RiCL Research in Corpus Linguistics

Reflexive metadiscourse in a corpus of Spanish bachelor dissertations in EFL

Noelia Navarro Gil
Universidad Complutense de Madrid · Universitat Internacional de Catalunya / Spain

Abstract – Academic English has often been described as a reader-oriented discourse, in which the structure, objectives and claims are made explicit and carefully framed. Metadiscourse markers help to build coherence and cohesion, and allow writers to guide their readership through their texts. Spanish EFL learners often transfer part of their L1 writing culture into their L2 texts. This is problematic because academic Spanish tends to show a slightly more reader-responsible style, and academic texts call for a high degree of disciplinarity: learners not only have to be aware of the conventions of the L2 regarding metadiscourse, but also of their own discipline. This article explores the use of reflexive metadiscourse in a learner corpus of bachelor dissertations written in English by Spanish undergraduates in medicine and linguistics, and compares the results with an expert corpus of research articles. The results show that overall both corpora contain similar frequencies of textual metadiscourse, but this is only true when we look at the results according to discipline. In spite of this quantitative similarity, there are cases of overuse and underuse in the learner corpus that highlight features of the bachelor dissertations genre, on the one hand, and EFL Spanish writing, on the other hand.

Keywords – reflexive metadiscourse, learner corpus, research articles, writer-reader interaction, disciplinary discourse, reader-oriented texts

1. INTRODUCTION

Metadiscourse (MD) is an umbrella term used in discourse analysis to describe a range of linguistic elements that, deliberately used by the writer, helps readers to navigate successfully through a text. For example, *our study*, *see Table 3* or *in other words* signal authorial involvement, an awareness of the reader and an awareness of the evolving text, respectively. Reader-oriented texts, i.e. those that contain metadiscursive markers to help readers to "organise, classify, interpret, evaluate and react" to the ideas presented (Vande Kopple 1985: 83) are found to be more convincing, comprehensible and more likely to be remembered (Crismore and Vande Kopple 1997). In this article I employ a reflexive model of MD (Mauranen 1993; Ädel 2006) to study MD features in the academic writing in English of Spanish undergraduate students of medicine and linguistics.

Hyland (2008: 548) points out that, "compared with many languages, academic texts in English tend to be more explicit about structure and purposes, to be less tolerant of digressions, to be more cautious in making claims, and to use more sentence connectors". For learners of English as a foreign language (EFL), as well as for novice writers, MD markers that help to achieve some of the writing goals mentioned above may be difficult to acquire and may even go unnoticed when reading a text (Low 1996; Hyland 2010). Learners' academic written production that lacks metadiscoursal devices can come across as too direct, digressive and sometimes unconvincing (Hinds 1987; Montaño-Harmon 1991). In contrast, an appropriate use of MD markers is often related to text quality, enhanced readability and even higher grades (Intaraprawat and Steffensen 1995; Cheng and Steffensen 1996; Hyland 1998; Dafouz 2003; Noble 2010;

Lee and Deakin 2016). Becoming acquainted with the many forms and functions MD markers can have is therefore of paramount importance for academic language writers.

As in most European universities, undergraduate students in Spain are required to write a bachelor dissertation (BD) at the end of their studies, and many of them do it partly (e.g. abstract, viva) or entirely in English. BDs are a major piece of scholarly work that allows students to adopt a scientific approach to explore a topic in depth and present it to experts in the field (i.e. a supervisor and the examining committee), and it is the academic project that most resembles a research paper. Academic writing courses, textbooks or style guides are sometimes provided to guide learners through this writing process. Many have argued, however, that these often take a 'one-size-fits-all' approach and group all needs, failing to notice differences across disciplines (Hyland 2008; Springer 2012). Moreover, some textbooks provide conflicting advice about the extent to which writers can intrude into their texts (Hyland 2001, 2002), and since MD tends to be considered secondary to the main objective, i.e. presenting information, little instruction on MD is provided (Martín-Laguna and Alcón 2015). EFL learners often fail to use sufficient metadiscursive markers, and may not be aware of the contribution these elements make to the full understanding of the text, or the differences between their L1 and L2 disciplinary discourses (Hyland 2000, 2005, 2012). To date, there are few studies that explore reflexive MD in Spanish EFL academic writing across disciplines (see Pérez-Llantada 2010; Mur Dueñas 2011). Corpus-based and corpus-driven studies that explore the MD dimension in EFL texts produced by Spanish undergraduate students are therefore needed.

The present study seeks to analyze the frequency and types of reflexive MD markers in a learner corpus of EFL Spanish undergraduates' BDs with the intention of highlighting rhetorical conventions of this genre and L2 writers' linguistic features regarding the use of MD. The corpus has been analyzed from a discipline variable, exploring BDs in medicine and linguistics, and also from a writer status variable, comparing the learner corpus with an expert corpus of published research articles (RAs) in the same discipline. The results of the present study will shed light on the use of reflexive MD in EFL academic writing, and stress the importance of teaching MD to L2 writers taking into account their specific discipline. This study also presents pedagogical implications, relevant for academic writing teachers who wish to equip their students with genre-sensitive metalinguistic devices. Finally, the present article provides a systematic basis for the analysis of reflexive MD markers in BDs and RAs, useful to design pedagogical material on MD that is corpus-informed and genre-sensitive.

2. A VIEW ON THE TRAJECTORY OF METADISCOURSE

In applied linguistics, the term 'metadiscourse' was first coined by Harris in 1970, but the concept gained traction with Williams' (1981) work, who defined it as "discourse about discourse" (1981: 47) or "writing about writing, [...] whatever does not refer to the subject matter being addressed" (1981: 212). Since its conception, there has been a distinction between *meta*discourse and *primary* discourse (i.e. propositional content) (Sinclair 1981; Williams 1981; Vande Kopple 1985; Crismore 1989). As aptly described by Toumi (2009: 66):

> [Metadiscourse] marks the writer's awareness of the current text as text or as language, of him/herself as writer, and of the potential reader as reader of this text. Metadiscourse supports propositional content, but remains separate from it. It is the means by which propositional content is made coherent, legible and persuasive to the reader in accordance with the writer's intentions.

In writing, metadiscursive elements can make reference to three dimensions: the evolving text (e.g. *in figure 1*, *secondly*, *as mentioned previously*), the writer of the text (e.g. as *I* said, *we* found, *our* study), and/or the imagined reader (e.g. *see* appendix 1, *you* may question, *we* will see how); these categories are not exclusive and markers can refer to one or more of these dimensions at the same time (Toumi 2009). In some cases, the second and third categories (i.e. writer and reader) are merged into one category only, called 'interpersonal' (Halliday 1973; Dafouz 2003; Ädel 2006; Toumi 2009; Bondi 2010; Mauranen 2010) or 'interactional' (Hyland 2017).

Since the early days of MD (Williams 1981; Vande Kopple 1985; Crismore 1989; Crismore and Farnsworth 1990; see Toumi 2009 for a comprehensive review) three differences have been made: metadiscourse from ideational content, textual from interpersonal elements and reflexive from attitudinal MD. The first difference, as mentioned earlier, has been the starting point of the discipline: distinguishing 'metadiscursive' elements from the 'ideational' content of the text. The characters in bold in (1) illustrate this difference:

(1) This can be accounted for two different principles**:** a weak one **– *also known as*** linguistic relativity **–** and a strong one (...) [from the learner corpus]

Also known as and two punctuation marks: colon and dashes in (1) do not add content but help the writer to give an explanation of the two principles (colon), add a commentary or aside (dashes) and provide a different term, perhaps a more scientific one, for one of the principles (*also known as*). Even though differentiating MD markers from content may seem an easy task to perform, in some cases there is no such a clear distinction. Consider, for example, the use of the deictic marker *Here* in (2):

(2) The other would be represented by a case in which commodity prices fall by the full extent of the degree of cost-cutting involved in technological progress. *Here* the effect on real wage rate is very simple to analyze. [from Toumi (2009: 70)]

It is ambiguous if the deictic marker *Here* refers to the current text (e.g. in this study), which would qualify as MD, or to the content (e.g. in that context or situation), in which case it could not be coded as MD. Examples like this make the nature of MD itself difficult to delimit and, as frequently described in the literature, fuzzy (Hyland 2017). One of the main contributions of Ädel (2006) and Mauranen (1993) is a set of criteria to help identify reflexive MD. This set of criteria has been taken into account in the present analysis and will be described in Section 4.

The second difference, 'textual' and 'interpersonal', gave birth to what have later been called 'broad' and 'narrow' approaches (Mauranen 1993; Ädel 2006; Toumi 2009). A broad approach to MD explores and includes both textual (e.g. *in section 1*, *in other words*, *in contrast*) and interpersonal (e.g. *we* can see, *our* study, *note* that) categories A narrow approach, in contrast, will focus on textual categories only (Mauranen 1993; Dahl 2004). However, this distinction has also been a source of disagreement. Some rhetoricians claim that, since all MD elements in some way or another take the reader into account – be it textual or interpersonal, the limits between the interpersonal and textual categories are also fuzzy (Hyland and Tse 2004) and propose a broader and more inclusive interpersonal perspective of MD called 'interactive', whose main representative is Ken Hyland (2017: 20). In this regard, Mauranen (1993) and Ädel (2006) distinguish two models of MD: the 'reflexive' model, also known as 'non-integrative' (Ädel 2006; Ädel and Mauranen 2010), and the 'non-reflexive' or 'integrative' model. These models are an attempt to bridge the gap between textual and interpersonal MD: markers to refer to the text, the writer and the reader are included in both of these models; this conceptualization also helps restrict the fuzzy notion of MD (Ädel 2010) and shares the idea that the main rhetorical strategy of MD is that of achieving persuasiveness. As Dafouz (2003: 32–33) aptly puts it, "metadiscourse categories, both textual and interpersonal, ultimately intend to convince readers of the validity of the arguments presented in the text [...] it is the perfect combination of these two elements that makes a text persuasive".

