Scientific writing by English and non-English Background students: A Systemic Functional Linguistics

Approach to text analysis

Saman Abraham

(Student ID: 815493)

Submitted in partial fulfillment of the requirements of the

Master of Applied Linguistics at the University of Melbourne

June 2018

Author Note

I declare that this thesis does not contain material which has been accepted for the award of any other degree or diploma in any university; nor does it contain material previously published or written by any other person, except where due reference is made in the text of the thesis.

Acknowledgements

I would like to provide my heartfelt gratitude and appreciation to the following people for their support.

This achievement would not have been possible without the kind, encouraging, and continuous support and guidance of my supervisor, Dr Janne Morton. Her advice and teachings were vital and of enormous benefit to me during the course of the project and the writing of this thesis.

The data for this thesis were very kindly provided by the participant students and their supervisors, and were central to its development. Their kind willingness to provide their manuscripts and make time for the surveys was of utmost importance for the progress of this thesis.

Lastly, I would like to express my deep appreciation to my wife and son, for their patience and support with time away from them.

Abstract

Associated with the rise of English as the global language of communication, researchers worldwide have been pressured to 'Publish in English or perish' in an internationally competitive academia (Crystal, 2007; Di Bitetti & Ferreras, 2017). A consequence of this has been intense research into the many locally influenced styles of English-medium research writing. The current study applies one discourse analytical approach, systemic functional linguistics (SFL), to investigate textual differences between scientific abstracts written by three groups of researchers: English L1 scientists (published/'ideal' abstracts), L1 (Australian) students (original drafts from less experienced writers), and EFL (Japanese) students (also original drafts from less experienced writers from a different L1 background). SFL is a theory of grammar that looks at the function of language in use (Halliday, 1985), and has been used effectively to diagnose student writing problems (e.g. Moore, 2007). Four main SFL tools were applied in the current study, including thematic progression, the choice of thematized participants, processes of transitivity, and finally, nominalization. Interesting differences were found between the three groups of abstracts, suggesting a role for experience and/or L1 background on certain writing features. A survey also revealed a further role for experience and exposure to L1 environments. Unlike previous work, this study uses original/non-edited student texts with the hope of providing useful insight on student writing features, with potential implications for pedagogy.

Keywords: English as a global language, scientific writing, systemic functional linguistics

Table of Contents

List of table and figures

List of Tables

List of Figures

Introduction

In the recent decades, advances in technology have further mediated the growing influence of English language on the global stage, as the primary language of international research communication, including in science (Crystal, 2007; Carter-Thomas, 2016). This spread of English throughout the world has led to the development of many local English varieties that have accumulated aspects of the local cultures and their linguistic expressions (Crystal, 2012; Firth, 1996; Seidlehofer, 2005). This has made publishing in English as a Foreign Language (EFL) contexts a core area of applied linguistics research (Carter-Thomas, 2016). In this field, scholars have tended to focus either on textual features or on factors extraneous to the text (Coffin and Donohue, 2012). Understanding extraneous influences and textual features of writing can help provide useful insight into English as a first language (L1) and EFL author writing styles (Coffin and Donohue, 2012).

Differences between L1 and EFL stylistic writing features are at times a major concern for EFL writers during the manuscript review process of publication (Gosden, 2003; Lillis & Curry, 2015; Paltridge, 2013). EFL writers feel that sometimes their manuscript is rejected because it may not have followed the standard 'native English speaker' writing style (as discussed by Lillis & Curry, 2015). Here, a major point of concern for EFL writers is that their manuscripts can even get criticized for idiomatic expression, a stylistic feature mainly of 'native' writers (Lillis & Curry, 2015). Often, during the review process correspondence, a direct request by the reviewers/assessors to EFL writers is to have their English checked by a 'native English speaker' (Lillis & Curry, 2015; Murugesan, 2014). This is even directly stated on some major publisher instructions for author websites, including high standard scientific and medical publishers. *Springer*, for example, states on their website, "If English is not your native language we strongly urge you to have the text of your paper checked by a native English speaker before submission" (Springer, 2018). In fact, in Japan, there is even a term to

refer to the English editing of manuscripts - "native-check" (Propenko, 2018). With this controversy surrounding the review process, and especially in science, where the process is not double-blinded, i.e. author names and origins/affiliations are revealed to assessors (e.g. Murugesan, 2014), it is important to understand the nature of differences in stylistic features of writing between EFL and L1 researcher written texts.

Systemic functional linguistics (SFL) can be used to study many aspects of textual features (Carter-Thomas, 2016; Coffin & Donohue, 2012), and is one form of discourse analysis that has been used extensively in the analysis of writing in the academy (e.g. Fontaine, 2003; Gao, 2012, Moore, 2007). It is a theory of grammar that sees language in terms of function and purpose of utterances (Halliday, 1985). Among various tools of SFL, thematic progression, thematized participants, processes of transitivity, and nominalization have had a great impact on understanding scientific writing (e.g. Fontaine, 2002; Gao, 2012; Zheng, Yang & Ge, 2014). Using these tools of SFL, followed by a survey to explore possible extraneous factors, this study compares the texts of three different groups: L1 accomplished scientists (also termed the 'ideal' group), and L1/EFL students-Australian/Japanese born and educated respectively. The scientist texts are published abstracts - hence, termed 'ideal', while the student texts are original/non-edited early drafts.

The thesis is organized as follows. It first reviews the literature on global English as the language of research and previous studies using SFL to explore academic discourse. It then describes the methods used, elaborating the SFL tools of analysis that are employed in the current study, before reporting the findings of the textual analysis and the survey results. Finally, it discusses methodological and pedagogical implications, and possible future research directions.

2. Literature review

2.1 Global English as the language of research

The last five decades have been the fastest in the growing predominance of English as the common language of communication and research amongst the worlds' speakers of various first languages (L1) (Crystal, 2007). The advancing technology has led to efficient modern means of communication, further aiding the spread of English, creating many local English varieties with their own culturally influenced linguistic expressions (Crystal, 2007; Seidlehofer, 2005; Smokotin, Alekseyenko, & Petrova, 2014). These developments have led to a fast growing internationalization of research, and made universities across the world compete for credibility globally (Flowerdew & Wang, 2015). There are now international comparisons of University research outputs, which are used to rank universities worldwide (Flowerdew & Wang, 2015). These factors make the publication of research results by academics as well as Masters and Doctoral research students in English a growing requirement by many universities (e.g. Canagarajah 2002; Flowerdew & Wang, 2015). Therefore, English is now the widely accepted lingua franca of research, across many fields, including the natural sciences (Carter-Thomas, 2016; Salagar-Meyer, 2008).

According to a 2010 report (Lillis & Curry, 2010), more than 95% and 90% of natural and social sciences journals respectively, publish in English, according to Institute for Scientific Information (ISI)–which has a database of the largest collection of journals in these areas (Flowerdew & Wang, 2015). This trend in English language publication of research journals has been progressively increasing in the last decades for natural and social sciences alike (see, for example, Crystal, 2007; Graddol, 1997; Swales, 2004). In as early as 1995, English accounted for 87.2% and 82.5% of natural and social sciences publications respectively (Ammon, 2001). A study by Benfield and Howard (2000) showed that the proportion of natural science publications in English increased from 72.2% to 88.6% from

1980 to 1996. In line with these developments, there has also been a significant increase in the proportion of English publication of PhD theses in EAL/EFL settings (see for example, Ammon & McConnell, 2002). One consequence of the dominance of English for publication of research results and of access to information is that, often publications in local languages may go unnoticed (Salagar-Meyer, 2008). Researchers from many EFL settings, especially those from the developing countries, are often referred to as 'peripheral' researchers, as some of their data may be published in their L1s, and hence their findings may not be recognized/noticed for an extended period of time (Salagar-Meyer, 2008). In one such case, Chinese researchers reported in 2004 that a deadly strain of avian flu was transferable from pigs to humans (as referred to by Panko, 2017). Despite its urgent importance and immediate impact on health, it was not given attention outside China until much later by the World Health organization (WHO), when it was quickly translated into English. It is argued that this was mainly due to the article being published in a small Chinese journal, which was also possibly deemed not significant (Panko, 2017). Similarly, a detailed study by Amano, González-Varo and Sutherland (2016) reports on many cases when language can become a barrier in the way of publication and access to important information. The researchers report that a significant number of scientific research findings (around 35%) are published in local languages, possibly leading to late discovery of important information. They also report that the availability of latest scientific research results, mainly in English, puts local practitioners and policy makers of many non-English speaking regions at a disadvantage. These factors have made the teaching and learning of English in research settings an important aspect of access to and publication of research results.

The rise of English as the global language of research and communication has had a great impact on researcher career success (see, for example, Ammon, 2001; Crystal, 2007; Graddol, 1997; Salager-Meyer 2008; Swales, 2004). Currently, the productivity of researchers

(including those in EFL/EAL settings)–and hence their ability to compete in academia and acquire research funding, is measured to a large extent by their publications, primarily in journals visible to the international research community–i.e. English medium journals (Hyland, 2004, p. 5). Under pressure from their institutes and the need for an international research profile, i.e. "publish in English or perish", including in science, effective writing for publication in English in EAL and EFL settings has become a strongly sought after skill (Di Bitetti & Ferreras, 2017; Salagar-Meyer, 2008). This has put emphasis on the challenging task of writing and publishing in English, especially for the less experienced students in EFL/EAL settings (Flowerdew & Wang, 2015).

The importance of EFL/EAL researchers to global knowledge (e.g. Flowerdew 2007), and the demand for proficient writing in English, has led to the emergence of a new field of English language for research, education and academia, termed 'English for Academic Purposes' (EAP) (Belcher, 2009). As Belcher (2009) discusses, EAP is tailored to learner English language needs at various levels of education. This field has now become a key field of education in many universities around the world with the aim of facilitating the acquisition of English language proficiency in academic settings, including research (Afful, 2007). Within this field, there is also strong emphasis on discovering new ways to facilitate the learning of English for the writing of research results–as this area has now attracted great attention due to high demands, termed 'English for Research Publication Purposes' (ERRP) (Belcher 2007; Flowerdew 2001). This area of EAP has now attracted the attention of many scholars, especially in EFL/EAL contexts (Belcher 2007). With the advent of global English, English is now also the lingua franca of science, and ERRP has attracted intense research and debates in scientific research communication (Tardy, 2004).

2.2 Using SFL tools to understand stylistic features of texts

Systemic Functional Linguistics (SFL) is a functional and semantic (meaning) view of grammar, based on the hypothesis that the function of language has evolved depending on use in context (Halliday, 1985). 'Systemic' refers to a unified group of utterances (expressions/statements) that together carry a message, and 'functional' refers to the function of language, i.e. what it has evolved to do, which is getting the message across in a meaningful way (Halliday & Matthiessen, 2004; Fontaine, 2013). It emphasizes a semantic perspective of grammar; the grammar used for everyday life communication in a stylistically meaningful and logical way. Unlike traditional grammar, which focuses on structure separate from meaning and context, e.g. the rules of how each individual sentence is made; SFL sees meaning and form in the context and purpose of communication, that is, how a group of utterances (e.g. sentences in a text) can together carry a message (as discussed in Fontaine, 2013). In the remainder of this section, a number of SFL tools are introduced and explained in terms of how they are understood in previous research and in the current study.

All SFL analytical tools operate with clauses/clause complexes as the unit of analysis (Halliday, 1985). That is, a complete clause/clause complex with all its grammatical items, sufficient for getting a unit of message across. This can be a single independent clause, i.e. a complete unit of meaning; or a clause complex-an independent clause together with dependent clause(s). So, a sentence in this sense can consist of one or more clauses/clause complexes, depending on how many separate independent units of meaning it can be divided into. For the purpose of analysis, the sentence gets divided into its separate independent units (clause/clause complexes) and each is analyzed independently because of the message they carry on their own. In the rest of this thesis, for simplicity purposes, an independent unit of meaning will be referred to as a clause, even if it is a clause complex. This will also be further clarified in the methods section.

One aspect of SFL, thematic progression, is concerned with the unity of text, especially with the flow of main ideas/messages from clause to clause (Danes, 1974; Halliday, 1985; McCabe, 2004). In this sense, a text is understood as a collection of clauses that together carry a unified meaning/message. Even if clauses in a text are all grammatically correct, if they are not in unity with each other in terms of how a message gets passed on from one clause to the next, there may still be problems with its comprehension (Carter-Thomas, 2016; Muranen,1996; Ventola & Muranen, 1991). This unity of text is called texture; that is, a unified text that functions within a context, and carries a main message in a cohesive and logical manner (Halliday & Hasan, 1976). This passing on of a message from clause to clause within texture is called thematic progression.

Another tool of SFL, closely related to thematic progression, is the choice of thematized participants (Halliday, 1994, p. 58). Thematized participants refer to the subject(s)/ideas that have become thematized, i.e. they have been made the main theme/idea of a text. As Forey (2002) argues, success in terms of the logical development of a topic, involves knowing where to put thematic participants within a clause. Thematic participants are the subject of the text and need to always occupy the subject position, within the beginning of the text, just up to but not including the finite verb. This part of the clause is hence called the theme, the part that carries the idea/message. Everything else in the clause, including the finite verb and onwards is called the rheme (Cummings, 2003, p. 133; Halliday, 2014, p. 89). The theme is known information, and the rheme is new information regarding the theme. To develop an idea in a logical manner, an author mentions the theme (topic) first (before the finite verb) followed by relevant new information in a clause (from the finite verb to the end). Thematic progression is, therefore, the progression of thematic participants from clause to clause, as they may get modified or changed, to develop an idea in a text.

