An Error Analysis of Japanese scientists' research articles

Leigh McDowell (B.App.Sci, M.Sci.Ed.)

Department of Linguistics

Faculty of Human Sciences

Macquarie University

April, 2016

This thesis is presented as a partial fulfilment of the requirements for the Masters of Research in Applied Linguistics at Macquarie University, Sydney, Australia.

Table of Contents

Abstract vi

Statement of Candidate vii

Acknowledgements viii

Chapter 1: Introduction 1

Chapter 2: Literature Review 4

2.1 English as a Global Language 5

2.1.1 The development of English Language Teaching. 7

2.1.2 English for Research Publication Purposes. 9

2.1.3 Japanese scientists and English for Research Publication Purposes. 10

2.2 Error Analysis 13

2.2.1 The development of Error Analysis. 13

2.2.2 Error Analysis methodology. 14

2.2.3 Limitations of Error Analysis. 17

2.2.4 Error Analysis and English Language Teaching and research. 18

2.3 Systemic Functional Linguistics 20

2.3.1 Rank scale and system networks. 21

2.3.2 Systemic Functional Linguistics and Error Analysis. 25

2.4 Corpus Linguistics 27

2.4.1 Corpus Linguistics and academic discourse. 27

2.4.1 Corpus Linguistics and Error Analysis. 28

2.5 Chapter Summary 29

Chapter 3: Methodology 30

3.1 Integrated Framework of Analysis. 31

3.2 Corpus Design 36

3.3 Elaborated Error Analysis Procedure 39

3.3.1 Recognition and reconstruction of errors. 39

3.3.2 Classification and quantification of errors in nominal groups. 42

3.3.2.1 Classification of errors by location and error type. 42

3.3.2.2 Errors with articles and plural -s. 43

3.3.2.3 Errors with preposition -of. 46

3.4 Chapter Summary 48

Chapter 4: Findings **49**

4.2 Overview of Errors in Nominal Groups **54**

4.2.1 Errors in complex and simple nominal groups 57

4.2.2 Errors in Pre-modifier, Head, and Post-Modifier. 58

4.3 Error Patterns in Nominal Groups **62**

4.3.1 Errors with articles and plural -*s*. 63

4.3.1.1 Singular referent errors with articles and plural -s. 71

4.3.1.2 Plural referent errors with articles and plural -s. 74

4.3.1.3 Mass referent errors with articles and plural -s. 77

4.3.1.4 Summary of errors with articles and plural -s. 79

4.3.2 Errors with preposition -*of* 81

4.3.2.1 Addition errors with preposition -of 86

4.3.2.2 Selection errors with preposition -of 87

4.3.2.3 Omission errors with preposition -of 90

4.3.2.4 Ordering errors with preposition -of 91

4.3.2.5 Summary of errors with -of 93

4.4 Chapter Summary **94**

Chapter 5: Conclusion **96**

References **99**

Appendices **107**

Appendix A: Ethics Approval Letter **107**

Appendix B: Participant Consent Form **109**

Appendix C: Participant Questionnaire and Responses **111**

Appendix D: 1,000-word Excerpt **116**

Appendix E: Nominal Group Denotation and Quantification **118**

Appendix F: Thesis Word Count **120**

Figures and Tables

Figure 2.1 Distribution in 2012 of foreign students in tertiary education by country of destination 8

Figure 2.2 Algorithm for the first stage of error analysis 15

Figure 2.3 Canonical hierarchy of Systemic Functional Linguistics 21

Figure 2.4 System network mapping the English personal pronoun system 24

Figure 3.5 System network for articles and plural -*s*. 35

Table 2.1 Top 20 countries based on papers published in Thomson Reuters-indexed journals from 1999-2009 11

Table 2.2 Error types and linguistic levels 16

Table 2.3 Rank scale at lexicogrammar stratum 22

Table 3.4 Pre-modification pattern of complex nominal groups 33

Table 3.5 Overview of Corpus of Japanese Scientific Discourse (COJSD) 38

Table 3.6 Overview of Nominal Group Subcorpus (NGS) 41

Table 3.7 Errors with articles and plural -*s* 44

Table 3.8 Matrix for analysis of errors with articles and plural -*s* 46

Table 4.9 Overview of error frequencies in the Corpus of Japanese Scientific Discourse (COJSD) 51

Table 4.10 Overview of error frequencies in the Nominal Group Subcorpus (NGS) 55

Table 4.11 Errors in Pre-modifier, Head, and Post-modifier as percentage of nominal group errors 60

Table 4.12 Errors with articles and plural -*s* 65

Table 4.13 Percentage distribution of error types in errors with articles and plural -*s*.... 68

Table 4.14 Relative frequencies of article and plural -*s* errors.. 70

Table 4.15 Error types within article and plural -*s* errors with singular referents............. 72

Table 4.16 Error types within article and plural -*s* errors with plural referents 75

Table 4.17 Error types within article and plural -*s* errors with mass referents 78

Table 4.18 Errors with preposition -*of*.. 82

Table 4.19 Proportion of error types in errors with -*of* constructions.................................... 85