The third and last difference, 'reflexive' and 'attitudinal' MD, is what separates the two models mentioned above: the 'interactive' approach includes the category of 'stance' as a unit of analysis, i.e. markers that show the writer's attitude, express certainty (such as boosters) or doubt (hedges) (e.g. *fortunately*, *clearly*, *might*). The 'reflexive' approach, on the other hand, excludes stance and focuses on the reflexive aspect of language, i.e. items used exclusively to refer to the finite world of the evolving text; stance is a non-reflexive feature of language because it reflects the state of mind of the writer, as an experiencer of the real world (Toumi 2009). However, a tendency of reflexivity and stance to co-occur in academic writing has been described in the literature (Dafouz 2003; Mauranen 2010) and often labeled as 'discourse collocations' (Mauranen 2010) (e.g. *our paper* has *clearly* shown). A view that defends a reflexive approach to MD comes from Mauranen (2010: 37), who argues that "if we opt for a very broad, embracing notion of metadiscourse [e.g. including stance, hedges or boosters], we risk losing sight of its collocability and interaction with other discourse phenomena". The reflexive model adopted in this study afforded the researcher a narrower approach to MD which, together with the text-internal criterion, facilitated the identification and selection of MD markers in the corpora. The taxonomy of reflexive MD used, together with the identification and tagging system will be described in Section 3.

These different approaches, broad and narrow, and interactive and reflexive, not only differ in the categories they explore, but also in the methodology they apply. There are two types of methodology that are often used in MD research, namely 'thin' and 'thick' (Bondi 2010). The first one is a corpus-based approach that consists in predefining a list of terms to be analyzed (e.g. comparing the frequency and types of MD markers between two corpora). It allows for cross-linguistic, cross-disciplinary and cross-generic comparisons of large corpora. The downside of the 'thin' method is that potentially metadiscursive items present in the texts but not included on the list will never be found (e.g. Vande Kopple 1985; Crismore et al. 1993; Hyland 2005). The 'thick' approach, on the contrary, relies on a corpus-driven methodology. In this contextualized form of analysis, the elements explored are based on and set by the data (i.e. no predefined list of terms). The main difference is that, as the analysis is mostly done manually, that is,

discovering and tagging markers actually present in the data, the units of analysis are often smaller (e.g. one category of MD markers such as 'self-mentions') than in the'thin' method (e.g. Mauranen 1993; Ädel 2006; Bondi 2010; Pérez-Llantada 2010). I have adopted a mixed-method approach (i.e. 'thin' and 'thick') by which each reflexive MD marker, belonging to a predefined set of categories (e.g. 'endophoric markers'), actually present in the texts has been manually tagged, to later calculate frequency counts for all the elements found.

Due to the fact that the quantity of elements that qualify as MD varies from one model to the other (e.g. hedges and boosters would be included in the analysis of 'interactive' MD, but excluded in a 'reflexive' approach to MD), the estimates about average proportion and range of MD markers in a given genre and discipline vary greatly in the literature: for example, following an interactional model, Hyland (2005) reported that 1 every 15 words in RAs was metadiscursive (an average of 370 occurrences per paper) and 1 every 21 words in postgraduate dissertations (Hyland 2010); Pérez-Llantada (2010), in contrast, reported that the quantity of reflexive metadiscourse represents a very low proportion compared to ideational content.

3. TAXONOMY OF REFLEXIVE MD MARKERS

The current study follows a reflexive model of MD drawing on Mauranen (1993) and Ädel (2006). Previous taxonomies have been taken into account as a point of departure but some adjustments have been made in order to render the proposed taxonomy more applicable for the RA and the BD genres. I have explored metatextual (MT) and interpersonal (IP) markers in both the learner and the expert corpus across disciplines (linguistics and medicine). These categories were manually analyzed and tagged in the texts as shown in Table 1:

CATEGORY	SUBCATEGORY	EXAMPLE	TAG
Metatext			_MD_MT_
References to the text			_RT_
	Full text	*this study* / *the current paper* / *our article*	_FT
	Part of the text	*this section* / *Appendix A* / *in this chapter*	_PT
	Semiotic modes	*Table 1* / *this diagram* / *Fig.1*	_SM
Endophoric markers			_EN_
	Anaphoric	*aforementioned* / *as previously discussed* / *as noted above*	_AN
	Cataphoric	*the following* / *as follows* / *next paragraph*	_CA
	Deictic	*here/ now/ so far*	_DE
Code glosses			_CG_
	Reformulators	*i.e.* / *that is,* / *in other words*	_RE
	Exemplifiers	*e.g.* / *for instance* / *such as*	_EX
	Parentheticals	*(inaccurate) translations* / *in a degenerative (vs. naïve) environment*	_PA
	Colons	*in the data:* / *three reasons:*	_CL
	Semicolons	*Pandora's box; hence* / *FI hours; however*	_SC
	Dashes	*categorical difference –i.e., between writer and the audience* / *paradigm of three pillars -- scaffolds, cells, signals* --	_DA
Linking Devices			_LD
	Additive	*in addition* / *also* / *furthermore*	_AD
	Contrastive	*however* / *in contrast* / *nevertheless*	_CN
	Consecutive	*therefore* / *as a result* / *thus*	_CO
	Organizers[1]	*firstly* / *second* / *third*	_OR
	Topicalizers	*regarding* / *as for* / *with respect to*	_TO

[1] In order to qualify as MD, these elements must function text-internally (i.e. signal transition in the world of discourse) and not text-externally (refer to real processes: e.g. *second*, we added the solution, and *then*, we removed the lid) (Ädel 2006; Mauranen 1993).

CATEGORY	SUBCATEGORY	EXAMPLE	TAG
Interpersonal			MD_IP
Writer oriented			-WO
	Self-mention	*I* / *our* / *(exclusive) we* / *the researcher* / *the author*	_SF
Reader oriented			_RO
	Directives	*see* / *consider* / *cf.*	_DI
	Rhetorical questions[2]	*if L2 proficiency alone cannot account for the incorrect meaning components, what are other possible explanations?*	_RQ
Participant oriented			_PO
	Inclusive we[3]	*Let's have a look* / *as we can see* / *if we take*	_IW

Table 1: Reflexive Metadiscourse: categories, subcategories, examples and tags

Examples of how MD markers were tagged are given in (3) and (4), in which the code MED (short for medicine) or LIN (linguistics) indicates the discipline, and BD (short for Bachelor Dissertation) or RA (Research Article) indicates the subcorpus the example belongs to:

(3) MED_BD02: *(e.g. see Appendix 1)*
Tagged text:
(e.g._MD_MT_CG_EX see_MD_IP_RO_DI Appendix_MD_MT_RT_PT 1) _MD_MT_CG_PA
Tags stand for:
*e.g.*_Metadiscourse_Metatext_Code Gloss_ Exemplifying
*see*_Metadiscourse_Interpersonal_Reader-oriented_Directive
*Appendix*_ Metadiscourse_Metatext_Reference to text_Part of the text
*)*_Metadiscourse_Metatext_Code Gloss_Parenthetical

(4) LIN_RA02: *For instance, let us take the PV show up with the following meaning sense distribution:*
Tagged text:
For instance_MD_MT_CG_EX , let us_MD_IP_PO_IW take the PV show up with the following_MD_MT_EN_CA meaning sense distribution:_MD_MT_CG_CL
Tags stand for:
*For instance*_Metadiscourse_Metatext_Code Gloss_Exemplifying
*let us*_Metadiscourse_Interpersonal_Participant Oriented_Inclusive We
*the following*_Metadiscourse_Metatext_Endophoric_Cataphoric
*:*_Metadiscourse_Metatext_Code Gloss_Colon

This reflexive model excludes stance markers (e.g. hedges and boosters) and also intertextual references (e.g. reporting verbs). As was mentioned earlier, the set of criteria developed by Mauranen (1993) and Ädel (2006) to help identify reflexive MD markers, namely (a) explicitness or self-awareness, (b) contextuality, (c) current text and (d) writer and reader, was applied during the selection process as follows:

a) explicitness or self-awareness: to qualify as reflexive MD, the writer had to make explicit reference to (i) the ongoing text, to (ii) her/himself as the write, and (iii) the reader of the text.

b) contextuality: according to this criterion, the rhetorical function of each MD marker refers only to its immediate discourse context (Ädel 2010). Thus, all items were analyzed in context to count reflexive elements only (e.g. the isolated word *author* could refer to the author of the text, to the author of any other text or to *authors* in general).

c) current text: from a reflexive perspective, the connection with the real world – e.g. propositional content, personal judgments and opinions, or intertextuality – does not qualify as reflexive MD. Only those markers that refer to the evolving text were counted.

d) writer and reader: only references to the writer and reader as immediate participants of the current text, and not as experiencers of the real world, qualify as MD (see e.g. specifications for 'inclusive *we*' mentioned previously).

After the identification process, 230 reflexive markers belonging to 21 different categories were found and tagged (see Appendix 4 for a complete list of markers).

