INTRODUCTION

Abstracts are short summaries of longer academic papers. They include references to the research topic, aims and methods, and often also indicate previous research, own results and conclusions. It is important how these pieces of information are constructed linguistically to ensure clarity and unambiguity, to show the author's intentions and to indicate the certainty of their assertions. Lorés (2004, p. 281) concludes that "there seems to be general agreement on three ways in which research abstracts differ from RAs [research articles]. They differ in their function, in their rhetoric structures and in their linguistic realizations". In the last forty years many researchers have analyzed the structural and linguistic contents of RAs and other published or unpublished academic papers. Studies which have compared papers across disciplines have pointed to some universality, but mainly discovered discipline-specific rhetorical features (e.g., Pho, 2008; San & Tan, 2012; Stotesbury, 2003).

Due to this specificity it is difficult to compare results based on corpora drawn from different fields, journals, authors and languages. Relatively few studies have analyzed research article abstracts (RAAs) in applied linguistics and less is known about the differences between abstracts written by expert and novice writers. According to the APA manual, a research article abstract which is "accurate, succinct,
quickly comprehensible, and informative will increase the audience and the future retrievability” of an article (APA, 2003, p. 15). Gate-keeping bodies, such as editors, reviewers, conference organizers, may decide on the dissemination of a research project based on a quick screening of the abstract, and published articles also find their readers through the abstracts that are often open-access even when the articles themselves are not. It may be argued that thesis abstracts serve a different purpose as they do not exist independently from the main text, and their audience is restricted to the supervisors and in-house reviewers. Nevertheless, students are expected to follow academic conventions and summarize their thesis in an abstract similar to those of published papers. As Koutsantoni (2006) rightly points out, RAAs and thesis abstracts may be considered two similar but distinct genres, especially due to the status of the authors in the academic discourse community of their fields. She also backs up this proposition by showing the power asymmetry between student writers and supervisors.

However, I would argue that a similar kind of power asymmetry exists between authors of RAAs who wish to publish and the editors or independent reviewers who may reject their papers. I believe that abstract writing is a marginal part of academic writing instruction at all levels, and both novice and experienced writers learn it implicitly while they read published articles. I assume that most authors pay less attention to abstracts compared to other sections, and that abstract writing remains the last step, often done with hustle due to final deadlines. For these reasons abstracts are good indicators of authors' target language academic writing skills and their ability to see and summarizes the key points of their papers by following rhetorical and linguistic conventions (Dorado, 2013a). Student writers are expected to follow these conventions in all sections of their papers, even when their writing reaches a very small audience and their experience as researchers and as academic text producers is very limited and often insufficient. For the reasons mentioned above, I believe that these two abstract genres are close enough to be comparable and their systematic review may shed light on the similarities and differences in the linguistic realization of abstracts written by professional and student writers. Applied linguistics papers have not been widely studied from this point of view, and a comparison of published papers and undergraduate EFL theses is hoped to serve as basis for a more explicit academic writing instruction of abstracts.

BACKGROUND

The next sections provide an overview of the literature concerning the following areas: (a) genres characteristics of abstracts and the differences between RAAs and thesis abstracts, (b) verb tense usage in RAAs and (c) hedging in academic discourse.
Abstracts and their structure

Abstracts serve an important role in academic discourse and are considered a distinct genre. Studies such as Swales (1990) and Askehave and Swales (2001) point out that genre analysis should include both bottom-up (text based) and top-down (discourse community practices, values, beliefs based) identifications. Based on their information content, structure and communicative purpose, two main types of abstracts are distinguishable, namely descriptive and informative (Martín-Martín, 2005). The second one is more detailed and provides information also about the results, outcomes and conclusions of studies. The rhetoric structure of research articles and abstracts has been broadly studied in the past. In 1981, Malcolm Swales designed the so called Create a Research Space (CARS) model which consists of the following main moves: establishing the territory, establishing the niche and occupying the niche (Swales, 1990). This CARS model has been applied by a large number of studies to investigate the rhetoric structure of RA introductions (e.g., Hirano, 2009; Ozturk, 2007; Samraj 2002, 2005; Sheldon, 2011). Apart from research on the introduction sections, attention has also been paid to the analysis of RA abstracts. Santos (1996) created a model for RA abstracts, which includes the following five moves: situating the research, presenting the methodology, summarizing the findings, and finally discussing the findings. This five-step model has formed the basis for the analysis of move structures of both published RAAs and theses written by student writers (e.g., Dorό, 2013a, 2013b; Lon, Tan & Abdullah, 2012; Pho, 2008; Tseng, 2011). A similar framework, referred to as the IMRD or IMRaD, has also been used by researchers (e.g., Bhatia, 1995; Lorés, 2004; Serholt, 2012; Swales, 1990). These acronyms stand for the introduction, methods, result and discussion sections of abstracts. The CARS and IMRD frameworks are almost identical, but the first two moves in the CARS model are combined in the IMRD framework; this ensures the comparability of research conclusions of studies that use either of the two models or any of their modified versions. Lorés (2004) analyzed abstracts published in applied linguistics journals and compared the thematic distribution of moves according to both the IMRD and the CARS models. She found that 61% of the abstracts followed the IMRD model, 30.5% resembled the CARS structure and 8% of them a combination of the two.