Table 4.20 Selection errors with preposition -*of*... 89

Abstract

For scientists in the twenty-first century, a mastery of written English for Research Publication Purposes (ERPP) is fundamental to professional success. For many Japanese scientists, this is a source of frustration given their use of English as a foreign language is subject to idiosyncrasies, which may be perceived by editors, reviewers, and readers as errors detracting from the impact of their research. As a basis for building data-driven needs-specific pedagogical support, this study investigates the major error patterns in Japanese scientists' written English. Participants in the study are 13 Japanese scientists working in the field of materials science. The primary data are the participants' scientific research article manuscripts (i.e., the research articles before publication). An elaborated corpus-assisted Error Analysis (EA) methodology is employed, investigating error patterns through the lens of Systemic Functional Linguistics (SFL). In short, the EA involves the collection of texts, identification of errors in the texts, and the classification, linguistic description, and quantification of those errors, with these classifications and descriptions elaborated through the SFL theory. For purposes of scope, the investigation focuses on errors in nominal groups—a key feature of scientific English. A total of 654 nominal group errors are identified and analysed according to the elaborated EA methodology. From the analyses, two major error patterns emerge: errors with articles and plural -*s*, and errors with preposition -*of*. Results highlight the difficulties Japanese scientists face with both the pre- and post-modification of complex nominal groups. Particularly, the omission of the indefinite article *a* with singular referents, and the mis-selection of embedded -*of* prepositional phrases in post-modification are revealed as dominant errors in this study. Detailed descriptions of the two major errors patterns and their implications along with recommendations for pedagogical intervention are presented.

Statement of Candidate

I certify that the work in this thesis entitled "An Error Analysis of Japanese scientists' research articles" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis was approved by Macquarie University Ethics Review Committee, reference number: Reference No: 5201500154 on the 18th May, 2015.

Leigh McDowell

Student ID: 43828507

Date: 26th April, 2016

Not everything that can be counted counts.

Not everything that counts can be counted.

Maxim ascribed to William Bruce Cameron, 1963

Acknowledgements

This study is concerned with the counting of errors. It is not to assert that counting errors is all that counts. What really counts is the learning, which cannot be counted. For guiding me so graciously through the learning, I offer my sincere gratitude to my supervisor, Dr Cassi Liardet. For supporting me all the way from Japan to Sydney and back, I thank my family. I would also like to thank Prof. Jim Martin for including me in his special lecture series at the University of Sydney; Lian Qian and the members of the Macquarie University Language Teachers and Learners Group for including me in their research community; and Macquarie University professors and staff Stephen Moore, Annabelle Lukin, Phil Chappell, Robyn Bishop, and Kylie Drake, for their kind assistance. Last but not least, I express gratitude to the Japanese scientists who participated in this study. My hope is that the learning from this thesis may help them along their paths.

Chapter 1: Introduction

This study examines the second language (L2) English errors in the scientific writing of Japanese scientists.[1] Specifically, this thesis presents an elaborated Error Analysis framework for the investigation of L2 errors, with the aim of revealing the most prevalent and frequent error patterns in a corpus of scientific research articles written by Japanese scientists. This introductory chapter outlines the motivations of the study, beginning with an overview of the research context.

English is firmly established in the twenty-first century as a global language and is the lingua franca in many institutional and disciplinary settings (Crystal, 2003; Kirkpatrick, 2007). One discipline where this is readily apparent is the sciences. In Japan, despite being a foreign language, English is the primary professional language for Japanese scientists, who are the second largest producers of scientific publications behind the United States in the recent decade between 1999–2009 (Thomson Reuters, 2009). Japanese scientists have traditionally favoured publishing their work in international English-medium journals due to the higher impact factors relative to Japanese-medium journals (Yamazaki, 1995). However, the subsequent linguistic demands of publishing in a foreign language can be a source of frustration leading many Japanese scientists to seek specialised language support.

The personal motivations for this study stem from my work in English Language Teaching (ELT) in Japan, where for the last six years I have taught English to graduate students in the materials science department of a national university. In addition to teaching, I provide English language support for PhD students and professors in their efforts to publish their scientific research articles in international journals. For many of these early-career scientists and established-career professors alike, the levels of sophistication and accuracy demanded in technical publications present significant challenges. The Japanese scientists I work with are conscious that when writing in English they are prone to using language that may be abstruse or perceived as errors, detracting from the quality of their work and the messages they wish to convey to their peers. This concern over errors prompts Japanese scientists, and often their reviewers

[1] In practice, these Japanese scientists use English as a foreign or additional language (i.e., EAL). *L2*, as it is used in this thesis, encompasses these distinctions.

and editors, to require extensive proofreading of their manuscripts prior to publication (Wiley & Tanimoto, 2012). In my role supporting early- and established-career materials scientists, I have proofread many manuscripts, and over the course of these activities, noticed patterns of errors occurring regularly across all texts. These activities motivate this investigation to firstly, elucidate the most prevalent errors in order to, secondly, develop a needs-specific, focused pedagogy to help Japanese scientists achieve the professional standards demanded in scientific publications.

