Hyland, 2006). It has been suggested that poorly written manuscripts tend to be rejected because the rectification of language difficulties is expensive and time-consuming (Burrough-Boenisch, 2003). Considering this, the motivation of this thesis was to contribute information that may be used in academic writing courses to improve NNSE practitioners' and writers' knowledge so that they can make informed choices that may render their written productions idiomatic and effective. In the specific case of RAs in English written by NNSE authors, the manuscript submitted for publication undergoes a long pre-publication process of revision which focuses on several aspects, the first of which is linguistic appropriateness in relation to the conventions of the genre (Burrough-Boenisch, 2003). The findings of this study may contribute to draw NNSE writers` and practitioners` attention to certain linguistic devices that may help improve their command of academic English so as to ensure that the language in their submitted manuscripts does not interfere with the reviewer's assessment of the manuscript's scientific content. What is more, the information provided should work towards helping NNSE writers make appropriate linguistic choices that allow them to project their own voice along with scientific

content into their written work. This requires skill manipulating language which has been identified as problematic for NNSE writers. It has been pointed out that the results of studies like the one in this thesis "are clearly applicable to the needs of those seeking to be socialized into that community" (Hunston, 2002, p. 204).

The studies in English for Specific Purposes show a tendency to become "narrower and deeper" in scope (Swales, 1990, p. 3). The research presented in this thesis is one of such studies for it focuses on the analysis of a specific feature and its use/s in the genre of the research article in the specific field of experimental sciences and intends to relate these to the communicative functions of the genre.

The findings of this thesis can be applied to the teaching of English for Academic Purposes:

a. contributing information on a specific type of vocabulary, -ly stance adverbs, used for the expression of stance in the RA in English.
b. relating the use of different stance adverb categories, Epistemic, Attitudinal, and Style of speaking stance, to the communicative purposes of the RA in English.
c. relating the use of –ly adverbs that express stance to the different sections (IMRD) of the RA.

The findings of the present study provided lists of stance –ly adverbs used in both corpora of RAs as complete documents and in the IMRD sections (Tables 5.3, 5.15, 5.18, 5.20, and 5.22). They also provided lists of stance adverbs grouped into categories as used by NSE and NNSE writers (Appendix C). Worldlists are considered to be "an effective way for learning a great deal of vocabulary" (Nation as cited in Ryan, 1997, p. 201). These lists can be used in the design of class material to raise NNSE writers' awareness of the variety of resources that the language offers to express different types of stance by focusing on those items that are not used by non-native writers. We should bear in mind that lists of items are important but they should be complemented with descriptions of the collocational and colligational patterns in which –ly adverbs as stance markers frequently occur, descriptions of the contexts where they occur and the purposes they help realize. Compiling lists of frequently used items for their inclusion in courses is considered a worthwhile strategy for informing classroom practice to help students vary their academic writing (Milton as cited in Hyland, 2006).

The inclusion of lists in writing courses to show the variety of items, particularly those used only by NSE writers, should be part of a methodological approach that favors explicit teaching, authentic language analysis, and discussion. Such an approach should take into consideration the analysis of concordance lines of specific –ly derivational adverbs, together with the analysis of the collocational and colligational patterns in which they occur. Granger and Tyson (1996) argue in favor of explicit teaching of items which are identified as problematic for NNSE users. They propose that one of the methods for doing this is the examination of the use, both semantic and syntactic, of these items in authentic texts. This kind of analysis can prompt the development of learner writers' pragmatic appropriacy. According to Flowerdew (2001), *pragmatic appropriacy* concerns the writer's attitude (*stance* in the present thesis) and its formulation in relation to the contexts of situation and culture, which are essential in determining appropriacy. Flowerdew (2001) suggests that there are certain types of activities that can contribute to the development of pragmatic appropriacy. For example, discussion-type activities on specific linguistic devices may enable students to assess the pragmatic appropriacy of such devices. In addition to this, learners may be presented with exercises designed to help identify the most common collocational and colligational patterns of a specific stance adverb.

Different authors (Hyland, 1994, 2002; Martínez, 2002; Swales, 1990) have emphasized the value of making writers aware of the rhetorical structure of genres and the variety of syntactic and lexical choices used to express rhetorical purposes effectively. The findings of the present study in relation to the use of stance adverbs in the writing of RAs in English may be valuable in this respect. Epistemic, Attitude and Style stance adverbs are used to contribute to the realization of different rhetorical functions in the RA and its sections and from this perspective they can be considered valuable resources for NSE and NNSE writers. The expression of stance is an important aspect of professional academic discourse for it can help writers to gain acceptance of their work. This has been noted by different authors (Martinez, 2005; Salager-Meyer, 1994) who have argued in favor of explicit instruction of problematic features so as to empower NNSE writers with the necessary tools that will enable them to make their writing more idiomatic and effective. It is clear that much is involved in the creation of an appropriate authorial stance for it is realized through a wide range of linguistic resources; however, the findings of this study are felt to contribute to enhancing the stance repertoire of NNSE writers and EAP/ESP instructors as well.

7.2. Strengths of the study

One of the strengths of this corpus-based thesis is the fact that it contributes to a specific area that deserves more attention, the use of stance adverbs in academic writing courses, because it is problematic for NNSE writers. The findings of this study shows that -ly stance adverbs are used differently by NSE and NNSE.

Although the scope of the study reported in this work is limited to a list of –ly stance adverbs, it has the advantage of allowing for comparisons with other studies which have focused on similar items, or which have focused on features that characterize the experimental RA genre. Also, the use of statistical information to describe the distribution and association patterns of the specific items studied adds value to the results of the present research since this makes it possible to make interpretations on the relationship of words that would not have been possible using raw frequency counts.

Granger (2002) states that contrastive corpus-based studies are useful in identifying features which are difficult for NNSE and that deserve attention. NNSE writing may be packed with features that make it sound unidiomatic. This study has identified some of these features as instances of underuse and overuse. The ones that were underused by the NNSE writers of RAs in English are *interestingly*, *presumably*, and *likely*, and the ones that were overused are *definitely*, *basically*, *usually*, and *mainly*. By drawing writers' attention to these items, it is possible to inform learners about choices available to increase their vocabulary.

This study provides information that may be used in academic writing materials design. Different authors have pointed out that corpus-based studies can provide information for materials designers that is empirically based rather than based on intuitions (Conrad, 1999; Hyland, 1994).

Corpus-based approaches can answer questions about language and provide information using concrete support and examples from real language in use. In this particular case the study has made a humble contribution to the way stance, by means of the use of –ly adverbs, is conveyed in the writing of RA in English.

7.3. Limitations of the study

The first limitation of this study is that this study focused on only one aspect of stance, although stance can be realized through a variety of features that are beyond the scope of this research. Considering this, this study describes a small proportion of the realizations of stance in RA writing and it probably excludes features, such as other –ly adverbs not present in the worldlists compiled, stance nouns and adjectives, and evaluative constructions.

Corpus-based studies can only describe what is attested in the data provided in the corpora used. There may be aspects of evaluation that are not overtly expressed, therefore, in such cases corpus-based studies like this one may not be able to provide a complete description of the category studied.

Despite these limitations, it is felt that this work is a small step toward a better understanding of how NNSE writers' production relates to NSE writing of experimental RA in English.

7.4. Suggestions for further research

In future works on the way stance is realized in the text of the research article in English it would be desirable:

- To perform a move analysis to see if these stance markers can be related to the different moves identified for the different sections of the research article.
- To study other linguistic features that express stance, for example nouns and adjectives, to see the way they combine, if so, to convey writers' opinions, attitudes, and evaluations.
- To identify and study other adverbs functioning as stance markers in the genre of RAs in English. This would be useful to complement the results provided in the present study.
- To study the other linguistic features pointed out by Biber et al. (1999) as realizing stance. This would give a better picture of how writers express their opinions, feelings and judgments in this specific register and genre.

7.5. Personal Comments

You can do anything in this world if you're prepared to take the consequences.

W. Somerset Maugham, *The Circle*

Carrying out the research work described in this thesis has been time-consuming and demanding but worthwhile for it has contributed to my academic growth as a researcher and as an ESP practitioner.

APPENDIX A

REFERENCE CORPUS

Albig, A.R., & Decker, C.J. (2001). The target of rapamycin signaling pathway regulates mRNA turnover in the yeast saccharomyces cerevisiae. *MolecularBiology of the Cell, 12(11)*, 3428-3438.

Alleman, A.R., Barbet,A. F., Bowie, M.V., Sorenson,H. L., Wong, S.J., & Bélanger, M. (2000). Expression of a gene encoding the major antigenic protein 2 homolog of ehrlichia Chaffeensis and potential application for serodiagnosis. *Journal of Clinical Microbiology, 38 (10)*, 3705-3709.

Almond, N., Jenkins, A., Jones, S., Arnold, C., Silverab, P., Kentc, K., Mills, K.H.G., & Stott, E.J. (1999). The appearance of escape variants in vivo does not account for the failure of recombinant envelope vaccines to protect against simian immunodeficiency virus. *Journal of General Virology, 80(9),* 2375-2382.

An,D.S., Wersto,R.P., Agricola,B.A., Metzger,M.E., Lu,S., Amado,R.G., Chen,I.S.Y., & Donahue, R.E.(2000). Marking and gene expression by a lentivirus vector in transplanted human and nonhuman primate CD34+ Cells. *Journal of Virology,74 (3)*, 1286-1295.

Anson, R.M., Huson, E., & Bohr, V. (2000). Mitochondrial endogenous oxidative damage has been overestimated. *The FASEB Journal, 14,* 355-360.

Balasundaram, D., Benedik, M.J., Morphew, M., Dang, V., & Levin, H.L. (1999). Nup124p is a nuclear pore factor of schizosaccharomyces pombe that is important for nuclear import and activity of retrotransposon Tf1. *Molecular and Cellular Biology,19(8),*5768-5784.

Baron, M.D., Foster-Cuevas, M. Baron, J., & Barrett, T. (1999). Expression in cattle of epitopes of a heterologous virus using a recombinant rinderpest virus. *Journal of General Virology,80(8)*, 2031- 2039.

Bedford, M.T., Frankel,A., Yaffe, M.B., Clarke,S., Leder,P., & Richard, S. (2000)Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *Journal of Biological Chemistry,275(21)*, 16030-16036

Beltman,J., Erickson, J.R., Martin,G.A., Lyons,J.F. & Cook, S.J. (1999). C3 toxin activates the stress signaling pathways, JNK and p38, but antagonizes the activation of AP-1 in rat-1 Cells. *Journal of Biological Chemistry, 274(6)*, 3772-3780.

Bergmann, D.J., Zahn, J.A., Hooper, A.B. & Dispirito, A.A. (19998). Cytochrome P640 genes from the Methanotroph Mthylococcus capsulatus Bath. *Journal of Bacteriology, 180(24)*, 6440-6445.

Berson,J.F., Harper,D.C., Tenza,D., Raposo,G., & Marks, M.S. (2001). Pmel17 Initiates premelanosome morphogenesis within multivesicular bodies. *Molecular Biology of the Cell, 12(11)*, 3451-3464.