[2] Research questions are excluded here.

[3] Only those cases in which *we* is used to refer to 'you and me', i.e. the author and the reader of the text, qualify as reflexive MD. As a rule of thumb, Noble (2010) suggests that those instances in which *we* can be replaced by the term *people* or *anyone*, as it is overtly general, do not qualify as MD.

4. CORPUS-BASED STUDIES ON METADISCOURSE

Regarding the use of MD in academic writing, four main patterns have been found in the literature:

(i) EFL learner writers tend to underuse certain categories of MD markers when compared to experts or native writers in the same discipline. Devices that signal authorial confidence such as 'self-mention' or 'elaboration' (Springer 2012), or refer to the evolving texts, such as 'endophoric markers' (e.g. in the *following* section), and 'reader-oriented markers' (e.g. *see* Table 3) are often underused, which has been attributed to students' "inexperience in structuring big texts" (Burneikaite 2008: 45) and to having a "low audience-awareness" (2008: 45) possibly due to a lack of exposure and explicit learning of MD markers.

(ii) Several studies contrasting the use of MD in L1 and L2 English, and L1 Spanish in different disciplines (medical sciences, social sciences and humanities), genres (research articles, textbooks, newspaper opinion articles) and contexts (international vs. national journals) suggest that academic texts in English are likely to contain quantitatively more MD (especially 'logical markers', 'code glosses', 'adversative connectors' and 'self-mentions') than academic texts in Spanish (Moreno 1997; Dafouz 2003; García Negroni 2008; Pérez-Llantada 2010; Mur Dueñas 201).

(iii) Differences are also found from an interdisciplinary perspective: research on the use of MD shows how different disciplinary communities have different conventions of MD. In fact, texts belonging to humanities (e.g. linguistics) are likely to contain quantitatively more MD devices than other disciplines (e.g. medicine) (Hyland 2001). This difference has been attributed to the need of human sciences to elaborate claims more since they are often based on qualitative methods (Hyland 2010); the nature of the topics itself – i.e. language being the subject matter of the linguistic discipline, also accounts for the discrepancies found (Salas 2015).

(iv) L2 learners or novice writers who do not use MD markers in their texts accurately (i.e. may not be aware of their disciplinary community conventions) tend to produce less persuasive and thus less successful texts: positive correlations between high-scoring essays and a higher frequency and range of MD devices have been found in the literature (Intaraprawat and Steffensen 1995; Noble 2010). Teaching MD explicitly seems to be both educationally and statistically significant in that learners improve their texts' quality and achieve higher scores (Cheng and Steffensen 1996).

As we have seen, academic writing is community situated (Hyland 2005: 142), and therefore not only language (e.g. English) but also discipline (e.g. medicine), mode (e.g. written), genre (e.g. research articles) and even part of the text (e.g. introduction) play a role in the choice of metadiscursive practices (Dafouz 2003; Hyland and Tse 2004; Bondi 2010; Pérez-Llantada 2010; Mur Dueñas 2011; Hyland 2012; Salas 2015). Writers who conform to specific disciplinary conventions, express ideas clearly and create a balanced textual persona that sounds familiar and convincing to their readers are more likely to succeed in the scientific communication realm (Intaraprawat and Steffensen 1995; Ivanič 2004). Hence, in order to investigate the production of reflexive MD by EFL undergraduate learners in different disciplines, three research questions have been established:

1. To what extent do Spanish undergraduate students use reflexive MD markers when writing in academic English? The frequency rates of all reflexive MD markers found in the corpora will be calculated, to later explore the different categories used.
2. Are there any differences across disciplines? We will look at interdisciplinary variation in the corpus (i.e. BDs and RAs in medicine and linguistics).
3. Is there overuse or underuse of reflexive MD markers when compared to an expert corpus of RAs? This analysis will help us to identify possible learner features in academic writing.

5. METHODOLOGY

5.1. Data collection

In order to carry out an interdisciplinary analysis of reflexive MD markers in medical and linguistic academic texts, two corpora were compiled, namely a learner corpus of 20 BDs written in English by Spanish undergraduate students in linguistics and medicine from two Spanish universities (103,971 words) and an expert corpus of 50 RAs published in medical and linguistic academic journals (see Appendix 1 for the list of journals) to match the discipline and (roughly) the topic of the BDs (258,223 words). The texts in both the learner (BDs) and the expert corpus (RAs) varied in length; for this reason, normalized values per 1,000 running words were calculated and added to the tables. The total number of texts, tokens and types in each corpus are shown in Table 2.

	BDs		RAs	
Discipline	Linguistics	Medicine	Linguistics	Medicine
No. texts	10	10	25	25
Tokens	65,180	38,791	177,041	81,1182
Types	5537	4656	9853	7553
Average text length	6518	3879	7081	3247
Total words	103,971		258,223	

Table 2: The learner and the expert corpora

5.2. Data analysis

I carefully read and scanned all reflexive MD markers in each text (see tagging system and identification criteria in Section 3) and only relevant examples – that is, reflexive and text-internal – were coded. This corpus-driven or, as previously described, 'thick' method used to retrieve instances of MD gave me a deeper view of the choices learners made, how textual and interpersonal interactions were realized, the most prevalent types of MD markers in each discipline, how they were distributed and how these patterns may have affected the whole structure of the text. Subsequently, the corpus analysis software AntConc was used to concordance all the different categories (searching by code, e.g. _MD_MT_RT_FT). Frequency rates were calculated, and the most remarkable differences on the use of reflexive MD markers were carefully studied.

6. RESULTS AND DISCUSSION

The results of the analysis of reflexive MD are reported on as follows: first, the overall differences between BDs and RAs production across disciplines are given. Second, the frequency counts of textual (MT) and interpersonal (IP) categories, and their subcategories in each corpus are presented. Finally, a second and more qualitative analysis across disciplines (i.e. linguistics vs. medicine) and writer status (learner vs. expert) is performed to explore cases of overuse and underuse – these terms are used in a quantitative sense, that is, to refer to the highest or lowest differences in frequency when comparing the learners' and the experts' production – to finally draw some pedagogical implications.

6.1. Overall frequency of reflexive MD markers

The overall frequency results of the two main types of MD markers (i.e. textual and interpersonal) across disciplines is shown in Table 3 and illustrated in Figure 1. Appendix 2 presents global results of all MD categories and subcategories explored, and it provides both raw and normalized results.

	LIN		MED	
	BDs	RAs	BDs	RAs
Total MT	32	32.5	25.4	24.6
Total IP	3.2	6.2	6.3	4
Total MD	35.2	38.8	31.7	28.6
Total MD %	3.5%	3.8%	3.1%	2.8%
Average markers per text	229	275	123	92

Table 3: Total production of reflexive MD in BDs and RAs according to discipline (per 1,000 words)

The analysis of textual markers in both corpora reveals that both learners and experts have used MD to a similar extent. However, this is only true if we look at the texts according to discipline (linguistics and medicine), which suggests that disciplinary conventions do play an important role in the choice of MD practices. This global results support frequency findings across disciplines reported in the literature (e.g. Hyland 2001, 2010; Hyland and Tse 2004; Salas 2015). As can be seen, linguistics contains more MD markers in general (RAs 38.8, BDs 35.2) than medicine (RAs 28.6, BDs 31.7). In fact, medical RAs contain the least amount of MD markers in all five measures: total metatext (MT), total interpersonal (IP), total metadiscourse (MD), percentage of MD (%) and average markers per text. It is interesting to note, however, that BDs in medicine contain the highest frequency of interpersonal markers than any other subcorpus in this study, being almost twice as frequent as in the medical RAs; this points towards a case of overuse that will be explored further in Section 6.5.

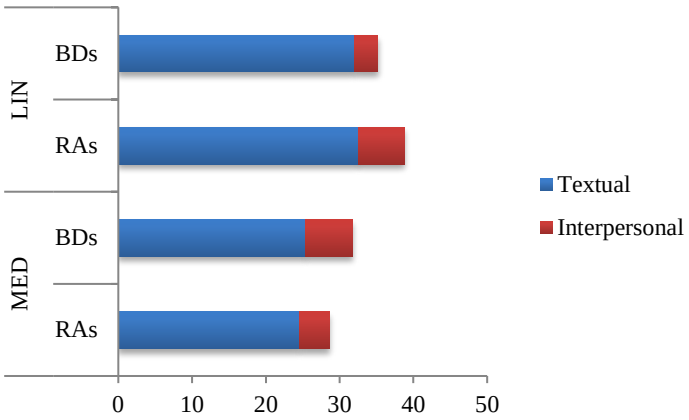

Figure 1: Metatext and interpersonal MD in linguistics and medicine

In Figure 1, we can see how textual markers have been used much more frequently than interpersonal markers (there are many more subcategories that belong to textual MD, which partly explains why this is so). This finding is also in line with previous research (Hyland 2001; Dafouz 2003; Hyland and Tse 2004; Hyland 2010; Pérez-Llantada 2010; Salas 2015). It is interesting, however, to remark on how both BDs and RAs in linguistics, and BDs and RAs in medicine have used textual MD to practically the same extent compared to one another (at least numerically). This could very well suggest that learners in this corpus are well aware of the textual MD practices of their discipline. Another possible explanation is the fact that many of these textual markers (e.g. use of connectors, exemplifiers, reformulators) are often taught in English language instruction in secondary or tertiary education, so EFL students may feel more confident when using them. In spite of this quantitative similarity, there is nevertheless an interesting difference in the choice of makers within this category that will be explored further in Section 6.2.