The theoretical motivations for this study derive from three linguistic approaches: Error Analysis (EA), Systemic Functional Linguistics (SFL), and Corpus Linguistics (CL). The first of these, EA, was developed in the 1960s–70s as a methodology for the investigation of L2 errors, with theoretical roots in early Second Language Acquisition (SLA) research (Corder, 1981, pp. 35-36). EA has met with criticisms regarding its limitations and in recent decades has been mostly abandoned by SLA researchers; however, it has found wider practical applications in ELT as a hands-on and data-driven approach to the investigation of a relevant and everyday professional concern (i.e., errors; James, 1998, p. x). Primary among EA's limitations is its reliance on traditional surface grammar for the linguistic description and understanding of errors (Hamilton, 2015; Schleppegrell, 2002). In order to overcome this limitation, this study integrates a Systemic Functional Linguistics (SFL) approach within the EA framework. SFL offers a comprehensive view of language as it functions in context as a meaning-making resource (Eggins, 2004, p. 3). In addition to enriching traditional grammar with delicate grammatical descriptions, SFL enables analyses along functional lines, and provides unique insight into the system-structure relationships of grammar (Martin, 2013; Matthiessen & Halliday, 2009). By integrating SFL within the EA framework, this study benefits from the decades of theoretical advances made in SFL since EA was developed.

While SFL provides the basis for fine-grained linguistic analyses, one consistent criticism is its limitation to a small number of texts (Butler, 1999). To overcome this limitation, the EA framework in this study is further elaborated by the integration of a Corpus Linguistics (CL) approach. CL enhances linguistic descriptions by providing an empirical basis for analyses through collections of real language texts, or corpora (Meyer, 2002). Corpora are designed to focus on a specific population (e.g., Japanese scientists) or a specific text type or genre (e.g., research articles), and in this way provide representative language samples. Corpus software tools, for example frequency and keyword analyses, can be employed for the efficient analyses of large bodies of texts (Evison, 2010). Additionally, a corpus-assisted approach may employ manual analyses of texts to deduce patterns not readily identified by software tools (Park, 2012). In this study, a specialised corpus is developed from the scientific research articles of Japanese scientists (i.e., the Corpus of Japanese Scientific Discourse), and a corpus-assisted approach is integrated within the EA framework, enabling frequency analysis and the identification of error patterns across multiple texts.

This thesis is organised in five chapters. Following from this introduction, Chapter 2 presents a background to the theoretical and analytical frameworks of the study with a review of the most relevant literature in the areas of EA, SFL, and CL. Chapter 3 describes the research design and methodology including an overview of the integrated framework of analysis and elaborated EA procedure. Chapter 4 presents the findings of the investigation of errors in the scientific writing of Japanese scientists, highlighting the major error patterns to emerge from the data. The results and discussions are combined here to facilitate an integrated reading. Finally, Chapter 5 concludes the investigation with a summary of the major results and contributions, along with the limitations and directions for future investigation.

Chapter 2: Literature Review

This chapter presents a background to the investigation of L2 English errors in the scientific writing of Japanese scientists. The background comprises a comprehensive review of the literature most relevant to the investigation and the knowledge integral in understanding the theoretical and analytical frameworks. The chapter is presented in four main sections. **Section 2.1** illustrates the predominance of English as a global language, and the subsequent development of the field of English Language Teaching (ELT), particularly the emergence of English for Research Publication Purposes (ERPP). **Section 2.2** traces the development of Error Analysis (EA), a methodology employed in the investigation of L2 errors, from its theoretical beginnings to its practical application. **Section 2.3** presents an overview of the Systemic Functional Linguistics (SFL) approach to language analyses, and the resources that the approach offers in the investigation of L2 errors. Finally, **Section 2.4** reviews the complimentary approach of Corpus Linguistics (CL), identifying its contribution to understanding academic discourse and investigating errors.

2.1 English as a Global Language

The dominance of English as a global language has been well documented. As early as 1997, English has been distinctly ranked first among the world's ten most influential languages (Weber, 1997).[2] In 2003, Crystal (2003) estimated the number of English speakers throughout the world to be 1.5 billion,[3] which given the world's population had reached six billion, indicated that one in four people in the world were able to use English "to a useful level" (Crystal, 2003, pp. 68-69). As of 2016, an online daily report shows English currently accounting for 53.6% of webpages on the Internet, with the next nearest languages, Russian and German, at 6.4% and 5.7%, respectively (W3Techs, 2016). Furthermore, in a study of three massive global language networks, Ronen, Gonçalves, Hu, Vespignani, Pinker and Hidalgo (2014) identify the centrality of English as a global language hub.[4] This analysis mapped millions of online and printed instances of language, revealing that, "the world's languages exhibit a hierarchical structure dominated by a central hub, English" (p. E5622).