Bobanovic, L.K., Laine, M., Petersen, C.C.H., Bennet, D., Berrigdge, M., Lipp, M., Ripley, S.J., & Bootman, M.D. (1999). Molecular cloning and immunolocalization of a novel vertebrate trp homologue from Xenopus. *Biochemistry* Journal, 340, 593-599.

Boring, l., Gosling, J., Monteclaro, F.S., Lusis, A.J., Tsou, C., Charo, I.F. (1996). Molecular cloning and functional expression of murine JE (Monocyte

Chemoattractant Protein 1) and murine macrophage inflammatory protein 1 Receptors. *Journal of Biological Chemistry 271 (13)*, 7551-7558.

Bunikis,J., & Barbour, A.G.(1999).Access of antibody or trypsin to an integral outer membrane protein (P66) of borrelia burgdorferi is hindered by Osp lipoproteins. *Infection and Immunity, 67(6),* 2874-2883.

Bunnell,S.C., Henry,P.A., Kolluri, R., Kirchhausen,T., Rickles, R.J., & Berg, L.J. (1996). Identification of Itk/Tsk Src homology 3 domain ligands. *Journal of Biological Chemistry, 271 (41)*, 25646-15656.

Butt, A. J., Harvey, N., Parasivam,G., & Kumar, S. (1998). Dimerization and Autoprocessing of the Nedd2 (Caspase-2) Precursor Requires both the Prodomain and the Carboxyl-terminal Regions. *Journal of Biological Chemistry, 273 (12)*, 6763-6768.

Caricchio, R., Reap, E.A., & Cohen, P.L. (1998). Fas/Fas ligand interactions are involved in ultraviolet-B-induced human lymphocyte apoptosis. *The Journal of Immunology, 161(1)*, 241-251.

Caron,J.M. Vega,L.R., Fleming,J., Bishop,R., & Solomon, F.(2001). Single Site - Tubulin Mutation Affects Astral Microtubules and Nuclear Positioning during Anaphase in Saccharomyces cerevisiae: Possible Role for Palmitoylation of – Tubulin. Molecular Biology of the Cell, 12 (9), 2672-2687.

Carroll,J.A., Cordova,R.M., & Garon,C.F. (2000). Identification of 11 pH-regulated genes in borrelia burgdorferi localizing to linear plasmids. *Infection and Immunity, 68(12),* 6677-6684.

Carteau,S., Gorelick,R.J., & Bushman, F.D. (1999). Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *Journal of Virology, 73(8)*, 6670-6679.

Cassatt,D.R., Patel,N.K., Ulbrandt,N.D., & Hanson, M.S. (1998). DbpA, but not OspA, is expressed by borrelia burgdorferi during spirochetemia and is a target for protective antibodies. *Infection and Immunity,66(11),*5379-5387.

Chattopadhyay,D., Ghosh,M.K., Mal,A. & Harter, M.L.(2001). Inactivation of p21 by E1A Leads to the Induction of Apoptosis in DNA-Damaged Cells. *Journal of Virology, 75(20),* 9844-9856.

Chen, D., Heath,V., O'Garra,A., Johnston, J., & McMahon, M. (1999). Sustained activation of the Raf-MEK-ERK pathway elicits cytokine unresponsiveness in T cells. *The Journal of Immunology, 163(11),* 5796-5805.

Cimini,D., Howella,B., Maddoxa,P., Khodjakovc,A., Degrassib,F. & Salmona, E.D. (2001). Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *The Journal of Cell Biology, 153(3),* 517-528.

Cleveland, S.M., Jones, T.D., & Dimmock, N.J. (2000). Properties of a neutralizing antibody that recognizes a conformational form of epitope ERDRD in the gp41 C-terminal tail of human immunodeficiency virus type 1. *Journal of General Virology,81(5)*, 1251- 1260.

Coffin, S.E., Clark, S.L, Bos,N.A., Brubaker, J.O., & Offit, P.A. (1999). Migration of antigen-presenting B cells from peripheral to mucosal lymphoid tissues may induce intestinal antigen-specific IgA following parenteral immunization. *The Journal of Immunology, 163(6)*, 3064-3070.

Comer,A., Ahern-Djamali,S.M., Juang,J., Jackson,P.D., & Hoffmann, F.M. (1998). Phosphorylation of enabled by the drosophila abelson tyrosine kinase regulates the in vivo function and protein-protein interactions of enabled. *Molecular and Cellular Biology,18(1),*152-160.

Cowley,J.A., Dimmock, C.M., Spann, K.M., & Walker, P.J. (2000). Gill-associated virus of Penaeus monodon prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *Journal of General Virology,81(6)*, 1473- 1484.

Davy., A.J., Dunsford, S.J., & Free, A.J. (1998). Acidifying peat as an aid to the reconstruction of lowland heath on arable soil: lysimeter experiments. *Journal of Applied Ecology, 35,* 649-659.

Déas, O., Dumont, C., MacFarlane., M., Rouleau, M., Hebib, Ch., Harper,F., Hirsch, F., Charpentier, B., Cohen, G., & Senik, A. (1998). Caspase-Independent Cell Death Induced by Anti-CD2 or Staurosporine in Activated Human Peripheral T Lymphocytes1. *The Journal of Immunology, 161,* 3375-3383.

Deshpande, S. S., Angkeow, P., Huang, J., Ozaki, M., & Irani1, K. (2000). Rac1 inhibits TNF--induced endothelial cell apoptosis: dual regulation by reactive oxygen species. The FASEB Journal, 14, 1705-1714.

Deveraux,Q., Roy,N., Stennicke,H.R., Van Arsdale1,T., Zhou,Q., Srinivasula, S., Alnemri,E., Salvesen, G.S., & Reed, J.(1998).IAPs block apoptotic events induced by caspase-8 and cytochrome c by direct inhibition of distinct caspases. *The EMBO Journal 17(8),* 2215-2223.

Dirnbach, E., Steel, D.G., & Gafni, A. (1999). Proline isomerizatiion is unblikely to be the cause of slow annealing and reactivation during the folding of alkaline phosphatase. *The Journal of Biological Chemistry, 274(8)*, 4532-4536.

Dnnenberg,M.S., Tacket,C.O., Losonsky,G., Frankel,G., Nataro,J.P. Dougan,G., & Levine, M.M. (1998). Effect of prior experimental human enteropathogenic escherichia coli infection o illness following homologous and heterologous rechallenge. *Infection and Immunity, 66(1),52-58.*

Domachowske, J.B., Bonville,C.A., Gao, J., Murphy, P.M., Easton, A.J., & Rosenberg, H.F. (2000). The chemokine macrophage-inflammatory protein-1 and its receptor CCR1 control pulmonary inflammation and antiviral host defense in paramyxovirus infection. *The Journal of Immunology, 165 (5),* 2677-2682.

Dorstyn, L., Colussi, P., Quinn. L.M., Richardson, H. & Kumar, S. (1999). DRONC, an ecdysone-inducible Drosophila caspase (apoptosis / caspase activation / caspase recruitment domain / development / metamorphosis). *Proceedings of the National Academy of Sciences, 96 (8)*, 4307-4312.

Doukas, J., Hoganson,D.K., Ong,M., Ying, W., Lacey, D., Baird,A., Pierce, G.F., &. Sosnowski, B.A. (1999). Retargeted delivery of adenoviral vectors through fibroblast growth factor receptors involves unique cellular pathways. *The FASEB Journal, 13,* 1459-1466.

Drevets, D.A. (1998). Listeria monocytogenes virulence factors that stimulate endothelial cells. *Infection and Immunity, 66(1),*232-238.

Dumont, F.J., Staruch,M.J., Fischer,P., DaSilva, C., & Camacho, R. (1998). Inhibition of T cell activation by pharmacologic disruption of the MEK1/ERK MAP kinase or calcineurin signaling pathways results in differential modulation of cytokine production. *The Journal of Immunology, 160(6),* 2579-2589.

Duncan,M.R., Frazier, K.S., Abramson, S., Williams,S., Klapper,H., Huang,X., & Grotendorst, G.R. (1999). Connective tissue growth factor mediates transforming growth factor ß-induced collagen synthesis: down-regulation by cAMP. *The FASEB Journal, 13(13),* 1774-1786.

Dunn,R., & Hicke, L. (2001). Domains of the Rsp5 ubiquitin-protein ligase required for receptor-mediated and fluid-phase endocytosis. *MolecularBiology of the Cell, 12(2)*, 421-435.

Dunstan, C., Salafranca, M.N., Adhikari, S., Xia, Y., Feng, L., & Harrison, J.K. (1996). Identification of two rat genes orthologous to the human interleukin-8 receptors. *Journal of Biological Chemistry 271 (51)*, 32770- 32776.

Ekert, P. G., Silke, J., & Vaux, D. L.(1999). Inhibition of apoptosis and clonogenic survival of cells expressing crmA variants: optimal caspase substrates are not necessarily optimal inhibitors. *The EMBO Journal, 18(2),* 330-338.

Ekert, P.G., Silkea, J., Hawkinsc, C.J., Verhagena, A. M., & Vauxa, D.L. (2001). Diablo promotes apoptosis by removing miha/xiap from processed caspase. *The Journal of Cell Biology, 152 (3)*, 2001 483-490.

Elkins,C., Totten,P.A., Olsen,B., & Thomas, C.E. (1998). Role of the haemophilus ducreyi ton system in internalization of heme from hemoglobin. *Infection and Immunity, 66(1), 151-160.*

Enslen,H., Raingeaud,J., & Davis, R. (1998). Selective Activation of p38 Mitogen-activated Protein (MAP) Kinase Isoforms by the MAP Kinase Kinases MKK3 and MKK6. *Journal of Biological Chemistry, 273 (3),* 1741- 1748.

Erhardt,P., Schremser,E.J., & Cooper, G.M. (1999). B-Raf inhibits programmed cell death downstream of cytochrome c release from mitochondria by activating the MEK/Erk pathway. *Molecular and Cellular Biology, 19(8)*,5308-5315.

Esposito, D., & Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *The EMBO Journal, 17 (19)*, 5832-5843.

Evans, B.A., Agar, L., Summers,. R.J. (1999). The role of the sympathetic nervous system in the regulation of leptin synthesis in C57BL/6 mice. *FEBS Letters, 444*, 149-154.

Evans, E. K., Kuwana1,T., Strum, S. L., Smith, J.J., Newmeyer1. D.D., & Kornbluth, S. (1997). Reaper-induced apoptosis in a vertebrate system. *The EMBO Journal, 16,* 7372-7381.

Exner,M.M., Wu,X., Blanco,D.R., Miller,J.N. & Lovett, M.A. (2000). Protection elicited by native outer membrane protein Oms66 (p66) against host-adapted borrelia burgdorferi: conformational nature of bactericidal epitopes. *Infection and Immunity, 68(5),2647-2654.*

Faleiro, L., Kobayashi,R., Fearnhead, H., & Lazebnik,Y.(1997). Multiple species of CPP32 and Mch2 are the major active caspases present in apoptotic cells. *The EMBO Journal, 16(9),*2271-2281.