Regarding the use of interpersonal markers, the learner corpus has yielded somewhat unexpected results: while BDs and RAs seem to agree in their use of textual markers according to discipline, the use of interpersonal markers varies greatly in all four subcorpora, as illustrated in Figure 1. BDs in linguistics have used half as many interpersonal markers (3.2) as RAs (6.2), and the opposite tendency occurs in BDs in medicine (6.3) compared with RAs in the same discipline (4.0). Although it is difficult to find the exact reason for these differences, a possible explanation could be related to the fact that BDs and RAs have different audiences: a BD displays knowledge to a supervisor and the evaluating committee, while RAs display knowledge to peers of more or less the same expertise. Mauranen (2001: 209) hypothesized that "those in a dominant position in any speech event will use more reflexive expressions". However, this is only true for the linguistic subcorpora and not for the medicine subcorpora, in which the learners have produced more MD in general than RA authors. In any case, I believe that the lack of explicit teaching on the use of writer, reader and participant-oriented mentions in different disciplines may account for this quantitative difference. Let us have a closer look at each of these categories (textual and interpersonal) across disciplines in order to see these differences in more detail.

6.2. Textual metadiscourse

Table 4 displays the categories and subcategories that belong to textual MD. The most significant differences in each subcategory are explained below.

	BDs		RAs	
	LIN	MED	LIN	MED
Reference to the text				
Full text	1.76	1.39	1.81	1.22
Part of the text	1.96	1.16	1.10	0.60
Semiotic modes	1.21	0.80	2.68	3.07
Total RT	4.94	3.35	5.59	4.89
Endophoric markers				
Anaphoric	1.38	0.46	0.90	0.59
Cataphoric	0.81	0.67	1.19	0.64
Deictic	0.20	0.00	0.75	0.02
Total EN	2.39	1.13	2.83	1.26
Code Glosses				
Reformulators	2.12	1.01	2.50	1.34
Exemplifiers	2.38	0.80	3.93	1.64
Parentheticals ()	4.08	7.53	3.49	5.57
Dashes (–)	0.34	0.00	0.23	0.16
Colons (:)	2.95	2.55	1.56	0.65
Semicolons (;)	1.03	0.62	1.22	1.68
Total CG	12.89	12.50	12.93	11.04
Linking Devices				
Additive	2.42	2.60	1.59	1.88
Constrastive	4.66	2.81	4.43	2.82
Consecutive	1.69	1.50	1.96	1.13
Organizers	2.33	1.16	2.19	1.39
Topicalizers	0.68	0.36	1.04	0.18
Total LD	11.78	8.43	11.22	7.42
Total METATEXT	32.00	25.42	32.57	24.60

Table 4: Frequency of reflexive metatext in BDs and RAs (per 1,000 words)

6.2.1. Reference to text

RAs in linguistics have included more references to the text (5.6) than any other subcorpus, followed by BDs in linguistics (4.9). According to these findings, authors in the field of linguistics tend to refer to the full text (e.g. our *paper*) and to parts of the text (e.g. the next *section*) more often than authors of other disciplines. On the other hand, RAs in medicine seem to contain more references to semiotic modes (e.g. see *Figure* 1); to be more precise, there is an average of 10 references to semiotic modes per paper (*Figure* is the first semiotic mode in medicine RAs), whereas in the medical BDs corpus there is an average of three references per text. Learners in this corpus do not refer to their semiotic modes (tables, figures, diagrams) as often as the RA authors. We will return to cases like this in Section 6.3.

6.2.2. Endophoric markers

As shown in Table 4, the linguistic subcorpora contain more endophoric markers than medicine. There is, however, a notable difference: BDs in linguistics have used anaphoric markers (e.g. *as mentioned previously*) more frequently (1.3) than the RAs (0.9). In contrast, RAs have used cataphoric markers (e.g. *as follows*) to tell the reader to look forward in the text, more often: cataphoric markers help foreground upcoming material, so the reader knows what is next and where to find that information. The frequent use of anaphoric markers by learners in linguistics (average of nine anaphoric references per text) may have made some parts of their texts a bit redundant. Another important observation here is the fact that the medicine subcorpus (both RAs and BDs) contain very few – or practically none – deictics (e.g. *here*, *now*).

6.2.3. Code glosses

Markers in this category are the most popular ones in the corpus. Exemplifiers (e.g. *for instance*) and parentheticals (e.g. *(see Table 2)*) abound in all four subcorpora. The former is one of the most frequent MD subtypes in RAs and BDs in linguistics (3.9 and 2.3 per 1,000 words, respectively). Authors of this discipline tend to provide the reader with many examples in order to illustrate their points. *Such as*, *e.g.* and *for example* are the top-3 markers that help authors to exemplify in their texts (see Appendix 3 for a list of the top-3 textual and interpersonal markers in each subcorpus). The latter, parentheticals, is one of the most

frequent markers in the medical BDs. Learners use parentheticals to refer the reader to different sections in their text or to specify the type of variable they have used, as in example (5) below:

(5) MED_BD01: Measured trough Charlson Comorbidity Index (Charlson/Deyo version*)_MD_MT_CG_PA* with data figuring in the clinical course (see Annex IV*) _MD_MT_CG_PA*. This variable will be categorized [...]

In the case of colons, they are used more frequently in BDs (LIN 2.9, MED 2.5) than in RAs (LIN 1.5, MED 0.6), and they often appear after the cataphoric marker *the following*, preceding examples or lists of concepts, as in (6):

(6) LIN_BD04: Some examples of epistemic modality are*: _MD_MT_CG_CL* "We may/might lose the elections / They must have won the elections"

In contrast, semicolons have been used much more frequently in medical RAs (1.6), especially before *and*, *however* and *therefore*, as illustrated in (7):

(7) MED_RA05: It may be presumed that physicians prescribe statins to patients who suffered more severe obesity*;_MD_MT_CG_SC* therefore, statin users could have been more likely to develop diabetes and diabetic complications.

Regarding the use of dashes, they were only found in BDs and RAs in linguistics, in particular before – *also known as*, – *and* and – *thus*; authors used single (–) double (--) or even triple (---) dash at the beginning, and sometimes also at the end of the commentary, as can be seen in (8) and (9):

(8) LIN_BD08: This can be accounted for two different principles: a weak one –*_MD_MT_CG_DA* also known as linguistic relativity– and a strong one –*_MD_MT_CG_DA* also known as linguistic determinism–.
(9) LIN_RA13: In the same vein, the Pidgin uses full-NPs to signal anaphoric --*_MD_MT_CG_DA* and thus by logical inference (22a) also cataphoric discontinuity.

6.2.4. Linking devices

Two subtypes – additive and contrastive – were the most popular ones in the corpus across disciplines. BDs and RAs in linguistics contain 11.7 and 11.2 linking devices per 1,000 words, respectively, whereas BDs and RAs in medicine contain notably fewer markers in this category (8.4 and 7.2, respectively). Within the linking devices category, contrastive markers are more frequent than additive markers, especially in BDs in linguistics (4.6) – almost twice as many as in medical BDs (2.8). *However* is the number one contrastive marker in all corpora, followed by *therefore* and *thus*. On the other hand, the most popular additive marker is *in addition*, followed by *moreover* and *furthermore*. It is also worth mentioning that there are two subtypes, namely organizers, illustrated in example (10), and topicalizers, in (11), that mainly appear in the linguistic corpus only. *In terms of*, *in the context of* and *with respect to* are the top-3 topicalizers in the corpora:

(10) LIN_BD03: *First_MD_MT_LD_OR*, an overview on what ToM means [...]. *Then_MD_MT_LD_OR*, different theories on which elements of language foster ToM development are explained [...]. *Finally_MD_MT_LD_OR*, the view of those who deny the role of language [...]
(11) LIN_RA12: *With respect to_MD_MT_LD_TO* vocabulary acquisition from a supportive reading context, the results showed that providing explicit clues can result in relatively high lexical gains [...]

6.3. Overuse and underuse of textual markers

If we look at the total production of textual MD according to discipline, as we did earlier, we see that BDs and RAs in linguistics (32 and 32.5) and BDs and RAs in medicine (25.4 and 24.6) contain quite a similar amount of textual markers. However, when we look in more detail at the type of MD markers used in each category, important differences emerge. It is interesting to note here that, in the case of textual MD, all cases of overuse and underuse are found in both subcorpora of BDs, regardless of their discipline, which could highlight learner-writing features as opposed to conventions of different disciplines, in this case.

First, BDs in general refer to parts of the text (e.g. in this *section*) more often (LIN 1.9, MED 1.1) than RAs (LIN 1.1, MED 0.6). This finding contrasts with Burneikaite (2008), who found that EFL learners in fact underused endophoric markers, producing somewhat unstructured texts. It could be argued that learners in this corpus have a higher audience-awareness: they indicate and inform the reader, perhaps too often, about the different sections of their texts. In contrast, however, the BDs do not include as many references to semiotic modes (LIN 1.2, MED 0.8) as the RAs (LIN 2.6, MED 3), even though they did include tables

and figures in their dissertations. This could suggest that learners do not guide the reader enough through the semiotic modes presented in their texts; it is up to the reader, in some cases, to understand and analyze the information presented. This could be indicative of transfer from their L1 (Spanish), a slightly more reader-responsible writing style (Hinds 1987) and thus worthy of pedagogical attention. Regarding the use of exemplifiers, the BDs seem to have underused this type (LIN 2.3, MED 0.8) compared to the RAs (LIN 3.9, MED 1.6). Students may lack confidence or may not know enough so as to give examples about certain topics. It is also possible that, having a supervisor who knows well the topic as the intended reader of their text, students may not feel the need of giving many examples in their dissertations. Another difference found in the analysis concerns the use of colons: BDs have used colons much more frequently (LIN 2.9, MED 2.5) than authors of the RAs (LIN 1.5, MED 0.6). In addition, and with regard to semicolons, it is important to mention that medical BDs contain very few semicolons (if at all) (0.6), which contrasts with the use of semicolons in published RAs in the same discipline (1.6). This finding suggests that learner writers need to revise the use of these two punctuation marks in academic writing.