A third relevant tool of SFL, processes of transitivity, describes how a writer expresses a process using the object/predicate of an action (Halliday & Matthiessen, 2014, p. 227). Halliday & Matthiessen (2014, p. 227) identify six ways by which presenters/writers can express themselves. These are used for different purposes in a text, including for describing an action (material process), analysis/sensing (mental process), saying (verbal process), comparing the relationship between two participants/ideas (relational), expressing if something exists (existential process), and laughing/crying/talking (behavioral process). According to MacDonald (1994) and Coffin (1997), these processes play a significant role in how a message is portrayed within a clause by a writer. If a writer uses an incorrect process to describe a message, then their intended message may not get across clearly. In sum, the correct use of processes of transitivity also contributes to the logical flow of a text.

The fourth application of SFL relevant to this study is Nominalization, which Halliday (1993) defines as the use of nouns/nominal groups to convey more messages in the same clause, thus making texts shorter, more lexically rich, and more abstract. Two different clauses carrying independent units of information have one finite verb each, and they can be combined by converting one of the finite verbs into a noun (Halliday & Martin, 1993; Halliday & Matthiessen, 2004, p.214). For example, instead of using a verb group (e.g. 'to be able to), a noun can be used (e.g. 'ability') (Liardet, 2016). This feature is 'essential', as Halliday (1994) states, in construction of abstract and lexically rich sentences, for example, to meet the demands of scientific journal publication standards, especially with their relatively low number of words compared to other fields (Halliday, 1993; Martin, 2008, p.803). In fact, it has been argued that scientific texts are the most nominalized of all discourses (Banks, 2005). Due to its importance in writing, nominalization is used to assess the effectiveness of student writing (Liardet, 2016), making it potentially a key feature in assessment and teaching of writing.

The different tools of SFL used in this study can be combined to make texture. For example, correct use of thematic progression, depends on the correct location of thematic participants- at the start of the clause as the subject, and the type of keywords used to describe the participants plays a significant role in correctly describing a theme (main topic) as it progresses from clause to clause (Danes, 1974; Halliday, 1985; McCabe, 2004). Similarly, the type of processes of transitivity used determines how a message is portrayed (e.g. Moore, 2007). The incorrect use of processes of transitivity can therefore contribute to disruption of progression of the main idea (theme) and logical flow. And lastly, nominalization helps determine the intended type of text: for example, the length (how concise?), the formality (how abstract?), and the flow (how logical?) (e.g. Gao, 2012). As a number of scholars have argued (e.g. Fontaine, 2003; Gao, 2012; Moore, 2007), understanding how to use these tools to diagnose student writing problems can be of great benefit, and has been the matter of intense scholarship.

2.3 Linguistic features of scientific texts written by L1 and EFL/EAL writers

The English language used for the writing of scientific articles (English for Scientific Purposes/ESP) has many similarities and yet distinct features to that of other fields (as argued by Halliday and Martin, 1993). Critical here is the idea that scientific writing may have undergone much more standardization as it requires much more preciseness and unambiguity, describing phenomena in literal and direct terms, with little room for compromise and contradictions (Halliday and Martin, 1993; Ortoney, 1993). This standardization of scientific writing has been facilitated by the evolving communication technology and advances in science, leading to changes in the linguistic structure of scientific texts, to allow for more concise writing for more efficient and direct/literal communication of more data (Banks,

2005). This makes the process of writing a scientific manuscript especially difficult for EAL/EFL writers.

Many studies have investigated differences in text structures in writing by L1 versus EAL/EFL scientific writers, with particular emphasis on nominalization (e.g. ElMalik & Nesi, 2008; Fontaine & Kodratoff; 2003; Mahbudi, A, Mahbudi, L, & Amalsaleh, 2014). In one study, Gao (2012) compared the rate of nominalization in abstracts of scientific articles written by Chinese (EFL) and English-background (L1) writers. Counting the number of nominal groups per word, he found that L1 writers used nominalization on average one out of every 15.8 words, while the Chinese writers used it one out of every 21.8 words. Nominalization can help in the construction of lexically rich abstract texts by facilitating the writing of ideational information of two or more clauses in a single participant (Ravelli, 1988). By calculating the number of lexical items per clause, Gao (2012) also found that as a result of higher frequency of nominalization, the L1 writers had sentences with higher lexical density. Other studies focusing on English L1 and EAL/EFL writers of various contexts have also reported similar findings (e.g. Flowerdew, 2006).

Another aspect of scientific writing, also an application of SFL, is the introduction of background and new information in sentences, called theme and rheme respectively (Danes, 1970). As described earlier, this feature of writing is important in structuring information flow in texts as background information is given and new information is introduced (Danes, 1974, Halliday, 1985). Here, text analysis can be carried out to investigate if every theme (background information) is stated for the purpose of introducing a rheme (new information). If there is no related rheme for a theme, then the logical flow of a text may be disrupted. Through theme and rheme connections, information can be built and discussed (Danes, 1974; Halliday, 1985; Hawes, 2015; Ventola & Mauranen, 1991). Gao (2012) also analyzed the progression of theme and rheme in the texts of Chinese and English L1 writers. He found one

pattern of progression—theme in clause two followed rheme in clause one (linear progression) was much more commonly used by the L1 speakers. He argued this may be due to the Chinese writers' difficulty in finding words similar to the rheme to describe the theme. Other studies have also found varying patterns of expression between L1 and EFL/EAL writers (Fontaine & Kodratoff; 2003; Hawes, 2015; Ventola & Mauranen, 1991).

While most of the studies above have made use of published—and possibly extensively edited data given the rigorous assessment for journal publications (Arunsirot, 2013; Flowerdew, 2006; Fontaine & Kodratoff; 2003; Gao, 2012), studies into non-edited scientific texts by researchers, especially students of EAL/EFL contexts using SFL tools are lacking. The studies above have also not clarified any influence of potential English-background co-authors and coworkers on the writing of the manuscripts by the non-English speaking participants. Therefore, analysis of non-edited writing may help us better understand writers' original scientific texts, especially those of students, which so far remain to be investigated. In addition, so far the scientific texts of EFL/EAL students have not been compared to that of L1 students and to that of published articles to investigate how similar/different they are from texts that have met journal standards (here, termed 'ideal'). Investigating how students from different backgrounds write may help understand student problems with linguistic expression and with meeting the demands of publication-level writing.

Therefore, this study aims to answer the following questions:

1) Can the use of SFL analytical tools, such as thematic progression, the processes of transitivity, and nominalization, identify and explain the differences between the early drafts of scientific texts written by students from Australia (L1) and Japan (EFL)?

2) What are these differences?

3) How do the student texts differ from 'ideal' scientific texts written by experts?

<center>**3. Methods**</center>

The purpose of this study is to identify and explain any differences in the linguistic features of three corpora of scientific abstracts using SFL tools of analysis. The first two corpora are non-edited original student texts from 1) EFL students (Japan) and 2) English L1 students (Australia). These were compared with a third corpus of published 'ideal' abstracts written by accomplished scientists and published in scientific journals. A short online survey of the students complemented the textual analysis. This section describes the student participants as well as the data collection and data analysis methods.

3.1 Participants

The participants chosen for this study were graduate students in the field of neuronal development and disease (student researcher's field of expertise), from Australia (L1) and Japan (EFL). As the main eligibility criteria for the EFL students, they must be all born, raised and educated in Japan, and where possible, have no significant experience of studying/staying in an English L1 country at the time of writing their manuscripts (Table 3.1).

Table 3.1: Participant demographics and backgrounds							
Participant	Country/ Institute	L1	Gender	Research area	Years studied English	Prior stay in an L1 country	Level of education
L1 student 1	Australia/ Institute 1	English	Female	Neurons and disease	Whole life	Whole life	PhD
L1 student 2	Australia/ Institute 1	English	Male	Neuronal development	Whole life	Whole life	Honours
L1 student 3	Australia/ Institute 2	English	Female	Neurons and disease	Whole life	Whole life	PhD
EFL student 1	Japan/ Institute 1	Japanese	Female	Neuronal development	Japanese schooling	None	PhD
EFL student 2	Japan/ Institute 2	Japanese	Male	Neurons and disease	Japanese schooling	None	PhD
EFL student 3	Japan/ Institute 3	Japanese	Male	Neuronal development	Japanese schooling and 4 years university	Three months, US	PhD

This is to limit the possibility that the results are confounded by possible English learning experiences in an L1 country. Similarly, participants from Australia must also be born, raised

<center>12</center>

and educated in Australia, with no extended periods of study overseas. All student participants were undertaking their graduate research studies (honours or PhD) at the time of writing their manuscripts. Of the six students, two were females and four were males. All were working at laboratories specializing in a similar field of science, i.e. neuronal development and disease, to allow easier comparisons across texts. It is, important to note that the students in the Japanese group are from very high standard universities, two of which require them to pass difficult English language tests for successful admission.

Participant recruitment was first initiated by directly contacting the students' supervisors using the laboratory emails provided online on their laboratory websites. After obtaining supervisors' permission and students' emails, the students were directly contacted and asked to provide non-edited texts of their manuscript drafts. A plain language statement explaining the content of the intended research study and a consent form to sign were forwarded to them. They were also asked to fill out a survey questionnaire of about 20 minutes regarding their educational and linguistic backgrounds.

3.2 Data collection

This study uses two main sources of data, scientific manuscript abstracts and online surveys. As mentioned, there were two main groups of scientific abstracts, those published in scientific journals and written by experts – labeled in this study as 'ideal' and abstracts written by students. The ideal abstracts were sourced from the United States National Library of Medicine (NCBI) website (NBCI, 2018). Both the ideal and the student abstracts were chosen based on a number of criteria. These include: 1) required word limit (by the journals), 2) style of abstract (i.e. single paragraph), 3) type of manuscript (research papers only), 4) authors: first and last author must be from the same L1 background, 5) area of research, 6) participants born, raised and educated in their L1 countries, and 7) minimal residence/study in

an English L1 country for EFL students at the time of writing their manuscript. Criteria 1 to 3 are related to linguistic expression and logical flow. For example, if the abstract is made up of smaller sections (criterion 2), i.e. objective, methods, results and conclusions (e.g. MacDonald et al. 2017), all of which require only one or two sentences, then the style of writing, lexis, and cohesiveness may be different to that of an abstract written as a one-paragraph piece (e.g. Stoppel et al. 2017). One can assume that in a one-paragraph abstract, the author is required to summarize all sections together in one piece while keeping with correct logical flow and repetition of ideas. With respect to criterion 3, review articles do not discuss new data, as they are designed to review all the known relevant literature, e.g. see *Nature* Reviews Neuroscience Instructions for Authors (*Nature* Reviews Neuroscience, 2018), while research articles discuss original new findings and integrate their significance with the relevant literature, e.g. see *Nature* Neuroscience Instructions for Authors (*Nature* Neuroscience, 2018). The criteria ensure that the justification and discussion of results follows a similar format, comparable for analysis. Along with these criteria, it was also important that the L1 of both the main author and the last author (supervisor) is that of the intended EFL/L1 setting. This is to ensure that the writing represents that of a student in the intended language setting without significant exposure to English (if in an EFL setting) and exposure to English (if in an L1 setting). The following table (Table 3.2) is a list of the abstracts and some of the criteria.

Table 3.2: List of abstracts and their criteria						
Abstract	Main author	Institute	Journal/ Year	Total number of words/ Journal word limit	Field	Code
Ideal 1	Larson RS	University of North Carolina	Female	127/150	Synapse function	Ideal1
Ideal 2	Stoppel JL	Massachusettes Institute of Technology	Male	149/150	Synapse and disease	Ideal2
Ideal 3	Sanders J	University of California San Diego	Female	147/150	Synapse and memory	Ideal3
L1 student 1	N/A	Australia/ Institute 1	Female	204/200	Synapse and behavior	L1-1
L1 student 2	N/A	Australia/ Institute 1	Male	190/200	Synapse and disease	L1-2
L1 student 3	N/A	Australia/ Institute 2	Male	198/200	Synapse and disease	L1-3
EFL student 1	N/A	Japan/ Institute 1	Female	249/250	Synapse function	EFL1
EFL student 2	N/A	Japan/ Institute 2	Male	256/250	Synapse and disease	EFL2
EFL student 3	N/A	Japan/ Institute 3	Male	224/250	Synapse and disease	EFL3
Notes: 1. For privacy reasons, pseudonyms are used instead of the names of student participants. 2. The institutions of student participants are not revealed. 3. All information is with regards to the situation at the time of writing the manuscripts.						

Given the small number of abstracts collected for analysis (three from each of L1 and EFL students), it is acknowledged that this study is exploratory in nature and that the data may not fully represent student writing in general in the corresponding settings. It is important to reiterate here that the main aim of the study is to understand if and to what extent the various SFL tools can be used successfully in analysis and diagnoses of student scientific writing in comparison to requirements by journal standards (ideal/published ones), an area so far not investigated.

A short online survey was emailed to students (see Appendix 1) to collect data to assist in the interpretation of the analysis of abstracts. The survey elicited information from the student writers, including, year of graduate study (i.e. level of experience), language used for the first draft of their abstract (i.e. directly in English, or Japanese followed by

translation), level of help and advice received during writing (i.e. is it their original writing?), prior experience in writing in English, prior residency in an English L1 country, and finally, any knowledge of SFL. The survey data was collated, and was referred to in the process of interpretation of the findings of the textual analyses.

3.3 Data analysis

SFL tools were used to compare the three corpora of abstracts. Five main types of analyses were conducted as follows: 1) thematic progression: in which the clause to clause progress through the text was identified via constant, linear or hypertheme progression; 2) theme and rheme composition, which included the number of words (proportion) used in the themes and rhemes of each clause; 3) thematic participants: involving the classification of the use of different types of keywords such as generic, subordinate, summarized, abbreviated, etc.; 3) transitivity: i.e. classification of messages (here clauses) into six different types of processes, of which five are relevant to this study; relational, material, mental, verbal and existential; 4) nominalization: i.e. identification of the use of nouns/nominal groups to combine two clauses into one, in order to make a text more concise and abstract. The textual data were combined with a survey questionnaire to understand potential extraneous factors.