Several studies have pointed out the variability in the number of moves and the order of these moves employed in RAAs. Not all moves are obligatory, but when major moves (aims, methods and results) are missing, the reader may feel that the abstract is incomplete or less informative. In an earlier study (Dorό, 2015a) I compared BA-level thesis abstracts (BTAs) written on applied linguistics, literature and culture topics. While some kind of reference was given to the research topic in all the theses abstracts and methodology was introduced in at least 90% of the papers in the three sub-corpora, move 4 (findings) showed a significant difference between linguistic and non-linguistic theses. This means that results were not summarized in more than half of the literature, culture and history theses, while...
90% of the authors of linguistics papers felt the need to indicate results in their abstract. The non-obligatory nature of referring to the findings in abstracts has also been reported in other studies (e.g., Doró, 2015b; Lon et al., 2012; Lorés, 2004; Santos, 1996).

**Verb tenses in academic research papers**

Compared to the boom in the move structure analysis of research articles, much less has been systematically reviewed in terms of the verb tense usage in research papers, and most of these analyses concentrated on a limited number of science articles. Malcolm (1987) reviewed studies from the 1970s and 1980s that point to the functional use of tenses in academic discourse. He highlights that the choice of tenses, on the one hand, reflects the genre traditions and, on the other hand, the rhetorical choices of authors to express their intentions. Moreover, he cites examples of obligatory and optional constraints of tense usage in various parts and sections of a paper. For example, references to past research are done in present simple, present perfect or simple past, while methods and results are explained either in the present or in the past. Research manuals often provide clear guideline as to which tenses and aspects should be used for certain sections or purposes. Nevertheless, they often fail to discuss the flexibility of tense usage found in research articles. For example, the fifth edition of the APA manual suggests the following:

> Use verbs rather than their noun equivalents and the active rather than the passive voice. Use the present tense to discuss results with continuing applicability or conclusions drawn; use the past tense to describe specific variables manipulated or tests applied (pp. 12-13).

Salager-Meyer (1992) pointed to a close relationship between rhetorical functions (moves) and finite verb forms and modals used in medical RAAs. He reported that the three most common tenses (simple present, past and present perfect) made up 89.8% of all finite verbs. Within these, past dominated with 51.4%, while the present constituted one-third of all verbs (32.8%). The purpose, methods, results and case presentation sections used mainly the past tense, while general truth, introductory notes, conclusions, recommendations and data synthesis were predominantly referred to in the present. The present perfect was used to refer to previous research. Tseng (2011) analyzed 90 applied linguistic abstracts and concluded that the present tense was applied mainly to present the background, aims and conclusions, while the past tense was used to discuss the methodology and results. This is in line with the results of the Salager-Meyer (1992) study. Tseng also reported some variation across the three journals he used. Authors accepted for *Applied Linguistics* referred to their methods and results more often in the present than in the past. While analyzing differences between native and non-native English speaking
first authors, Tseng concluded that native speakers used the present tense more for aims and results than did non-natives.

Li and Ge (2009) reviewed the structural and linguistic changes in medical articles in a 20-year time span and found that the frequency of tenses in the various sections of articles had significantly changed. The overall frequency of the four most frequent tenses had not changed much, but their roles had shifted. Simple past increased in the presentation of new research, while the use of simple present increased in the discussion sections.