Finally, and again, in the learner subcorpora, additive markers are used much more frequently in the BDs (LIN 2.4, MED 2.6) than in the RAs (LIN 1.5, MED 1.8). Spanish L1 writers of English have preferred to add ideas to their argument to achieve credibility, which is a common practice in academic literature written in Spanish, rather than including pros and cons of the discussed topic, or contrasting findings and different perspectives on the matter, which is a common practice in academic literature written in English. This finding is in line with previous studies (Dafouz 2003; Perez-Llantada 2010) that suggest that writers may retain part of their Spanish L1 writing style when writing in English. More pedagogical attention should therefore be given to culture-driven preferences in general, and to the use of linking devices in academic texts in particular.

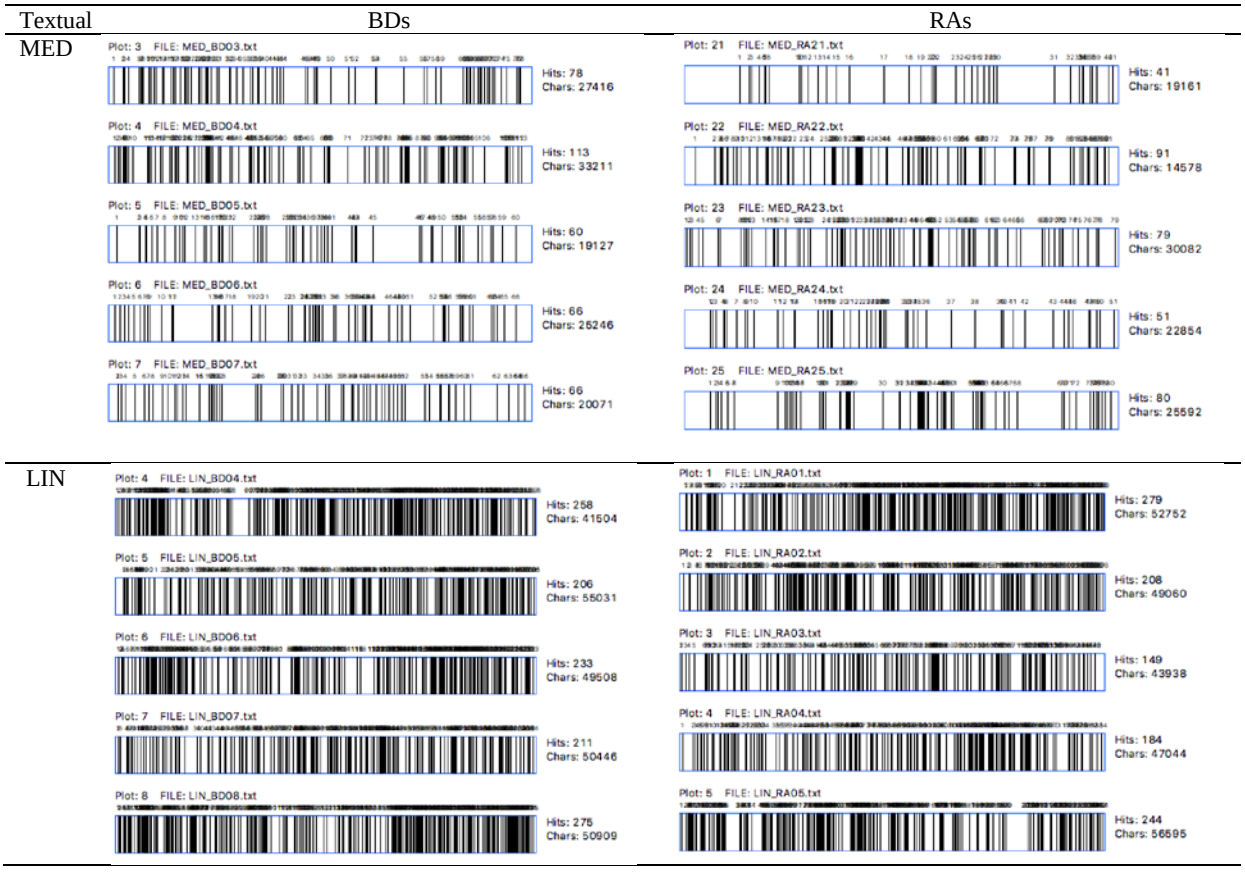

Figure 2: Concordance plot of the use of textual markers

Figure 2 shows where exactly textual markers (of all subtypes) occur along the texts and how frequently. Five random texts of each subcorpus have been selected to illustrate the plot. We can observe how there is a similar dispersion (distribution of vertical lines) of textual markers across potentially different sections (e.g. introduction, method, conclusion) and texts (e.g. LIN_BD04, MED_BD09, MED_RA14), but also a different density (thicker lines represent higher frequency) according to discipline. This frequency and distribution of textual markers contrasts very much with the use of interpersonal markers, which can be seen in Figure 3 in Section 6.5.

6.4. Interpersonal reflexive metadiscourse

Turning now to the use of interpersonal MD markers, we can see some remarkable differences: as illustrated in Table 5, and as mentioned earlier, BDs in medicine contain the highest frequency of interpersonal markers – especially self-mention (6.3 markers per 1,000 words) – than any other subcorpus analyzed in this study.

	BDs		RAs	
	LIN	MED	LIN	MED
Writer oriented				
Self-mention	1.83	5.13	4.10	3.82
Reader oriented				
Directives	0.43	0.90	1.13	0.17
Rhetorical questions	0.05	0.00	0.07	0.00
Total RO	0.48	0.90	1.20	0.17
Participant oriented				
Inclusive *we*	0.89	0.34	0.98	0.04
Total INTERPERSONAL	3.19	6.37	6.27	4.03

Table 5: Frequency of interpersonal markers in BDs and RAs (per 1,000 words)

If we look closely at each category, writer-oriented (i.e. self-mention) markers are more popular than reader- or participant-oriented markers in all subcorpora, regardless of their discipline: *we* is the most preferred marker of self-mention – even for single-authored texts, which sometimes serves a hedging purpose (Hyland 2001), followed by *our*, *I* (only in the case of linguistics) and *us* (see Appendix 3). Also important to note here is that directives are much more frequent in RAs in linguistics (1.1), compared to the other subcorpora, as can be seen in Table 5; the way authors prefer to direct to the reader is by means of the imperative *see*, as in (12):

(12) MED_RA11: one that does not in itself create the potential for contamination of the environment in which it is used (*see_MD_IP_RO_DI* Box 1 for experimental details).

As for rhetorical questions (see (13) and (14)), they were infrequent in general and only present in BDs and RAs in linguistics:

(13) LIN_BD05: What does this mean*?_MD_IP_RO_RQ* According to the interpretations provided before
(14) LIN_RA22: What lessons for syllabus design can one draw from these findings*?_MD_IP_RO_RQ* As far as prepositional postmodifiers are concerned

6.5. Overuse and underuse of interpersonal markers

In the case of interpersonal markers, it is important to mention that the cases of overuse apply to medical BDs only, whereas the cases of underuse apply to BDs in linguistics: medical BD writers produced self-mention (15), directives (16) and inclusive *we* (17) (5.1, 0.9 and 0.3, respectively) much more frequently than RA writers (3.8, 0.1 and 0.04, respectively), and in some cases even more frequently than BDs and RAs in linguistics.

(15) MED_BD02: *We_MD_IP_WO_SF* expect to observe the existence of additional benefits, not explained by the weight loss alone,
(16) MED_BD07: *Take_MD_IP_RO_DI* the high number of atypical squamous cells of unknown significance [...] detected by Pap test for instance,
(17) MED_BD05: about the incidence of schwannoma as reference, *we_MD_IP_PO_IW* could find that the incidence of the vestibular schwannoma (VS) has been stabilized

With regard to underuse, self-mention appears much less frequently in linguistic BDs (1.8) than in RAs (4.1). In addition, directives also seem to have been underused by BD writers in linguistics (0.4) compared to RAs (1.1). These findings thus have an important implication, namely that undergraduate students need more explicit instruction on the use of interpersonal markers taking into account their specific field of study.