The origins of the dominance of English can be traced back to British colonialism and more recently, its expansion into new territories, especially by the United States (Phillipson, 1992). The subsequent rise and successes of American capitalism, political and military power, along with the proliferation of its media, sports, and technology has furthered the spread of English to all corners of the globe (Blake, Lass & Romaine, 1992, p. 6; Pennycook, 2014). The spread of English throughout the world has led to its use as the lingua franca of many professions. For example in diplomacy, the League of Nations was one of the first major international organisations to adopt English to conduct its proceedings (Crystal, 2003, p. 86). English has been adopted as the de facto lingua

---

[2] Languages were ranked via a point system with points assigned to each language after weighing six factors: (a) number of primary speakers, (b) number of secondary speakers, (c) number and population of countries where used, (d) number of major fields using the language internationally, (e) economic power of countries using the languages, and (f) socio-literary prestige. The top ten languages in order of ranking were: (a) English, (b) French, (c) Spanish, (d) Russian, (e) Arabic, (f) Chinese, (g) German, (h) Japanese, (i) Portuguese, and (j) Hindi (Weber, 1997).

[3] Approximately 750 million First- and Second-Language speakers, and another 750 million speakers of English as a Foreign Language

[4] Controlled for income and number of speakers, the three global language networks were (a) book translations, (b) multiple language editions of Wikipedia, and (c) multilingual tweets on Twitter.

franca of the Association of South-East Asian Nations (ASEAN) and is now the primary working language of European Union organisations, the United Nations, and International Olympic Committee (Ammon, 2006, p. 321; Hagen, 2016; Kirkpatrick, 2007, p. 169). Similarly in business, multinational companies are increasingly instituting English as "the common corporate language" (Neely, 2012, p. 116).[5] Furthermore, in marine transport and civil aviation, English is the basis of Seaspeak and Airspeak, respectively, two controlled natural languages designed for simple and effective international communication (Glover, Johnson, Strevens, & Weeks, 1988; Robertson, 1988). Other professions in which lingua franca English is widely employed include law, medicine, nursing, pharmacology, engineering, computers and IT, tourism and hospitality (Paltridge & Starfield, 2013, pp. v-vi).

One profession where English distinctly occupies a central role is in the sciences. Building on the early successes of British scientists, the rise of the US in the twentieth-century boosted English to the forefront of scientific languages, with the backlash against the German language after the world wars cementing its position (Ammon, 2001; Ferguson, 2007; Gordin, 2015). In specific scientific disciplines, English has long been the predominate language of communication; for example, as early as 1997, 98% of German physicists claimed English as their de facto working language (Graddol, 1997).[6] Moreover, in 1995, English accounted for 87.2% of journal publications in the natural sciences and 82.5% in the social sciences (Ammon, 2003, p. 244). More recently, more than 95% of natural science journals and 90% of social science journals are published all, or in part, in English (Flowerdew, 2013a, p. 301).

In this way, English is firmly established in the twenty-first century as a global language and lingua franca in many disciplinary and institutional settings. This widespread use of English has necessitated further developments in ELT and specifically, in researching the ways in which English is used within various and distinct *discourse communities*.[7]

---

[5] Neely's (2012) examples include companies such as, Airbus, Daimler-Chrysler, Fast Retailing, Nokia, Renault, Samsung, Technicolor, Microsoft-Beijing, and Rakuten (p. 116).

[6] Closely followed by German chemists (83%), biologists (81%), and psychologists (81%).

[7] Swales (1990) defines *discourse communities* as sociorhetorical networks with common goals, in which specific language, or genres, are used to communicate and realise those goals (p. 9).

2.1.1 The development of English Language Teaching.

The pre-eminence of English as a professional and scientific lingua franca has led to the development of a sub-field of Applied Linguistics and specialised branch of ELT known as *English for Specific Purposes* (ESP). Dudley-Evans and St John (1998) define ESP in terms of "absolute" and "variable" characteristics, asserting that ESP is designed to meet specific learner needs (pp. 4-5).[8] ESP prepares learners to enter target discourse communities by familiarising them with the distinct language practices that define those communities (Basturkmen, 2006, p. 88). The central feature of ESP is a focus on the specialised language-learning needs of purpose-oriented learners, which is typically identified through needs analyses (e.g., Hutchinson & Waters, 1987; Long, 2006). Increasingly, these specialised needs, and the challenges for ESP professionals to meet them, are arising in "noncenter countries" where English is used as a foreign or additional language (i.e., EAL; Belcher, 2006, p. 150).[9]

Alongside the growth in English as a global language has come increased mobility in education and study abroad. Over the last 40 years, the number of students studying abroad for tertiary education has increased more than five-fold from 750,000 in 1975 to more than 4.5 million in 2012 (OECD, 2014). Significantly, the majority of these students are choosing to study in English-speaking countries, particularly the USA and UK. As illustrated in **Figure 2.1**, the collective percentage of students studying in English-speaking countries in 2012 is 44%.

---

[8] According to Dudley-Evans and St Johns' (1998) *absolute characteristics* of ESP, (a) ESP is defined to meet specific needs of the learners; (b) ESP makes use of underlying methodology and activities of the discipline it serves; (c) ESP is centred on the language appropriate to these activities in terms of grammar, lexis, register, study skills, discourse and genre. According to the *variable characteristics* of ESP, (a) ESP may be related to or designed for specific disciplines; (b) ESP may use a different methodology from that of General English; (c) ESP is likely to be designed for adult learners, either at a tertiary level institution or in a professional work situation; (d) ESP is generally designed for intermediate or advanced students.