The following illustrates the analytical procedures with examples from one of the ideal texts (Ideal3):

1. The abstracts were divided into clauses:

As mentioned in Section 2.2, SFL tools use clauses as the unit of analysis (Halliday, 1985). First, a diagram was made which includes the whole text of an abstract divided into clauses (Figure 3.3).

16

Figure 3.3: An example of the analyses carried out using SFL tools

Elimination of Dendritic Spines with Long-Term Memory Is Specific to Active Circuits
of Experience-Dependent Plasticity by Presynaptic NMDA Receptors

A ↓ **Hypertheme**

Clause1 *Structural changes* in brain circuits active during learning are thought to be important for long-term memory storage.

B ↓

Clause2A If these changes support long-term information storage, they might be expected to be present at distant time points after learning,

Clause2B C → as well as, to be specific to the circuit activated with learning,

Clause2C D → and sensitive to the contingencies of the behavioral paradigm.

E ↓

Clause3 *Here,* we show such changes in the hippocampu as a result of contextual fear conditioning.

F → Clause4 There were significantly fewer spines specifically on active neurons of fear- conditioned mice.

G → Clause5 This spine loss did not occur in homecage mice or in mice exposed to the training context alone.

H → Clause6 Mice exposed to unpaired shocks showed a generalized reduction in spines.

I ↓

Clause7A These learning-related changes in spine density could reflect a direct mechanism of encoding,

J ↓

Clause7B or alternately could reflect a compensatory adaptation to previously described enhancement in transmission due to glutamate receptor insertion.

As can be seen from Figure 3.3 above, the text consists of 10 clauses, but seven sentences, as sentence 2 and 7 are divided into three and two clauses respectively, all connected via conjunctions. Each clause carries a separate unit of information.

2. The theme and rheme in each clause were identified:

The theme is the beginning of the message that gives background information, i.e. all the words up to but not including the finite verb (colored in blue). The rheme is new information related to the theme and includes everything else, including the finite verb, i.e. the exit of the clause (colored in green) (Halliday, 1985).

3. The participant keywords were identified:

The main focus (message/idea) of abstract Ideal3 is "elimination of dendritic spines", i.e. loss of synapses/neuronal connections, which is a "structural change" in the brain as synapses (tiny structures) are lost. The keywords that identify this main participant are underlined in blue and correctly located in the themes of each clause of the abstract (e.g. Moore, 2007). They start from very specific (title) and go to general (clause1), modified (clause 2A, i.e.

"changes" modified by "these"), textual conjunctions (clauses 2B, 2C and 7B), and word group experimental participant (clause 6).

4. The different types of thematic progression were identified:

There are three simple types of thematic progression (Danes, 1974): illustrated in Figure 3.3. as hypertheme (boxed red arrows), constant (blue arrows) and linear (green arrows). Hypertheme is when the theme of one clause is derived from the title of the abstract (red arrow, marked 'A'), the whole study (red boxed arrow, marked 'E'), and/or summarizes a few preceding clauses (red boxed arrow, marked 'I'). The first hypertheme ('A') uses the general keyword "structural changes" at the beginning of clause 1, which is related to the specific keyword "elimination of dendritic spines" (i.e. loss of neuronal connections). The boxed arrow ('E') refers to the whole study using the summative (summarizes) keyword "here", while the summative keyword from the arrow marked 'I' ("these learning related changes") summarizes the themes of clauses 4, 5 and 6. As for constant progression, the blue arrow marked 'B' uses a keyword from the preceding theme ("changes") modified by an article ("these"), and for linear theme, one example is clause 3 to 4 (marked 'F'), i.e. "there were" (textual keyword) refers to "such changes".

In addition to these progression types, there are two higher order progressions, constant split and linear split (McCabe, 2004). Constant split is when a theme (e.g. clause 2A) gets divided into more than one sub-theme (the themes of clause 2B and 2C), i.e. in this case, the textual keywords "as well as" and "and" (underlined), progressions 'C' and 'D' respectively, refer to the keyword "these changes". Similarly, split linear is the division of a rheme into separate themes, as in the division of clause 3 to 4, 5 and 6.

5. The processes of transitivity were identified:

Transitivity is the system for interpreting and expressing the world inside and outside of the speaker/writer (Halliday & Matthiessen, 2014, p. 227). These processes are classifying experience of events in the world around the speaker/writer into a small number of types, so we can identify each type. The processes of transitivity can be divided into six types, five of which are relevant to this study (see Table 3.3A).

Table 3.3A The processes of transitivity			
Process	**Meaning/message**	**Example**	**Explanation**
Material	action/doing	Sensory experience orchestrates the development of cortical circuitry. (Source: Ideal1)	"orchestrates" refers to 'doing'
Mental	Perception/sensing/ analyzing	Structural changes in brain circuits active during learning are thought to be important for long-term memory storage. (Source: Ideal3)	"are thought to be" is analyzing
Behavioral	Behavior	N/A	Not used in this study
Verbal	Saying	*However, the mechanisms underlying these experience-dependent modifications <u>are not reported</u>. (Source: Ideal1)	"reported" refers to saying
Existential	Exisiting	However, the mechanisms underlying these experience-dependent modifications remain elusive. (Source: Ideal1)	"remains elusive" refers to a knowledge that doesn't exist yet.
Relational	Being and having, here, to relate two participants	This spine loss did not occur in homecage mice or in mice exposed to the training context alone. (Source: Ideal3)	the two participants from the previous clauses compred are "homecage mice" and "mice exposed to the training context"

* The same example from Ideal 1 is altered to show an example for the verbal proces, as none of the Ideal abstracts use this process type, and the student abstracts were not used for examples to protect their privacy.

6. Nominalization frequency and lexical density were calculated:

As discussed in Section 2.3, a definition of nominalization (relevant to this study) is the use of nouns/nominal groups instead of verbs to make texts concise, by joining ranking clauses, i.e. the small constituents of a clause that contain a finite verb (Halliday and Matthiessen, 2004) (Table 3.3B).

Table 3.3B: Calculation of nominalization frequency and lexical density							
	Examples	Ranking clauses	Total words	Nominalizations	Nominalozation frequency	Lexical items	Lexical density
Nominalized clause	A) **These learning related changes** in spine density <u>could reflect</u> a direct mechanism of **encoding**	1	14	2	7	9	9.00
Denominalized/not nominalized version2	B) These changes in spine density that <u>are related</u> to learning, <u>could reflect</u> a mechanism of **encoding**.	2	16	1	16	8	4.00

* A clause can be unpacked into smaller constituents, each of which must contain a finite verb

In table 3.3B, an example is used from Ideal3 to illustrate the process of calculating nominalization frequency and lexical density for each clause. Here, the original nominalized version (example A) has two nominalizations (marked in bold) and only one finite verb (underlined). The total number of words in the clause is divided by the total number of nominalizations to get the nominalization frequency. So, the resultant nominalization frequency of 7 means that there is one nominalization event per every 7 words, and hence, a total of 2 nominalizations (in a total of 14 words). The number of lexical items (nouns, verbs, adverbs and adjectives; all highlighted in blue) can be calculated and divided by the number of ranking clauses (Halliday and Matthiessen, 2004). As can be seen from the table, de-nominalizing the clause complex (as in example B, by reducing one nominalization event) introduces more finite verbs into the clause complex increasing the number of ranking clauses. This leads to a decrease in lexical density. During the data analysis, the findings were also discussed with a researcher with expertise using SFL as an analytical tool in the area of academic discourse.

4. Results

This chapter will present the results of the analyses carried out using SFL tools on the 'ideal'- published abstracts of accomplished English L1 scientists, and the abstracts of L1 - Australian born and educated students, and EFL - Japanese born and educated students. Comparing L1 students to their L1 accomplished scientist counterparts allowed investigation of the role of experience in writing, while comparing EFL students to L1 students allowed examining the potential influence of L1 background. The results of the analyses are reported in the following order: thematic progression patterns and types in each group; the theme and rheme length (number of words in each), the use of keywords to describe participants/ideas; the use of processes of transitivity for each group; nominalization frequency and lexically density; and a survey to understand student linguistic and educational backgrounds.

4.1 Analysis of thematic progression types and patterns:

A tendency to use different types of thematic progression in texts is thought to depend on experience and L1 background in many disciplines (Ebrahimi & Ebrahimi, 2012; Lu, 2013; Rorvik, 2012) however, in scientific writing, this remains unclear. As mentioned earlier, progression in texts from clause to clause can be via *simple constant*, where participant keywords from the theme of one clause are linked to the theme of a preceding clause; *simple linear*, where the participant keywords of the theme of a clause are connected to keywords (new information) in the rheme of a preceding clause; and *hypertheme*, where the theme is connected to the abstract of the title and/or concepts in the whole study. These thematic relations are used for maintaining the same focus while adding information, developing an idea and summarizing main ideas, respectively. Analysis of the three corpora revealed interesting progression types in each groups of texts. On average, the ideal abstracts

used linear progression more often, whereas the L1 student and EFL student texts tended to use constant progression more often (Table 4.1, Figure 4.1, see also Appendix 2 to 4 for all samples of each group). This is consistent with previous studies investigating published medical abstracts of L1 and EFL scientists (Gao, 2012).

Table 4.1: Percentage of thematic progression types used by each group

A) Thematic progression types in ideal texts

	Hypertheme	Constant	Linear	Rupture
Ideal1	22.2	11.1	66.7	0
Ideal2	44.4	44.4	11.1	0
Ideal3	33.3	33.3	33.3	0

C) Thematic progression types in EFL student texts

	Hypertheme	Constant	Linear	Rupture
EFL1	25.0	37.5	18.8	18.8
EFL2	26.7	53.3	20.0	0
EFL3	30.8	38.5	23.1	7.7

B) Thematic progression types in L1 student texts

	Hypertheme	Constant	Linear	Rupture
L1-1	33.3	33.3	33.3	0
L1-2	37.5	62.5	0.0	0
L1-3	33.3	44.4	11.1	11.1

B) Average thematic progression types in all groups

	Hypertheme	Constant	Linear	Rupture
Ideal	33.3	29.6	37.0	
L1	34.7	46.8	14.8	3.7
EFL	27.5	43.1	20.6	8.8

Figure 4.1: Diagrammatic examples of thematic progression types for one text of **each of the ideal, L1 student and EFL student samples**

A: An example of thematic progression in ideal texts. Both split linear and split constant progressions can be seen.

B: An example of thematic progression in L1 student texts. Simple linear, simple constant, and linear JUMP and constant JUMP progressions can be seen.

C: An example of thematic progression in EFL student texts. Simple linear, simple constant, and linear JUMP and ruptured progressions can be seen.

Note: For diagrams of all groups for each text see appendix figure 4.1

The thematic progression types in the corpora under investigation in the current study were further clarified into more elaborate progressions (McCabe, 2004). These are: split

theme (elaboration of a simple theme), where a theme can be divided into two or more subthemes; and split rheme (elaboration of a simple rheme) where a rheme can be divided into a set of sub-ordinate ideas (themes). While, as can be seen from Table 4.1 and Figure 4.1, some ideal abstracts also make frequent use of constant progressions, they tend to use them to make a set of subordinate ideas, i.e. to go from one theme to a set of sub-themes (Figure 4.1A, splitting blue arrows). The ideal abstracts also used split-rhemes to make a set of thematic subordinate ideas (Figure 4.1A, splitting green arrows). These types of progressions were very rare in L1 student and EFL student samples (Figure 4.1, B and C), with only one EFL student splitting a theme into two subordinate ideas.

In addition to the above theme and rheme types, cases of 'jump' and 'rupture' connections have also been reported in the literature (e.g. Fontaine, 2003; Gao, 2012). The former happens when a writer introduces a new participant in a theme that links to a theme (constant jump) or rheme (linear jump) a few preceding themes earlier in the text (hence, it jumps over a few themes or rhemes). The latter happens when a theme does not connect to any preceding themes or rhemes (hence, the progression of idea in the text is ruptured). Analysis of progression types in the samples showed no cases of linear or constant jumps in the ideal samples; however, these were common in the L1 and EFL student samples. The data suggests, therefore, even if the students were able to use linear progressions, they used them to refer to a few clauses earlier, resulting in jumps in the cohesion of ideas. In addition, there were ruptures in two EFL student texts (EFL1, three ruptures, and EFL3, one rupture) and one L1 student text (L1-2, one rupture). For diagrammatic representation of all samples, please see Appendix 2 to 4 (Pages 51-53)

4.2 Analysis of theme and rheme length in the different corpora

Studies on thematic progression analysis have so far focused on the types of progressions used in different samples, with no reports on the total length of themes and rhemes in texts; or possible changes in theme and rheme length from clause to clause, as the texts progress from introduction to conclusion (e.g. Fontaine, 2003; Gao, 2012). Counting the number of words for each theme and rheme of the texts showed intriguing results in the current study. The total number of words for all themes and rhemes of each text showed interesting differences in the proportion of theme and rheme lengths (Figure 4.2A, and Table 4.2A). While those of L1 students was only marginally lower than the ideal texts, those of EFL students, especially samples 1 and 2 showed a much lower difference between the length of themes and rhemes (Figure 4.2A, red lines with double arrows, and Table 4.2A).