In a recent study de Waard and Pander Maat (2012) examined how readers classify statements taken from various sections of biology research articles (problems, hypothesis, methods, results and implications). Their aim was to see what happens if verb tenses are modified in these sentences. Their results demonstrate that verb forms serve as markers of intention and segment types of research papers, and by changing verb tenses the reader may classify the same sentence differently. For example, fact statements and hypotheses used with the past tense are interpreted as results. Result sections, in contrast, are understood by and large as results either in the present or in the past. The authors of this study found that modal auxiliaries also have a crucial role, as hypothesis-type statements without modals are understood as facts or results, whereas results with modal auxiliaries may be interpreted as hypotheses or weak statements.

**Hedging in academic discourse**

While academic writing is often referred to as impersonal and objective, especially when it reports on facts and results, research papers go beyond the mere recitation of definitions or general truths. Authors analyze previous research and interpret their own results. While doing so they take personal responsibility for their claims, show their attitudes or the strength of their claims. Hyland (1994, p. 240) states that “effective academic writing actually depends on interactional elements which supplement propositional information in the text and alert readers to the writer’s opinion”. This does not mean a highly individual or subjective voice, but is treated as a natural part of academic writing. Authors provide tentative analyses through the use of epistemic modality which refer to their (lack of) confidence in their propositions. The degree of confidence is most often expressed through hedges (e.g., *suggest, seem to, likely*) and boosters (e.g., *strongly, clearly*). Hedges are one form of metadiscourse that authors use to modify their arguments in accordance with the needs and expectation of possible readers (Hyland, 2004). Indeed, hedges have been found to be the most frequent category of metadiscourse in a large corpus of doctoral and masters level dissertations (Hyland, 2000). Hyland (1996a, 1996b, 2000) warns us that the usage of hedges is troublesome for most authors, especially non-native researchers, who do not clearly feel the meaning and function of hedges or boosters. He states the following:
Effective academic writing always carries the individual's point of view. Writers also need to present their claims cautiously, accurately and modestly to meet discourse community expectations and to gain acceptance for their statements. Such pragmatic aspects of communication however are vulnerable to cross-cultural differences and L2 students are rarely able to hedge their statements appropriately. (2000, p. 477)

Milton and Hyland (1999) compared native and non-native students' essays written in English and concluded that non-natives used epistemic modifiers (such as usually, actually, may) less often and that 75% of such words were restricted to the ten most often occurring ones. On the contrary, Koutsantoni (2006), while analyzing engineering journal articles and theses, found that student writers hedge more often than experienced authors, but also documented that students use different types of and a more restricted repertoire of hedges. Abdollahzadeh (2011) compared the conclusion sections of applied linguistic RAs published in English by native and non-native, Iranian researchers. He pointed to a "remarkable tendency by both writer groups towards hedging their propositions" (p. 288), but found no significant differences in the overall number of hedges between the two groups. Nevertheless, the Iranian researchers used fewer attitude markers (e.g., I feel it is even more important, it is interesting) and emphatics (e.g., certainly, truly). He concluded that the "differences are attributed to the degree of rhetorical sensitivity to and awareness of audience, purpose, cultural leanings, and the proclivities of the disciplinary genre" (p. 288). Variability in the reviewed research results are partly due to the categorization problems of metadiscourse markers, the proficiency level of the authors, and the section of articles under review in the different studies. Another important reason could be the fact that most studies concentrate on predetermined lists of markers (e.g., Hyland, 2000, 2010; Serholt, 2012), while others do a manual search of their corpora and include all signs of hedges or boosters (e.g., Abdollahzadeh, 2011).

THE STUDY

Research questions

Based on the literature reviewed above the following main research questions were formulated:

1. What is the proportion of finite verb tenses used in published and unpublished applied linguistics research papers?
2. What are the hedges used in these abstracts and what is the frequency of their occurrence?
3. Are there significant differences between applied linguistics research article abstracts and BA thesis abstracts in terms of verb tenses and hedging?
Corpus and methods