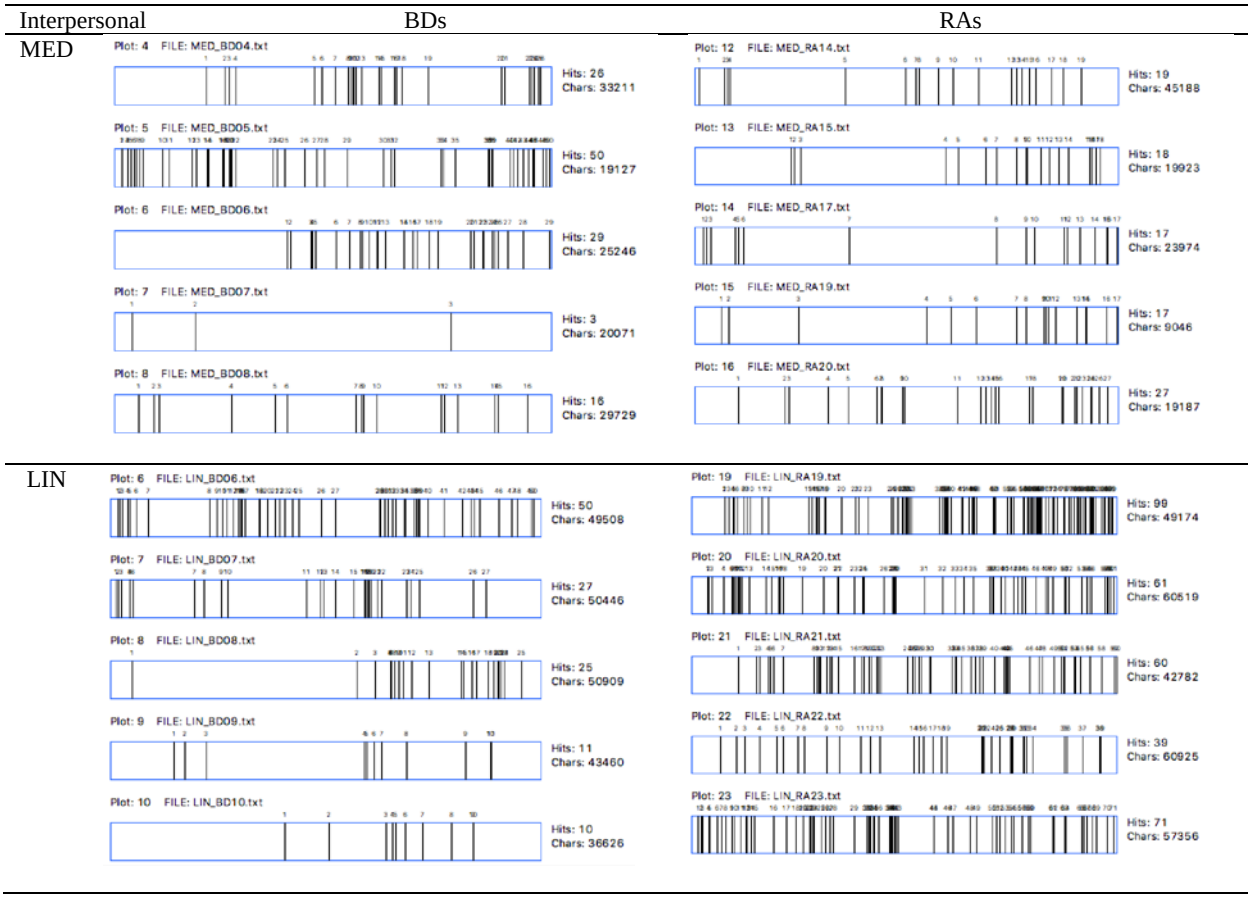

Figure 3: Concordance plot of the use of interpersonal markers

Figure 3 shows the concordance plot of interpersonal markers. Two interesting points arise here: first, the density is clearly much less prominent than that of textual markers (see Figure 2 above), and, second, the dispersion of these markers in BDs and RAs is different: if we take a closer look at the plot, we can see how RAs use interpersonal markers (especially self-mention) mostly at the beginning and towards the end of the text (which could represent the introduction or methods, and discussion or conclusion sections); such a pattern is not found in BDs, in which interpersonal markers are used elsewhere. In addition, and in terms of density of interpersonal markers, it differs in both BDs and RAs, in both disciplines, so we can say that learners' use of interpersonal markers does not approximate the use of these by experienced writers.

6.6. Summary

This corpus-driven study has yielded results on the frequency as well as the usage patterns of reflexive MD markers produced by learners and expert writers. One of the first objectives of the present study was to find out the extent to which Spanish undergraduate students use reflexive MD in their academic texts. The results show that MD represents an average of 3.1 percent (BDs in medicine) and 3.5 percent (BDs in linguistics) of the total texts. Comparisons with the expert corpus show that overall learners use MD to a similar extent (MD represents 2.8 percent in RAs in medicine, and 3.8 percent in RAs in linguistics). We may therefore say that EFL Spanish undergraduate students produce reflexive MD to an appropriate extent in terms of frequency. The second objective was to detect differences across disciplines: the analysis shows that BDs and RAs in linguistics contain more MD in general than BDs and RAs in medicine (except for self-mention markers in medical BDs). This result supports previous findings reported in the literature about different conventions of MD across disciplines (Hyland 2001, 2010; Mur Dueñas 2011; Salas 2015). Finally, the third objective was to see if there were any differences according to writer status – i.e. learners vs. experts. The analysis reveals that there is an extremely similar frequency of textual MD in both BDs and RAs, which suggests that learners in this corpus are aware of their readership and have guided their readers appropriately through their texts. However, comparisons with the expert corpus have also allowed me to find cases of overuse and underuse of certain MD markers, which, surprisingly, apply to the entire learner corpus, regardless of their discipline. Moreover, and in terms of interpersonal markers, BDs neither

approximate the use of self-mention, inclusive *we* or directives in RAs, nor are they comparable to one another. Some of these findings could be indicative of a different genre (e.g. BDs display knowledge to a supervisor); they could also denote a more reader-responsible writing style, as a culture-driven preference (L1 transfer) or even be due to the conflicting advice on the use of self-mention devices in academic writing textbooks.

6. Conclusion

This article has analyzed the density and range of reflexive textual and interpersonal MD markers present in two corpora, namely a learner corpus of BDs written in English by Spanish undergraduate students in two different disciplines (linguistics and medicine), and an expert corpus, consisting of RAs published in English-medium academic journals. The quantitative and qualitative analysis performed shows that, overall, BDs and RAs in the same disciplines contain a similar amount of textual MD markers, which may indicate that EFL Spanish undergraduate students are aware of the textual MD conventions of their community of practice, at least in terms of frequency of use. Under a closer look, however, a qualitative analysis shows that BD learner writers use references to parts of their text, colons and additive linking devices significantly more often than expert writers. On the other hand, learners seem to underuse references to their semiotic modes, exemplifiers and semicolons. These cases of overuse and underuse of textual MD markers are present in both corpora, regardless of their discipline, which may highlight features of the BD genre on the one hand, and of EFL Spanish undergraduate students' writing on the other hand. Regarding the use of interpersonal MD, the learner corpus in this study has yielded interesting results: learners' use of interpersonal markers does not approximate that of more experienced writers at all: BDs in linguistics seem to underuse self-mentions and directives compared with RA writers, and the opposite tendency occurs in BDs in medicine, in which writers refer to themselves and engage the reader much more frequently than RA writers. And neither does their use of interpersonal markers approximate one another in the same genre (BDs). These cases of overuse and underuse are therefore worthy of pedagogical attention.

It is important, however, to expose the limitations of this study. The first limitation is related to the corpus size (362,194 total words) and the number of participants (20 undergraduate students): the manual analysis and tagging of MD markers in the corpora was very time-consuming and did not allow me to include more texts in the corpus; using a larger corpus would certainly help to make these findings more representative and generalizable. Second, the comparison of MD markers across corpora was done from a word-level scope (normalizing values per 1,000 words). It has been argued, however, that T-units may be a more appropriate basis for calculating density than words, since MD markers typically have a clause-level scope (Intaraprawat and Steffensen 1995). Calculating the mean length of T-units in the corpus and using it as a basis for comparison between two corpora could provide different results. This study has also looked at *inter*disciplinary variation (linguistics vs. medicine) but not at *intra*disciplinary variation (e.g. texts on second language acquisition vs. texts on learner corpus research). Performing an intradisciplinary analysis to explore the differences in the use of MD within texts in the same discipline, but on different topics, would be something worth investigating. Finally, this study could be improved by using a classical contrastive analysis of parallel corpora (e.g. BDs in L2 English and in L1 Spanish) to study interlanguage, which could help to detect transferred MD practices from an L1 more accurately.

The results of this study have attempted to shed light on the types and frequency of reflexive MD makers in two somewhat similar genres (BDs and RAs) across two vastly different disciplines (linguistics and medicine). This analysis has also provided a comprehensive list of 230 MD markers in 21 different subcategories that may be of interest to EFL learner writers, and also to academic writing teachers and material developers, who are interested in teaching the use of MD in these two specific disciplines. To conclude, the findings of the present research corroborate the need for more explicit teaching and corpus-informed materials on MD: more pedagogical attention should be given to MD, especially in EFL academic writing, taking into account the writers' L1, genre and field of study, so that MD practices are taught and learnt in agreement with each disciplinary community.

REFERENCES

Ädel, Annelie. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.

Ädel, Annelie. 2010. *Just to give you kind of a map of where we are going*: a taxonomy of metadiscourse in spoken and written academic English. *Nordic Journal of English Studies* 9/2: 69–97.

Ädel, Annelie and Anna Mauranen. 2010. Metadiscourse: diverse and divided perspectives. *Nordic Journal of English Studies* 9/2: 1–11.

Bondi, Marina. 2010. Metadiscursive practices in introductions: phraseology and semantic sequences across genres. *Nordic Journal of English Studies* 9/2: 99–123.