Fassati, A.& Goff, S.P. (2001). Characterization of Intracellular Reverse Transcription Complexes of Human Immunodeficiency Virus Type 1. *Journal of Virology, 75 (8)*, 3626-3635.

Feldman, M., Bryan,R., Rajan,S., Scheffler, L., Brunnert,S., Tang,H., & Prince, A. (1998). Role of flagella in pathogenesis of pseudomonas aeruginosa pulmonary infection. *Infection and Immunity, 66(1),43-51.*

Feng,S., Hodzic,E., Stevenson,B., & Barthold, S.W. (1998). Humoral immunity to borrelia burgdorferi N40 decorin binding proteins during infection of laboratory mice. *Infection and Immunity, 66(6),* 2827-2835.

Feranchak, A.P., Berl, T., Capasso, J., Wojtaszek, P.A., Han, J., & Fitz, J.G.(2001). p38 MAP kinase modulates liver cell volume through inhibition of membrane Na+ permeability. *The Journal of Clinical Investigation,108(10),*1495-1504.

Feranchak, A.P., Roman, R.M., Schwiebert, E.M., & Fitz, J.F. (1998). Phosphatidylinositol 3-kinase contributes to cell volume regulation through effects on ATP release. *Journal of Biological Chemistry, 273 (24),* 14906-14911.

Ferguson, E.L. & Anderson, K.V. (1992). Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the drosophila embryo. *Cell, 71*, 451-461.

Foster, C., Lehrnbecher, T., Mol, F., Steinberg, S.M., Venzon, D. J., Walsh, T. J., Noack, D., Rae, J., Winkelstein, J.A., Curnutte, J.T., & Chanock, S.J. (1998). Host defense molecule polymorphisms influence the risk for immune-mediated complications in chronic granulomatous disease. *Journal of Clinical Investigation, 102 (12)*, 2146-2155.

Foster, L.C, Wiesel,P., Huggins, G.S., Pañares, R., Chin, M.T., Pellacani, A., & Perrella, M.A. (2000). Role of activating protein-1 and high mobility group-I(Y) protein in the induction of CD44 gene expression by interleukin-1ß in vascular smooth muscle cells. *The FASEB Journal, 14(2),* 368-378.

Fraser, A. G., & Evan, G. (1997). Identification of a Drosophila melanogaster ICE/CED-3-related protease, drICE. The EMBO Journal, 16 (10), 2805-2813.

Fraser, A.G., McCarthy, N. J., & Evan, G. (1997). drICE is an essential caspase required for apoptotic activity in Drosophila cells. *The EMBO Journal, 16 (20)*, 6192-6199.

Fukunaga, R., & Hunter, T. (1997). MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *The EMBO Journal* 16(8), 1921-1933.

Gallay,P., Hope,T., Chin,D., & Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway (nuclear import / human immunodeficiency virus). *Proceedings of the National Academy of Science, 94(18),* 9825-9830.

Gao, J. & Murphy, P. (1995). Cloning and differential tissue-specific expression of three mouse chemokine receptor-like genes, including the gene for a functional macrophage inflammatory protein-1 receptor. *Journal of Biological Chemistry, 270 (29),* 17494-17501.

George, K.M., Barker,L.P., Welty,D.M., & Small, P. L. C. (1998). Partial purification and characterization of biological effects of a lipid toxin produced by mycobacterium ulcerans. *Infection and Immunity, 66(2)*, 587-593.

Gilmore Jr., R.D., & Mbow, M. (1998). A monoclonal antibody generated by antigen inoculation via tick bite is reactive to the borrelia burgdorferi rev protein, a member of the 2.9 gene family locus. *Infection and Immunity, 66(3),* 980-986.

Gilmore Jr., R.D., & Mbow, M. (1999). Conformational nature of the borrelia burgdorferi B31 outer surface protein C protective epitope. *Infection and Immunity, 67(10),* 5463-5469.

Gordon, D.M., & FitzGibbon, F. (1999). The distribution of enteric bacteria from Australian mammals: host and geographical effects. *Microbiology, 145*, 2663-2671.

Goyal, L., McCall, K., Agapite, J., Hartwieg, E., & Steller , H. (2000). Induction of apoptosis by Drosophila reaper, hid and grim through inhibition of IAP function. *The EMBO Journal, 19(4),* 589-597.

Gruber, A. Kan-Mitchell, J., Kuhen, K.L., Mukai,M. & Wong-Staal; F. (2000). Dendritic cells transduced by multiply deleted HIV-1 vectors exhibit normal phenotypes and functions and elicit an HIV-specific cytotoxic T-lymphocyte response in vitro. *Blood, 96(4)*, 1327-1333.

Gyetko, M.A., Sud, S., Kendall, T., Fuller, J.A. Newstead, M.W., & Standiford, T.J. (2000). Urokinase receptor-deficient mice have impaired neutrophil recruitment in response to pulmonary pseudomonas aeruginosa infection. *The Journal of Immunology, 165(3),* 1513-1519.

Hall, D.M., Oberley,T. D., Moseley,P.M., Buettner,G.R., Oberley, L.W., Weindruch, R. & Kregel, K.C.(2000). Caloric restriction improves thermotolerance and reduces hyperthermia-induced cellular damage in old rats. *The FASEB Journal, 14(1),* 78-86.

Hanson,M.S., Cassatt,D.R., Guo,B.P., Patel,N.K., McCarthy,M.P., Dorward,D.W., & Höök, M. (1998).Active and passive immunity against borrelia burgdorferi decorin binding protein A (DbpA) protects against infection. *Infection and Immunity, 66(5),* 2143-2153.

Harvey, N.L., Daish, T., Mills,K., Dorstyn,L., Quinn,L.M., Read,S., Richardson,H., & Kumar, S. (2001). Characterization of the Drosophila Caspase, DAMM. *Journal of Biological Chemistry, 276 (27)*, 25342-25350.

Hefty,P.S., Jolliff,S.E., Caimano,M.J., Wikel,S.K., Radolf,J.D., & Akins, D.R.(2001). Regulation of OspE-related, OspF-related, and Elp lipoproteins of borrelia burgdorferi strain 297 by mammalian host-specific signals. *Infection and Immunity, 69(6),* 3618-3627.

Henning,S.W., Galandrini,R., Hall, A., & Cantrell, D.A. (1997).The GTPase Rho has a critical regulatory role in thymus development. *The EMBO Journal, 16 (9)*, 2397-2407.

Hildebrandt, E.R., & Hoyt, M.A. (2001). Cell cycle-dependent degradation of the saccharomyces cerevisiae spindle motor cin8p requires APCCdh1 and a bipartite destruction sequence. *MolecularBiology of the Cell, 12(11)*, 3402-3416.

Hill,C.L., Bieniaszb,P.D., & McClure, M.O. (1999. Properties of human foamy virus relevant to its development as a vector for gene therapy. *Journal of General Virology,80(8)*, 2003- 2009.

Hogaboam, C.A., Gallinat,C., Taub,D., Strieter, R.M., Kunkel, S.L., & Lukacs, N.W. (1999). Immunomodulatory role of C10 chemokine in a mrine model of allergic bronchopulmonary aspergillosis. *The Journal of Immunology, 162(10),*6071-6079.

Hogan, S.P., Mishra, A., Brandt, E.B., Foster,P.S., & Rothenberg, M.E.(2000). A critical role for eotaxin in experimental oral antigen-induced eosinophilic gastrointestinal allergy. *Proceedings of the National Academy of Science, 97 (12),* 6681-6686.

Hooper, D.C., Scott, G.S., Zborek, A., Mikheeva, T., Kean, R.B., Koprowski, H., & Spitsin, S.V.(2000). Uric acid, a peroxynitrite scavenger, inhibits CNS inflammation, blood–CNS barrier permeability changes, and tissue damage in a mouse model of multiple sclerosis. *The FASEB Journal, 14(5) , 691-698.*

Horn, D., & Cross, G.M. (1997). Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. *The EMBO Journal, 16(24)*, 7422-7431.

Hosking, S.L., Craig, J.E., & High, N.J. (1999).Phase variation of lic1A, lic2A and lic3A in colonization of the nasopharynx, bloodstream and cerebrospinal fluid by Haemophilus inluenzae type b. Microbiology, 145, 3005-3011.

Howell, M., & Hill1, C. S. (1997). XSmad2 directly activates the activin-inducible, dorsal mesoderm gene XFKH1 in Xenopus embryos. *The EMBO Journal, 16,* 7411-7421.

Hozak,R.R., Manji,G.A., & Friesen, P.D. (2000). The BIR motifs mediate dominant interference and oligomerization of inhibitor of apoptosis Op-IAP. *Molecular and Cellular Biology,20(5),* 1877-1885.

Hunter, S., Indik, Z. K., Kim, M., Cauley, M. D., Park, J., & Schreiber, A.D. (1998). Inhibition of Fc receptor-mediated phagocytosis by a nonphagocytic Fc receptor. *Blood, 91(5)*, 1762-1768.

James, E., Wu,P.G., Stites, W., & Brand, L. (1992). Compact denatured state of a staphylococcal nuclease mutant by guanidinium as determined by resonance energy transfer. *Biochemistry, 31*, 10217-10225.

Jenkins,M., Hanley, M.B., Moreno, M.B., Wieder,E., & McCune, J.M. (1998). Human immunodeficiency virus-1 infection interrupts thymopoiesis and multilineage hematopoiesis in vivo. *Blood, 91 (8)*, 2672-2678.

Johnston,J.C.,Gasmi,M., Lim,L.E., Elder,J.H., Yee,J., Jolly,D.J., Campbell,K.P., Davidson,B.L., & Sauter, S. L.(1999). Minimum Requirements for Efficient Transduction of Dividing and Nondividing Cells by Feline Immunodeficiency Virus Vectors. *Journal of Virology,73(6)*, 4991-5000.

Johnstone, R., Cretney,E. & Smyth, M.J. (1999). P-Glycoprotein Protects Leukemia Cells Against Caspase-Dependent, but not Caspase-Independent, Cell Death .*Blood, 93 (3)*, 1075-1085.

Jones, M.H., He, X., Giddings, T.H., & Winey, M. (2001). Cell Biology. Yeast Dam1p has a role at the kinetochore in assembly of the mitotic spindle. *Proceedings of the National Academy of Science, 98(24),* 13675-13680.

Jones, M.K., Sasaki, E., Halter, F., Pai,R., Nakamura, T., Arakawa, T., Kuroki, T., & Tarnawski, A.S.(1999). HGF triggers activation of the COX-2 gene in rat gastric epithelial cells: action mediated through the ERK2 signaling pathway. *The FASEB Journal, 13(15), 2186-2194.*

Junger,W.G., Hoyt, D.B., Davis,R.E., Herdon-Remelius,C., Namiki,S., Junger,H., Loomis, w. & Altman, A. (1998). Hypertonicity Regulates the Function of Human Neutrophils by Modulating Chemoattractant Receptor Signaling and Activating Mitogen-activated Protein Kinase p38. *Journal of Clinical Investigation, 101 (12),* 2768-2779.