For this study English studies BA level theses, written and defended by students at a large Hungarian university in the years 2011 and 2012, were considered. Since a previous study using these papers (Doro, 2013a) found results similar to those voiced by other authors, namely that marked differences exist between research papers written in various disciplines, only one field of studies was chosen. Applied linguistics paper abstracts were found to adhere more to the rhetoric structures and linguistic conventions of research papers than theses written in other fields (Doro, 2013a) and showed the most consistency, therefore they were chosen to be the ones used in parallel with published RAAs. The first corpus for this study consists of 30 applied linguistics BA thesis abstracts and the second one of 30 RAAs. The 30 student papers cover the majority of theses written on various applied linguistic topics in these two academic years. The RAAs were published in 2012 and randomly retrieved from three journals (Applied Linguistics, Journal of Second Language Writing and System). All three are peer reviewed and prestigious, but cover slightly different topics and include different authors. These selection criteria resulted in a thesis corpus (referred to as BTA) of 6,046 words and a RA abstract corpus (referred to as RAA) of 5,535 words. This shows that, on average, the student abstracts are slightly longer. All student authors used English as a foreign language and wrote their papers as a final step for the completion of their 3-year BA English studies. As the native/non-native dichotomy was not the target of this research, it was not a selection criterion, although most of the published articles have at least one non-native author.

The two corpora were first manually screened for finite verb phrases and were categorized in the following tense and aspect combinations: present simple, past simple, present perfect, past perfect, present continuous, past continuous and future will. Cases of unusual or infrequent verb form choices were also screened for the move functions they perform in order to better understand the authors' intentions. As a second step the corpus was searched electronically for twenty items which commonly perform hedging functions in academic writing (see Table 1 for the list) using AntConc, a text analysis and concordance program. With the help of the concordance function, all instances were carefully analyzed individually to ensure that they were performing metadiscoursal functions. Results were normalized per 1,000 words to allow comparison across the two corpora of slightly different sizes. Final figures were calculated as number of occurrences, percentages, and also as proportions.
Table 1. Linguistic devices of hedges used in this study

<table>
<thead>
<tr>
<th>Based on Hyland (2000) and Serholt (2012)</th>
<th>Taken from the BAT and RAA corpora</th>
</tr>
</thead>
<tbody>
<tr>
<td>can</td>
<td>tend</td>
</tr>
<tr>
<td>may</td>
<td>try</td>
</tr>
<tr>
<td>could</td>
<td>seek</td>
</tr>
<tr>
<td>might</td>
<td>relatively</td>
</tr>
<tr>
<td>seem</td>
<td>sometimes</td>
</tr>
<tr>
<td>appear</td>
<td>mostly</td>
</tr>
<tr>
<td>assume</td>
<td>generally</td>
</tr>
<tr>
<td>believe</td>
<td></td>
</tr>
<tr>
<td>speculate</td>
<td></td>
</tr>
<tr>
<td>indicate</td>
<td></td>
</tr>
<tr>
<td>likely</td>
<td></td>
</tr>
<tr>
<td>possible/possibly</td>
<td></td>
</tr>
<tr>
<td>probable/probably</td>
<td></td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

Verb tenses

The frequency counts show the ruling position of simple present, followed by simple past. This is in line with the research reviewed above (e.g., de Waard & Pander Maat, 2012; Li & Ge, 2009). Table 2 shows the different patterns of verb use. Simple present and simple past together accounted for 90% of the finite verbs in the BTA corpus and 98% of the RAA corpus. Five published articles and one student paper used the present tense exclusively. Present perfect was identified much less often (accounting for 5% and 3%, respectively) and the other tenses occurred only rarely. There is, however, difference in the number of finite verb phrases and the distribution of verb forms between the two corpora. The BTAs used between 9 and 27, overall 453 (75/1000) finite verbs, while the RAAs employed between 7 and 19, overall 366 (69/1000) finite verbs. This suggests that on average, student writers did not only write longer summaries, but while doing so they used fewer non-finite verb phrases and subordinate sentences than the RAAs.
When taking a closer look at the functions of these verbs, results indicate that students use tenses logically and systematically and show good knowledge of the academic writing conventions of abstracts. This was also found in an experimental study (Doró, 2015a) in which ten applied linguistics student abstracts were compared with non-linguistic abstracts. No unclear verb choices were detected in the RAAs and only a few cases are problematic in the BTAs. These include the sudden switch from simple past to simple present or present perfect in the results move, and the use of present perfect to refer to methods (see sample sentences 1 and 3). Tense mixing may cause confusion in the reader, especially if it becomes unclear whether the reference is made to previous investigations or to the project in question, as in samples 1 to 3. Sudden changes may be the result of students’ shift between general English language conventions and academic writing practices or the influence of some noun phrase/verb phrase collocations used as fixed expressions (e.g., studies have shown) without considering their functions in the specific sections of the abstracts. These uncertainties are more typical of students who are weaker in their English language proficiency.