Burneikaite, Nida. 2008. Metadiscourse in linguistics master's theses in English L1 and L2. *Kalbotyra* 59/3: 38–47.

Cheng, Xiaoguang and Margaret Steffensen. 1996. A technique for improving student writing. *Research in the Teaching of English* 30/2: 149–181.

Crismore, Avon. 1989. *Talking with readers: metadiscourse as rhetorical act*. New York: Peter Lang.

Crismore, Avon and Rodney Farnsworth. 1990. Metadiscourse in popular and professional science discourse. In Walter Nash ed. *The writing scholar: studies in academic discourse*. Newbury Park, CA: SAGE, 118–136.

Crismore, Avon, Raija Markkanen and Margaret Steffensen. 1993. Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Written Communication* 10: 39–71.

Crismore, Avon and William J. Vande Kopple. 1997. Hedges and readers: effects on attitudes and learning. In Raija Markkanen and Hartmut Schröder eds. *Hedging and discourse: approaches to the analysis of a pragmatic phenomenon in academic texts*. Berlin: Walter de Gruyter, 83–114.

Dafouz, Emma. 2003. Metadiscourse revisited: a contrastive study of persuasive writing in professional discourse. *Estudios Ingleses de la Universidad Complutense* 11: 29–52.

Dahl, Trine. 2004. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36/10: 1807–1825.

García Negroni, María Marta. 2008. Subjetividad y discurso científico-académico: acerca de algunas manifestaciones de la subjetividad en el artículo de investigación en español. *Revista Signos* 41/66: 5–31.

Halliday, M.A.K. 1973. *Explorations in the functions of language*. London: Edward Arnold.

Harris, Zellig. 1970. *Papers in structural and transformational linguistics*. Dordrecht: D. Reidel.

Hinds, John. 1987. Reader versus writer responsibility: a new typology. In Ulla Connor and Robert B. Kaplan eds. *Writing across languages: analysis of L2 text.* Reading, MA: Addison-Wesley, 141–152.

Hyland, Ken. 1998. Boosting, hedging and the negotiation of academic knowledge. *Text* 18/3: 343–382.

Hyland Ken. 2000. *Disciplinary discourses: social interactions in academic writing*. Harlow: Longman.

Hyland, Ken. 2001. Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes* 20: 207–226.

Hyland, Ken. 2002. Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics* 34/8: 1091–1112.

Hyland, Ken. 2005. *Metadiscourse: exploring interaction in writing.* London: Continuum.

Hyland, Ken. 2008. Genre and academic writing in the disciplines. *Language Teaching* 41/4: 542–562.

Hyland, Ken. 2010. Metadiscourse: mapping interactions in academic writing. *Nordic Journal of English Studies* 9/2: 125–143.

Hyland, Ken. 2012. *Disciplinary identities.* Cambridge: Cambridge University Press.

Hyland, Ken. 2017. Metadiscourse: what is it and where is it going? *Journal of Pragmatics* 113: 16–29.

Hyland, Ken and Polly Tse. 2004. Metadiscourse in academic writing: a reappraisal. *Applied Linguistics* 25/2: 156–177.

Intaraprawat, Puangpen and Margaret Steffensen. 1995. The use of metadiscourse in good and poor ESL essays. *Journal of Second Language Writing* 4/3: 253–272.

Ivanič, Roz. 2004. Discourses of writing and learning to write. *Language and Education* 18/3: 220–243.

Lee, Joseph J. and Lydia Deakin. 2016. Interactions in L1 and L2 undergraduate student writing: interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing* 33: 21–34.

Low, Graham. 1996. Intensifiers and hedges in questionnaire items and the lexical invisibility hypothesis. *Applied Linguistics* 17/1: 1–37.

Martín-Laguna, Sofia and Eva Alcón. 2015. Do learners rely on metadiscourse markers? An exploratory study in English, Catalan and Spanish. *Procedia–Social and Behavioral Sciences* 173: 85–92.

Mauranen, Anna. 1993. *Cultural differences in academic rhetoric: a textlinguistic study*. Frankfurt: Peter Lang.

Mauranen, Anna. 2001. Reflexive academic talk: observations from MICASE. In Rita C. Simpson and John Swales eds. *Corpus linguistics in North America: selections from the 1999 symposium*. Ann Arbor, MI: University of Michigan Press, 165–178.
Mauranen, Anna. 2010. Discourse reflexivity – a discourse universal? The case of ELF. *Nordic Journal of English Studies* 9/2: 13–40.
Montaño-Harmon, María Rosario. 1991. Discourse features of written Mexican Spanish: current research in contrastive rhetoric and its implications. *Hispania* 74/2: 417–425.
Moreno, Ana I. 1997. Genre constraints across languages: causal metatext in Spanish and English RAs. *English for Specific Purposes* 16: 161–179.
Mur Dueñas, Pilar. 2011. An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics* 43/12: 3068–3079.
Noble, Wendy. 2010. Understanding metadiscoursal use: lessons from a 'local' corpus of learner academic writing. *Nordic Journal of English Studies* 9/2: 145–169.
Pérez-Llantada, Carmen. 2010. The discourse functions of metadiscourse in published academic writing: issues of culture and language. *Nordic Journal of English Studies* 9/2: 41–68.
Salas, Millaray D. 2015. Reflexive metadiscourse in research articles in Spanish: variation across three disciplines (linguistics, economics and medicine). *Journal of Pragmatics* 77: 20–40.
Sinclair, John. 1981. Planes of discourse. In S.N.A Rizvi eds. *The two-fold voice: essays in honour of Ramesh Mohan*. Salzberg: Salzberg University Press, 70–89.
Springer, Philip Ernest. 2012. *Advanced learner writing* (1st ed.). Amsterdam: Vrije Universiteit.
Toumi, Naouel. 2009. A model for the investigation of reflexive metadiscourse in research articles. *Language Studies Working Papers* 1: 64–73.
Vande Kopple, William J. 1985. Some exploratory discourse on metadiscourse. *College Composition & Communication* 26: 82–93.
Williams, Joseph M. 1981. *Style: ten lessons in clarity and grace*. Glenview, IL: Scott, Foresman.

Corresponding author
Noelia Navarro Gil
Institute for Multilingualism
Josep Trueta, s/n
08195 Sant Cugat del Vallès
e-mail: nnavarrog@uic.es

received: June 2018
accepted: November 2018

Appendix 1: List of academic journals used to compile the expert corpus

MED JOURNALS
BMJ Quality & Safety
European Journal of Clinical Investigation
Journal of International Medical Research
Journal of Investigative Medicine
Journal of the Canadian Association of Emergency Physicians
Lancet Neurol
Nursing Older People
Regenerative Medicine
The New England Journal of Medicine
Tissue Engineering

LIN JOURNALS
Applied Linguistics
Computer Learner Corpora
Second Language Acquisition and Foreign Language Teaching
Corpora and Language Teaching
English for Specific Purposes
Journal of Second Language Writing
Language Teaching Research
Lingua
Linguistics and the Human Sciences
TESOL Quarterly
Text: Interdisciplinary Journal for the Study of Discourse

Appendix 2: Global results for metadiscourse categories in the learner and the expert corpus (raw and normalized results per 1000 words)

Reflexive metadiscourse		BDs				RAS			
Discipline		LIN	LIN norm	MED norm	MED	LIN	LIN norm	MED norm	MED
Tokens		65,180			38,791	177,041			81,182
Types		5,537			4,656	9,853			7,553
METATEXT	Tags								
Reference to the text									
Full text	_MD_MT_RT_FT	115	1.76	1.39	54	321	1.81	1.22	99
Part of the text	_MD_MT_RT_PT	128	1.96	1.16	45	194	1.10	0.60	49
Semiotic modes	_MD_MT_RT_SM	79	1.21	0.80	31	475	2.68	3.07	249
Total RT		322	4.94	3.35	130	990	5.59	4.89	397
Endophoric markers									
Anaphoric	_MD_MT_EN_AN	90	1.38	0.46	18	159	0.90	0.59	48
Cataphoric	_MD_MT_EN_CA	53	0.81	0.67	26	210	1.19	0.64	52
Deictic	_MD_MT_EN_DE	13	0.20	0.00	0	132	0.75	0.02	2
Total EN		156	2.39	1.13	44	501	2.83	1.26	102
Code Glosses									
Reformulators	_MD_MT_CG_RE	138	2.12	1.01	39	442	2.50	1.34	109
Exemplifiers	_MD_MT_CG_EX	155	2.38	0.80	31	696	3.93	1.64	133
Parentheticals	_MD_MT_CG_PA	266	4.08	7.53	292	618	3.49	5.57	452
Dashes (-)	_MD_MT_CG_DA	22	0.34	0.00	0	40	0.23	0.16	13
Colons (:)	_MD_MT_CG_CL	192	2.95	2.55	99	277	1.56	0.65	53
Semicolons (;)	_MD_MT_CG_SC	67	1.03	0.62	24	216	1.22	1.68	136
Total CG		840	12.89	12.50	485	2289	12.93	11.04	896
Linking Devices									
Additive	_MD_MT_LD_AD	158	2.42	2.60	101	282	1.59	1.88	153
Constrastive	_MD_MT_LD_CN	304	4.66	2.81	109	785	4.43	2.82	229
Consecutive	_MD_MT_LD_CO	110	1.69	1.50	58	347	1.96	1.13	92
Organizers	_MD_MT_LD_OR	152	2.33	1.16	45	388	2.19	1.39	113
Topicalizers	_MD_MT_LD_TO	44	0.68	0.36	14	184	1.04	0.18	15
Total LD		768	11.78	8.43	327	1986	11.22	7.42	602
Total METATEXT		2086	32.00	25.42	986	5766	32.57	24.60	1997
INTERPERSONAL	Tags								
Writer oriented									
Self-mention	_MD_IP_WO_SF	119	1.83	5.13	199	725	4.10	3.82	310
Reader oriented									
Directives	_MD_IP_RO_DI	28	0.43	0.90	35	200	1.13	0.17	14
Rethorical questions	_MD_IP_RO_RQ	3	0.05	0.00	0	12	0.07	0.00	0
Total RO		31	0.48	0.90	35	212	1.20	0.17	14
Participant oriented									
Inclusive *we*	_MD_IP_PO_IW	58	0.89	0.34	13	173	0.98	0.04	3
Total INTERPERSONAL		208	3.19	6.37	247	1110	6.27	4.03	327
Total METADISCOURSE		2294	35.19	31.79	1233	6876	38.84	28.63	2324
Total MD %			3.52%	3.18%			3.88%	2.86%	