Jürgensmeier,J., Xie,Z., Deveraux,Q., Ellerby,L., Bredesen,D. & Reed, J.(1998).Bax directly induces release of cytochrome c from isolated mitochondria. *Proceedings of the National Academy of Sciences,95(9),* 4997-5002.

Kang,J.J, Schaber,M.D., Srinivasula,S. Alnemri,E., Litwack,G. Hall,D., & Bjornsti, M.(1999).Cascades of Mammalian Caspase Activation in the Yeast Saccharomyces cerevisiae. *Journal of Biological Chemistry, 274(5),* 3189-3198.

Katz, E. , Deehan, M. R., Seatter, S., Lord, C., Sturrock, R. & Harnett, M. (2001). B Cell Receptor-Stimulated Mitochondrial Phospholipase A2 Activation and Resultant Disruption of Mitochondrial Membrane Potential Correlate with the Induction of Apoptosis in WEHI-231 B Cells. *The Journal of Immunology, 166,* 137-147.

Kenna,M.A., Brachmann,C.B., Devine,S.E. & Boeke, J.D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. *Molecular and Cellular Biology,18(2),*1115-1124.

Kitchen, S.G., Korin, Y.D., Roth,M.D., Landay,A., & Zack1,J.A.(1998). Costimulation of naive CD8+ lymphocytes induces CD4 expression and allows human immunodeficiency virus type 1 infection. *Journal of Virology, 72 (11)*, 9054-9060.

Kooby, D.A., Carew, J.F., Halterman, M.W., Mack, J.E., Bertino, J.R., Blumgart, L.H., Federoff, H., & Fong, Y. (1999). Oncolytic viral therapy for human colorectal

cancer and liver metastases using a multi-mutated herpes simplex virus type-1 (G207). *The FASEB Journal, 13(11),1325-1334.*

Korin, Y.D., & Zack, J.A. (1998). Progression to the G1b phase of the cell cycle is required for completion of human immunodeficiency virus type 1 reverse transcription in T cells. *Journal of Virology, 72 (4),* 3161-3168.

Kovala, A.T., Harvey, K.A., Mcglynn, P., Boguslawski, G., Garcia. J.G.N., & English, D. (2000). High-efficiency transient transfection of endothelial cells for functional analysis. *The FASEB Journal, 14(15),2486-2494.*

Kovarik,P., Stoiber, D., Eyers,P.A., Menghini, R., Neininger,A., Gaestel,M., Cohen,P., & Decker, T. 1999. Stress-induced phosphorylation of STAT1 at Ser727 requires p38 mitogen-activated protein kinase whereas IFN- uses a different signaling pathway. *Proceedings of the National Academy of Science, 96 (24),* 13956-13961.

Krauer, K.G., Buck, M., & Tom Sculley (999). Characterization of the transcriptional repressor RBP in Epstein–Barr virus-transformed B cells. *Journal of General Virology,80(12),* 3217-3226.

Kubo, H.., Morgenstern, D., Quinlan, W.M., Ward, P.A., Dinauer, M.C., & Doerschuk, C.M. (1996). Preservation of Complement-induced Lung Injury in Mice with Deficiency of NADPH Oxidase. *Journal of Clinical Investigation, 97 (11),* 2680-2684.

LaCount,D.J., Hanson,S.F., Schneider,C.L., & Friesen, P.D. (2000). Caspase inhibitor P35 and inhibitor of apoptosis Op-IAP block in vivo proteolytic activation of an effector caspase at different steps. JBC *275 (21),* 15657-15664

Laird, S.M., Graham, A., Paul, A., Gould, G.W., Kennedy, C., & Plevin, R. (1998). Tumor necrosis factor stimulates stress activated protein kinases and the inhibition of DNA synthesis in cultures of bovine aortic endothelial cells. *Cellular Signalling, 10(7),* 473-480.

Lampson,M.A., Schmoranzer,J., Zeigerer,A., Simon,S.M., & McGraw, T.E. (2001). Insulin-regulated release from the endosomal recycling compartment is regulated by budding of specialized vesicles. *MolecularBiology of the Cell, 12(11,* 3489-3501.

Lee, C.G., Ramachandra,M, Jeang, K.T., Martin, M. A., Pastan, I. & Gottesman, M.M. (2000). Effect of ABC transporters on HIV-1 infection: inhibition of virus production by the MDR1 transporter. *The FASEB Journal, 14 (3),* 516-522.

Lee, Raymond, Y.N., Lobel, L., Hengartner, M., Horvitz, H.R., & Avery. L. (1997). Mutations in the α1 subunit of an L-type voltage-activated Ca^{2*} channel cause myotonia in Caenorhabditis elegans. *The EMBO Journal, 16(20),*6066-6076.

Lee,B., Bi,L., Garfinkel,D.J., & Bailis, A.M. (2000). Nucleotide excision repair/TFIIH helicases Rad3 and Ssl2 inhibit short-sequence recombination and Ty1 retrotransposition by similar mechanisms. *Molecular and Cellular Biology,20(7),* 2436-2445.

Levy, B.D., Fokin,V.V., Clark, J.M., Wakelam, M.J.O., Petasis, N.O. & Serhan, C. (1999). Polyisoprenyl phosphate (PIPP) signaling regulates phospholipase D activity: a `stop' signaling switch for aspirin-triggered lipoxin A4. *The FASEB Journal. 13(8),*903-911.

Lewis, P. J., van Drunen Littel-van den Hurk1, S., & Babiuk, L. A. (1999). Induction of immune responses to bovine herpesvirus type 1 gD in passively immune mice after immunization with a DNA-based vaccine. *Journal of General Virology, 80(11),* 2829-2837.

Lowe, E.D., Noble, M.E.M, Skamnaki, V.T., Oikonomakos, N. G., Owen, D.J., & Johnson, L.N. (1997). The crystal structure of a phosphorylase substrate recognition. *The EMBO Journal, 16(22)*, 6646-6658.

Macartney,K.K., Baumgart,D.C., Carding,S.R., Brubaker,J.O., & Offit,P.A. (2000). Primary murine small intestinal epithelial cells, maintained in long-term culture, are susceptible to rotavirus infection. *Journal of Virology, 74(12),* 5597-5603.

Manji, G.A. & Friesen, P.D. (2001). An apical, p35-insensitive caspase mediates programmed cell death in insect cells. *Journal of Biological Chemistry, 276 (20)*, 16704-16710.

Martins, L. M., Kottke, T., Mesner, P.W., Basi, G.S., Sinha, S., Frigon, N. Jr., Tatar, E. Tung, J.S., Bryant, K., Takahashi, A., Svingen. P.A., Madden, B.J., McCormick, D.J., Earnshaw, W.C., & Kaufmann, S.H. (1997). Activation of multiple interleukin-1 converting enzyme homologues in cytosol and nuclei of hl-60 cells during etoposide-induced apoptosis. *Journal of Biological Chemistry 272(11),* 7421-7430.

Matsukawa, A., Hogaboam, C.M., Lukacs, N.W., Lincoln, P.M., Strieter, R.M., & Kunkel, S.L. (1999). Endogenous monocyte chemoattractant protein-1 (MCP-1) protects mice in a model

Matthews, A.N., Friend,D.S., Zimmermann, N., Sarafi, M., Luster, A.D., Pearlman, E., Wert, S.E., & Rothenberg, M. (1998). Eotaxin is required for the baseline level of tissue eosinophils. *Proceedings of the National Academy of Science, 95 (11),* 6273-6278.

Maurice,M., Mazur,S., Bullough,F.J., Salvetti,A., Collins,M.K., Russell,S.J., & Cosset, F. (1999). Efficient gene delivery to quiescent interleukin-2 (IL-2)-dependent cells murine leukemia virus-derived vectors harboring IL-2 chimeric envelopes glycoproteins. *Blood, Vol. 94 (2)*, 401-410.

McBride, J. W., Corstvet, R. E., Breitschwerdt, E., & Walker, D. H. (2001). Immunodiagnosis of ehrlichia canis infection with recombinant proteins. *Journal of Clinical Microbiology, 39 (1)*, 315-322.

Mcdonald,M.C., Mota-Filipe,H., Paul,A., Cuzzocrea,S., Abdelrahman,M., Harwood,S., Plevin,R., Chatterjee,P.K., Yaqoob, M.M., & Thiemermann, C. (2001) Calpain inhibitor I reduces the activation of nuclear factor-B and organ injury/dysfunction in hemorrhagic shock. *The FASEB Journal, 15 (1)*, 171- 186

McKisic, M.D., & Barthold, S.W. (2000). T-cell-independent responses to borrelia burgdorferi are critical for protective immunity and resolution of lyme disease. *Infection and Immunity, 68(9),* 5190-5197.

Meier, P., Silke, J., Leevers, S.L. & Evan, G. (2000). The Drosophila caspase DRONC is regulated by DIAP1. *The EMBO Journal, 19(4)*, 598-611.

Melby,P.C., Yang,Y., Cheng,J., & Zhao, W. (1998). Regional differences in the cellular immune response to experimental cutaneous or visceral infection with leishmania donovani. *Infection and Immunity, 66(1),*18-27.

Meynell, H.M., Thomas, N.W., James, P.S., Holland, J. Taussig, M., & Nicoletti, C. (1999). Up-regulation of microsphere transport across the follicle.-associated epithelium of Peyer's patch by exposure to Streptococcus pneumoniae R36a. *The FASEB Journal, 13, 611-619.*

Mishra, A., Hogan, S.P., Brandt,E.B., & Rothenberg, M.E. (2000). Peyer's patch eosinophils: identification, characterization, and regulation by mucosal allergen exposure, interleukin-5, and eotaxin. *Blood, 96, 4*, 1538-1544.

Mishra,A., Weaver, T.E., Beck, D.C., & Rothenberg, M.E. (2000). Interleukin-5-mediated allergic airway inflammation inhibits the human surfactant protein C

promoter in transgenic mice. *Journal of Biological Chemistry, 276 (11)*,8453-8459.

Mitchell,T., Kappler,J., & Marrack,P.(1999). Bystander virus infection prolongs activated T Cell survival. *The Journal of Immunology,162(8)*,4527-4535.

Moore, S., Rincke,L.A., & Garfinkel, D.J. (1998). A Ty1 integrase nuclear localization signal required for retrotransposition. *Molecular and Cellular Biology,18(2),* 1105-1114.

Morgan,H., Smith,M., Burke, Z., & Carter, D. (2000).The transactivation-competent carboxyl-terminal domain of AF-9 is expressed within a sexually dimorphic transcript in rat pituitary. *The FASEB Journal, 14 (9)*, 1109-1116.

Morgan. A.L., & Katzman, M. (2000).Subterminal viral DNA nucleotides as specific recognition signals for human immunodeficiency virus type 1 and visna virus integrases under magnesium-dependent conditions. *Journal of General Virology, 81(3), 839-849.*

Morton, L.F., Peachery, A.R., Knight, C.G., Farndale, R.W., & Barnes, M.J. (1997). The platelet reactivity of synthetic peptides bsed on the collagen III fragment α1(III)CB4. *The Journal of Biological Chemistry, 274(5)*, 11044-11048.