[9] "Noncenter countries" correlates with the "outer" and "expanding" circles in Kachru's *three-circles model* of English as it is used throughout the world (see for example, Kachru, 1992).

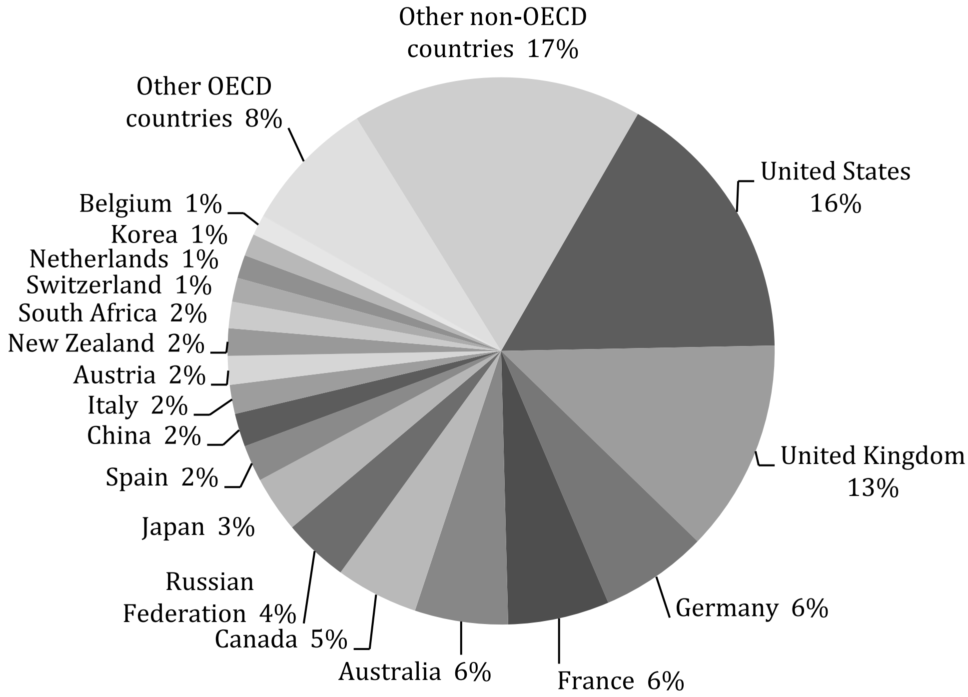

Figure 2.1 Distribution in 2012 of foreign students in tertiary education by country of destination

Note. Adapted from OECD (2014, p. 367). In this chart, the English-speaking countries are the US, UK, Australia, Canada, New Zealand, and South Africa.

This increase in student mobility has resulted in millions of students receiving tertiary education in English, giving English "a central place in the process of constructing, disseminating, and legitimizing knowledge" (Canagarajah, 2002, p. 6).

Coinciding with the boom in English-medium education, *English for Academic Purposes* (EAP) has emerged from the larger field of ESP as a branch of Applied Linguistics in its own right. In contrast with ESP, where the focus is on learners in specific occupations, EAP focuses wholly on academic contexts (Jordan, 1997, p. 3). According to Hyland & Hamp-Lyons (2002),

> The modern-day field of EAP addresses the teaching of English in the academy at all age and proficiency levels, and it draws on a range of interdisciplinary influences for its research methods, theories and practices. It seeks to provide insights into the structures and meanings of academic texts, into the demands placed by academic contexts on communicative behaviours, and into the pedagogic practices by which these behaviours can be developed. (p. 3)

In this way, EAP focuses on the diverse academic genres with their specialised linguistic patterns distinct from everyday social interactions. For example, research in EAP shows that academic discourse is characterised as more lexically dense and authoritative, with information presented as static, synoptic entities in cause-and-effect relational networks (Hyland, 2009, p. 7; Schleppegrell, 2004, p. 43-45). The mastery of this form of English is essential to those who desire to participate effectively in academic discourse communities, and has given rise to not only EAP, but also an emerging sub-field known as *English for Research Publication Purposes* (ERPP).

2.1.2 English for Research Publication Purposes.

In conjunction with the spread of English as a global language, the twenty-first century has seen a rise in scholarly publication. As of 2010, there were an estimated 5.5 million scholars, 2,000 publishers, and 17,500 research or higher education institutions worldwide contributing to academic writing for publication (Lillis & Curry, 2010, p. 1). The number of refereed academic journals listed in Ulrich's Periodicals Directory rose from 14,694 in 2001 to 25,864 in 2009, and 67% of these academic periodicals were published in part or fully in English (Mabe, 2003, p. 192; Lillis & Curry, 2010, p. 10). Flowerdew (2013b) attributes this increase in publications to (a) more universities, more faculty, more research students; (b) universities competing globally to produce more; (c) publication becoming a requirement for degrees; and (d) English as a lingua franca. The last point particularly leads Lillis and Curry (2010) to note, "the ever-growing status of the journal and journal articles is paralleled by the ever-growing use of English as the medium of such articles" (p. 9).

This burgeoning enterprise of research publication in English and the support from ELT professionals that it demands has led to a further specialised field of inquiry in Applied Linguistics: ERPP. Cargill and Burgess (2008) identify ERPP as a branch of EAP, "addressing the concerns of professional researchers and post-graduate students who need to publish in peer-reviewed international journals" (pp. 75-76). Social constructivism and situated learning are the main theoretical approaches being applied in this research, where the focus is on writers' academic literacy practices and experiences (Flowerdew, 2013a, pp. 314-315). This contrasts to work in EAP, which tends to focus more on the textual features of academic English (Curry & Lillis, 2004, p. 664). Research in ERPP has found that language is not always the major impediment for publication acceptance (e.g., Belcher, 2007; Rozycki & Johnson, 2013), leading ERPP practitioners to focus specialised support in English beyond the discourse level, in areas such as study design, literature review, and acculturation into academic communities.

2.1.3 Japanese scientists and English for Research Publication Purposes.

While remaining on the periphery for English medium communication, Japan has established itself as central in the fields of science and technology. For example, Japan consistently ranks among the top countries in terms of patents granted worldwide (WIPO, 2014), and as **Table 2.1** illustrates, between 1999-2009, was the second-largest producer of published scientific papers following the US.

Table 2.1 Top 20 countries based on papers published in Thomson Reuters-indexed journals from 1999-2009

Rank	Field	Papers
1	USA	2,974,344
2	Japan	788,650
3	Germany	766,162
4	England	682,018
5	China	649,689
6	France	548,046
7	Canada	424,562
8	Italy	403,588
9	Spain	305,430
10	Australia	276,622

Note. Source: Thomson Reuters (2009) Science Watch

A comparison of Japanese- and English-medium journals reveals that the h5-index[10] of the top-100 Japanese-medium journals ranges from 5–12 citations, much lower than the 109–377 range for the top-100 English-medium journals.[11] For this reason, Japanese scientists have traditionally preferred to submit their papers to English-medium journals, and in particular, to the high-impact journals (Yamazaki, 1995). In this way, for many Japanese scientists, ERPP is an essential part of their professional life.

[10] The h5-index measures the number of publications, by an individual or organisation, and the number of citations per publication.

[11] As indexed by Google Scholar, 2016 April 5th, https://scholar.google.com.au/citations?view_op=top_venues&hl=en

Chapter 2: Literature Review

The challenges that arise for Japanese scientists using English as both a professional language and a foreign one have motivated various ERPP studies in recent decades. For example, Yamazaki (1995) examines the refereeing systems of the journals most preferred by Japanese scientists, reporting a rejection rate of one in two papers (i.e., 49%; p. 126). Similarly, Gosden (2003) examines the linguistic exchanges between Japanese scientists, journal editors and referees, and highlights interactional deficiencies in Japanese researchers' manuscripts as a major concern. In another study, Willey and Tanimoto (2012; 2013) identify that Japanese scientists regularly seek proofreading of their English manuscripts before submitting them to international journals for publication; and in investigating the strategies that Japanese scientists employ to overcome the difficulties in mastering scientific discourse in English, Okamura (2006) found that in addition to reading academic texts within their field, direct attention to mastering English language seems to "pay off in the long run" for Japanese scientists (p. 77). One area where this "direct attention" may be applicable is within the analytical framework of Error Analysis (EA), a branch of Applied Linguistics dedicated to identifying discrete areas of weakness within a particular genre or discourse.

2.2 Error Analysis

EA is a methodology developed early within the field of Second Language Acquisition (SLA) to help researchers refine their theoretical focus on the construct that came to be known as *interlanguage* (IL). IL is the learner's language at a given point in their learning; it is systematic, with rules that may approximate but differ from the target language (Selinker, 1972, 1992). IL's significance in SLA has been established in the characterisation of IL as the "disciplinary beginnings" of SLA (Ortega, 2013, p. 2). During these beginnings, the search for a comprehensive theory of second language acquisition was the research theme bringing the field together, and it was from this investigation that the EA methodology was born.

2.2.1 The development of Error Analysis.

The development of EA in the 1960–70s was concurrent with a shift in ELT from what Candlin (in the Preface of Richards, 1974) called "teacher as controller" towards a more "learner-centred view" (p. iii). Based on the hypothesis that *errors* in learners' L2 utterances are systematic, EA was developed for SLA researchers to uncover the systems underlying L2 errors (Corder, 1967, p. 163). With influence from Chomsky, a distinction was made between "errors of competence" and "errors of performance" (Corder, 1967, p. 166-167). Errors of competence are systematic, not readily self-corrected, and "provide evidence of the system of the language that [the learner] is using at a particular point in the course" (Corder, 1967, p. 167). In contrast, errors of performance are random, readily self-corrected and can be distinguished as *mistakes*, which are of no concern in EA. Reflecting on the development of EA, James (1998) synthesised two decades of EA research to derive the following definition for errors:

> Errors cannot be self-corrected until further relevant (to that error) input (implicit or explicit) has been provided and converted into intake by the learner. In other words, errors require further relevant learning to take place before they can be self-corrected. (p. 83)[12]

[12] One earlier distinction denotes errors at the level of pragmatics as *infelicities* (Austin, 1962). Pragmatic infelicities are outside the scope of EA and this review.

As SLA research moved away from contrastive analysis,[13] EA was subsequently developed to focus wholly on the L2 and the "built-in syllabus" of the learner, reflecting as Corder (1967) noted, a move "away from a preoccupation with teaching towards a study of learning" (p.163).

2.2.2 Error Analysis methodology.

The EA methodology was first introduced by Corder (1967, 1971) and subsequently elaborated by others (e.g., Abbott, 1980; Dulay, Burt & Krashen, 1982; James, 1998; Levelt, 1978; Richards, 1974). After data elicitation (i.e., the collection of L2 texts), most EA analyses can be mapped along the following three stages: (a) recognition and reconstruction of errors, (b) description of errors, and (c) explanation of errors. In the first stage, errors are identified (i.e., *recognised*) as idiosyncratic or unsuccessful language and revised (i.e., *reconstructed*) into well-formed sentences based on plausible interpretation.[14] An algorithm proposed by Corder (1971) for this first stage is reproduced in **Figure 2.2**.

[13] Prior to the development of EA, another methodology known as *contrastive analysis* (CA) was the dominant paradigm in the search to explain second language acquisition. Based on the *language transfer hypothesis* (i.e., linguistic systems similar in the L1 and L2 are easier to learn, and conversely, those different are more difficult to learn), CA was used to predict L2 acquisition. These L1-based predictions were subsequently used to inform language syllabi and ELT practices. However, the failure of CA to produce accurate and informative predictions, along with its link to the discredited behaviourism theory, led SLA researchers to abandon the L1 contrast.

[14] *Plausible interpretation* of errors can be taken directly from the L2, or back-translation from the L1 (Corder, 1981, pp. 21-24).

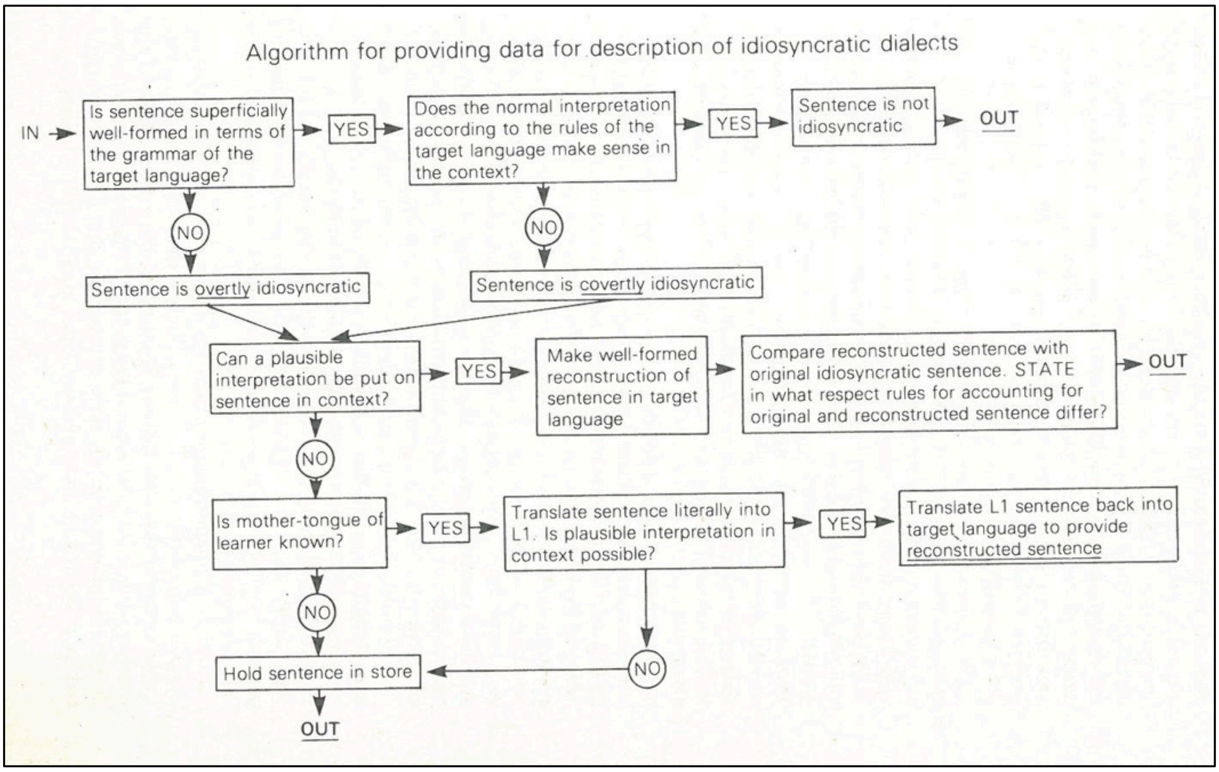

Figure 2.2 Algorithm for the first stage of error analysis

Note. This algorithm was originally published in Corder (1971); however the version re-produced here is taken from a re-print of the original article in Corder (1981, p. 23).

In this algorithm and generally in EA, the sentence is the unit of analysis. Every sentence is regarded as idiosyncratic until shown to be otherwise (Corder, 1981, p. 21).

For the second stage of EA, *the description of errors*, a variety of error classification systems have been developed. The initial classification is in terms of the linguistic level (e.g., phonological, grammatical, lexico-semantic) and the ways in which the errors occur (e.g., omission, addition, etc.), as illustrated in **Table 2.2**.

Table 2.2 Error types and linguistic levels

	Graphological Phonological	Grammatical	Lexico-semantic
Omission			
Addition			
Selection			
Ordering			

Note. Reproduced from Corder (1981, p. 36), this framework serves only as visual representation; it is not a complete EA model.

The first column of **Table 2.2** identifies four *error types*. Dulay, et al. (1982) validated these error types as the four principal ways in which IL and L1 diverge, in what they termed the *surface structure taxonomy*. James (1998) later elaborated the four error types, re-labelling and adding a fifth, in what he called the *target modification taxonomy*: (a) omission, (b) over-inclusion, (c) mis-selection, (d) mis-ordering, and (e) blends (p. 111). This initial level of classification is simply the "starting point for systematic analysis", as further linguistic description is applied to indicate what is being omitted, added, mis-selected, or mis-ordered (Corder, 1981, p. 36). For example, at the grammatical level, errors are further analysed "in terms of systems, such as tense, number, mood, gender, case, and so on" (Corder, 1981, p. 37).[15]

The third stage of the EA methodology is *the explanation of errors*. This was the theoretical raison d'être of the method**.** In other words, it was the explanation of errors, or more specifically the explanation of the systems underlying the errors, that motivated SLA researchers to pursue this method towards an overall understanding of IL. In addition to an earlier distinction between *interlingual* and *intralingual* errors,[16] one enduring framework for the explanation of the possible processes underlying IL errors

[15] A number of other error taxonomies have been developed. For example, Burt and Kiparsky's (1972) error taxonomy distinguishes interlingual, intralingual, local and global errors.

[16] From CA, *interlingual* errors derive from L1 transfer to the L2, while *intralingual* errors are the result of systems within the L1 (Richards, 1971).

is Richards' (1971): (a) overgeneralisation, (b) ignorance of rule restriction, (c) incomplete application of rules, (d) false hypothesis. The theoretical understandings that these explanations fostered were central to the development of the field of SLA. However, the approach was not without criticism and limitations.

2.2.3 Limitations of Error Analysis.

While EA was integral to advancing SLA theory in the early stages of the field, it was soon met with contention and limitations. One early criticism of the EA approach to understanding the processes of SLA is *avoidance* (Schachter, 1974; Schachter & Celce-Murcia, 1977). Learners' avoidance of L2 forms cannot be accounted for by EA since it is not present in the data. As James (1998) notes, the first error type in the target modification taxonomy, omission, could actually be "non-acquisition" (p. 107). Another criticism of EA is the *comparative fallacy hypothesis* (Bley-Vroman, 1983, p. 15). This hypothesis asserts that IL is a complete language with its own set of rules, and not simply a defective version of the L2. From this perspective, EA is rejected on the grounds of its comparison with the target language (see also the *target deviation hypothesis*; Klein, 1998). Furthermore, while Corder's methodology set the standard for EA, technical difficulties with its descriptive basis created issues of validity (Corder, 1981, p. 24). Even the very name "error analysis" itself became contentious.[17]

Despite these short-comings, EA has always had a more "practical task" in guiding language pedagogy and learning in the classroom (Strevens, 1969, p. 6). As Corder (1981) explains:

> It is now generally recognized that that branch of applied linguistic activity which is usually called error analysis has two functions. The first is a theoretical one and the second is a practical one. [...] The practical aspect of error analysis is its function in guiding the remedial action we must take to correct an unsatisfactory state of affairs for learner or teacher. (p. 45)

In the decades following its development informing theoretical and descriptive linguistics, this practical application of EA has been utilised widely in ELT.

---

[17] In order to avoid the "target-language based" associations, Corder (1971) actually preferred the term, *idiosyncratic dialect analysis* (p. 61). Hammarberg (1974) moved to re-brand the approach as *performance analysis*, taking into account errors and *non-errors*.

2.2.4 Error Analysis and English Language Teaching and research.

While EA was initially established as a method for developing SLA theory, it has always appealed to ELT practitioners as a "methodology for dealing with data" (Cook, 1993, p. 22). In other words, EA offers teachers a "hands on" approach to the investigation of errors, a relevant professional and everyday concern (James, 1998, p. x). This aspect of EA led to its widespread application in ELT. For example in Taiwan, in order to measure the pre- and post-treatment effects of a computer-assisted-instruction program on learners' grammar skills, Chen (2006) performed an EA of L1 Taiwanese written essays. The EA procedure followed four stages and classified the errors into 15 categories.[18] While the results indicated no significant difference between the experimental and control groups, the EA enabled the identification of the most salient error categories for pedagogical intervention.

In Iran, Khodabandeh (2007) used an EA approach in the investigation of Persian cross-linguistic influence on newspaper headline translations. In this study, errors were investigated at three linguistic levels: grammatical, lexico-semantic, and discoursal. An extensive framework of error classification was adopted from earlier work (i.e., Burt and Kirparsky, 1972; