The transition from university to publication: Register and interactional metadiscourse features in immunology research written in Catalan and English

Ana Pujol Dahme and Moisés Selfa Sastre
Universitat de Barcelona and Universitat de Lleida (Spain)
apujoldahme@ub.edu & mselfa@didesp.udl.cat

Abstract

The ability to use the register and the rhetorical conventions of the research article is important to achieving academic success and professional development in researchers' careers. Numerous studies have focused on research articles across different disciplines and cross-culturally. However, little research has been carried out into the students' research report, from a developmental perspective and in a different language from English. To address this gap, we are reporting on a longitudinal study that aimed to characterize the transition of the academic register and the interactional function from university to scientific publication. The research focus is twofold: (1) it examines the academic register by means of lexical diversity, syntactic complexity and lexical density, and (2) it examines and compares the distribution of stance and engagement markers across stages. The data ($N = 16$) consists of university master's theses written in Catalan (Romance language) and published articles in English, in the discipline of immunology, written by the same eight subjects. As discipline-specific writing conventions are an integral aspect of determining writing proficiency, overall findings suggest that students have not yet acquired writing proficiency, either in academic register or in writer-reader interactions.

Key words: research articles, register, metadiscourse, Catalan, English.

Resumen

La transición de la universidad a la publicación: El registro lingüístico y el metadiscurso interaccional en trabajos de investigación de inmunología en catalán e inglés

La habilidad para usar de forma adecuada el registro lingüístico y las convenciones retóricas del artículo de investigación es importante para lograr el éxito académico y el desarrollo profesional en la carrera de los investigadores. Numerosos estudios se han centrado en artículos de investigación a través de diferentes disciplinas e interculturalmente. Sin embargo, pocas investigaciones se han ocupado del análisis de la escritura del trabajo de investigación de estudiantes, desde una perspectiva de desarrollo y en un idioma diferente al inglés. Para llenar este vacío, se presenta un estudio longitudinal que tiene como objetivo caracterizar la transición del registro académico y la función interaccional desde la universidad a la publicación científica. El foco de la investigación es doble: (1) se examina el registro académico por medio de la diversidad léxica, la complejidad sintáctica y la densidad léxica, y (2) se examina y compara la distribución de los marcadores de postura (stance) y afiliación (engagement) a través de los niveles universitario y publicaciones. Los datos (*N* = 16) proceden de Trabajos Finales de Máster escritos en catalán (lengua románica) y artículos publicados en Inglés, en la disciplina de la inmunología, escritos por los mismos ocho sujetos. Las convenciones de escritura específicas para cada disciplina son un aspecto integral de la determinación de la competencia de escritura. Los resultados generales sugieren que los estudiantes aún no han adquirido el dominio de escritura, ni en el registro académico ni en las interacciones escritor-lector.

Palabras clave: artículo de investigación, registro, meta-discurso, catalán, inglés.

1. Introduction

Academics construct knowledge as members of professional groups and they communicate their findings usually through research articles, which are also part of the process allowing the integration of new scholars into different discourse communities (Hyland, 2008a). Students in their transition from university into the professional research community have to master the academic register of the research article, as there is an increased pressure to communicate research through this genre. They have to learn not only the linguistic register for communicating properly in academic English, but also the rhetorical conventions that their discourse community considers

convincing (Berkenkotter, Huckin & Ackerman, 1991). The function of the research article (henceforth RA) is to report on investigations through publication and therefore it is an important channel for sharing knowledge in the academic culture (Swales, 1990). We understand genre as a social act containing discourse community's forms of knowing and acting (Miller, 1984). When students, for instance, learn to use the RA, they need to assimilate specific linguistic forms to communicate knowledge and also the practices and norms of the scientific community (Bawarshi & Reiff, 2010). These impose social, cognitive and linguistic demands, especially for students, whose native language is not English, as it means fitting in with the Anglo-American academic writing tradition (Hyland & Hamp-Lyons, 2002; Rienecker & Stray Jörgensen, 2003).

From the perspective of Contrastive Rhetoric, academic writing is viewed as culturally determined and two major styles of writing can be distinguished: formal-oriented cultures (e.g. English) and content-oriented cultures (e.g. German, Spanish, Catalan) (Clyne, 1994; Cuenca, 2003). English is the international language for dissemination of research, especially in Biomedical Sciences (Swales, 2004). Proficiency in academic English is an important matter for students. In the Spanish academic context, for instance, Martín, Rey-Rocha, Burgess and Moreno (2014) reported that for scholars in Medical Science, communicating in English is certainly a hurdle and they are concerned with learning the strategies to enhance the contribution of their research.

Moreover, there is a broad consensus of the dialogic and interactive nature of specialized academic discourse (Flowerdew, 2014). In written academic texts, such as RAs, writers seek to present their arguments dialogically. They have to situate their language use to express their authorial voice and guide the readers of the discourse community to a particular interpretation, so as to persuade them of the claims. This could be accomplished by means of interactional discourse markers, such as stance and engagement ones (Hyland, 2008a). Although presenting an awareness of self and audience is important to achieve academic success and professional development, it is challenging for students to properly employ the interactional resources of their discourse community. Non-native English-speaking students usually experience difficulties to manage their authorial voice (Henderson & Barr, 2010; Nelson & Castelló, 2012). Native English-speaking students show considerable variations of forms, structures and functions of lexical bundles in comparison with published academic writing (Cortes, 2002; Hyland, 2008b).

Therefore, providing evidence to explain how writers use the academic register, lexico-grammatical and discourse features is crucial to improving writing skills towards an expertise stage. Whereas the importance of writer and reader representation in academic register is well-established by developmental research on writing (Bereiter & Scardamalia, 1987; Kellogg, 2008), in different disciplines (Kuo, 1999; Cortes, 2004; McGrath & Kuteeva, 2012) and cross-culturally (Mur, 2007; Vázquez, 2010), very little is known about how these features develop through different stages and from one language to another.

The present study addresses this gap in two ways: (1) it analyzes the academic register through grammatical and lexical measures, such as lexical diversity, syntactic complexity and lexical density in biology texts, specifically, in the immunology field, in university master's theses written in Catalan and published articles in English written by the same subjects. (2) It also analyzes and compares the distribution of stance and engagement markers across these stages.

Below, we first focus on the characteristics of theses and RAs, likewise we provide a theoretical overview of the lexico-grammatical and discourse features of the academic register. Then, we examine these features in a corpus-based study, in order to provide some quantitative and qualitative data. Finally, we discuss the results and present some conclusions.

2. Master's theses and research articles: Different genres

Students' assignments are usually considered as a less valued genre since they are normally read only by teachers. The concept of genre has been defined commonly as a set of events sharing particular communicative purposes (Swales, 1990). So, genre is defined as a social action. Bazerman (1988: 62) understands genre as "a socially recognized, repeated strategy for achieving similar goals in situations socially perceived as being similar". In this sense, following Miller (1984: 165) "genre serves as keys to understanding how to participate in the actions of a community". Master's theses and RAs are not the same genre. However, the goal of most academic writing is to develop the ability to write professional texts like, for instance, scientific texts (Russell & Cortes, 2012; Gardner & Nesi, 2013). As stated by Gardner & Nesi (2013: 28) "the RA is a model for the sort of writing that students aspire to produce".