Moser, C. A., Cookinham, S., Coffin, S.E., Clark, H.F., & Offit, P.A. (1998). Relative importance of rotavirus-specific effector and memory b cells in protection against challenge. *Journal of Virology, 72 (2),* 1108-1114.

Moser, M., Speaker,T.J., & Offit, P.A. (1998). Effect of water-based microencapsulation on protection against EDIM rotavirus challenge in mice. *Journal of Virology, 72 (5),* 3859-3862.

Mould,A.W., Ramsay,A.J., Matthaei, K.I., Young, I.G., Rothenberg, M. & Foster, P.S.(2000).The effect of IL-5 and eotaxin expression in the lung on eosinophil trafficking and degranulation and the induction of bronchial hyperreactivity. *The Journal of Immunology,164(4),*2142-2150.

Murphy, S.M., Preble,A.M., Patel,U.K., O'Connell,K.L., Prabha Dias,D., Moritz,M., Agard,D., Stults,J.T., & Stearns,T. (2001). GCP5 and GCP6: two new members of the human Î³-tubulin complex. *Molecular Biology of the Cell*, 12(11), 3340-3352.

Murphy,K.L., Dennis, a.P., & Rosen, J.M.(2000).A gain of function p53 mutant promotes both genomic instability and cell survival in a novel p53-null mammary epithelial cell model. *The FASEB Journal, 14 (14)*, 2291-2302.

Neil,S., Martin,F., Ikeda,Y. & Collins, M. (2001). Postentry Restriction to Human Immunodeficiency Virus-Based Vector Transduction in Human Monocytes. *Journal of Virology, 75(12)*, 5448-5456.

Nelson, E.L., Strobl,S., Subleski, J., Prieto, D., Kopp, W.C., & Nelson, P.J. (1999). Cycling of human dendritic cell effector phenotypes in response to TNF-: modification of the current `maturation' paradigm and implications for in vivo immunoregulation. The *FASEB Journal,13(14),* 2021-2030

Nelson, G.W., Stohlman, S.A., & Tahara, S.M. (2000). High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *Journal of General Virology,81(1),*181-188.

Newell, G. R. (1998). Characterization of vegetation in an Australian open forest community affected by cinnamon fungus (Phytophthora cinnamomi): implications for faunal habitat quality. *Plant Ecology, 137*, 55-70.

Nibbs, R.J., Salcedo, T.W., Campbell,J.D., Yao,X., Li,Y. Nardelli,B., Olsen,H.S., Morris,T.S., Proudfoot, A.E., Patel, V.P., & Graham, G.J. (2000). C-C chemokine receptor 3 antagonism by the ß-chemokine mcrophage inflammatory

protein 4, a property strongly enhanced by an amino-terminal alanine-methionine swap. *The Journal of Immunology,164(3),*1488-1497.

Nibbs, R.J., Wylie,S.M., Pragnell, I.B., & Graham, G.J.(1997). Cloning and characterization of a novel murine chemokine receptor, D6 comparison to three other related macrophage inflammatory protein-1 receptors, CCR-1, CCR-3, AND CCR-5. *Journal of Biological Chemistry, 272 (19),* 12495-12504.

Norris, C.F., Pricop, L., Millard, S. S., Taylor, S.M., Surrey, S., Schwartz, E., Salmon, J. & McKenzie, S.M. (1998). A naturally occurring mutation in FCRIIA: A Q to K127 change confers unique IgG binding properties to the R131 allelic form of the receptor . *Blood, 91(2),* 656-662.

of acute septic peritonitis: cross-talk between MCP-1 and leukotriene B4. *The Journal of Immunology, 163(11),* 6148-6154.

Oxenius, A., Martinic,M.M., Hengartner,H., & Klenerman, P. (1999). CpG-Containing oligonucleotides are efficient adjuvants for induction of protective antiviral immune responses with T-Cell peptide vaccines. J*ournal of Virology, 73 (5),* 4120-4126.

Ozaki, M., Deshpande, S. S., Angkeow, P., Bellan, J., Lowenstein, C.J., Dinauer, M.C., Goldschmidt-Clermont, P., & Irani, K. (2000). Inhibition of the Rac1 GTPase protects against nonlethal ischemia/reperfusion-induced necrosis and apoptosis in vivo. *The FASEB Journal, 14(2),* 418-429.

Padgett, P.E., & Allen, E.B. (1999). Differential responses to nitrogen fertilization in native shrubs and exotic annuals common to Mediterranean coastal sage scrub of California. *Plant Ecology,* 144, 93-101.

Paluh,J.L., Nogales,E., Oakley,B.R., McDonald,K.Pidoux, A.L., & Cande, W. Z.(2000). A Mutation in -Tubulin Alters Microtubule Dynamics and Organization and Is Synthetically Lethal with the Kinesin-like Protein Pkl1p. *Molecular Biology of the Cell, 11(4),* 1225-1239.

Panaretou, B., Sinclair, K., Prodromou, C., Johal, J., Pearl, L., & Piper, P.W. (1999). The Hsp90 of Candida albicans can confer Hsp90 functions in Saccharomyces cerevisiae: a potential model for the processes that generate immunogenic fragments of this molecular chaperone in C. albicans infections. *Microbiology, 145,* 3455-3463.

Parry, D., Mahony, D., Wills, K., & Lees, E. (1999). Cyclin D-CDK subunit arrangement is dependent on the availability of competing INK4 and p21 class inhibitors. *Molecular and Cellular Biology,19(3),* 1775-1783.

Pedroza-Martins,L., Gurney,K.B., Torbett,B.E., & Uittenbogaart, C.H. (1998). Differential tropism and replication kinetics of human immunodeficiency virus type 1 isolates in thymocytes: coreceptor expression allows viral entry, but productive infection of distinct subsets is determined at the postentry level. *Journal of Virology,72(12),* 9441-9452.

Ponath, P.d., Qin, s., Ringler, D., Clark-Lewis,I., Wang, J., Kassam, N., Smith, H., Shi, X., Gonzalo, J., Newman, W., Gutierrez-Ramos, J., & Mackay, C. (1996). Cloning of the human eosinophil chemoattractant, eotaxin expression, receptor binding, and functional properties suggest a mechanism for the selective recruitment of eosinophils. *Journal of Clinical Investigation, 97(3),* 604-612.

Prikhod'ko, E.A., & Miller, L.K. (1999). The baculovirus PE38 protein augments apoptosis induced by transactivator IE1. *Journal of Virology, 73 (8),* 6691-6699.

Prikhod'ko,G.G., Wang,Y., Freulich,E., Prives,C., & Miller,L.K. (1999). Baculovirus p33 binds human p53 and enhances p53-mediated apoptosis. *Journal of Virology, 73(2),* 1227-1234.

Quinn,L. M., Dorstyn,L., Mills,K., Colussi,P., Chen, P., Coombe,M., Abrams,J., Kumar, S., & Richardson, H., (2000). An essential role for the caspase dronc in developmentally programmed cell death in drosophila. *Journal of Biological Chemistry,* 275(51), 40416-40424.

Raich, J.W, Russell, A. E,. & Vitousek, P.M. (1997). Primary productivity and ecosystem development along an elevational gradient on Mauna Loa, Hawaii. *Ecology, 78(3)*, 707-721.

Rawlings,S.L., Crooks,G.M., Bockstoce,D, Barsky,L.W., Parkman,R., & Weinberg, K.I. (1999). Spontaneous apoptosis in lymphocytes from patients with Wiskott-Aldrich Syndrome: correlation of accelerated cell death and attenuated Bcl-2 expression. *Blood, 94 (11)*,3872-3882.

Rieseberg, L.H. & Linder, C.R. (1999). Hybrid classification: insights from genetic map-based studies of experimental hybrids. *Ecology, 80(2)*, 361-370.

Riley,J.L., Levine,B.L., Craighead,N., Francomano,T., Kim,D:, Carroll,R.G., & June, C.H. (1998). Naïve and memory CD4 T cells differ in their susceptibilities to human immunodeficiency virus type 1 infection following CD28 costimulation: implications for transmission and pathogenesis. *Journal of Virology,72 (10),*8273-8280.

Roy, N., Deveraux, Q.L., Takahashi, R., Salvesen, G.S., & Reed, J.C. (1997). The c-IAP-1 and c-IAP-2 proteins are direct inhibitors of specific caspases. *The EMBO Journal, 16(23)*, 6914-6925.

Ruefli, A., Smyth, M.J., & Johnstone, R.K. (2000). HMBA induces activation of a caspase-independent cell death pathway to overcome P-glycoprotein-mediated multidrug resistance. *Blood, 95(7)*, 2378-2385.

Sawyers, C.L., Callahan, W., & Witte, O.N. (1992). Dominant negative MYC blocks transformation by ABL oncogenes. *Cell, 70*, 901-910.

Schafer,P., Wang,L., Wadsworth,S., Davis, J.E., & Siekierka, J.(1999). T cell activation signals up-regulate p38 mitogen-activated protein kinase activity and induce TNF-production in a manner distinct from LPS activation of monocytes. *The Journal of Immunology, 162(2),*659-668.

Schwartz,K., Richards,K., & Botstein, D. (1997). BIM1 Encodes a Microtubule-binding Protein in Yeast. *Molecular Biology of the Cell, 8(12),* 2677-2691.

Shang,E.S., Champion,C., Wu,X., Skare,J.T., Blanco,D.R., Miller,J.M., & Lovett, M.A. (2000). Comparison of protection in rabbits against host-adapted and cultivated borrelia burgdorferi following infection-derived immunity or immunization with outer membrane vesicles or outer surface protein A. *Infection and Immunity, 68(79),* 4189-4199.

Sharp,D.J., Brown,H.M., Kwon,M., Rogers,G.C., Holland,G., & Scholey, J.M. (2000). Functional coordination of three mitotic motors in drosophila embryos. *MolecularBiology of the Cell, 11(1)*, 241-253.

Shaw,S.L., Yeh,E., Maddox,P., Salmon,E.D., & Bloom, K. (1997). Astral microtubule dynamics in yeast: a microtubule-based searching mechanism for spindle orientation and nuclear migration into the bud. *The Journal of Cell Biology, 139(4)*, 985-994.

Shaw-White, J.R., Bruno, M.D., & Whitsett, J.A. (1999). GATA6 activates transcription of Thyroid transcription factor-1. *The Journal of Biological Chemistry, 274(5)*, 2658-2664.

Sherman,M.P., Noronha,M.M.C de, Pearce,D.& Greene, W.C. (2000). Human Immunodeficiency Virus Type 1 Vpr Contains Two Leucine-Rich Helices That

Mediate Glucocorticoid Receptor Coactivation Independently of Its Effects on G2 Cell Cycle Arrest. *Journal of Virology, 74(17)*, 8159-8165.