Figure 4.2A Total theme and rheme length (in words) in ideal, L1 student, and EFL student texts
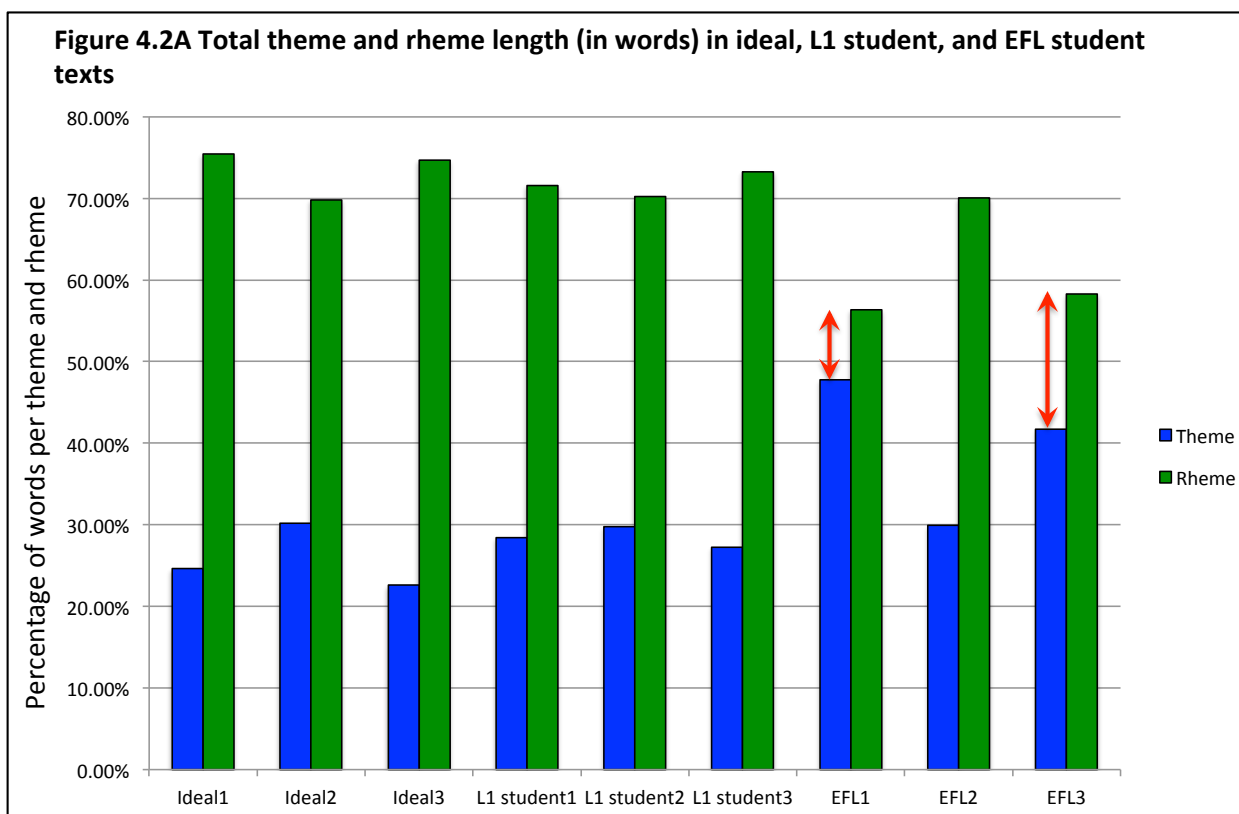
Table 4.2A Total word number (%) for the themes and rhemes in ideal, L1 student, and EFL student texts

	Ideal1	Ideal2	Ideal3	L1 student1	L1 student2	L1 student3	EFL1	EFL2	EFL3
Theme	24.6	30.2	22.6	28.4	29.8	27.3	47.8	30.0	41.7
Rheme	75.4	69.8	74.7	71.6	70.2	73.2	56.3	70.0	58.3

From this set of data, it was not clear, however, if there were any intra-textual differences in theme and rheme length. To do this, the total number of words in the themes and rhemes was measured individually and plotted on a graph against the clauses of the text in progression from the first clause of the text through to the final clause (Figure 4.2B).

Figure 4.2B: Theme and rheme length (in words, Y-axis) per clause of each text (X-axis) from beginning (left) to end (right) of abstract
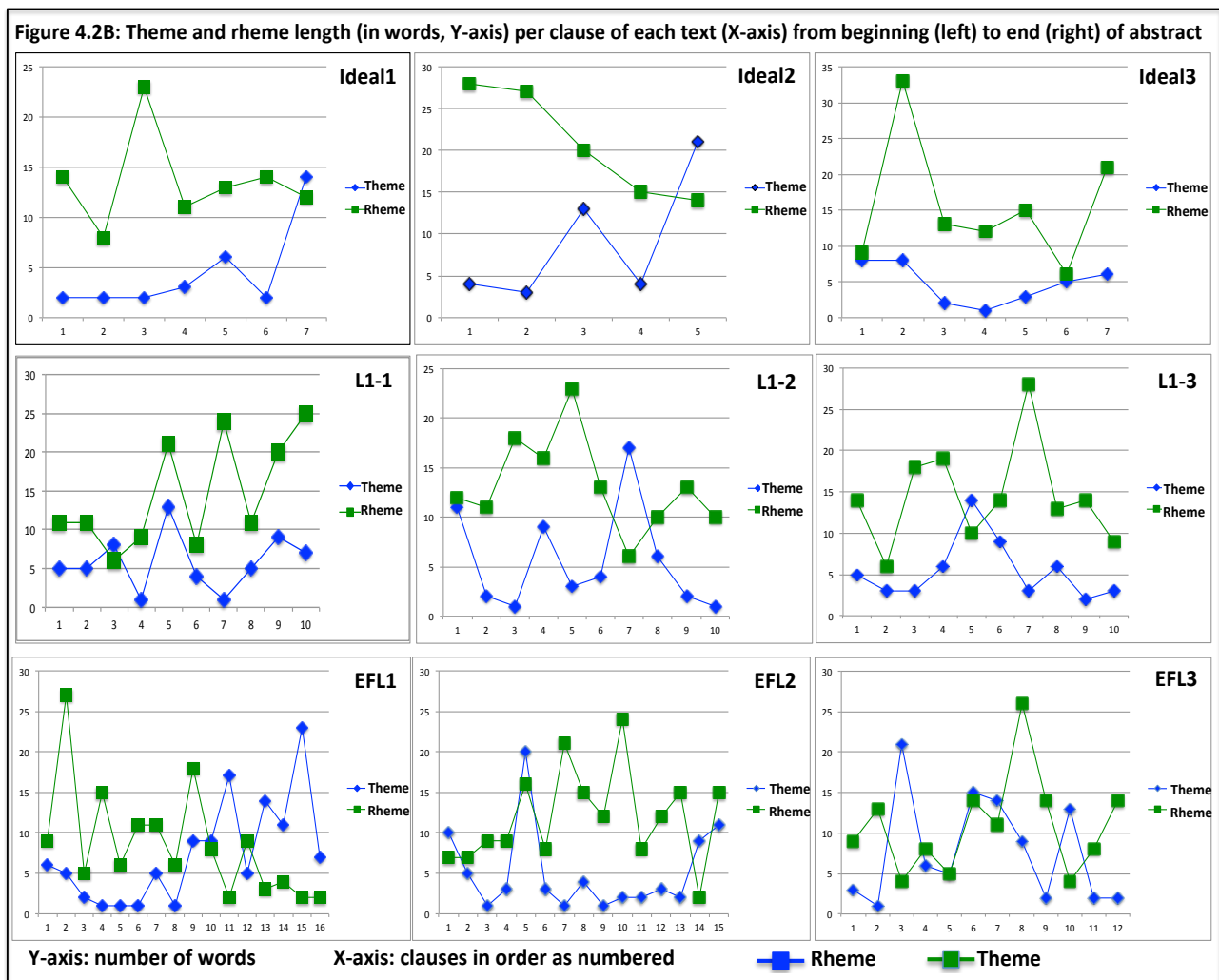
These graphs showed an interesting pattern for the different samples. The ideal samples seemed to start with short themes in the beginning of the abstracts and mostly (very few exceptions) stayed low until the end. However, theme length seemed to change randomly in the student samples with progression through the text. Here, EFL student texts showed much more randomness than the L1 student texts.

4.3 The choice of participant keywords as a possible cause of the differences in thematic progression and theme and rheme composition

While many studies have investigated the use of different word group types for thematic participants in themes and rhemes (e.g. Green, Christopher, & Lam, 2000; Hu, 2008; Qian, Andrés Ramírez, & Harman, 2007; Schleppegrell, 2004), none have suggested any link between keyword choice and its possible influence on the length of themes and rhemes. To further understand possible causes of the differences in thematic progression and theme/rheme length, the location and types of thematic participant keywords were investigated. As mentioned in earlier chapters, thematic participants should occupy words in the theme (e.g. Moore, 2007), and many different types of keywords can be used to describe them. Although all samples, with the exception of one occurrence in EFL1 text, put the thematic participants within the theme, analysis of theme composition revealed major differences in the use of participant keywords by the different groups. It was interesting to see that the use of keyword types was similar between L1 student and EFL student texts. Figure 4.3 shows an example from one ideal and one EFL student text.

Figure 4.3: An example of participant keyword progression through themes from one ideal and one EFL sample
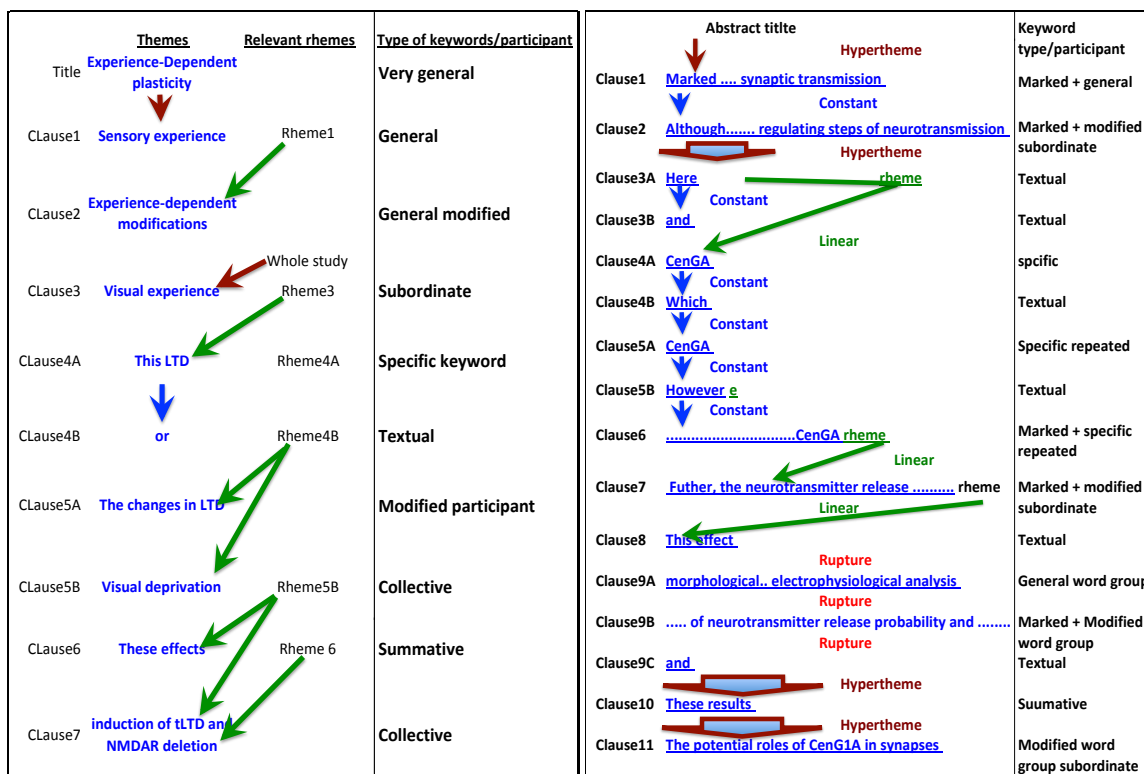
As shown in Figure 4.3, the ideal text (Ideal1) uses keywords that flow in a logical manner taking the focus from general, to modified (to be relevant to a certain area), to subordinate and then specific. The keywords then become summative-summarizing key concepts, and collective-bringing different participants together into one word group. However, as apparent from the diagram, the EFL student (and similar for L1 student but marginally better) does not make a logical transition. The text goes back and forth between different keyword types and includes a high proportion of textual participants. These findings suggest that keyword use may have an impact on thematic progression and theme composition. For a detailed diagrammatic representation together with a tabulated description of participant keywords for all Ideal samples, please see Appendix 5 to 7 (page 54-56).

4.4 Analysis of the use of processes of transitivity by the different groups

Processes of transitivity-the types of utterances used by authors to express the message/purpose of each clause complex are important in helping writers to express themselves clearly and appropriately (Halliday & Matthiessen, 2004, p.214). Each of the processes is used for a specific utterance: relational for describing two participants in the same clause complex; mental for analytical purposes (e.g. perception, sensing), material for actions; verbal for saying, and existential for existing. Another process of transitivity, behavioral did not become relevant to this study. Studies investigating transitivity processes have also found differences in the use of the different processes in medical writing (Zheng, Yang & Ge, 2014). However, they looked at corpora including whole articles, and potential differences in different sections of articles remain unknown.