Students' assignments, such as master's theses, differ from research genres in various aspects. They vary not only in the targeted audience, but also in terms of purpose and the requirement of skills and knowledge. While students' assignments aim to demonstrate the acquisition of required skills, RAs aim to report new experiments and persuade the reader of the validity and importance of new findings (Gardner & Nesi, 2013). Despite these differences, there are some similarities between both types. They are results of a complete piece of a research process. Both include research aim/question, investigation, links and relevance to other research in the field. Furthermore, these kinds of texts share the fixed structure of the RA in their organizational level (macro-structure): Introduction, Method, Results and Discussion section (IMRD), which follows the steps of the research process and gives coherence at the rhetorical level. Students and professionals have shared a typified way of acting within the inquiry process, namely, they have written their research (Mauranen, 1993). So, in both cases, the final stage of the inquiry process is a written product expected to communicate knowledge. Therefore, because of their similarities, we have decided to analyze master's theses and compare them with the RAs, which could offer us an insight into the development of writing in the specific discipline.

3. Academic register

A register is the configuration of lexical and grammatical features that characterizes particular uses of language (Halliday & Hasan, 1989; Schleppegrell, 2001). As the point of departure of our study derives from psycholinguistic research on later language development, we conceive mastering of register as the ability to use linguistic forms properly in different communicative contexts (Ravid & Tolchinsky, 2002; Berman, 2004). This not only entails linguistic knowledge, in this case, of specialized scientific terms, but also "communicative competence" (Hymes, 1972). We assess register by means of the following lexico-grammatical features: lexical diversity, syntactic complexity and lexical density. The academic written register generally has a high lexical diversity, which refers to the different words used in a text (Chafe & Danielewicz, 1987) and it has been shown to be characteristic of advanced proficiency (Malvern, Richards, Chipere & Durán, 2004). Lexical diversity differs according to age and genre (Berman & Verhoeven, 2002; Johansson, 2008). Further, academic register is

syntactically more complex than colloquial language; in the sense of structurally "compressed". This observation is important, since it changes the stereotypical notion that complexity in academic writing lies in the high use of subordinate clauses (Biber & Gray, 2010). Recent corpus studies on professional RAs have shown that syntactic complexity relies on phrasal (non-clausal) modifiers embedded in noun phrases, rather than on clausal subordination (Biber, Gray & Poonpon, 2011). Density is also a characteristic of academic language and it is associated with more propositional content. It is understood as the proportion of content words per total words (Schleppegrell, 2001), which are usually shown using a nominalized structure (Halliday, 1993).

These characteristics contribute to the complexity of the academic register and usually make it more difficult to read (Ventola, 1996). However, although this complexity goes against "reader-friendliness", greater lexical diversity and more complex syntax may be an indication of more skilled writing production. In fact, if proficiency is judged by sophisticated production rather than easy readability, then higher quality is related to these features (McNamara, Crossley & McCarthy, 2010). These features have been studied in high school students' persuasive essays (Uccelli, Dobbs & Scott, 2013) and essays written by freshman college students (McNamara, Crossley & McCarthy, 2010), and in both studies they found them to be predictive of writing quality. Despite the evidence that, for instance, syntactic complexity correlates with text quality, other studies show inconsistency when trying to relate both features (Beers & Nagy, 2009). Based on a review of the literature, these authors suggest that this inconsistency could be due to the variety of measures, genres and ages across studies.

Lexical density characterizes academic register (Schleppegrell, 2001). This is usually achieved through nominalizations, especially in the scientific community, for instance, in medical written journals (Guillén Galve, 1998). This could lead to the belief that dense texts, with high use of nominalizations, are more formal and prestigious and therefore students would aspire to acquiring this rhetorical effect (Baratta, 2010). However, as Ventola (1996) highlights, in order to write successful texts, writers should keep a balance in their use of lexical density. Accurate and conscious linguistic choices are more important than high lexical density. An excessive use of nominalizations gives the sensation of a hard text without fluency (Ventola, 1996). Moreover, we have to take into account, as pointed by Baratta (2010: 1018), that "there may indeed be discipline-specific

writing conventions with regard to how frequently nominalizations are used".

Few studies have analyzed the use of nominalization (Colombi, 2002; Baratta, 2010), from a developmental perspective. Colombi (2002) focused on how bilingual university students developed their academic register in Latino Spanish essays. The findings suggest that as students' writing skills develop, their lexical density and nominal structure grow while grammatical intricacy decreases. Baratta (2010) analyzed in students' English essays the frequencies and function of nominalizations across an undergraduate program. Results reveal proficient development with regard to nominalization usage within their academic community. Some studies which have attempted to relate lexical density with writing quality in L2 students (e.g. Engber, 1995) reveal that lexical density has no relation with quality.

Furthermore, our study has a different purpose from those studies, as our motivation is not to judge writing quality, but to characterize the academic register of the RAs written by students in university and as professional researchers, because a connection between register and literacy has been demonstrated (Ravid & Berman, 2009). Advanced literacy refers to meaning-making abilities in university and these are required for participation in professional or social institutions (Schleppegrell & Colombi, 2002).

4. Interactional metadiscourse

Metadiscourse, defined as "how writers situate their language use to include a text, a writer and a reader" (Hyland & Tse, 2004: 167), gives at an advanced level of writing an account of how writers negotiate propositional information properly to a specific discourse community. Further, Snow and Uccelli (2009), in their proposal for a research agenda, point out that using the taxonomy of metadiscourse markers might prove relevant for the study of academic language from a developmental perspective. Within this model of metadiscourse, the interactional dimension is linked to the dialogic function in academic writing and comprises stance and engagement resources. Stance can be defined as the expression of attitude, feelings, judgments or commitment concerning the propositional content of the message by means of grammatical and lexical features (Biber & Finegan, 1989). Engagement, on the other hand, refers to linguistic features used by the writer to involve the reader, to guide the reader to the interpretation of

the argument and to include him as a participant of the discourse (Hyland, 2008a).

Within stance, the linguistic resources which have been found to be useful for the interpersonal function are self-mention markers, as they highlight the authorial presence (Martínez, 2005; Mur, 2007) and the epistemic modality markers, boosters and hedges, since their function is to express the writer's commitment or lack of commitment to the truth of the proposition (Vázquez & Giner, 2008).

Regarding the authorial presence, academic texts are often described as detached or informational as opposed to involved (Biber, 1993). The informational dimension is marked by frequent occurrences of nouns, adjectives, prepositional phrases, etc., whereas the involvement dimension is expressed through 1st and 2nd personal pronouns, questions, and hedges (Biber, 1993). Although involvement, and therefore the use of personal pronouns, is prototypical of face-to-face conversations (Tannen, 1985), when writing RAs, researchers must emphasize the importance of their findings (Kuo, 1999). They have to construct their social identity, as researchers of the scientific community (Berkenkotter & Huckin, 1995), and this is discursively constructed usually through first person pronouns (Kuo, 1999).