1. ... Differences between gender groups and faculties have been examined more specifically. The results show evidence for the initial hypothesis according to which age is the most determining variable in the Hungarian address system. Numerous differences have been found between...

2. Color-naming has been investigated among...

3. ... I have identified and revealed numerous beliefs and practices .... More precisely, I have showed the differences ... The results showed that ...

**Hedges**

The frequency of the various types of hedges is illustrated in Table 3. The 20 epistemic modifiers that indicate uncertainty appear 50 times in the BTA corpus, while only 36 times in the RAA corpus. Only three of them do not occur in the two corpora. Their distribution is very much uneven, with the auxiliary can leading both...
abstract groups (with 44% for BTAs and 28% for RAAs), of which the past form could is also used by students, but not by researchers. The word may occupies second position, with might appearing only once. Altogether the four most frequent hedges are can, may, indicate and tend.

Table 3: Occurrences of hedges in the two corpora

<table>
<thead>
<tr>
<th>List of hedges</th>
<th>BTAs</th>
<th>RAAs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>can</td>
<td>22</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>may</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>could</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>might</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>seem</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>appear</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>assume</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>believe</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>speculate</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>indicate</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>likely</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>possible/possibly</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>tend</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>try</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>seek</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>should</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>relatively</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>sometimes</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>mostly</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>generally</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50</td>
<td>36</td>
<td>86</td>
</tr>
</tbody>
</table>

Hedges are frequently used to indicate results. Some illustrations for the cautious interpretation of results in the thesis abstracts are provided below. Examples 5 to 7
are more academic and elegant and could also come from published abstracts, while the use of can is superfluous in sentence 4.

4. Findings of this paper can be useful not only in ...
5. Findings presented in this paper may contribute to further research ...
6. This might be the reason why ...
7. Results seem to indicate that ...

To compare, Salager-Meyer (1992) reported the leading position of may in medical RAs; other modals had a marginal position in his corpus. In contrast, the four most frequent hedges in the Serholt (2012) Swedish student abstract corpus are might, seem, could and suggest. These differences in the results also point to the importance of using small, specialized corpora to illustrate linguistic choices in different genres, specific fields and various author groups. Large corpora may yield impressive results, but mask the individual and group differences.

CONCLUSION AND IMPLICATIONS

This study analyzed two small, specialized applied linguistics corpora to reveal the linguistic realization of abstracts written by expert writers and novice, BA student writers. It focused on the similarities and differences found in the linguistic realization of 30 abstracts published in applied linguistics journals and 30 BA level applied linguistics thesis abstracts written by Hungarian students of English. The results indicate that thesis abstracts incorporate more finite verb clauses, but students are aware of the genre conventions. The few instances of misuse were discussed in context and in terms of the rhetoric functions they have. The concordance analysis of 20 epistemic modifiers revealed that modals have a major role in indicating uncertainty in both groups of abstracts, while other hedges appear only marginally.

The literature has called for the need for a careful selection of texts to be compared, and these two groups of abstracts, although having different functions, proved to be good choices. The analyses of verb tenses and hedges in humanities papers have not been in the center of attention of researchers, therefore this study aimed to fill this gap. A future direction of research may be the extension of analysis to non-linguistic thesis and to the interview and questionnaire methods used by Hyland (2000) and de Waard and Pander Maat (2012). Hyland (2000) rightly pointed out that the analysis of specialized corpora is a valuable tool to assist novice writers and teachers to take control over disciplinary-sensitive writing practices. Only this way can students fully explore the meaning of word choices and learn the acceptable and appropriate means to engage with their readers, to mark their attitudes and to appear confident and expert writers. As there is no time for the explicit teaching of all tricks and steps of good academic writing in the few hours of academic writing classes or thesis consultations, students should be shown
ways and tools to extract academic texts written about the topics they wish to explore, analyze them with easy-to-use tools such as AntConc or the Complete Lexical Tutor, so that they can check word choices in context and discover solutions that are typical of experienced writers. This way it is hoped that student writers do not plagiarize sections of texts merely due to language shortcomings, but find good alternatives for academic vocabulary boosters.

REFERENCES


AntConc. Corpus analyzer kit designed by Laurence Anthony. http://www.antlab.sci.waseda.ac.jp


Compleat Lexical Tutor. Language learning and research tools designed by Tom Cobb http://www.lexutor.ca/vp