Out of the six processes of transitivity, the relational process was much more commonly used in the ideal texts and its usage was only marginally lower in the L1 student texts (Table 4.4). However, that of the EFL students was considerably lower. Further

Table 4.4: Processes of transitivity in ideal (A), L1 student (B), and EFL student texts (C)

A)	Relational	Mental	Material	Verbal	Existential
Ideal1	5/9	1/9	1/9	0/9	1/9
Ideal2	4/9	2/9	2/9	0/9	0/9
Ideal3	7/9	2/9	0/9	0/9	0/9
Total	16/27	5/27	1/9	0/27	1/27
Proportion per 10 clauses	6	1.9	1.1	0	0.3

B)	Relational	Mental	Material	Verbal	Existential
L1-1	4/10	3/10	3/10	0/10	0/10
L1-2	5/10	4/10	1/10	0/10	0/10
L1-3	3/10	3/10	4/10	1/10	0/10
Total	12 12/30	10/30	3 8/30	1 1/30	0/30
Proportion per 10 clauses	5	3.3	1.1	0.3	0

C)	Relational	Mental	Material	Verbal	Existential
EFL1	2 2/16	3 6/16	5 7/16	1 1/16	0/16
EFL2	8 4/15	6 4/15	2 5/15	0/15	2 2/15
EFL3	1/13	2/13	6/13	2/13	2/13
Total	7/45	12/45	18/45	3/45	4/45
Proportion per 10 clauses	1.6	2.7	4	0.7	0.9

Table 4.5D: Average use of processes of transitiviy in ideal, L1 student, and EFL student contexts per 10 words of text

	Relational	Mental	Material	Verbal	Existential
Ideal	6	1.9	1.1	0	0.3
L1 student	5	3.3	1.1	0.3	0
EFl studnet	1.6	2.7	4	0.7	0.9

investigation of the process types and numbers showed almost a proportionate increase to the lower usage of the relational process, in the material process in the EFL student samples. There was also a marginal increase in the use of the mental process, but most of that came from EFL1. So, it seems that, the tendency for the use of material versus relational processes was a characteristic specific to the EFL students and not to L1 students. These results are similar to those of the study carried out by Moore (2007) investigating the writing of a non-English background student (in arts criticism), a mature-age Australian L1 student (in sociology) and another L1 student returning to study after an extended period of full-time employment (in history), all of whom were students at an Australian university.

4.5 Nominalization frequency and lexical density

Nominalization, here, the use of nouns/nominal groups instead of verbs, is a strong characteristic of scientific writing, and shown to be less frequently used by EFL writers (e.g. Gao, 2012). As Gao (2012) explains, nominalization makes texts very abstract and lexically rich (also see Halliday & Martin, 1993); a characteristic that also makes them difficult to understand, especially in EFL contexts. While Gao carried out his study using published articles (possibly edited) of scientists (possibly more experienced than students), original writings of students have so far not been analyzed. In the samples in this study, there were clear differences in nominalization frequency between all groups. These were calculated as
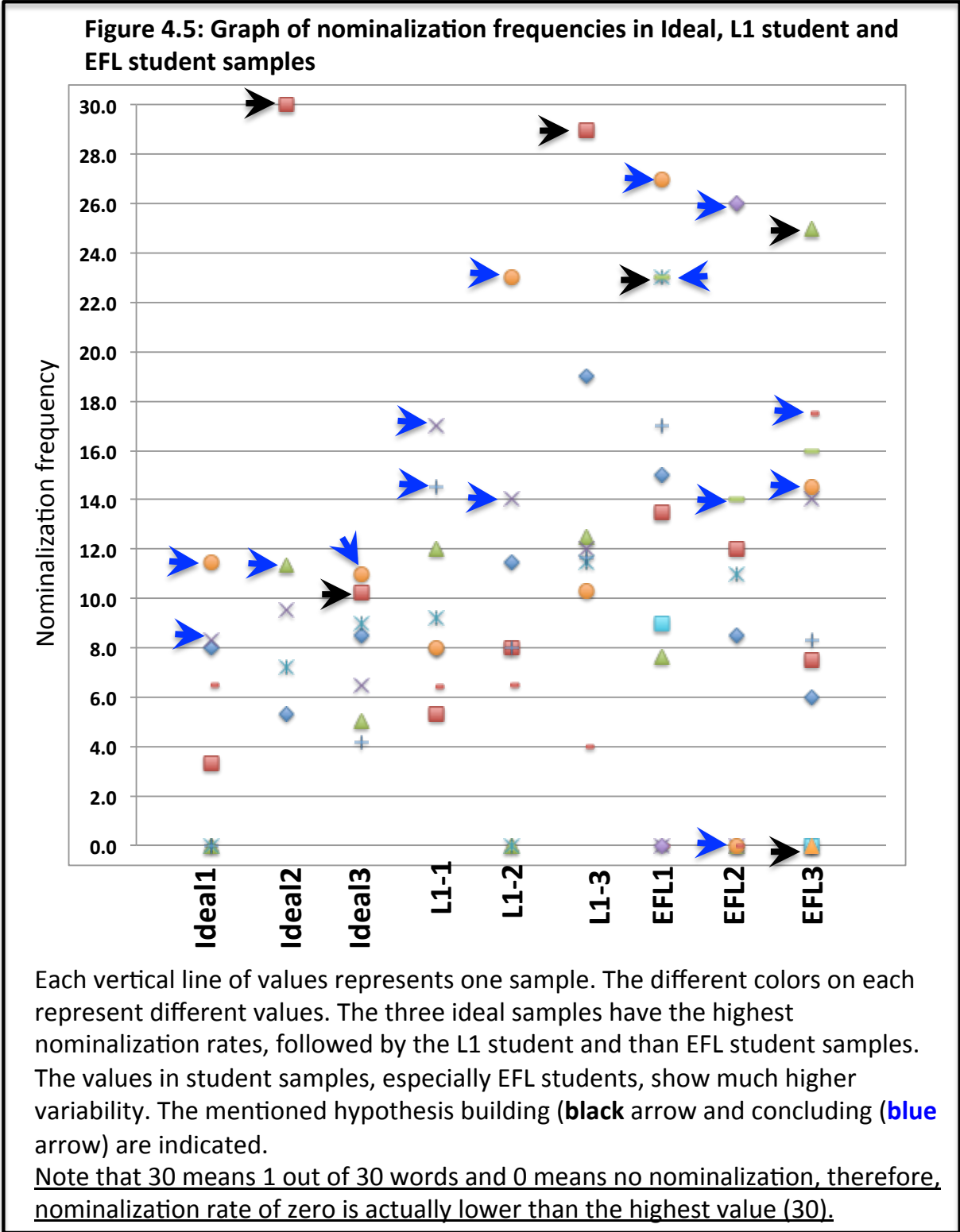
explained in the methods section, i.e. by dividing the total number of words to the total number of nominalizations. So, for example, a nominalization frequency of 9.3 (Table 4.5A, Ideal1) actually means 1 nominalization event per every 9.3 words in a text, therefore, a lower number (e.g. 9.3, i.e. 1out of every 9.3 words) represents a higher nominalization frequency than a higher number (e.g. L1-2, 12.7, i.e.1 out of every 12.7 words). Ideal to L1 student to EFL student group nominalization frequencies were high to medium (relative to ideal) to low (relative to L1 student and ideal) respectively (Table 4.5A).

Table 4.5: Nominalization frequency (A) and lexical density (B) in Ideal, L1 student, and EFL texts									
A)	Ideal1	Ideal2	Ideal3	L1-1	L1-2	L1-3	EFL1	EFL2	EFL3
Total words	130	151	144.0	204	190	198.0	252	257	224.0
Total nominalizations	14	17	21.0	22	15	17.0	12	10	14.0
Nominalization frequency	9.3	8.9	6.9	9.3	12.7	11.6	21	25.7	16.0
B)	Ideal1	Ideal2	Ideal3	L1-1	L1-2	L1-3	EFL1	EFL2	EFL3
Lexical items	87	100	88	137	133	104	129	147	160
Ranking clauses	12	15	14	13	14	16	21	25	23
Leixical density	7.3	6.7	6.3	10.5	9.5	6.5	6.1	5.9	7.0

To see if nominalization frequencies also lead to differences in lexical density, the total number of lexical items (nouns, verbs, adjectives, and adverbs) and non-lexical items were counted in each text. Although nominalization frequency was lower in the L1 student samples than the ideal samples, their lexical density was higher (Table 4.5A and B). Further analysis of the lexical density revealed that the ratio of non-lexical to lexical density in the student samples was higher than that of the ideal samples (0.65 and 0.55 respectively).

Another intriguing question regarding nominalization frequency was whether it fluctuates from clause to clause in the text. Studies have so far only investigated nominalization frequencies in whole texts, leaving potential fluctuations in frequency unknown in clause-to-clause progression through abstracts. A closer look at each clause of the texts of the ideal, L1 student and EFL student samples, revealed differences in nominalization frequency as the texts progressed from introduction to conclusion. Of particular note were the clauses that were used to develop hypothesis/ideas and making experimental conclusions.

Surprisingly, in these clauses nominalization frequency was the lowest (e.g. 1 out of 30 words), some with no nominalization (e.g. 0, blue and black arrows, EFL2 and 3 respectively) as shown in Figure 4.5 below.

Figure 4.5: Graph of nominalization frequencies in Ideal, L1 student and EFL student samples

Each vertical line of values represents one sample. The different colors on each represent different values. The three ideal samples have the highest nominalization rates, followed by the L1 student and than EFL student samples. The values in student samples, especially EFL students, show much higher variability. The mentioned hypothesis building (**black** arrow and concluding (**blue** arrow) are indicated.

Note that 30 means 1 out of 30 words and 0 means no nominalization, therefore, nominalization rate of zero is actually lower than the highest value (30).

4.6 Understanding possible extraneous factors of influence on writing features

Studies investigating features of writing may use SFL to analyze the texts of different writers and/or directly investigate writer sociolinguistic backgrounds and environments (e.g. using surveys and interviews) to understand how/why certain stylistic writing features may be more common than others (Coffin and Donohue, 2012). To ensure, at least to some extent, that knowledge of writer background and environmental factors are considered in the analysis of the data in this study, an online survey questionnaire was sent to the students (Appendix 1). The survey results showed that, as expected, all L1 students were exposed to English language communication as their first language from birth till the time of writing their manuscripts, with no extended stays in a non-English speaking country. The EFL students were similar; they were all raised and educated in their L1 (Japanese) with little exposure to English language education, except for that in the Japanese schooling system. Out of the three EFL students, only EFL3 had stayed in an English speaking country for a period of three months, and had prior experience in scientific English writing (a masters thesis, a few years earlier). Interestingly, EFL3 text was more nominalized (close but not equal to those of L1 students) and had a much higher lexical density (close to ideal texts) compared to the other two EFL students. This suggests that prior experience in scientific writing and possibly, even a short study experience in an English speaking country may influence student writing styles.

5. Discussion

In this section, potential corelations between the outcomes of the different tools of SFL will be made and compared to results of previous studies. First, the results of thematic progression analysis will be discussed, and potential corelations will be made between the use of progression types and patterns by L1 versus L2 writers, especially, with regards to the role of experience in scientific writing. Then, correlations will be drawn between the choice of keywords for thematic participants and their influence over/by thematic progression patterns. The discussion will continue on to relate the length of theme and rheme to patterns of participant keywords. This will be followed by nominalization frequency and how it may relate to not just lexical density but also the type of utterance, i.e. topic introduction, hypotheis building, as well as a potential relation to progression types. The survey results will then be discussed to include potential extraneous factors that may have influenced some of the results. As will be discussed in the rest of this section, all the above tools of SFL identified differences in stylistic writing features in the texts of L1 and EFL students on the one hand, and all student and Ideal texts on the other.

5.1 Experience-gained use of thematic progression types and patterns

In the previous chapter, the proportion of use of each thematic progression type, and their pattern-sequence of use, was reported upon. This revealed differences in the preference/tendency of each group for the different progression types as the abstracts progressed from introduction to conclusion. The preferential use of linear progression over constant progression by L1 scientists is consistent with previous studies investigating medical research texts (e.g. Gao, 2012). Gao (2012) showed that L1 writer use of linear progression was higher than that of Chinese (L2) writers. Whether this was due to gain of experience and/or writer L1 background (Ebrahimi & Ebrahimi, 2012; Rorvik, 2012; Lu, 2013) was not

clear, as the study did not include less experienced L1 and L2 writers (e.g. students) of the same background for comparison. It was also not clear if the texts of the Chinese writers were edited, as they were already published. In the current study, the original texts of both L1 and EFL student groups had a lower rate of usage of linear progression, and a higher rate of constant progression. The latter, was also identified by Belmonte and Mccabe-Hidalgo (1998) as a common problem in student texts. Unlike previous studies in other fields, in this study, it seems that preferential use of linear progression in scientific writing, at least to some extent, is an experience-gained writing characteristic. In addition, as the texts had minimal editing, the more pronounced difference in usage of constant versus linear progression may reflect the original writing tendency of the students. Future studies with larger sample sizes would be helpful in clarifying the role of experience versus L1 background.

Another main difference in progression patterns was that of the use of split themes and rhemes (Danes, 1974; McCabe, 2004). As mentioned earlier, splitting a rheme or theme allows the division of one main idea to a number of subordinate ideas in the themes of the following clause complexes. This characteristic of writing is considered to be of a higher order and potentially difficult for less experienced writers, especially students (e.g. Rustipa, 2010). The ideal texts in this study used splitting more often than the the students. So, they were either making frequent use of linear/split linear theme progression, or if they used constant progression a few times in a row, then they used it to split one theme into multiple subordinate ones. Consistent with previous studies in other fields (e.g. Rustipa, 2010), none of the L1 students used this pattern. Although one EFL student used split theme to split an idea into two, whether this was intentional is not clear, especially since the survey results revealed that this student wrote the abstract in Japanese first followed by translation into English. Therefore, if confirmed, the results suggest that this pattern of progression may also have some dependence on experience.