Further, the use of epistemic modality markers to express certainty (boosters) and doubt (hedges) have been the focus of much research as it is central to academic writing (Crismore, Markkanen & Steffensen, 1993; Salager-Meyer, 1994). However, boosters have been studied less extensively than hedges, although the two concepts are equally interesting topics of research (Vázquez & Giner, 2008). It seems a general claim that members of the discourse community cannot make categorical statements about their findings. They should balance their claims with an appropriate degree of commitment (Lafuente, 2008). That means to express possibility or project an image of humility (Swales, 1990; Hyland, 1996). Beyond this, the fact is that the use of certainty markers has to do with some constraints established in scientific professional research. A general assumption is that more robust results require fewer hedges, although the results in the use of these two markers could vary, especially in scientific RAs, according to the degree of industrial sponsorship or journal impact factor (Gross & Chesley, 2012).

Hedging is important in academic writing and clarifying the motivation for its use is interesting, especially in the field of English for Specific Purposes

(Markkanen & Schröder, 1997; Lewin, 2005), as even proficient L2 students have difficulties in hedging (Hyland, 1996). By means of hedging, writers can present a proposition as a possibility, an opinion, rather than as a factual argument and this might help them gain the reader's acceptance (Hyland, 1998). Hedges are not intrinsic to texts, their meanings are realized through writer-reader interaction and this is linked to the specific discourse community (Markkanen & Schröder, 1997).

From a cognitive developmental perspective, the use of hedges implies the consideration of multiple perspectives and this is a factor in shaping the development of a proficient academic writer (Berman, 2004). Prior research on metadiscourse found that, whatever the discipline, proficient academic writers use more hedges (Hyland & Milton, 1997; Aull & Lancaster, 2014). Other studies found this to be a predictor for the writing quality of persuasive essays in English (Uccelli, Dobbs & Scott, 2013). The use of more hedges itself is not always an indicator of writing quality, as students could have learned hedges as decontextualised lexical formulae and apply them in an indiscriminate way (Hyland & Milton, 1997). Despite this, writers in all disciplines, including the "hard" ones such as biology, use hedges (Hyland, 2005).

We have to bear in mind that most general academic writing manuals in Spanish (e.g. Montolío, 2000) suggest that in academic writing, references to the writer or reader should not appear. Also, Mendiluce (2005), in his review of scientific manuals, concluded that in general they do not deal with discursive markers, such as boosters or hedges. In contrast, students at university are requested increasingly to adopt an explicit position on the issues on which they write in order to build their own voice (Castelló et al., 2010). Therefore, it represents a challenge for students to cope with the rhetorical ways in which scholars communicate their findings.

To summarize, it is suggested that lexical diversity, syntactic complexity, lexical density and the rhetorical markers are worth investigating, assuming that these are features that characterize the academic register and the interactional strategy used by the specific community. In addition, by analyzing these features in the transition from one context to another, from university to the professional world, it is expected that information can be found on developing writing abilities in a specific discourse community.

5. Method

5.1. Description of the corpora

Data for this study consist of master theses in immunology in Catalan (n=8) from the TARBUC corpus and published RAs from immunology journals in English, collected from PuBMed (n=8). The TARBUC corpus (Treballs Acadèmics de Recerca de Batxillerat i Universitat en Català) is constituted by 114 academic research reports. It comprises 60 reports, of the disciplines of biology and history, written by twelfth-grade high school students (age 17-18) and 54 master's theses produced at the University of Barcelona (Spain). These disciplines were chosen because they represent different methodologies and research traditions (Cortes, 2004). Biology is based on experimental-quantitative methodology and history on qualitative or primary source research to present evidence. The study reported here focuses only on master's theses in Catalan and published RAs in English in the discipline of immunology (see the list of titles in appendix 1).

5.2. Preparation of the corpora

The original RAs were collected in digital format (pdf) and two different versions of the corpus were created: a plain text version (.txt) without figures, graphs and format, and a second plain text version, which was morphologically tagged (see example below) using FreeLing 3.1 (Padró & Stanilovsky, 2012).

```
En en SPS00
les el DA0FP0
cèllules cèllules NCFP000
hematopoètiques hematopoètic AQ0FP0
els el DA0MP0
sindecans sindecans NCMS000
també també RG
juguen jugar VMIP3P0
un un DI0MS0
paper paper NCMS000
```

Figure 1. Morphologically tagged output from FreeLing 3.1.

As illustrated, the output contains three columns: word-lemma-part of speech (PoS). The first version was used to analyze syntactic complexity, measured in terms of words per sentence, and lexical diversity and to identify metadiscourse markers of stance and engagement. The second version was used to analyze syntactic complexity indexed by number of words before the main verb, and lexical density.

5.3. Categories of analysis: Lexico-grammatical features

To examine changes in the academic register of RAs indexed according to a selection of lexico-grammatical features (lexical diversity and density, syntactic complexity) data were analyzed using a shell script for the Linux Operating System. This free and open source software enables script design (a collection of commands stored on a file) *ad hoc*, that is, the analysis of specific linguistic features within any text. It is worth mentioning that all the analysis tools we have employed are open source; this enables the reproducibility of our study. With Linux the following lexical and syntactic measures were generated.

Text Length: calculated as the number of sentences per text. A sentence refers to the orthographic unit that is inserted between full stops. Although most of the studies reported above have used the clause as a text segmentation unit, our segmentation unit is the sentence. We consider the sentence as a feasible unit of written language (Berman & Ravid, 2009), because it has a meaningful and complete syntactic relationship (Szmrecsanyi, 2004).

Lexical diversity: measured through the "subsample variety" with the Textable open source data analysis tool (Xanthos, 2014), using the average of word types in subsamples of a given size randomly drawn from the data. We used this measurement, because many other measures, such as type/token ratio have been known to be sensitive to the length of the text (McCarthy & Jarvis, 2010) and "subsample variety" is a simple but robust way to measure lexical diversity (Xanthos, 2014).

Syntactic complexity (1): calculated as the number of words per sentence. Sentence length was used as one of the two indexes of syntactic complexity. Scientific writing has little verbosity, as shown in the example (a) from the master's theses, therefore we decided that it was not necessary to sort out roundabout phrasing.

> a) El tractament consisteix en l'administració d'antibiòtics d'ampli espectre de seguida que se sospita el shock, suport vital intensiu (degut a la hipotensió cal aplicar fluïdoteràpia i fàrmacs vasoactius) i neutralització de les toxines bacterianes amb la injecció d'immunoglobulines intravenoses.
>
> The treatment involves the administration of broad-spectrum antibiotics immediately when a shock is suspected, intensive life support (due to hypotension, fluid therapy and vasoactive drugs should be applied) and

neutralization of bacterial toxins with injection of intravenous immunoglobulin.

Syntactic complexity (2): Another measure of syntactic complexity was calculated using the number of words before the first finite verb, and according to the PoS tagger output with a shell script.

Lexical density: calculated as the frequency of content words (nouns, adjectives, verbs and some adverbs) in a text as a ratio of total sentences. Adverbs not considered content words are intensifiers, delimiters, connectives and deictic adverbs, whereas those which have referential meaning are considered content words. Lexical density has been measured, by first applying the *FreeLing 3.1* open source language analysis tool (Padró & Stanilovsky, 2012). It marks up all the words in a text as part-of-speech (PoS), that is, it assigns its lexical category to each word. Once this was done, a script for Linux was used to compute all the content words in a sentence as a ratio of total sentences. Two results were obtained: content words with adverbs and those without, as the PoS tagger does not discriminate between different adverbs.