Shiels, A., Mackay,D., Bassnett, S., Al-Ghoul, K., & Kuszak, J. (2000). Disruption of lens fiber cell architecture in mice expressing a chimeric AQP0-LTR protein. The *FASEB Journal, 14(14)*, 2207-2212.

Shin,I., Bakin,A.V., Rodeck,U., Brunet,A., & Arteaga, C.L. (2001). Transforming growth factor Î² enhances epithelial cell survival via Akt-dependent regulation of FKHRL1. *MolecularBiology of the Cell, 12(11)*, 3328-3339.

Simbulan-Rosenthal, C., Rosenthal, D.S., Iyer, S., Boulares, A.H., & Smulson, M.A.(1998).Transient poly(ADP-ribosyl)ation of nuclear proteins and role of poly(ADP-ribose) polymerase in the early stages of apoptosis. *Journal of Biological Chemistry, 273 (22),* 13703-13712.

Smith, C.M., Steitz, J.A. (1998). Classification of gas5 as a Multi-Small-Nucleolar-RNA(snoRNA) host gene and a member of the 5'-Terminal Oligopyrimidine Gene Family Reveals Common Features of snorRNA host genes. *Molecular and Celular Biology, 18(12)*, 6897-6909.

Smith,J.J., Evansa,E.K., Murakamib,M., Moyerc,M.B. Moseleyc,M.A., Wouded,G.V. & Kornblutha , S. (2000). Wee1-regulated apoptosis mediated by the Crk adaptor protein in xenopus egg extracts. *The Journal of Cell Biology, 151(7)*, 1391-1400.

Spector, M.P., DiRusso, C.C., Pallen, M.J., Garcia del Portillo., F., Dougan, G., Finlay, B. (1999). The médium-/long-chain fatty acyl-CoA dehydrogenase (fado) gene of Almonella typhimurium is a phase 1 starvation-stress response (SSR) locus. *Microbiology, 145*, 15-31.

Spina, C.A., Prince,H.E., & Richman, D.D. (1997). Preferential replication of HIV-1 in the CD45RO memory cell subset of primary CD4 lymphocytes in vitro. Journal of Clinical Investigation, 99 (7), 1774-1785.

Standridge, M., Alemzadeh, R., Zemel, M., Koontz, J. & Moustaid-Moussa, N. (2000). Diazoxide down-regulates leptin and lipid metabolising enzymes in adipose tissue of Zucker rats. *The FASEB Journal, 14*, 455-460.

Stapleton, R.D., Savage, D.C., Sayler, Gary, & Stacey, G. (1998). Biodegradation of aromatic hydrocarbons in an extremely acidic environment. *Applied and Environmental Microbiology, 64(11)*, 4180-4184.

Steinhauser, M. L., Hogaboam, C. M., Matsukawa, A., Lukacs, N. W., Strieter, R. M. & Kunkell, S. L. (2000). Chemokine C10 Promotes Disease Resolution and Survival in an Experimental Model of Bacterial Sepsis. *Infection and Immunity, 68 (11)*, 6108-6114.

Steinhauser, M., Hogaboam, C., Kunkel,S., Lukacs, N.W., Strieter, R.M., & Standiford, T.J. (1999). IL-10 Is a major mediator of sepsis-induced impairment in lung antibacterial host defense. *The Journal of Immunology, 162(1),* 392-399.

Steinhauser, M.L., Hogaboam, C., Lukacs, N.W., Strieter, R.M., & Kunkel, S.L. (1999). Multiple Roles for IL-12 in a Model of Acute Septic Peritonitis. *The Journal of Immunology, 162(9)*, 5437-5443.

Stellato,C., Collins, P., Ponath,P.D., Soler,D., Newman,W., La Rosa,G., Li,H., White, J., Schwiebert,L.M., Bickel,C., Liu,M., Bochner,B.S., Williams, T., & Schleimer, R.P. (1997).Production of the Novel C-C Chemokine MCP-4 by Airway Cells and Comparison of its Biological Activity to Other C-C Chemokines. *Journal of Clinical Investigation, 99(5),* 926-936.

Stewart, S.A., Poon,B., Song, Y.J & Chen, I.S.Y. (2000). Human Immunodeficiency Virus Type 1 Vpr Induces Apoptosis through Caspase Activation. *Journal of Virology, 74(7)*, 3105-3111.

Stinson,M.W., McLaughlin,R., Choi,S. H.,. Juarez,Z.E. & Barnard, J. (1998). Streptococcal histone-like protein: primary structure of hlpA and protein binding to lipoteichoic acid and epithelial cells. *Infection and Immunity, 66(1),* 259-265.

Straus, D.B. & Weiss, A. (1992). Genetic evidence fo rthe involvement of the Ick tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell, 70*, 585-593.

Sunkin,S.M., Linke,M.J., McCormack,F.X., Walzer,P.D., & Stringer, J.R. (1998). Identification of a putative precursor to the major surface glycoprotein of pneumocystis carinii. *Infection and Immunity, 66(2)*,741-746.

Sutton,R.E., Reitsma,M.J., Uchida,N., & Brown, P. (1999). Transduction of human progenitor hematopoietic stem cells by human immunodeficiency virus type 1-based vectors is cell cycle dependent. *Journal of Virology, 73(5)*, 3649-3660.

Tacket,C.O., Taylor,R.K., Losonsky,G., Lim,Y., Nataro,J.P., Kaper,J.B., & Levine, M.M. (1998). Investigation of the roles of toxin-coregulated pili and mannose-sensitive hemagglutinin pili in the pathogenesis of vibrio cholerae O139 infection. *Infection and Immunity, 66(2)*, 692-695.

Teague, T.K., Hildeman,D., Kedl,R.M., Mitchell,T., Rees, W., Schaefer, B.C., Bender, J., Kappler,J., & Marrack,P. (1999). Activation changes the spectrum but not the diversity of genes expressed by T cells. *Proceedings of the National Academy of Science, 96 (22),* 12691-12696.

Teixeira,M.M., Wells,T.N., Lukacs,N.W., Proudfoot,A.E., Kunkel, S.L., Williams,T.J., & Hellewell, P.G. (1997).Chemokine-induced eosinophil Recruitment evidence of a role for endogenous eotaxin in an in vivo allergy model in mouse skin. *Journal of Clinical Investigation, 100(7),* 1657-1666.

Thomas, C.J., Gooday,G.W., King, L.A., & Robert D. Possee (2000). Mutagenesis of the active site coding region of the Autographa californica nucleopolyhedrovirus chiA gene. *Journal of General Virology, 81(5)*, 1403 – 1411.

Thress, K., Henzell, W., Shillinglaw, W., & Kornbluth, S. (1998). Scythe: a novel reaper-binding apoptotic regulator. *The EMBO Journal, 17*, 6135-6143.

Troxell,C.L., Sweezy,M.A., West,R.R., Reed,K.D., Carson,B.D., Pidoux,A., Zacheus Cande,W. & McIntosh, J.(2001). pkl1+and klp2+: two kinesins of the kar3 subfamily in fission yeast perform different functions in both mitosis and meiosis. *Molecular Biology of the Cell*, 12(11), 3476-3488.

Tumpey,T.M., Cheng,H., Cook,D.N., Smithies,O., Oakes,J.E., & Lausch, R.N. (1998). Absence of macrophage inflammatory protein-1 prevents the development of blinding herpes stromal keratitis. *Journal of Virology,72(5)*, 3705-3710.

Unver, A., Rikihisa, Y., Ohashi, N., Cullman,L.C., Buller, R., & Storch, G.A. (1999). Western and dot blotting analyses of ehrlichia chaffeensis indirect fluorescent-antibody assay-positive and -negative human sera by using native and recombinant e. Chaffeensis and e. Canis antigens. *Journal of Clinical Microbiology, 37 (12)*, 3888-3895.

Vella,A.T., Dow,S., Potter,T.A., Kappler,J., & Marrack, P.(1998). Cytokine-induced survival of activated T cells in vitro and in vivo (interleukin 4 / interleukin 2 / interleukin 7 / interleukin 15). *Proceedings of the National Academy of Science, 95 (7),* 3810-3815.

Verhagen, A.M., Silke,J., Ekert,P., Pakusch,M., Kaufmann, H., Connolly,L.M., Day, C.L., Tikoo, A., Burke, A., Wrobel,C., Moritz, R., Simpson, R.J., & Vaux,

D.L.(2002). HtrA2 promotes cell death through its serine protease activity and its ability to antagonize inhibitor of apoptosis proteins. *Journal of Biological Chemistry, 277(1)*, 445-454.

Vucic,D., Kaiser,W.J., & Miller, L.K. (1998). Inhibitor of apoptosis proteins physically interact with and block apoptosis induced by drosophila proteins HID and GRIM. *Molecular and Cellular Biology, 18(6)*, 3300-3309.

Warny, M., Keates, A.C., Keates, S., Castagliuolo,I., & Jeff, K. (2000). p38 MAP kinase activation by Clostridium difficile toxin A mediates monocyte necrosis, IL-8 production, and enteritis. *Journal of Clinical Investigation, 105 (8),* 1147-1156.

Waskiewicz,A., Flynn,A., Proud, C.G., & Cooper,A.(1997).Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *The EMBO Journal, 16 (8),* 1909-1920.

Waters, J.C., Chen, R., Murray, A.W. & Salmon, E.D. (1998). Localization of Mad2 to Kinetochores Depends on Microtubule Attachment, Not Tension. *The Journal of Cell Biology, 141(5)*, 1181-1191.

Weighardt, H., Feterowski, C., Veit, M., Rump, M., Wagner, H., & Holzmann, B. (2000). Increased resistance against acute polymicrobial sepsis in mice challenged with immunostimulatory CPG oligodeoxynucleotides is related to an enhanced innate effector cell response. *The Journal of Immunology, 165(8)*, 4537-4543.

White, P., Harris, S. & Smith, G.C. (1995). Fox contact behaviour and rabies spread: a model for the estimation of contact probabilities between urban foxes at different population densities and its implication for rabies control in Britain. *Journal of Applied Ecology, 32*, 693-706.

Widmer, F., Shaffer, B.T., Porteous, L.A., & Seidler, R., J. (1999). Analysis of nifH gene pool complexity in soil and litter at a Douglas Fir forest site in the Oregon cascade mountain range. *Applied and Environmental Microbiology, 65(2)*, 374-380.

Wigge,P.A., Jensen,O., Holmes,S., Souès,S., Mann, M., & John V. Kilmartin (1998). Analysis of the Saccharomyces Spindle Pole by matrix-assisted laser sesorption/Ionization (MALDI) mass spectrometry. The Journal of Cell Biology, 141 (4), 967-977.

Yeh,E., Yang,C., Chin,E., Maddox,P., Salmon, E. D., Lew,D.J., & Bloom K.(2000). Dynamic Positioning of Mitotic Spindles in Yeast: Role of Microtubule Motors and Cortical Determinants. *Molecular Biology of the Cell, 11(11),* 3949-3961.