Appendix 3: Top-3 textual and interpersonal markers in each corpus[4]

	LIN		MED	
TEXTUAL	BDs	RAs	BDs	RAs
Reference to the text				
Full text	(this) paper (64)	(current, this) study (213)	(this) study (43)	(current, present) study (77)
	(this) study (21)	(this) paper (75)	(this) project (4)	(this) paper (8)
	(this, final) project (10)	(this) article (23)	(this) document (2)	(our) trial (5)
Part of the text	(in this) section (45)	(in this) section (89)	(see) appendix (15)	(in) appendix (19)
	(in) appendix (20)	(see) appendix (17)	(see) annex (10)	(in) sections (14)
	(in the) introduction (6)	(in the) discussion (14)		(see) annex (2)
Semiotic modes	(in) table (27)	(in) table (149)	table (11)	figure (83)
	figure (17)	(in) figure (63)	figure (7)	table (48)
	in (x) (15)	in (x) (39)	diagram (4)	image (4)
Endophoric markers				
Anaphoric	(explained, stated) above (14)	(noted, listed) above (75)	above (8)	(described) previously (16)
	(the) latter (11)	(the) latter (21)	(as) mentioned (3)	(described) above (9)
	(the) previous (7)	(as) mentioned (12)		(as) mentioned (5)
Cataphoric	(in the) following (22)	(are the, in the) following (86)	(the) following (14)	(the) following (17)
	(as) follows (8)	(discussed) below (64)	(as) follows (7)	(as) follows (10)
	(described) below (3)	next (section) (12)		
Deictic	here (we) (9)	(adopted, used) here (94)	N.A.*	here (1)
	now (we) (3)	(let us) now (26)		
		so far (8)		
Code Glosses				
Reformulators	i.e. (21)	i.e. (114)	especially (12)	specifically (32)
	(defined, known, referred to) as (19)	especially (55)	defined as (8)	defined as (26)
	that is, (17)	particularly (50)	specifically (6)	especially (14)
Exemplifiers	such as (38)	e.g. (243)	such as (13)	such as (66)
	(for) instance (22)	such as (138)	e.g. (7)	for example (36)
	e.g. (20)	for example (135)	for instance (4)	e.g. (22)
Parentheticals	refer to sections (7)	list examples (13)	refer to sections (23)	refer to semiotic modes (68)
	list examples (4)	refer to sections (5)	specify type of variable (7)	
	refer to semiotic modes (4)	cataphoric markers (3)		
Dashes (-)[5]	- also known as (2)	- and (4)	N.A	- and (5)
	- e.g. (2)	- thus (2)		- for example (2)
	- i.e. (2)	- that is, (2)		
Colons (:)	for example: (8)	the following: (6)	are: (3)	as follows: (4)
	are: (6)	categories: (5)	for example: (2)	
	as follows: (4)	research question: (3)		
Semicolons (;)	; the (13)	; and (23)	; however (3)	; and (16)
	: and (7)	; however (12)		; however (10)
	; in (6)	; see (7)		; therefore (4)
Linking Devices				
Additive	moreover (32)	in addition (57)	moreover (16)	in addition (29)
	furthermore (23)	moreover (30)	furthermore (15)	additionally (29)
	another (12)	another (27)	in addition (11)	furthermore (18)
Contrastive	however (76)	however (247)	however (54)	however (89)
	whereas (49)	although (109)	although (5)	although (46)
	although (30)	while (75)	nonetheless (6)	while (20)
Consecutive	thus (35)	thus (184)	therefore (29)	therefore (42)
	therefore (30)	therefore (103)	thus (12)	thus (26)
	hence (16)	hence (24)	consequently (5)	as a result (10)
Organizers	finally (24)	(the) second (127)	respectively (7)	respectively (25)
	on the one hand (13)	finally (44)	(the) second (5)	finally (15)
	first (11)	third (42)	then (5)	then (14)
Topicalizers	in the (case, context) of (13)	in (terms, the case, the context) of (63)	in terms of (6)	with respect to (8)
	regarding (9)	with (respect, regard) to (45)	regarding (4)	in the context of (5)
	as far as (x) is concerned (2)	regarding (21)	as for (2)	with regard to (2)

[4] The function *Cluster* in the text analysis software AntConc has been used to identify the top-3 markers in each category; a minimum range of 2 was set (i.e. markers had to be present in at least two different texts to be included in the top-3 list).

[5] In the case of dashes, colons and semicolons, we provide the words that followed or preceded these marks more frequently. As for parentheticals, we indicate three of the most frequent functions they perform in all texts – i.e. contain lists of examples, refer to semiotic modes or to parts of the text.

	LIN		MED	
INTERPERSONAL	BDs	RAs	BDs	RAs
Writer oriented				
Self-mention	we (44) (have, can, found)	we (410) (will, have, examined)	we (will, have, expect) (133)	we (194) (found, used, examined)
	I (40) (would like to)	our (188) (study, data, investigation)	our (study, results) (55)	our (112) (study, findings, knowledge)
	our (20) (findings, analysis)	I (69) (have, will, would)	(allows) us (6)	
Reader oriented				
Directives	see (21)	see (118)	see (32)	see (15)
	consider (1)	cf. (30) consider (12)		
Rhetorical questions	N.A.	N.A.	N.A.	N.A.*
Participant oriented				
Inclusive *we*	we (can see, have seen) (45)	we (can see, need) (129)	we (can, need) (13)	we (should) (3)
	(let) us (6)	(gives, helps, let) us (22)		

*Non-Applicable

Appendix 4: List of reflexive metadiscourse markers found in the corpora

Textual Metadiscourse

Full text	Part of the text	Semiotic modes	Anaphoric	Cataphoric	Deictic	Reformulators	Exemplifiers
my paper	analytical framework	diagram	above	as follows	here	also known as	an example
our investigation	annex	extract	abovementioned	below	now	at the same time	and so on
our research	annexes	fig.	aforementioned	follows with	so far	defined as	and so on and so forth
our study	appendices	figure	Again	further on	up to this point	especially	as in
our work	appendix	fragment	as it has been mentioned	in the following		generally speaking	be it
the current article	in the analysis	graph	as mentioned	in the next		i.e.	e.g.
the current study	in the results	in (	as noted above/earlier	later on		in other words	for example
the present paper	section		as seen in	subsequent		in particular	for instance
the present research	sections		the first of	the following		known as	like
the present study	subsection		the former			more accurately	such as
the study	the conclusion		the latter			more specifically	
this article	the discussion		this first			namely	
this essay	the introduction		this second			particularly	
this investigation	the methodology					put differently	**Parentheticals**
this paper	the theoretical					put it	()
this project						put it simply	
this study						referred to as	**Colons**
this trial						simply put	:
this work						so-called	
						specifically	**Semicolons**
						that is to say	;
						that is,	
						to be more precise	**Dashes**
							–

Interpersonal metadiscourse

Adding	Contrasting	Consecutive	Organizers	Topicalisers	Self-mention	Directives	inclusive we
additionally	alternatively	as a consequence	(a) (b) (c)	as far as	I	cf.	brings us
Also	although	as a result	(i) (ii) (iii)	as for	me	consider	if we
And	and yet	consequently	1) 2) 3)	as regards	our	if you look at	let us
Another	But	hence	Afterwards	concerning	the authors'	note that	we can
apart from	but still	So	all in all	in regard to	us	one can	we find
as well as	by contrast	therefore	eventually	in terms of	we	one could	we may
Besides,	contrarily	thus	finally	in that light		one could	we might
furthermore	contrastively		first of all	in the case of		one might	we see
in addition	conversely		first,	in the context of		see	we should
moreover	despite		firstly	regarding		view	we will
Other	even though		in the first place	turning to			
Similarly	however		Last	with regard to			
	in any case		lastly	with respect to			
	in contrast		next,				
	in spite of		on the one hand				
	in turn		overall				
	instead		pn the other hand				
	nevertheless		respectively				
	nonetheless		Second,				
	notwithstanding		secondly				
	on the contrary		The first				
	on the other hand		The first of				
	otherwise		the last				
	Rather		the second				
	Still		Then				
	though		then,				
	unlike		third				
	whereas		thirdly				
	While		to begin with				
	whilst		To conclude				
	Yet		to sum up				