5.4. Categories of analysis: Interactional markers

Based on an approach used by Hyland (2005) the texts were coded for the metadiscoursal functions that account for stance and engagement. The following grammatical categories account for this function: lexical verbs in conditional mood, modal verbs, verbal periphrasis, adjectives, adverbs and personal pronouns.

Stance: this refers to an attitudinal dimension. It is the writer's textual "voice" and concerns how the writer presents himself and conveys his judgments, opinions and commitments. It is mainly realized through four resources: Hedges: they indicate the writer's reluctance to express a complete commitment to a proposition (e.g. *possible, might, perhaps*). Boosters: they signal the writer's certainty and involvement in a proposition (e.g. *sure, prove, definitely*). Attitude markers: they indicate the writer's affective attitude to a proposition, conveying surprise, importance and so on (e.g. *important, surprisingly, interestingly*). Self-mention markers: they explicitly refer to the writer, to the authorial identity, by means of personal pronouns and possessive articles (e.g. *we, our*).

The following examples were coded as stance markers in the present study: (1) hedges, (2) self-mention and boosters, and (3) attitude.

1. *Sembla* que activin STAT1...

   They *seem* to stimulate…

2. Cosa que *ens* porta a *afirmar* que ...

   Which leads *us* to *state* that ...

3. És *important* destacar que si l'anàlisi ...

   It is *important* to stress that if the analysis...

Engagement: this refers to a dimension where writers include readers as discourse participants by anticipating their possible objections and guide them towards interpretations. Engagement is mainly realized by reader pronouns, which are markers that signal the inclusion of the reader as a member of the discipline and guide readers through an argument as if they were participating in the discourse with shared goals (e.g. inclusive *we*). Directives: they are markers that instruct the reader to look at another part of the text (e.g. *see Figure 1*), how to carry out some action (e.g. *open the valve*) or interpret an argument (e.g. *note*). Appeals to shared knowledge: they are markers that give readers a participating role in the construction of the argument, allowing them to recognize the knowledge as familiar (e.g. *well-known*, *obviously*). Questions: they are markers that invite engagement.

The following examples accounted for some engagement markers: (4) reader pronouns, (5) directives, (6) appeals to shared knowledge.

4. La ressenya científica *ens* indica que la resposta...

   The scientific review indicates to *us* that the response…

5. *En la figura 6 veiem* els nivells...

   In Figure 6 we see levels ...

6. *Se sap* que malgrat...

   It is known that even...

Using this classification is a viable option for analyzing corpora, as is comparing the frequency of use of markers in different languages, such as Catalan and English, because although the linguistic forms differ, the interactional function is mostly the same.

The texts were explored with a freeware concordance program, AntConc 3.2.4w (Anthony, 2011), using a two-step procedure. First, a closed list of the linguistic forms which accounted for writer- and reader-oriented function was created (e.g. writer-oriented: El *nostre* objectiu general: *our* main goal; reader oriented: *Se sap* que: *it is known* that). Second, a more qualitative analysis was undertaken to check the functions were fulfilled by the linguistic forms. For this purpose we used the concordance program, AntConc 3.2.4w, to determine the context in which the closed list of forms appeared and to decide their specific function according to that context.

As shown in (7a) and (7b), the same linguistic form (*mostra*) may function as a referential term (7a) or as an interactional marker (7b), where the linguistic form addresses the reader with a directive marker.

7a Els microlitres de *mostra* que fóssin necessaris per a tenir 200mg

   The sample microliters needed to achieve 200mg

7b Tal i com es *mostra* a la figura 3.

   As *shown* in figure 3.

Furthermore, the same linguistic form (*indica*) may account for two different functions. In (8a), it functions as a directive marker to the reader whereas in 8b it functions as a booster marker, to signal the writer's certainty.

8a Tal com *s'indica* en la taula.

   As *indicated* in the table.

8b Fet que *indica* que s'ha dut amb èxit el procés.

   That fact *indicates* the process has been successful.

An independent rater, a degree holder in biology and linguistics who was familiar with the categories, checked 30% of the linguistic forms and decided their specific function according to that context. The comparison between the rater's coding and ours revealed an inter-rater reliability of 0.86 (Kappa), indicating a high level of agreement. To solve the cases of disagreement, a second rater, a linguist, discussed the cases with us to reach a consensus.

6. Results

Descriptive statistics were estimated for the corpus' lexical-grammatical features and distribution of interactional markers in university and in publications. Due to the data, a non-normal distribution, the Wilcoxon signed-rank test was applied for testing the effect of academic level on interactional markers and lexical-grammatical features. To estimate the effect size, the following equation has been used in which z is the z-score and N is the number of total observations that were made (Field, 2009).

Table 1 shows the statistical comparisons, medians, z-scores and effect size, for the lexico-grammatical features in university master theses and publications. Results indicate that university texts have significantly more words than publications. Publications showed significantly higher lexical diversity than university texts. When measuring syntactic complexity, as the mean number of words in a sentence, texts produced at university were more complex than publications. However, for the second measure of syntactic complexity, the mean number of words before the first finite verb in a sentence, the opposite is true: publications were more complex than university texts. University texts showed significantly higher lexical density, with adverbs, than publications.

Lexico-gramatical features	University (n = 8) Mdn	Journals (n = 8) Mdn	Z	r
Number of words	6,275	4,219	−2.38**	−.20
Number of sentences	202.50	165	−1.68	ns
Lexical diversity	76.00	78.00	−2.41**	−.21
Syntactic complexity (mean number of words in a sentence)	30.50	25.00	−2.38**	−.20
Syntactic complexity (mean number of words before the first finite verb)	5.00	10.50	−2.55**	−.22
Lexical density (with adverbs)	17.00	14.00	−2.38**	−.20

Table 1 Statistical comparisons for lexico-grammatical features in university master theses and publications, medians, z-scores and effect size **$p < .01$.
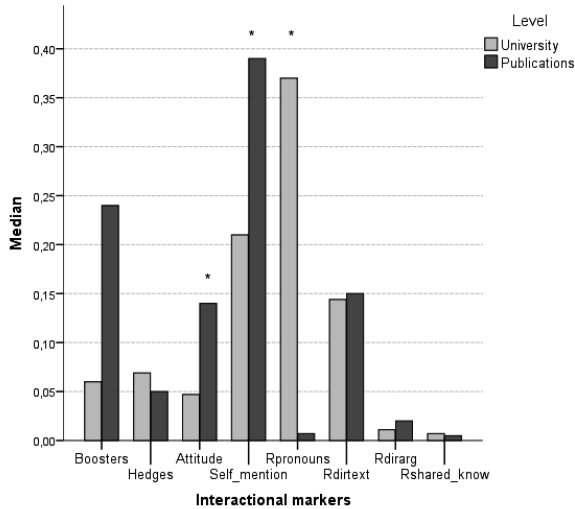
Figure 2. Medians of proportion of each interactional marker in university RAs and publications *p<.05, **p<.01.

At the discourse level, results reveal a higher use of stance markers (Mdn = 0.82, Mdn = 0.41) in publications and university, respectively, than engagement markers (Mdn = 0.17, Mdn = 0.20). It is shown that the difference between levels for stance markers is significant z = –2.52, p = .01, r = –.22 and that the difference in the use of engagement markers is not significant z = –.84. Within the stance and engagement markers (Figure 2) it is shown that publications (Mdn = .14) have significantly more stance attitude markers than university texts (Mdn = .05), z = –2.36, p = .01, r = –.20. Publications (Mdn =. 39) showed a significantly higher use of stance self-mention markers than university texts (Mdn = .21), z = –2.10, p = .03, r = –.18. University texts have significantly more engagement reader pronouns (Mdn = .03) than publications (Mdn = .003) z = –2.20, p = .02, r = –.19.

7. Discussion

We analyzed a variety of lexico-grammatical and discourse features, which are assumed to be characteristic of the academic register and specifically of the research article. Our study extends findings of recent developmentally oriented research on academic register (e.g., Uccelli, Dobbs & Scott, 2013)

but in a Romance language, Catalan. An innovative feature of our analysis concerns the transition made by eight subjects, who wrote their texts at university in Catalan and in a professional setting, in English. Our analysis focused on lexical diversity, syntactic complexity and lexical density, and interactional markers as a window on developing writing abilities in a specific discourse community.

We found differences between levels in the length of the texts, measured as number of words, but this result could be explained because of the task instructions. At university, texts are asked to be limited to 30 pages. Our value of lexical diversity is close to those reported and measured with VocD for English written narratives (VocD = 70-80) (McNamara, Crossley & McCarthy, 2010; Uccelli, Dobbs & Scott, 2013) and expository texts (VocD = 80-90) (Berman & Verhoeven, 2002). Higher lexical diversity is considered to be a characteristic of proficient writers (Malvern et al., 2004) and this seems to be consistent with our findings, since publications showed more lexical diversity than university RAs. It is worth noting here that in content-oriented cultures such as Spanish or Catalan, the general trend is an avoidance of lexical repetition, contrary to formal-oriented cultures, like English, which tend towards repetition (Cuenca, 2003). The lower use of lexical diversity in university texts, however, might also be based on the characteristic of the register. Informational registers, such as academic register, are characterized by the use of the same technical terms (Biber, 1993). Despite this, there are some other resources to create diversity in texts, for instance, substitutions of adjectives or deriving adverbs from adjectives. It seems that students do not make use of these techniques. The findings for syntactic complexity, measured as words in a sentence, from our study are consistent with the average of 24-28 words per sentence reported for English scientific prose (Bazerman, 1988; Gross, Harmon & Reidy, 2002). Our result, 30 words per sentence in university, is similar to the average reported for written science in Italian, another Romance language. The difference shown with the publications in English might be due to the syntax in Romance languages, such as Italian or Catalan, which allows a greater structural complexity than English (Scarpa, 2007). For the second measure of syntactic complexity, the mean number of words before the first finite verb in a sentence, the differences might be explained by language typology. Spanish and Catalan allow one to change the order of a sentence, while English is stricter regarding the order of the constituents (Vázquez, Fernández & Martí, 2000).

Furthermore, we found that university texts showed significantly higher lexical density than publications. This difference might be attributed to language differences (Johansson, 2008) or to the fact that the students relate dense texts to those written in a formal register (Baratta, 2010). It can be further conjectured that the students do not balance their use of lexical density properly. The following examples illustrate lexical density and also syntactic complexity (measured as words in a sentence). Consider the first line of a sentence, example (b) of the introduction section of a master's theses, and example (c) the introduction section of a publication. In this master's theses (b) there are 5 nouns (NN), 3 adjectives (JJ) and 3 prepositional (IN) phrases functioning as noun modifiers, before the comma. Whereas in a publication (c) only 3 nouns, 3 adjectives and 1 prepositional phrase to express more or less the same idea.

b) L'asma és una malaltia respiratòria crònica, de caràcter al·lèrgic en la

NN   NN   JJ   JJ   IN NN   JJ   IN

majoria de casos,

NN   IN NN

que té una alta prevalença en la nostra societat Europea actual.

c) Allergic asthma is a common inflammatory disease of the airway,

JJ   NN   JJ   JJ   NN IN   NN

and long-term therapy is aimed at counteracting episodes of bronchospasm and reducing allergic inflammation.

If we analyze these examples in the light of Biber, Gray and Poonpon's (2013) developmental stages of complexity we could say that the students' example (b) is more complex (stage 5: extensive phrasal embedding in the NP: multiple prepositional phrases as postmodifiers, with levels of embedding) than publications (c) (stage 4: more phrasal embedding in the NP: attributive adjectives, nouns as premodifiers). In this sense, we would argue that students do not fit the requirement of the discipline specific conventions.

Despite this, we found in a previous study (Pujol Dahme & Selfa Sastre, in revision) that lexical density increased markedly when comparing Catalan high school---to university research reports. These findings are consistent with the findings of previous studies showing that lexical density increases

with the development of writing skills (Colombi, 2002; Baratta, 2010). Further, with the statement that high lexical density is related to scientific writing (Halliday & Martin, 1993).

A second goal of the study was to examine the distribution of stance and engagement markers across levels. Our finding for the difference of the total writer-oriented markers between levels suggests that mastering the rhetorical discourse form of the specific discourse community is a lengthy process. Linguistic, cognitive and social factors interact in the protracted transition from a native speaker/writer to a skilled speaker/writer (Berman, 2004). Nevertheless, it is not only a developmental or a skill issue; it is also a matter of knowing and adequacy to academic writing. In fact, Hyland's (2005) analysis of 240 RAs in eight disciplines showed that in all disciplines stance markers are more common than engagement features. In our sample university students use a significantly higher use of reader pronouns, which are engagement markers, compared with publications, which show a greater proportion of self-mention markers. The lack of an authorial voice, in students' texts, point at fundamental differences in the purposes of master's theses and publications. To recap, students have to demonstrate the acquisition of required skills (Gardner & Nesi, 2013). It is not suggested that students have failed to adhere to the demands of the rhetorical conventions of the community. What is manifested as more reader-oriented could correspond to an engagement with the target audience, in other words, students have to align their discourse with the accepted knowledge and therefore involve the reader/professor.

Our findings show that there are no statistically significant differences in the proportions of hedges and boosters between levels, although the proportion of boosters in publications was almost four times higher. Whether hedges and boosters have been misused or do not fit the rhetorical section they are in, especially at university level, needs an additional and deeper analysis.

We had expected, taking into account the findings from Hylands' study (2005), that hedges would be more frequent than boosters in publications. We found the contrary; boosters were more frequent than hedges. These could be partly explained bearing in mind the high level of competiveness in biomedical sciences, so more robust results require fewer hedges (Gross & Chesley, 2012). Further, our results showed a significant difference in publications regarding attitude markers. Hyland (1998) found in his analysis of biology that attitudes markers were the less frequently used markers of stance. In contrast we found

that after self-mention and boosters, attitude markers were used more than hedges. For example, in one published RA we found an attitude marker expressing assessment 5 times, such as in the example below.