Ying, S., Meng, Q., Zeibecoglou,K., Robinson,D.S., Macfarlane,A., Humbert,M., & Kay, A.B. (1999). Eosinophil chemotactic chemokines (eotaxin, eotaxin-2, RANTES, monocyte chemoattractant protein-3 (MCP-3), and MCP-4), and C-C chemokine receptor 3 expression in bronchial biopsies from atopic and nonatopic (Intrinsic) asthmatics. *The Journal of Immunology, 163(11),* 6321-6329.

Yu, X., Crocquet-Valdes, Cullman, L.C., V.L., Popov, & Walker, D.H. (1999). Comparison of ehrlichia chaffeensis recombinant proteins for serologic diagnosis of human monocytotropic ehrlichiosis. *Journal of Clinical Microbiology, 37 (8),* 2568-2575.

Zheng, R., Jenkins, T.M., & Craigie, R. (1996). Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings of the National Academy of Science, 93 (24),* 13659-13664.

Zhou, Y., Kurihara, T., Ryseck, R., Yang, Y., Ryan, C., Loy,J., Warr, G., & Bravo, R. (1998).Impaired macrophage function and enhanced T cell-dependent immune response in mice lacking CCR5, the mouse homologue of the major HIV-1 coreceptor. *The Journal of Immunology,160(8),*4018-4025.

APPENDIX B

CORPUS OF LEARNER ENGLISH MANUSCRIPTS FOR PUBLICATION PURPOSES

Agostini, E., S. Milrad de Forchetti, & H. A. Tigier (2000). Peroxidases from cell suspension cultures of *Brassica napus*. *Biocell, 24* (2), 133-138. (TEXT 014)

Alba, C. M., S. Milrad de Forchetti, M. A. Quesada, V. Valpuesta, & H. A. Tigier, (1998). Localization and general properties of developing peach seed coat and endosperm peroxidase isoenzymes. *Plant Growth Regulation, 17*, 7-11. (TEXT 03)

Andrade, A., Vigliocco, A., Alemano, S., Miersch, O. Botella, M. & Abdala G. (2005). Endogenous jasmonates and octadecanoids in hypersensitive tomato mutants during germination and seedling development in response to abiotic stress. *Seed Science Research, 15*, 309–318. (TEXT 029)

Barros, G., Chiotta, M.L., Torres, A., & Chulze, S. (2006). Genetic diversity in *Aspergillus parasiticus* population from the peanut agroecosystem in Argentina. *Letters in Applied Microbiology, 42* (6), 560-566. (TEXT 026)

Bluma, R & Etcheverry, M. (2006). Influence of *Bacillus spp* isolated from maize agroecosystem on grown and aflatoxin B_1 production by *Aspergillus* section *Flavi. Pest Management Science, 62, 242-251.* (TEXT 027)

Bollo, M., Bonansea, S. & Machado, E. (2006). Involvement of Na+/H+ exchanger in the calcium signaling in epimastigotes of Trypanosoma cruzi. *FEBS Letters , 580*, 2686–2690. (TEXT 033)

Bonvillani, A.G., Di Renzo, M.A. & Tiranti, I.N. (2000). Genetic polymorphism of milk protein loci in Argentinian Holstein cattle. *Genetics and Molecular Biology, 23*, 819-823. (TEXT 017)

Cenzano,A., Vigliocco,A., Kraus, T. & Abdala, G. (2003). Exogenously applied jasmonic acid induces changes in apical meristem morphology of potato stolons. *Annals of Botany, 91*, 915-919. (TEXT 019)

Criado, S., A. Soltermann & N. Garcia (1998). Sensitized photooxidation of Di-and Tri-Peptides of Tyrosine. *Photochemistry and Photobiology, 68* (4), 453-458. (TEXT 06)

Di Renzo, M. A., M. A. Ibáñez, N. C. Bonamico, & M. M. Poverene (2000). Estimation of repeatability and phenotypic correlations in *Eragrostis curvula*. *Journal of Agricultural Science, 134,* 207-212. (TEXT 05)

Díaz, M., Luiz, M., Bertolotti, S., Miskoski, S. & García Norman. (2004) Scavenging of photogenerated singlet molecular oxygen and superoxide radical anion by sulpha drugs — Kinetics and mechanism. *Canadian Journal of Chemistry, 82*, 1752–1759. (TEXT 031)

Giordano, W., & c. E. Domenech (1999). Aeration affects acetate destination in *Gibberella fujikuroi*. *FEMS Microbiology Letters, 180*, 111-116. (TEXT 011)

Ibáñez, M. A., M. M. Poverene, C. V. Echenique, D. I. Puecher, & M. A. Di Renzo (1998). Response and stability in lovegrass cultivars (*Eragrostis curvula* (Schrad.) Nees) under controlled environmental conditions. *Seed Science and Technology*, 26, 97-106. (TEXT 04)

Luiz, M., A. Biasutti, A. T. Soltermann, & N. A. Garcia (1999). Interaction of UV photostabilizers with singlet molecular oxygen([$O_2(1\Delta_g)$]).The case of 1′-hydroxy-2′-acetonaphthone. *Polymer Degradation and Stability,63*, 447-453. (TEXT 02)

Marchesini, N., Santander, V., & Doménech, E. M. (1998). Diacyglycerol pyrophosphate: a novel metabolite in the *Trypanosoma Cruzi* phosphatidic acid metabolism. *FEBS Letters, 436*, 377-381. (TEXT 07)

Martori, R., Aun, L., Gallego, F. & Rozzi Giménez, C. (2005) Temporal variation and size class distribution in a herpetological assemblage form Córdoba, Argentina. *Cuad. Herpetol. 19 (1): 35-52* (TEXT 015)

Massad, W. Bertolotti, S. and N. García (2004) Kinetics and mechanism of the Vitamin B_2-sensitised-photooxidation of Isoproterenol *Photochemistry and Photobiology, 79(5),* 428-433. (TEXT 08)

Massimelli, M. J., Beassoni, P.R., Forrellad, M.A., Barra J.L., Garrido, M. N., Doménech C.E., & Lisa, A. T. (2005). Identification, Cloning, and Expression of Pseudomonas aeruginosa Phosphorylcholine Phosphatase Gene. *Current Microbiology, 50*, 251-256. (TEXT 018))

Milrad de Forchetti, S. & H. A. Tigier (1990). Indole-3-acetic acid oxidase and syringaldazine oxidase activities of eroxidase isozymes in soybean root nodules. Physiologia Plantarum 78, 1-14. (TEXT 01)

Montaña,M.P., Miskoski,S., Pappano,N., Debattista, N., & García, N.A. (2003). Vitamin B2-sensitized photooxidation of structurally related dihydroxyflavonoids. *Dyes and Pigments, 58*, 113-120. (TEXT 016)

Pedranzani, H., Racagni, G., Alemano,S., Miersch,O., Ramírez,I., Peña-Cortés, H., Taleisnik,E., Machado-Domenech, E., & Abdala,G. (2003). Salt tolerant tomato plants show increased levels of jasmonic acid. *Plant Growth Regulation, 41*, 149-158. (TEXT 021)

Posadaz, A., E. Sánchez, M. I. Gutiérrez, M. Calderón,. S. Bertolotti, M. A Biasutti & N.A. García, (2000). Riboflavin and rose bengal sensitised photooxidation of sulfathiazole and succinylsulfathiazole Kinetic study and microbiological implications. *Dyes and Pigments, 45*, 219-228. (TEXT 013).

Principe, R. E. & Corigliano, M. (2006). Benthic, drifting and marginal macroinvertebrate assemblages in a lowland river: temporal and spatial variations and size structure. *Hidrobiología, 553,* 303–317. (TEXT 028)

Ramirez, M.L., Reynoso, M.M., Farnochi, M.C., & Chulze, S. (2006). Vegetative compatibility and mycotoxin chemoptypes among *Fusarium graminearum* (*Gibberella zeae*) isolates from wheat in Argentina. *European Journal of Plant Pathology, 115(*(2), 139-148. (TEXT 032)

Reinoso, H., Sosa, L., Ramírez, L, and V.Luna (2004). Salt-induced changes in the vegetative anatomy *Prosopis strombulifera* (Leguminosae). *Canadian Journal of Botany, 82*, 618-628. (TEXT 022)

Rojo, F. G., Reynoso, M.M, Ferez, M., Chulze, S. N, & Torres, A.M (2006). Biological control by Trichoderma species of Fusarium solani causing peanut brown root rot under field conditions. Crop protection, 26, (4), 549-555. (TEXT 025)

Salvano, M. A., & C. E. Domenech (1999). Kinetic properties of purified *pseudomonas aeruginosa* phosphorylcholine phosphatase indicated that this enzyme may be utilized by the bacteria to colonize in different environments. *Current Microbiology, 39*, 1-8. (TEXT 09)

Sosa, L., Llanes, A. Reynoso, H. Reginato, M and V. Luna (2005) Osmotic and Specific Ion Effects on the Germination of Prosopis strombulifera. *Annals of Botany, 96*, 261-267. (TEXT 010)

Tordable, M. del C., E. Lorenzo, & M. E. Doucet (2003). Histopathology of Asgrow 5435 RG Soybean Roots Induced by *Heterodera glycines* Race 1, in Córdoba, Argentina. *Nematologia Brasileira, 27* (1), 55-60. (TEXT 012)

Tripole, E. S. & Corigliano, M. del C. (2005). Acid Stress Evaluation Using Multimetric Indices In The Carolina Stream (San Luis –Argentina). Acta Limnol. Bras. *Acta limnologica brasiliensia , 17(1)*, 101-114. (TEXT 023)

Vigliocco, A., Bonamico, B., Alemano, S., Miersch, O. & Abdala, G. (2002). Stimulation of jasmonic acid production in *Zea mays L.* infected by the maize rough dwarf virus- Rio Cuarto. Reversion of symptoms by saliculic acid. *Biocell, 26(3),* 369-374. (TEXT 020)

Villasuso, A. L. & Machado-Domenech, E. (2004). Gibberellin signaling pathway: role of phospholipids in barley aleurone. *Plant Growth Regulation, 43*, 101–108. (TEXT 024)

Wevar Oller, A.L., Agostini, E., Talano, M. A., Capozucca, C., Milrad, S.R., Tigier, H.A., & Medina, M.I. (2005). Overexpression of a basic peroxidase in transgenic tomato (Lycopersicon esculentum Mill. cv. Pera) hairy roots increases phytoremediation of phenol. *Plant Science, 169,* 1102–1111. . (TEXT 030)

APPENDIX C

Table C.1: Stance adverbs used according to categories

	Common to RC and CLEMPP	RC only	CLEMPP only
EPISTEMIC	1. actually 2. apparently 3. approximately 4. certainly 5. clearly 6. essentially 7. evidently 8. generally 9. largely 10. likely 11. mainly 12. obviously 13. possibly 14. presumably 15. primarily 16. probably 17. really 18. simply 19. strictly 20. typically 21. undoubtedly 22. usually	23. admittedly 24. arguably 25. broadly 26. conceivably 27. decidedly 28. formally 29. fundamentally 30. hypothetically 31. ideally 32. nominally 33. ostensibly 34. reportedly 35. roughly 36. seemingly 37. superficially 38. surely 39. theoretically 40. truly	1. basically 2. definitely
ATTITUDE	1. appropriately 2. curiously 3. fortunately 4. importantly 5. interestingly 6. remarkably 7. unexpectedly	1. correctly 2. incorrectly 3. inevitably 4. reasonably 5. surprisingly 6. unfortunately	
STYLE	1. briefly	1. accordingly	

REFERENCES

Adams Smith, D. (1984). Medical discourse: Aspects of author's comment. *The ESP Journal, 3,* 25-36.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Ed.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Barlow, M. (2005). Computer-based analyses of learner language. In R. Ellis & G. Barkhuizen (Eds.), *Analysing Learner Language*. (pp. 335-357). Oxford: Oxford University Press.

Bazerman, C. (1988). *Shaping written knowledge*. London: University of Wisconsin Press.

Beaugrande, R. de (2001). Large corpora, small corpora. And the learning of "language". In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. 3-28). Amsterdam: John Benjamins B.V.

Bhatia, V. (1993). *Analysing genre: Language in use in professional settings.* London: Longman.

Bhatia, V. (2001). Analysing genre: Some conceptual issues. In M. Hewings (Ed.), *Academic writing in context: Implications and applications* (pp. 79-92). Birmingham, UK: University of Birmingham Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes, 11*, 1-34.

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

Bloor, T., & Bloor, M. (1995). *The functional analysis of English. A Hallidayan Approach*. London: Edward Arnold.

Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing, 12*, 223–243.

Channell, J. (2000). Corpus-based analysis of evaluative lexis. In S. Hunston, & G. Thompson (Eds.), *Evaluation in text: authorial stance and the construction of discourse* (pp.38-55). Oxford: Oxford University Press.

Charles, M. (2003). "This mystery…": a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2,* 313-326.

Chen, Q. & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26,* 502–514.

Coates, R. Strurgeon, B., & Bohannan, J. (2002). Language and publication in cardiovascular research articles. *Cardiovascular Research, 53*, 279-285.

Collins Wordbanks*Online* English http://www.collins.co.uk/corpus/CorpusSearch.aspx)

Connor, U., Precht, K., & T. Upton (2002). Business English: Learner data from Belgium, Finland and the U.S. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 175-194). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System, 27*, 1-18.

Conrad, S. & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston, & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56-73). Oxford: Oxford University Press.

Coxhead, A. (2000). A new Academic Word List. *Tesol Quarterly, 34*, 213-238.

Dudley-Evans, T. (1994). Genre analysis: An approach to text analysis for ESP. In M. Coulthard (Ed.), *Advances in Written Text Analysis*. London: Routledge, 219-228.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19(1),* 61-74.

Ellis, R. (1994). *The study of second language acquisition.* Hong Kong: Oxford University Press.

Ellis, R. & Barkhuizen, G. (2005). *Analysing Learner Language.* Oxford: Oxford University Press.

Flowerdew, J. (2002). *Academic Discourse.* New York: Longman.

Flowerdew, L. (2001). The Exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.) *Small Corpus Studies and ELT* (pp. 363-379). Amsterdam: John Benjamins Publishing Co.

Gledhill, C.J. (2001). Knowledge & Discourse Conference Website: *The phraseology of rhetoric, collocations and discourse in cancer research abstract.* Retrieved on April 14, 2001 from http://www.st-and.ac.uk/~cjg6/.

Gosden H. (1992). Discourse functions of marked theme in scientific research articles. *English for Specific Purposes, 2,* 207- 224.

Granger, S. (1993). Cognates: an aid or a barrier to successful L2 vocabulary development? *ITL Review of Applied Linguistics, 99-100,* 43-56.

Granger, S. (1996). Learner English around the World. In S. Greenbaum (Ed), *Comparing English World-wide* (pp. 13-24). Oxford: Clarendon Press.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granger S. (2004). Practical Applications of Learner Corpora. [Abstract]. In Lewandowska-Tomaszczyk B. (ed.) *Practical Applications in Language and Computers (PALC 2003).* Frankfurt: Peter Lang.

Granger, S., Meunier, F., & Tyson, S. (1994). New insights into the learner lexicon: A preliminary report from the International Corpus of Learner English. In L. Flowerdew & K. K. Tong (Eds.), *Entering Text* (pp. 102-113). Hong Kong: The Hong Kong University of Science and Technology.

Granger, S. & Swallow, H. (1988). False friends: a kaleidoscope of translation difficulties. *Langage et l'Home, 23(2)*, 108-120.

Granger, S. & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15(1)*, 17-27.

Halliday, M.A.K (1993). On the language of Physical Science. In M.A.K. Halliday and J. R. Martin, *Writing research. Literacy and discursive power* (pp. 54-68). London: The Palmer Press

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Hiltunen, T. (2006). Coming-to-know verbs in research articles in three academic disciplines. In Perez-Llantada Auría, María Carmen, Ramón Plo Alastrué & Neuman, C. P. (Eds.) Proceedings of the 5th International AELFE Conference. Zaragoza: Prensas Universitarias de Zaragoza. Retrieved March 25, 2007, from http://www.unizar.es/aelfe2006/ALEFE06/1.discourse/35..pdf

Hunston, S. (1994). Evaluation and organization in a sample of written academic discourse. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 191-218). London: Routledge.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston S., & Francis, G. (2000). *Pattern Grammar. A corpus-driven approach to the lexical grammar of English.* Amsterdam/Philadelphia: John Benjamins B. V.

Hunston, S., Francis, G., & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal, 51(3),* 208-216.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes, 13(3),* 239-256.

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In H. Candlin & K. Hyland (Eds.), *Writing: Texts, Processes and Practices* (pp.99-121). London: Longman.

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

Hyland, K (2002). Options of identity in academic writing. *ELT Journal, 56 (4)*, 351-358.

Hyland, K. (2004). Perspectives on genre. In K. Hyland (2004). *Genre and second language writing* (pp. 24-50). Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2006). *English for academic purposes. An advanced resource book.* New York: Routledge.

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6 (2),* 183-206.

Hyland, K. & Tse, P. (2005). Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes, 24,* 123-139.

Hyland, K., & Tse, P. (2007). Is there an Academic Vocabulary?. *Tesol Quarterly, 41(2)*, p. 235-253.

Hyon, S. (1996). Genre in three traditions. *TESOL Quartely, 30(4)*, 693-719.

Johns, A. (1997). Genre knowledge and socioliteracies: What readers and writers may share. In A. Johns (Ed.), *Text, role and context. Developing academic literacies (pp.20-37).* Cambridge: Cambridge University Press.

Johns, A. M. (2002). Introduction: Genre in the Classroom. In A. M. Johns (Ed.), *Genres in the Classroom. Multiple Perspectives* (pp. 3-13). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Journal Citation Report. (1995). Institute of Scientific Information. Philadelphia: Thomson Reuters.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24,* 269-292.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt, & M. McCarthy

(Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 140-180). Cambridge: Cambridge University Press.

Luzon Marco, M.J. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes, 19,* 63-86.

Martin, J. R. (1992). *English text. System and structure.* Philadelphia: John Benjamins B.V.

Martin, J. R., & Rose, D. (2003). *Working with Discourse. Meaning beyond the clause.* London: Continuum.

Martín-Martín, P. & Burgess, S. (2004). The rhetorical management of academic criticism in research article abstracts. *Text, 24(2),* 171-195.

Martínez, I. (2000). Discrepancias entre las concepciones de hispanoparlantes del artículo de investigación en inglés y las características del mismo según el análisis del género. *Actas del II Congreso Iberoamericano de Educación en Ciencias Experimentales* [CD-Rom publication]. Córdoba, Argentina.

Martinez, I. (2002). Developing Genre Awareness in Nonnative-English-Speaking Writers of Experimental Research Articles: A Collaborative Approach. In J. Crandall, & D. Kaufman (Eds.), *Content-Based Instruction in Higher Education Settings* (pp. 79-92). Virginia: TESOL.

Martínez, I. (2003). Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes, 2,* 103-123.

Martínez, I. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing, 14,* 174-190.

Martínez, I., Beck, S. & Panza, C. (2008). *Academic vocabulary of research articles of the agricultural sciences: a corpus-based study.* Manuscript submitted for publication.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 84-102). Cambridge: Cambridge University Press.

Meunier, F (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics, 10 (1),* 1-35.

Myers, G. (1990). The Social Construction of Popular Science: The Narrative of Science and the Narrative of Nature. In G. Myers (Ed.), *Writing Biology: texts in the construction of scientific knowledge* (pp. 141-192). Wisconsin: The University of Wisconsin Press.

Nation, P. (2001). Using small corpora to investigate learner needs. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. 31-45). Amsterdam: John Benjamins B.V.

Ooi, V. B.Y. (2001). Investigating and teaching genres using the World Wide Web. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. 175-203). Amsterdam: John Benjamins B.V.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Grammar of Contemporary English.* London: Longman Group Limited.

Ryan, A. (1997). Learning the orthographical form of L2 vocabulary – a receptive and a productive process. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 181-227). Cambridge: Cambridge University Press.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13(2),* 149-170.

Schmitt, R. (1997). Vocabulary learning strategies. In Schmitt, N. & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Schmitt, N. & McCarthy, M. (1997). Glossary. In Schmitt, N. & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 327-331). Cambridge: Cambridge University Press.

Scott, M. (2004). *WordSmith Tools (Version 4)* [Computer program]. Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual Patterns. Key words and corpus analysis in language education*. Amsterdam: John Benjamins B.V.

Sharwood Smith, M. (1994). *Second language learning: Theoretical foundations.* London: Addison Wesley Longman Ltd.

Silver, M. (2003). The stance of stance: a critical look at ways stance is expressed and modelled in academic discourse. *Journal of English for Academic Purposes, 2*, 359-374.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sinclair, J. (2005). Corpus and Text - Basic Principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*. Oxford: Oxbow Books: 1-16. Retrieved November 28, 2007, from http://ahds.ac.uk/linguistic-corpora/

Sinclair, J.M. (2001) Preface. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. vii-xv). Amsterdam: John Benjamins B.V.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes, 2,* 327-341.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Glasgow: Cambridge University Press.

Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 156- 180). Cambridge: Cambridge University Press.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive in two astrophysics journal papers. With extensions to other languages and other fields. *English for Specific Purposes, 17(1)*, 113-132.

Thompson, D. (1993). Arguing for experimental "facts" in science. A study of research article results sections in Biochemistry. *Written Communication, 10(1)*, 106-128.

Thompson, G. & Hunston S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text. Authorial Stance and the Construction of Discourse* (pp. 1-27). New York: Oxford University Press.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: John Benjamins Publishing Co.

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes, 20(1)*, 83-102.

Wang, S.-p. (2005). Corpus-based approaches and discourse analysis in relation to reduplication and repetition. *Journal of Pragmatics, 37,* 505-540.

Widowson, H. G. (1996). *Linguistics.* Hong Kong: Oxford University Press.