5.2 Progression of the length of theme and rheme through the abstracts and their relation with participant keyword use

An interesting finding of this study was that there was a difference in the use of thematic participant keywords (i.e. the main focus/participants of each clause) and that this correlated with the length of theme and rheme (i.e. the number of words in each) as the text progressed. So far, studies have not investigated theme and rheme length and its potential correlation with the use of participant keywords. While in the ideal texts there was a gradual change in the theme length from low to high and vice versa for the rheme (as in Figure 4.3, page 25), that of the students changed randomly as the texts progressed from introduction to conclusion. A more careful look at the composition of theme and rheme in each group revealed that the difference was related to writer expression of message through participant keywords. In general, this was corelated with the type of keywords used for the thematic participant in each theme. While the ideal texts used short general keywords in the beginning and ended with more collective keywords (putting more than one participant in the same word group) describing a set of experiments at the end, the students tended to end the text with summative keywords (e.g. "this study", "these findings", etc.). For example, Ideal3 used a nominalized collective keyowrd to conclude the abstract, i.e. "These learning-related changes in spine density could reflect a direct mechanism of encoding…". As this author investigated the role of learning on changes in spine density, mentioned in all preceding clauses, this is a collective keyword that combines them together. On the other hand, L1-2 used, "These results clearly show that neuronal morphology is changed….", a summative keyword that summarizes the whole study. This difference suggests that the Ideal texts use collective/longer keywords for clarity, to effectively get their final message across, as collective keywords mention the previous keywords clearly in one word group while summative keywords are general and whole study related.

With regards to experience versus L1 influence, it seems that both experience and L1 background may play a role in keyword choice and the length of rheme and theme, as L1 students (the same L1 background) also differ from accomlished scientists. In addition, it is important to note that the randomness of the theme and rheme length was much more pronounced in the EFL texts. This was consistent with their much higher use of wordgroup adjuncts to modify participant kewords (compared to rather shorter subordinate/specific keywords instead in the ideal samples), overuse of conjunctions and pronoun textual participants, as well as circumstantial adjunct keywords. The misuse of keywords also lead to ruptures - where the theme did not connect to a preceding theme or rheme; this occurred rarely in the L1 student and scientist texts. Consistent with this, these features have been identified to be overused by EFL writers in other fields (Hu, 2008; Qian, Andrés Ramírez, & Harman, 2007; Schleppegrell, 2004). While it is difficult to make a clear conclusion with the small sample size in this study, it would be interesting to see if the corelation is a general feature of student and L1 scientific texts in future studies.

5.3 The use of processes of transitivity

Studies so far have shown that correct use of the six processes of transitivity is important for text clarity and logical flow, as it helps the writer to correctly express the purpose of an utterance (Halliday & Matthiessen, 2004, p.214). Similar to other studies (e.g. Zheng, Yang & Ge, 2014), five processes types were used in the texts, i.e. all with the exception of the behavioral process. While the processes used in order from highest to lowest, were: relational, material, mental, verbal and existential in the ideal texts; Zhang et al. (2014) saw a higher proportion of material to relational process types with all sections of articles combined and vice versa in the results and discussion sections - although, the differences were low (i.e. relational to material was only 2 to 4% higher). Here, only the abstract sections

were investigated, a section not investigated by their study. That could explain the much more pronounced difference (i.e. 41% higher use of relational to material process). In addition, they investigated published medical articles, so, it is not clear if they selected for English background authors. This could also explain the lower differences in the relational and material processes, as non-English background writers may use more of the material process (e.g. Moore, 2007) and the difference between the relational and material processes could be lower if they are included, due to their high usage of the material process.

An alternative and interesting explanation, which may also be likely, is that the much higher level of the relational process in the current study is a stylistic feature specific to abstracts; where all data from an article is introduced, summarised, discussed and concluded. Perhaps, the use of the relational process, i.e. for putting different participants in the one clause complex, can help with use of less words. This may have evolved to help with writing in detail, yet concisely, within a short number of words; similar to the suggested role for nominalization (Golden, 1993; Halliday & Martin, 1993; Trimble, 1985). For example, Ideal3 text summarizes experimental outcomes related to two different settings in one clause complex using the relational process, i.e. "This spine loss did not occur in homecage mice or in mice exposed to the training context alone". Here, the author has put two participants - "homecage mice"/control group, and "mice exposed to the training context"/experimental group, into the same context, therefore, using the same clause complex to descibe the two.

As mentioned earlier, the different process types are used for different purposes. For example, the material process is used for simple description/narration of messages (MacDonald, 1994; Coffin, 1997). Comparing all groups, consistently, all ideal and L1 student samples used the relational process very frequently, as they were usually describing one participant in relation to another. On the other hand, similar to a study by Moore (2007), the use of the material processes was much higher and the relational process much lower in

the EFL student samples compared to that of the L1 student and scientist samples. On the basis of the small corpora in the current study, it seems that writing analytically comparing different participants was challenging for the EFL students, who tended to write more simple, descriptive sentences.

5.4 Nominalization frequency and its corelation with lexical density and expression of messages in the texts

Nominalization, in simple terms, the use of nominal groups, makes texts more lexically rich and concise (Halliday & Martin, 1993). As scientific writing has evolved under strong word limit pressure, and has become very formal/abstract, nominalization has become an important aspect, yet, it makes texts difficult to understand, especially for readers of EFL backgrounds (Gao, 2012; Golden, 1993; Halliday & Martin, 1993). So far, the frequency of nominalization has not been related to types of utterances in a text. One can assume that the level of self expression of authors and their direct/indrectness - influenced strongly by nominalization in addition to other factors, may change from clause to clause in a text. It would therefore be interesting to see if this is the case in scientific writing and which clause complex/utterance types would use more or less nominalization frequency rates.

In this study, utterance types were categorized according to their message, i.e. research topic introduction (the first clause complex), hypothesis building/stating the aim (usually the second to fourth clause complexes), experiment concluding (towards the end of the abstract), and finally, summative conclusion of all results and significance (at the end of the abstract). When the rate of nominalization was checked against the different utterance types, it was evident that in all samples, even in the EFL student texts, there was a much lower tendency to use nominalization in two utterance types, the hypothesis building/stating aims, and the experimental conclusion utterances. In some cases, the frequency of nominalization in these

utterance types was zero or close to zero. Therefore, the findings of the current study support the idea that despite the evolution of scientific writing to use the highest frequency of nominalization among all disciplines (Gao, 2012; Golden, 1993; Halliday & Martin, 1993), its use in core aims and concluding clauses of scientific abstracts has remained low. It would be interesting to see if this is due to writers' aim to remain direct and make their intentions clear by emphasising it more in direct verbs.

Finally, previous research shows that nominalization frequency and lexical density are strongly correlated (e.g. Gao, 2012). The samples in this study showed an intruiging pattern. Nominalization frequency was the highest in the ideal texts, followed by L1 student texts and then the EFL student texts. However, unexpectedly, lexical density was higher in the L1 student group despite their lower nominalization rates than that of the L1 scientist abstracts. Non-lexical/lexical item ratio, calculated by dividing the number of non-lexical items by that of the lexical items, was much lower in the L1 student samples. Checking the type of participant keywords used in the clauses with very low lexical density, it was apparent that keyword choice played a major role in increasing lexical density, despite the lower nominalization frequency in the L1 student samples. On the other hand, the L1 scientist participant keywords included more non-lexical items such as adjuncts and conjunctions. These results suggest that a higher lexical density can sometimes be a diagnostic tool for lexis type used within a text. That is, a high lexical density may not always be a favorable stylistic feature in scientific writing. Rather, there may be an optimal level of lexical density, as judged by that of the ideal texts. Therefore, a relatively high lexical density can suggest that the author may be using less non-lexical words that would otherwise modify the participants and add to their proper description within a word group, e.g. by using conjunctions and adjuncts.

5.5 Using an online survey to understand potential extraneous factors of influence

When investigating stylistic features of texts from writers of different backgrounds, it is important to include potential extraneous factors of influence (Coffin, Donohue, 2012). In this study, a survey was carried out to understand writer level of experience, and educational and linguistic backgrounds. The experience background of one writer in particular was of note, EFL3, who spent time in an L1 country (US), and had written a masters thesis in English prior to the writing of the abstract. Interestingly, EFL3 text was much more nominalized than the other two EFL student texts, and close but not equal to the L1 student samples. This result suggests that prior experience in scientific writing and studying at an L1 country (in this case, three months) may have a positive influence on nominalization frequency. It would be interesting to clarify to what extent prior writing experience in EFL setting or studying English in an L1 country may influence the rate of nominalization. This result also suggests that experience may have a stronger role in nominalization, at least in the Japanese EFL context than in L1.

6. Summary and conclusions

This study compared Australian (L1) and Japanese (EFL) graduate student texts to published abstracts of English L1 scientists (i.e. 'ideal'/'they have met journal writing standards') using various SFL tools. It provides evidence that all the SFL analytical tools revealed interesting and useful differences between the L1 and EFL texts on the one hand and between the student and ideal texts on the other. While the Ideal asbtracts showed preference for linear thematic progression and the frequent use of of split themes and rhemes, these were rare in the L1 student and EFL student samples, as instead, they tended to overuse constant progression, a previously identified common problem in student writing (Belmonte & Mccabe-Hidalgo, 1998). These differences seemed to be at least partially due to experience in writing, as both student groups showed a similar tendency. The students also tended to use generic keywords with occasional introduction of a new specific or generic participant resulting in ruptures, jumps and/or weak linkage of ideas compared to the use of subordinate and/or specific keyword transitions in the ideal samples. Sudent use of keywords also accounted, to some extent, for their corelation of theme and rheme length, especially, that of EFL student texts, where instead of a low to high transition of length of theme from beginning to end of the abstract, their theme and rheme lengths were random throughout the text.

Analysis of processes of transitivity used for utterance of messages revealed a lower use of relational and higher use of material processes in student samples. This suggests that the students had not yet made the transition from narration to correlation aspects of research writing, a more difficult concept, requiring more experience (Coffin, 1997; Halliday, 1994, p.19; MacDonald, 1994). Classification of utterance messages into introducing, hypothesis building/aims, and experimental outcomes/conclusions, also revealed a strong corelation between nominalization frequency and the type of message a clause complex/utterance

carries. The strong correlation bewteen low nominalization with hypothesis building and experiment concluding clause complexes indicates that these clause types have been resistant to evolution of high frequency nominalization in scientific writing, probably to keep them clear, easy to understand and direct. In addition, nominalization frequency did not correlate well with lexical density in the L1 student samples. This was caused, at least in part, by the use of keywords that did not include non-lexical items in the student texts.

The use of SFL tools in discourse analysis arguably has some limitations. One limitation could be in the type of analysis carried out; that is, it may be considered purely textual with insufficient inclusion of extranous factors that could also influence writing (as discussed by Coffin and Donohue, 2012). To address this limitation, at least in part, a survey was carried out to understand linguistic, educational, L1 exposure and experience related implications in the L1 and EFL student contexts. Future studies using follow up and more detailed surveys, or interviews could further address this limitation. The other limitation of the current study is that the sample size is small, making the application of these findings to the general population difficult. For these results to be representative, they would need to be repeated in future studies with larger samples sizes. However, some of the findings are intriguing and suggest new concepts, such as the role of experience versus L1 background, keyword use and thematic progression patterns, keyword use and lexical density, and correlations between nominalization and clause message types. These findings, especially in scientific writing, may prove useful for pedagogical purposes if confirmed in future studies. In addition, unlike any previous study investigating scientific writing in published abstracts of accomplished L1 and L2 scientists, the work here compared original non-edited drafts of two different groups with the published drafts of a third group, rather than a binary study. This allowed the study of original student writing, with minimal editing, a deviation from previous studies, which gives it more relevance to student teaching and learning.

References:

Afful, J. B. (2007). Academic literacy and communicative skills in the Ghanaian university: A proposal. *Nebula, 4*(3), 141-159.

Amano, T., González-Varo, J.P., Sutherland, W.J. (2016). Languages are still a major barrier to global science. *PLoS Biology*, *14*(12), e2000933. https://doi.org/10.1371/journal.pbio.2000933

Ammon, U. (2001). *The dominance of English as a language of science: Effects on other languages and language communities* (Vol. 84). Berlin: Mouton de Gruyter.

Ammon, U., & McConnell, G. (2002). *English as an academic language in Europe: a survey of its use in teaching.* Frankfurt: Peter Lang.

Arunsirot, S. (2013). An analysis of textual metafunction in Thai EFL students writing. *Novitas-ROYAL*, *7*(2),160-174.

Banks, D. (2005). On the historical origins of nominalized process in scientific texts. *English for Specific Purposes*, *24*(3), 347-357.

Belcher, D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing, 16*(1), 1-22.

Belcher, D. (2009). *English for specific purposes in theory and practice: What ESP is and can be: An introduction.* Michigan: Ann Arbor.

Belmonte, I.A., & McCabe- Hidalgo, A. (1998). Theme-Rheme patterns in L2 writing. *Didactica*, *10*, 13- 31.

Benfield, J., & Howard, K. (2000). The language of science. *European Journal of Cardio Thoracic Surgery, 18*, 642-648.

Canagarajah, A. S. (2002). *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press.

Clarke, J. (2015). The value of understanding students' prior writing experience in teaching undergraduate science writing. *Critical Studies in Teaching and Learning*, *3*(1), 21-43

Coffin, C., & Donohue, J. (2012). Academic Literacies and systemic functional linguistics: How do they relate?. *Journal of English for Academic Purposes*, *11*, 64–75. 10.1016/j.jeap.2011.11.004.

Carter-Thomas, S. (2016). English as a Scientific and Research Language. *Journal of English as a Lingua Franca*, *5*(2), 377-383.

Coffin, C. (1997). 'Constructing and giving value to the past: An investigation into secondary school history', in F. Christie and J.R. Martin (eds) *Genre and Institutions: Social processes in the workplace and school* (pp. 196–230). Cassell: London.

Coffin, C., & P. Donohue, J. (2012). Academic Literacies and systemic functional linguistics: How do they relate?. *Journal of English for Academic Purposes. 11*, 64–75. 10.1016/j.jeap.2011.11.004.

Crystal, D. (2007). *English as a Global Language.* Cambridge: Cambridge University.

Crystal, D. (2012). A Global Language. In C. david, & S. a. Swann (Ed.), *English in the world: history, diversity, change.* Abingdon: Routledge.

Cummings, M. (2003). The role of theme and rheme in contrasting methods of organization of texts. In C. Butler (Ed.), *Dynamics of language use: Functional and contrastive perspectives* (pp. 129-154). Amsterdam/Philadelphia: John Benjamins.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In: F. Daneš. (ed.), *Papers on functional sentence perspective* (pp. 106–128). Praha: Academia.

Di Bitetti, M.S., Ferreras J.A. (2016). Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. *Ambio, 46*(1), 121-7. http://doi.org/cdz8

Ebrahimi, S. F., & Ebrahimi, S. J. (2012). Markedness in writing: A case of EFL students. *Theory and Practice in Language Studies*, *2*(4), 773-777.

ElMalik, A.T., & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes*, *7*(2), 87-96.

Firth, A. (1996). 'The discursive accomplishment of normality. On "lingua franca" English and conversation analysis'. *Journal of Pragmatics, 26*, 237–59.

Flowerdew, J. (2001). Attitudes of journal editors to non-native speaker contributions. *TESOL Quarterly,* 35, 121- 150.

Flowerdew, J. (2006). Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics, 11*(3), 345-362.

Flowerdew, J. (2007). The non-anglophone scholar on the periphery of scholarly publication. In A. Carli and U. Ammon (eds.), *Linguistic inequality in scientific communication today*. *AILA Review* 20 (pp. 14–27). Amsterdam: John Benjamins.

Flowerdew, J. (2013). *The Handbook of English for Specific Purposes.* In E.b. Starfield (Ed.), Boston: Wiley-Blackwell.

Flowerdew, J., & Wang, Ho. (2015). Identity in academic discourse. *Annual Review of Applied Linguistics*, *35*, 81-99. 10.1017/S026719051400021X.

Fontaine, L., & Kodratoff, Y. (2003). The role of thematic and concept texture in scientific text: comparing native and non-native writers of English. The English version of "Comparison du rôle de la progression thématique et de la texture conceptuelle chez des scientifiques anglophones et francophones s'exprimant en Anglais. Special Issue Asp, GERAS". Retrieved Jun 6, 2018 from: http://www.cardiff.ac.uk/encap/fontaine/thematicconcept.pdf

Fontaine, L. (2013). *Analysing English grammar: A systemic functional introduction*. Cambridge, UK: Cambridge University Press.

Forey, G. (2002) Aspects of theme and their role in workplace texts. PhD thesis, University of Glasgow. Retrieved, June 1, 2018 from: http://www.isfla.org/Systemics/Print/Theses/ForeyThesis/

Gao, W. (2012). Nominalization in medical papers: A comparative study. *Studies in Literature and Language*, *4*(1), 86-93.

Gosden, H. (2003). 'Why not give us the full story?': functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes, 2*(2), 87-101.

Graddol, D. (1997). *The future of English?: A guide to forecasting the popularity of the English language in the 21st century*. London: British Council.

Green, C.F., Christopher, E.R., & Lam, K.M. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes, 19*, 99-113.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Arnold.

Halliday, M. A. K. (1993). *Writing Science: Literacy and Discursive Power*. Pittsburgh: University of Pittsburgh Press.

Halliday, M.A.K. (1994). *An Introduction to Functional Grammar* (2nd edition). London: Edward Arnold.

Halliday, M.A.K. (2014). *Halliday's introduction to functional grammar* (4th edition). Revised by C.M.I.M. Matthiessen. New York: Routledge.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M.A.K., & Martin, J.R. (1993). *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Halliday, M.A.K., &. Matthiessen, C.M.I.M. (2004). *An Introduction to Functional Grammar* 3rd Edition. London: Arnold.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2014). *Halliday's Introduction to Functional Grammar* (4th Edition). Abingdon: Routledge.

Hawes, T. (2015). Thematic progression in the writing of students and professionals, *Ampersand*, *2*, 93–100.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor: University of Michigan Press.

Larsen, R.S., Smith, I.T., Miriyala, J., Han, J.E., Corlew R.J., Smith, S.L., Philpot, B.D. (2014). Synapse-specific control of experience-dependent plasticity by presynaptic NMDA receptors. *Neuron, 20*;83(4):879-93. doi: 10.1016/j.neuron.2014.07.039.

Liardet, C.L. (2016). Grammatical metaphor: Distinguishing success. *Journal of English for Academic Purposes*, *22,*109e118.

Lillis, T., & Curry, M.J. (2015). The politics of English, language and uptake: the case of international academic journal article reviews. *AILA Review*, *28*, 127–150.

Lu, A. (2013). A functional grammar approach to analysing Asian students' writing. *American Journal of Educational Research*, *1*(2), 49-57.

MacDonald, S.P. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale: Southern Illinois University Press.

MacDonald, M.L., Alhassan, J., Newman, J.T., Richard, M., Gu, H., Kelly, R.M., Sampson, A.R., Fish, K.N., Penzes, P., Wills, Z.P., Lewis, D.A., Sweet, R.A. (2014). Selective Loss of Smaller Spines in Schizophrenia. *Am. J. Psychiatry*, *174*, 586–594.

Mahbudi, A., Mahbudi, L., & Amalsaleh, E. (2014). A Comparison between the Use of nominalization in medical papers by English and Iranian writers. *International Journal of Applied Linguistics and English Literature*, *3*(6), 1-6.

Martin, J. R. (2008). Incongruent and proud: de-vilifying 'nominalization'. *Discourse and Society*, *19*(6), 801e810.

Mauranen, A. (1996). "*Discourse competence: Evidence from thematic development in native and non-native texts," in Academic writing: Intercultural and textual issues*, In E. Ventola & A. Mauranen (Eds.). Amsterdam/Philadelphia: John Benjamins.

McCabe, A. (2004). Thematic progression patterns and texts types in history textbooks. In D. Banks (ed.), *Text and Texture: Systemic Functional viewpoints on the nature and structure of text* (pp. 215-237). Paris: L'Harmattan.

Moore, T. (2007). The 'Processes' of Learning: On the use of Halliday's transitivity in academic skills advising. *Arts & Humanities in Higher Education*, *6*(1), 50–73.

Murugesan, R. (2014). Publishing a journal in English: tips for journal editors who are non-native English speakers. *Science Editing*, *1*(1): 46-48. https://doi.org/10.6087/kcse.2014.1.46

Nature Neurosciecne (2018). Nature Neuroscience instruction for authors. Retrieved June 6, 2018 from: http://mts-nn.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions

Nature Reviews Neurosciecne (2018). Nature Reviews Neuroscience instruction for authors. Retrieved from: https://media.nature.com/full/nature-assets/reviews/pdf/naturereviews.pdf

NBCI, (2018). US National Library of Medicine. Retrieved June 6, 2018 from https://www.ncbi.nlm.nih.gov/pubmed/

Ortoney, A. (1993). *Metaphor and thought*. Cambridge. Cambridge University Press.

Paltridge, B. (2013) "Learning to review submissions to peer reviewed journals: how do they do it?". *International Journal for Researcher Development*, *4*(1), 6-18. https://doi.org/10.1108/IJRD-07-2013-0011

Panko, B. (2017). English is the language of science. That isn't always a good thing: How a bias toward English-language science can result in preventable crises, duplicated efforts and lost knowledge. *Smithsonian magazine*. Retrieved on June 6, 2018 from: https://www.smithsonianmag.com/science-nature/english-language-science-can-cause-problems-180961623/#mTCTmqVq6cQyFstG.99

Propenko, M. (2018). Native speaker check. *Essays on English in Japan*. Retrieved from; http://www.essayengjp.com/en/essays/93-native-speaker-check.html

Qian, Y., Andrés Ramírez, J., & Harman, R. (2007). EFL Chinese students and high stakes expository writing: A theme analysis. *Colombian Applied Linguistics Journal*, *9*, 99-125.

Ravelli, L. (1988). Grammatical metaphor: An initial analysis. In E.H. Steiner & R. Veltman (Ed.), *Pragmatics, discourse and text: Some systemically-inspired approaches*. London: Pinter.

Rørvik, S. (2012). Thematic progression in learner language. In S. Hoffmann, P. Rayson, & G. Leech (Eds.). *English corpus linguistics: Looking back, moving forward* (pp.165-177). Amsterdam: Rodopi.

Salager-Meyer, F. (2008). Scientific publishing in developing countries: challenges for the future. *Journal of English for Academic Purposes*, *7*, 121–32.

Sanders, J., Cowansage, K., Baumgaˆrtel, K., & Mayford, M. (2012). Elimination of dendritic spines with long-term memory is specific to active circuits. *The Journal of Neuroscience*, *32*(36),12570 –12578

Schleppegrell, M. J. (2009). Grammar for generation 1.5: A focus on meaning. In M. Roberage, M. Siegal & L. Harklau (Eds.), *Generation 1.5 in college composition: Teaching academic writing to U.S.-educated learners of ESL* (pp. 221-234). New York: Routledge.

Sanders, J., Cowansage, K., Baumgartel, K., & Mayford, M. (2012). Elimination of dendritic spines with long-term memory is specific to active circuits. *The Journal of Neuroscience*, *32*(36),12570 –12578.

Seaton, I. (1997). Linguistic Non-Imperialism. *English Language Teachers Journal , 51* (4), 81-82 .

Seidlhofer, B. (2005). English as a lingua franca, *ELT Journal*, *59*(4), 339–341. https://doi.org/10.1093/elt/cci064

Smokotin, V., & Alekseyenko, A. &. (2014). The Phenomenon of linguistic globalization: English as the Global Lingua Franca (EGLF). *Procedia-Social Social and Behavioral Sciences, 154*, 509-513.

Springer Publisher (2018). Manuscript guidelines journals: instruction for authors. Retrieved June 6, 2018 from: https://www.springer.com/authors/manuscript+guidelines?SGWID =0-40162-6-795142-0

Stoppel, L.J., Benjamin D.A, R.K., Senter, R.K, Preza, A.R., Lefkowitz, R.J., & Bear, M.F. (2017). β-arrestin2 couples metabotropic glutamate receptor 5 to neuronal protein synthesis and is a potential target to treat fragile X, *Cell Reports*, *18*(12), 2807–2814. doi:10.1016/j.celrep.2017.02.075.

Swales, J. M. (2004). *Research genres: Exploration and applications*. New York, NY: Cambridge University Press.

Tardy, C. (2004). The role of English in scientific communication: lingua franca or Tyrannosaurus Rex? *Journal of English for Academic Purposes*, *3*,247–69.

Trimble, L. (1985). *English for science and technology: A discourse approach*. Cambridge: Cambridge University Press.

Ventola, E., and Mauranen, A. (1991). *Non-native writing and native revising of scientific articles, in Functional and systemic linguistics: approaches and uses*, In E. Ventola (Ed.). Berlin: Mouton de Gruyter.

Rustipa, K. (2010). Theme-rheme organization of learners' texts. *Dinamika Bahasa dan Ilmu Budaya, 4*(2), 1-17.

Zheng, S., Yang, A., and Ge, G. (2014). Functional stylistic analysis: Transitivity in English-medium medical research articles. *International Journal of English Linguistics*, I(2): 12-25.

Survey results for L1 and EFL students

	Abstract Code	Year of graduate study	The language used for the first draft	Level of difficulty to write in English scaled 1-10	Time requirement compared if writing was done in students' L1 scaled 1-10	Level of help/advice recieved for the writing scaled 1-10	Prior expereice in writing in English, including 1) research proposals, 2) Review articles, 3) Thesis, 4) Undergraduate essays, and 5) Other and how often (scaled 1-10)	Prior English writing study/training	Prior residence in an English L1 country	Prior work with English L1 colleeagues	Have you ever heard of Systemic funcitonal linguistics, e.g. thematized participants and/or nominalization
L1 student group	L1LA-204	PhD-final year	English	1	N/A	1	Yes, 1 to 5	Yes, throughout schooling	Whole life	N/A	No
	L1SH-190	Honours	English	1	N/A	1	Yes, 1, 2 and 4	Yes, throughout schooling	Whole life	N/A	No
	L1KL-198	PhD-final year	English	3	N/A	1	Yes, 1 to 5	Yes, throughout schooling	Whole life	N/A	No
EFL student group	EFLHO-249	PhD-2nd year	English	8	8	1	Yes, 1 Scale: 1	Yes, only through Japanese education system	Short oversease study trip	No	No
	EFLIS-256	PhD-2nd year	Japanese	7	7	5	Yes, 5-poster for presentation Scale: 1	Yes, only through Japanese education system	No	Yes-but we talk using a mix of Japanese and English	No
	EFLMA-224	PhD-final year	English	6	3	5	Yes, 3-masters thesis	Yes, only through Japanese education system	Yes, short trip-3 months	No	No

Appendix 2 Figure 4.1A: Diagrammatic representation of thematic progression types in the 'Ideal' texts
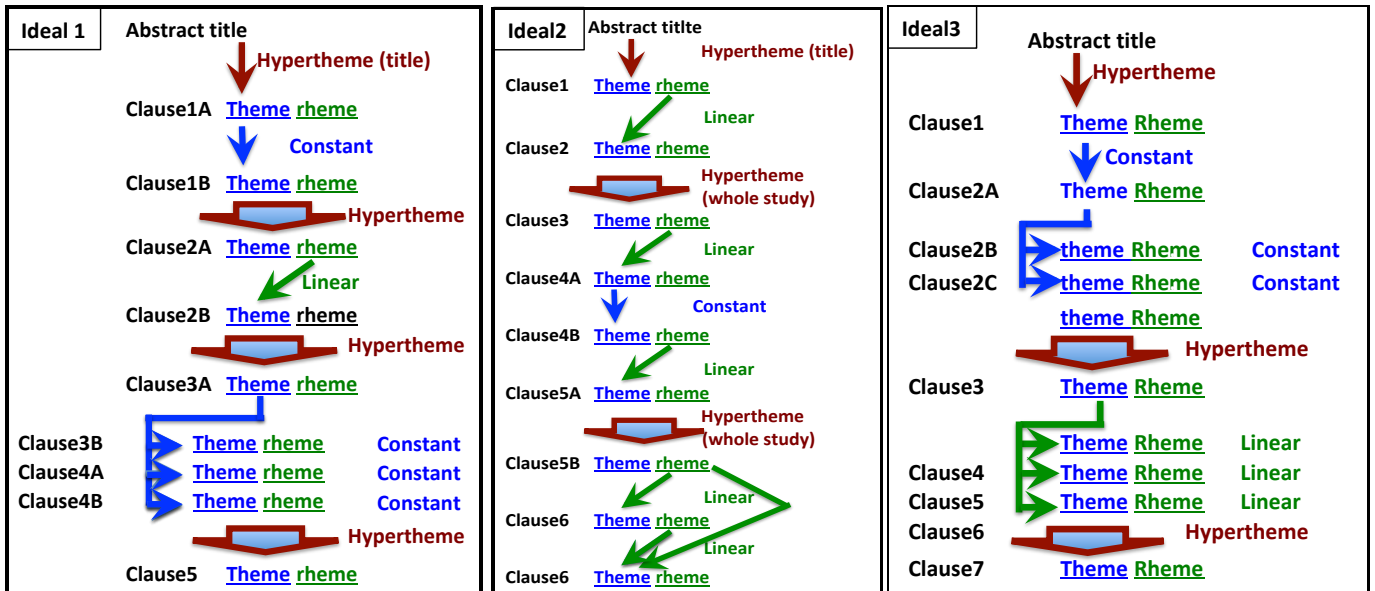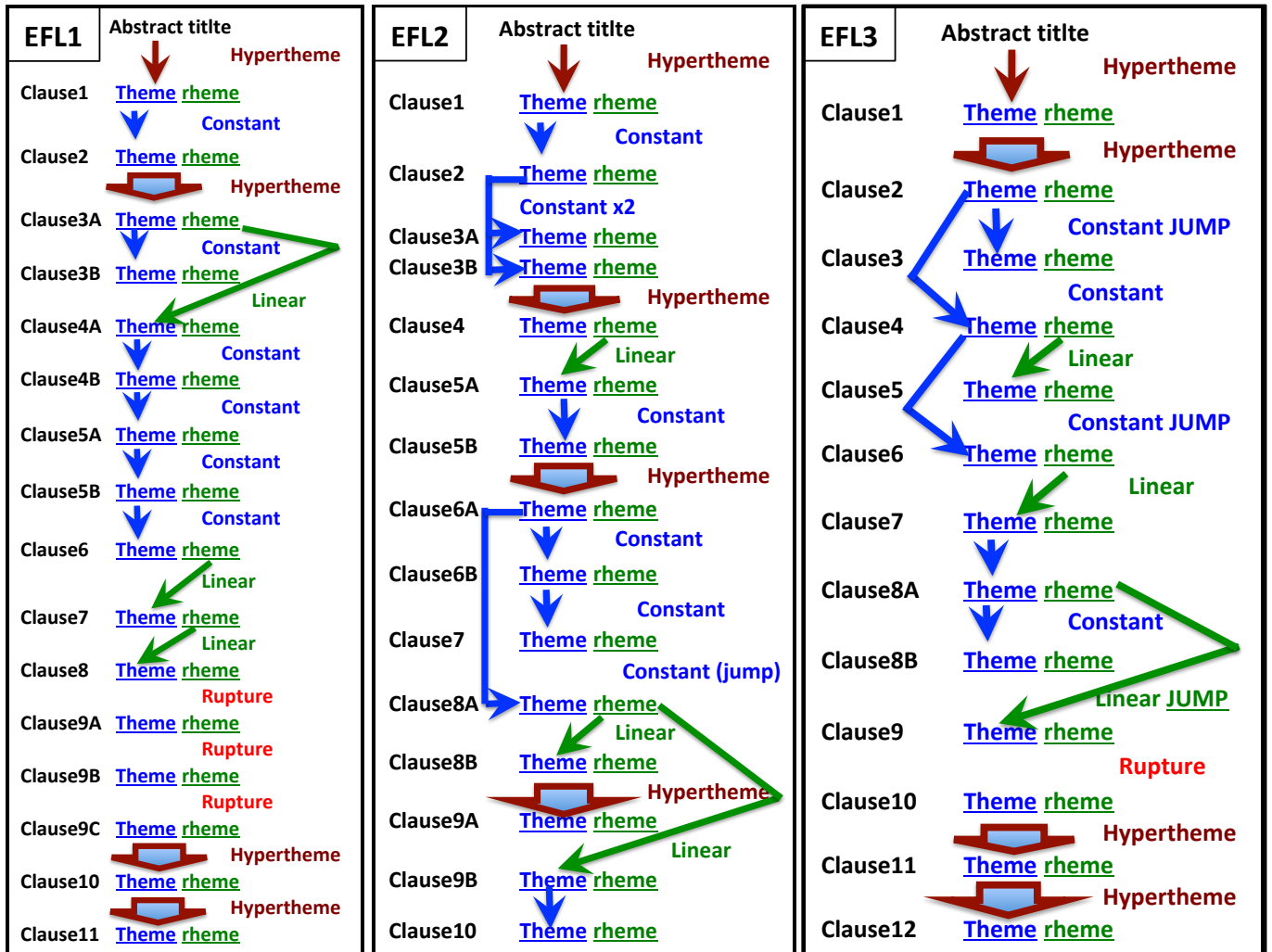
Appendix 3 Figure 4.1B: Diagrammatic representation of thematic progression in the

L1 student texts
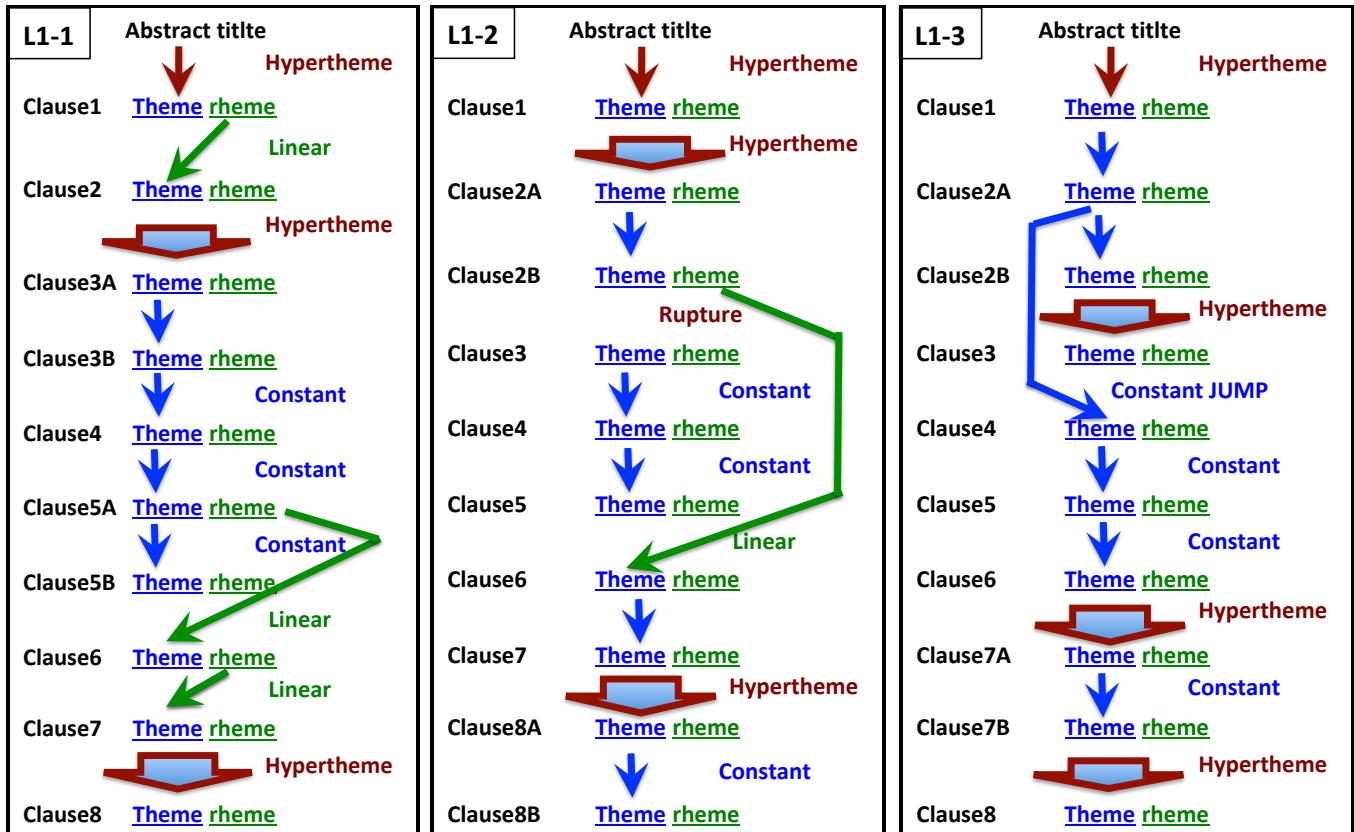
Appendix4 Figure 4.1C: Diagrammatic representation of thematic progression in the EFL student texts

Appendix 5 Figure 4.3A: Diagrammatic representation and tabulated details of the thematic participant keyword progression for Ideal1

	Themes	Relevant rhemes	Type of keywords/participant	Purpose
Title	Experience-Dependent plasticity		Very general	To introduce the topic in the abstract
CLause1	Sensory experience	modifying neuro-transmission	General	To introduce the intended research area
CLause2	Experience-dependent modifications		General modified	To add sub-area of research (i.e. not just the study of sense (how we sense) but also the study of how that may relate to changes (modifications) in the brain
CLause3	Visual experience	Whole study / timing-dependent long-term depression (tLTD)	Subordinate	To narrow down the topical theme to one that they want to focus on
CLause4A	This LTD	maintained	Specific abbreviated keyw	Change from idea building to experimental outcomes, i.e. to do LTD experiments and investigate its effects on visual experience
CLause4B	or	Reinstated	Textual	To avoid repeating "LTD" again, i.e. this LTD ....or this LTD.....
CLause5A	The changes in LTD		Modified participant	To describe LTD-related experimental outcomes
CLause5B	Visual deprivation	enhances both tLTD and glutamate release.	Collective	To summarize related outcomes
CLause6	These effects	NMDA	Summative	To describe and conclude outcomes from multiple experiments
CLause7	induction of tLTD and NMDAR deletion		Collective	

Arrows indicate links in thematic progression   Relevant rhemes are those that are used to make thematic connections (linear)

Appendix 6 Figure 4.3B: Diagrammatic representation and tabulated details of the thematic participant keyword progression for Ideal2

Clauses	Themes	Relevant rhemes	Type of keywords/participant	Purpose
Title	Synaptic protein synthesis.... Hypertheme (title)		Very General (directly from abstract	To introduce the general topic in the title
Clause1A	Synaptic protein synthesis	rheme 1A	Genera	To introduce the intended research area
Clause1B	and	rheme 1B	Textual	To add further information while mainting focus (i.e. on the intended research area)
Clause2A	Neural activity	(mGlu5)	General-derrived from whole topic	To use whole topic to introduce new participant (i.e. mGlu5)
Clause2B	yet, how mGlu5 couples to	rheme 2B	Specific	The specific focus of area research
Clause3A	Here, we ... that β-arrestin2	rheme 3A	Specific (i.e. the specfici gene β-arrestin2)	The specific experimental participant (gene) of interest. So, they are interested in mGlu5 mechanism and hypothesize a role for β-arrestin2
Clause3B	and genetic reduction of β-arrestin2	rheme	Specific-modified 1	To describe β-arrestin2-related experimental outcomes (1)
Clause4A	Importantly, reducing β-arrestin2	rheme	Specific-modified 2	To describe further β-arrestin2-related experimental outcomes (2)
Clause4B	and	rheme	Textual	To describe further β-arrestin2-related experimental outcomes (3)
Clause5 Thus,.. mGlu5-stimulated protein synthesis, these data .. β-arrestin2		rheme	Textual summarizing + collective summative (thus) + summative +specific	In order: "Thus" refers to all outcomes + two different participants combined + theses data (summative) + specific. Everything combined to conclude on the main experimental participant (β-arrestin2)

Arrows indicate links in thematic progression   Relevant rhemes are those that are used to make thematic connections (linear)

Appendix 7 Figure 4.3C: Diagrammatic representation and tabulated details of the thematic participant keyword progression for Ideal3

Clauses	Themes	Relevant rhemes	Type of keywords/participant	Purpose
	Elimination of Dendritic Spines			
Title			General (area of focus)	To introduce the general are of interest in the title
Clause1	*Structural changes*	rheme 1	Very general	To introduce the topic in the abstract
Clause2A	these changes	rheme 2A	Modified	and hypothesize
Clause2B	as well as,	rheme 2A	Textual	To further introduce main topic and hypothesize
Clause2C	and .	rheme 2C	Textual	To further introduce main topic and hypothesize
Clause3	*Here,*	changes in the	Textual/Summative	introduce aim/experimental outcome summary
Clause4	There were significantly	rheme 4	Textual	To describe outcomes (1)
Clause5	This spine loss	rheme 5	Modified general participant	To describe outcomes (2)
Clause6	Mice exposed to unpaired shocks	rheme 6	Modified new participant	To state the outcome using an experimental (new) participant
Clause7A	These learning-related changes in spine density	rheme 7A	Collective	Puts two different participants together, i.e. "learning-related changes" and "spine density".
Clause7B	or alternately........	rheme 7B	Textual	To add further information

Arrows indicate links in thematic   Relevant rhemes are those that are used to make thematic connections (linear)