> "Interestingly", TREX2 expression is largely reduced in the mice lacking IKKa, which is required to maintain skin homeostasis and prevent skin cancer. (Published article: Increased Susceptibility to Skin Carcinogenesis in TREX2 Knockout Mice)

The most common attitudinal markers that we found, and this is in line with Hyland's findings (1998), refer to issues the writer sees as important or interesting, as shown in the following example from our corpus.

> "Importantly", and in contrast to previous trials, our patients were HIV-1-infected individuals who were demonstrably immunosuppressed, having failed to make serological and CD4 cell immune responses to vaccination with at least one of the three selected antigens (hepatitis A, hepatitis B and tetanus toxoid). (Published article: The reconstitution of the thymus in immunosuppressed individuals restores CD 4 -specific cellular and humoral immune responses)

The use of attitude markers is based on disciplinary values and even different cultural backgrounds, such as English and Spanish, are overridden by shared disciplinary conventions (Mur, 2007). Since our sample of publications reveals a high use of attitudinal markers, using these markers might be a distinguishing mark of proficiency, in the sense of adopting a discipline-specific writing convention. Therefore we suggest that the difference observed between publications and master's theses in the use of these markers might be due to the lack of mastery of disciplinary rhetorical conventions.

Interestingly, our findings of the differences of self-mentions confirm the RA genre as a text involving a highly rhetorical discourse replacing a detached stance (Bazerman, 1988; Hunston, 1994; Swales, 2004). Through self-mention writers project an authorial voice. This personal projection, when used to an appropriate degree, that is, in the way used by the specific community, allows the writer to construct his identity by emphasizing his individual contribution. This, in turn, helps to construct his professional authority and credibility (Hyland, 2004). Our results of the high use of self-mention in publications together with a lower use of hedging could be interpreted as an effort to project an authorial imprint. The use of self-

mentions differs according to genre, discipline and culture. A cross-cultural study of personal pronouns between Business Management RAs published in English and in Iberian Spanish has shown that the frequency of self-mentions is higher in English than in Spanish RAs. Our results are in line with these findings, and could be partly explained by considering the author's cultural background, since Spanish culture seems to emphasize involvement relations (Mur, 2007). An additional explanation to the observed differences is related to the prescription of academic style manuals, which suggest that in academic writing references to the writer or reader should not appear.

These results must be viewed in light of some limitations. It is a preliminary analysis with a small sample size, and moreover, caution is required when comparing lexico-grammatical features in two different languages. Further research on these features should be undertaken, to see if the differences between Catalan and English texts are due to the language typology.

8. Concluding remarks

This study sheds light on some pedagogically relevant components of academic writing, from two perspectives: academic register and writer-reader interactions. Academic register has been analyzed by means of lexical diversity, syntactic complexity and lexical density. The writer-reader interactions have been analyzed by means of stance and engagement markers.

We have provided a snapshot view of the challenge that is posed with the transition from a university to a professional setting, with the aim to understand how students participate in the actions of a community. Although master's theses and RAs are not the same genre, the goal of most students' assignments is to develop the ability to write professional texts. Within these, the RA genre serves as a model for the discipline specific writing. On the assumption that discipline-specific writing conventions are components of determining writing proficiency, it is suggested that these students have not yet acquired writing proficiency, either in academic register or in writer-reader interactions. As such, the overall picture that emerges from this study is a reflection of the complexity of the protracted path to mastery, in which linguistic, cognitive, cultural factors and the constraint of genre and discourse community are linked processes. Further, it may seem that register and writer-reader interactions are, usually, fundamental elements in contemporary teaching. Nevertheless, we contend that a rethink of how

writing is addressed in the curriculum for the specific university disciplines may be necessary.

Acknowledgements

References

Anthony, L. (2011). *Antconc* (Version 3.2.4w) [Computer Software]. Tokyo, Japan: Laurence Anthony, University of Waseda. URL: http://www.laurenceanthony.net/

Aull, L.L. & Z. Lancaster (2014). "Linguistic markers of stance in early and advanced academic writing: A corpus based-comparison". *Written Communication* 31: 151-183.

Baratta, M. A. (2010). "Nominalization development across an undergraduate academic degree program". *Journal of Pragmatics* 42: 1017-1036.

Bawarshi, A.S. & M.J. Reiff (2010). *Genre: An Introduction to History, Theory, Research, and Pedagogy*. West Lafayette, Indiana: Parlor Press.

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison, WI: The University of Wisconsin Press.

Beers, S.F. & W. Nagy (2009). "Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?". *Reading and Writing: An Interdisciplinary Journal* 22: 185-200.

Bereiter, C. & M. Scardamalia (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum.

Berkenkotter, C., T. Huckin & J. Ackerman (1991). "Social context and socially constructed tests: The initiation of a graduate student into a writing research community" in C. Bazerman, & J. Paradis (eds.), *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*, 191-215. Madison, Wisconsin: The University of Wisconsin Press.

Berkenkotter, C. & T. Huckin (1995). *Genre Knowledge in Disciplinary Communities*. Hillsdale: Lawrence Erlbaum.

Berman, R.A. (2004). "Between emergence and mastery: The long developmental route of language acquisition" in R.A. Berman (ed.), *Language Development across Childhood and Adolescence*, 9-34. Amsterdam: Benjamins.

Berman, R.A. & D. Ravid (2009). "Becoming a literate language user: Oral and written text construction across adolescence" in D.R. Olson, & N. Torrance (eds.), *The Cambridge Handbook of Literacy,* 92-111. Cambridge: Cambridge University Press.

Berman, R.A. & L. Verhoeven (2002). "Cross-linguistic perspectives on the development of text-production abilities: Speech and writing". *Written Language and Literacy* 5: 1- 43.

Biber, D. (1993). "Using register-diversified corpora for general language studies". *Computational Linguistics* 19: 219-242.

Biber, D. & E. Finegan (1989). "Styles of stance in English: Lexical and grammatical marking of evidentiality and affect". *Text* 9: 93-124.

Biber, D. & B. Gray (2010). "Challenging stereotypes about academic writing: Complexity, elaboration, explicitness". *Journal of English for Academic Purposes* 9: 2-20.

Biber, D., B. Gray & K. Poonpon (2011). "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?" *TESOL Quarterly* 45: 5-35.

Castelló, M., G. Bañales & A. Vega (2010). "Enfoques en la investigación de la regulación de escritura académica: Estado de la cuestión". *Electronic Journal of Research in Educational Psychology* 8: 1253-1282.

Chafe, W. & J. Danielewicz (1987). "Properties of spoken and written language" in R. Horowitz & S.J. Samuels (eds.), *Comprenhending Oral and Written Language*, 83-113. San Diego, California: Academic Press.

Clyne, M. (1994). *Inter-cultural Communication at Work. Cultural Values in Discourse*. Cambridge: Cambridge University Press.

Colombi, M.C. (2002). "Academic language development in Latino Students' writing in Spanish" in M. Schleppegrell & M.C. Colombi (eds), *Developing Advanced Literacy in First and Second Languages: Meaning with Power*, 67-86. Mahwah: Lawrence Erlbaum.

Cortes, V. (2002). "Lexical bundles in freshman composition" in R. Reppen, S.M. Fitzmaurice & D. Biber (eds.), *Using Corpora to Explore Linguistic Variation*, 131-146. Philadelphia, PA: John Benjamins.

Cortes, V. (2004). "Lexical bundles in published and student disciplinary writing: Examples from history and biology". *English for Specific Purposes* 23: 397-423.

Crismore, A., R. Markkanen & M.S. Steffensen (1993). "Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students". *Written Communication* 10: 39-71.

Cuenca, M. (2003). "Two ways to reformulate: A contrastive analysis of reformulation markers". *Journal of Pragmatics* 35: 1069-1093.

Engber, C. (1995). "The relationship of lexical proficiency to the quality of ESL compositions". *Journals of Second Language Writing* 4: 139-155.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London, UK: SAGE.

Flowerdew, J. (2014). "Foreword" in L. Gil-Salom, & C. Soler-Montreal (eds.), *Dialogicity in Written Specialised Genres*, IX-XVI. Amsterdam: John Benjamins.

Gardner, S., & Nesi, H. (2013). "A classification of genre families in university student writing". *Applied Linguistics* 34: 25-52.

Gross, A. G. & P. Chesley (2012). "Hedging, stance and voice in medical research articles" in K. Hyland & C. Sancho Guinda (eds.), *Stance and Voice in Written Academic Genres*, 85-100. New York: Palgrave Mcmillan.

Gross, A. G., J. E. Harmon & M. Reidy (2002). *Communicating Science: The Scientific Article from the 17th Century to the Present*. New York: Oxford University Press.

Guillén Galve, I. (1998). "The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English". *Journal of Pragmatics* 30: 363-385.

Halliday, M.A. (1993). "Some grammatical problems in scientific English" in M A. Halliday & J. R. Martin (eds.) *Writing Science: Literacy and Discourse Power*, 69-85. Pittsburg, PA: University of Pittsburgh Press.

Halliday, M.A. & R. Hasan (1989). *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective* (2nd ed.). Oxford: Oxford University Press.

Henderson, A. & R. Barr (2010). "Comparing indicators of authorial stance in psychology students' writing and published research articles". *Journal of Writing Research* 2: 245-264.

Hunston, S. (1994). "Evaluation and organisation in a sample of written academic discourse" in M. Coulthard (ed.), *Advances in Written Text Analysis*, 191-218. London: Routledge.

Hyland, K. (1996). "Writing without conviction? Hedging in scientific research articles". *Applied Linguistics* 17: 433-454.

Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.

Hyland, K. (2004). "A convincing argument: Corpus analysis and academic persuasion" in U. Connor & T. A. Upton (eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics*, 87-112. Amsterdam: John Benjamins.

Hyland, K. (2005). "Stance and engagement: A model of interaction in academic discourse". *Discourse Studies* 7: 173-192. doi:10.1177/1461445605050365.

Hyland, K. (2008a). "Persuasion, interaction and the construction of knowledge: Representing self and others in research writing". *International Journal of English Studies* 8: 1-23.

Hyland, K. (2008b). "Academic clusters: Text patterning in published and postgraduate writing". *International Journal of Applied Linguistics* 18: 41-62.

Hyland, K. & L. Hamp-Lyons (2002). "EAP: Issues and directions". *Journal of English for Academic Purposes* 1: 1-12.

Hyland, K. & J. Milton (1997). "Qualification and certainty in L1 and L2 students' writing". *Journal of Second Language Writing* 6: 183-205.

Hyland, K. & P. Tse (2004). "Metadiscourse in academic writing: A reappraisal". *Applied Linguistics* 25: 156-177.

Hymes, D.H. (1972). "On communicative competence" in J.B. Pride, & J. Holmes (eds.), *Sociolinguistics. Selected Readings*, 269-293. London, UK: Penguin.

Johansson, V. (2008). "Lexical diversity and lexical density in speech and writing: A developmental perspective". *Working Papers* (Lund University) 53: 61-79.

Kellogg, R.T. (2008). "Training writing skills: A cognitive developmental perspective". *Journal of Writing Research* 1: 1-26.

Kuo, C.H. (1999). "The use of personal pronouns: Role relationships in scientific journal articles". *English for Specific Purposes* 18: 121-138.

Lafuente, E. (2008). "Epistemic and approximative meaning revisited: The use of hedges, boosters and approximators when writing research in different disciplines" in S. Burgess and P. Martín Martín (eds.), *English as an Additional Language in Research Publication and Communication*, 65-82. Bern: Peter Lang.

Lewin, B. (2005). "Hedging: An exploratory study of authors' and readers' identification of 'toning down' in scientific texts". *Journal of English for Academic Purposes* 4: 163-178.

Malvern, D., B. Richards, N. Chipere & P. Durán (2004). *Lexical Diversity and Language Development: Quantification and Assessment*. New York: Palgrave Macmillan.

Markkanen, R. & H. Schröder (eds.). (1997). *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Berlin: Walter de Gruyter.

Martín, P., J. Rey-Rocha, S. Burgess & A. I. Moreno (2014). "Publishing research in English-language journals: Attitudes, strategies and difficulties of multilingual scholars of medicine". *Journal of English for Academic Purposes* 16: 57-67.

Martínez, I.A. (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.

Mauranen, A. (1993). "Cultural differences in academic discourse: Problems of a linguistic and cultural minority" in L. Löfman, L. Kurki-Suonio, S. Pellinen & J. Lehtonen (eds.), *The Competent Intercultural Communicator*, 157-174. Vol. 51. Tampere, Finland: AFinLA Series.

McCarthy, P. M. & S. Jarvis (2010). "MTLD, Vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment". *Behavior Research Methods* 42: 381-392. doi: 10.3758/BRM.42.2.381.

McGrath, L. & M. Kuteeva (2012). "Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices". *English for Specific Purposes* 31: 161-173.

McNamara, D. S., S.A. Crossley & P.M. McCarthy (2010). "Linguistic features of writing quality". *Written Communication* 27: 57-86.

Mendiluce, G. (2005). "La matización asertiva en el artículo biomédico: Una propuesta de clasificación para los estudios contrastivos inglés-español". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 10: 63-90.

Miller, C.R. (1984). "Genre as social action". *Quarterly Journal of Speech* 70: 151-167.

Montolío, E. (Coord.). (2000). *Manual Práctico de Escritura Académica* (tomo III). Barcelona: Ariel.

Mur, P. (2007)." '*I/we focus on*…': A cross-cultural analysis of self-mentions in business management research articles". *Journal of English for Academic Purposes* 6: 143-162.

Nelson, N. & M. Castelló (2012). "Academic writing and authorial voice" in M. Castelló & C. Donahue (eds.), *University Writing: Selves and Texts in Academic Societies*, 33-52. Bingley: Emerald.

Padró, L. & E. Stanilovsky (2012). "FreeLing 3.0.: Towards wider Multilinguality" in *Proceedings of the Language Resources and Evaluation Conference* (LREC). Istanbul, Turkey.

Pujol Dahme, A. & M. Selfa Sastre (in revision). "From high school up to publication: Developing register and interactional metadiscourse in biology academic writing". *Journal of Writing Research*.

Ravid, D. & R.A. Berman (2009). "Developing linguistic register across text types. The case of modern Hebrew". *Pragmatics and Cognition* 17: 108-145.

Ravid, D. & L. Tolchinsky (2002). "Developing linguistic literacy: A comprehensive model". *Journal of Child Language* 29: 417-447.

Rienecker, L. & P. Stray Jörgensen (2003). "The (im)possibilities in teaching university writing in the Anglo-American tradition when dealing with continental student writers" in L. Björk, G. Bräuer, L. Rienecker & P. Stray Jörgensen (eds.), *Teaching Academic Writing in European Higher Education*, 101-112. Dordrecht, Netherlands: Kluwer Academic.

Russell, D.R. & V. Cortes (2012). "Academic and scientific texts: The same or different communities" in M. Castelló & C. Donahue (eds.), *University Writing: Selves and Texts in Academic Societies*, 3-17. Bingley: Emerald.

Salager-Meyer, F. (1994). "Hedges and textual communicative function in medical English written discourse". *English for Specific Purposes* 13: 149-170.

Scarpa, F. (2007). "Using an Italian diachronic corpus to investigate the 'core' patterns of the language of science" in K. Ahmad & M. Rogers (eds.), *Evidence-based LSP. Translation, Text and Terminology*, 75-95. Bern, Switzerland: Peter Lang.

Schleppegrell, M.J. (2001). "The linguistic features of the language of schooling". *Linguistics and Education* 12: 431-459.

Schleppegrell, M.J. & M.C. Colombi (2002). "Theory and practice in the development of advanced literacy" in M.J. Schleppegrell, & M.C. Colombi (eds.), *Developing Advanced Literacy in First and Second Languages: Meaning with Power*, 1-19. Mahwah: Lawrence Erlbaum.

Snow, C.E. & P. Uccelli (2009). "The challenge of academic language" in D.R. Olson & N. Torrance (eds.), *The Cambridge Handbook of Literacy,* 112-133. Cambridge: Cambridge University Press.

Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. New York: Cambridge University Press.

Swales, J.M. (2004). *Research Genres: Explorations and Applications*. New York: Cambridge University Press.

Szmrecsanyi, B. (2004). "On operationalizing syntactic complexity" in G. Purnelle, C. Fairon & A. Dister (eds.), *Le Poids des Mots. Proceeding of the 7th International Conference on Textual Data Statisctical Analysis* 2, 1032-1039. Louvain-la-Neuve: Presses Universitaires de Louvain.

Tannen, D. (1985). "Relative focus on involvement in oral and written discourse" in D. Olson, D., N. Torrance & A. Hilyards (eds.), *Literacy, Language, and Learning*, 124-147. Cambridge: Cambridge University Press.

Uccelli, P., C. Dobbs & J. Scott (2013). "Mastering academic language: Organization and stance in the persuasive writing of high school students". *Written Communication* 30: 36-62.

Vázquez, G., A. Fernández & M. A. Martí (2000). "Dealing with lexical semantic mismatches between English and Spanish" in *Proceedings of the International Conference of Knowledge Based Computer Systems*, 308-319. Mumbai, India.

Vázquez, I. (2010). "A contrastive analysis of the use of modal verbs in the expression of epistemic stance in Business Management research articles in English and Spanish". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 19: 77-96.

Vázquez, I. & D. Giner (2008). "Beyond mood and modality: Epistemic modality markers as hedges in research articles. A cross-disciplinary study". *Revista Alicantina de Estudios Ingleses* 21: 171-190.

Ventola, E. (1996). "Packing and unpacking of information in academic texts" in E. Ventola & A. Mauranen (eds.), *Academic Writing. Intercultural and Textual Issues*, 153-194. Philadelphia: Benjamins.

Xanthos, A. (2014). "Textable: Programmation visuelle pour l'analyse de données textuelles" in *Actes des 12èmes Journées Internationales d'Analyse Statistique de donées textuelles* (JADT).

Ana Pujol Dahme is an advanced doctoral student in Theoretical, Computational and Applied Linguistics at the University of Barcelona, and member of the Research Group for the Study of Linguistic Repertoire (GRERLI). She is also an adjunct lecturer at the Department of Teacher

Training in Language and Literature (University of Barcelona). Her research interests include later language development, academic writing and corpus linguistics.

Moisés Selfa Sastre is Doctor in Philosophy (University of Lleida, 2002). Currently, he is Professor of Teaching Language and Literature (University of Lleida) in the Department of Specific Didactics. His research interests revolve around the teaching of writing in the context of primary, secondary and students of the University. He is part of the Research Group (GRERLI. Grup de Recerca per a l'Estudi del Repertori Lingüístic) of the University of Barcelona. His recent publications deal with the teaching of writing literary texts and interpretation of these.

NOTES

[1] Catalan is spoken in four Spanish autonomous communities (Aragon, the Balearic Islands, Catalonia and the Valencian Community), in the French region of Roussillon, and in the city of Alghero, on the Italian island of Sardinia. In Catalonia, Catalan and Spanish are equally recognized as official languages, but Catalan is the language of education.

Appendix List of titles

List of master's theses in the discipline of immunology

1. Efecte d'un nou probiòtic en un model d'asma al·lèrgica en el ratolí
2. Estudi de la resposta al superantigen SEB a nivell de citocines i fosforilació d'STATs en un model de shock sèptic murí.
3. Estudi dels efectes coestimuladors de la interacció CD26-ADA en l'expressió de marcadors d'activació de la cèl·lula T
4. Estudi d'una doble estratègia per a induir i expandir el repertori de cèl·lules t basada en l'administració d'hormona de creixement i vacuna en pacients amb infecció pel vih
5. Disminució dels nivells de TCR en la membrana dels limfòcits T CD4+ humans per efecte de sindecà-2
6. Generació de cèl·lules dentrítiques tolerogèniques per terapia cel·lular en esclerosi múltiple
7. Modificació de l'expressió de cd36 i cd206 per lligands de receptors toll-like (tlrs) i citocines
8. Estudi de la influència dels pèptids naturals presentats per MHC de classe II en autoimmunitat tiroidea

List of published articles

1. Activity of the cyclooxygenase 2-prostaglandin-E prostanoid receptor pathway in mice exposed to house dust mite aeroallergens, and impact of exogenous prostaglandin E2
2. Increased Susceptibility to Skin Carcinogenesis in TREX2 Knockout Mice
3. Increased a-Defensins 1-3 Production by Dendritic Cells in HIV-Infected Individuals Is Associated with Slower Disease Progression
4. The reconstitution of the thymus in immunosuppressed individuals restores CD 4 -specific cellular and humoral immune responses
5. Syndecan-2 can promote clearance of T-cell receptor/CD3 from the cell surface
6. Specific T-cell proliferation to myelin peptides in relapsing-remitting multiple sclerosis
7. Functional consequences of CD36 downregulation by TLR signals
8. The peptide-binding motif of HLA-DR8 shares important structural features with other type 1 diabetes-associated alleles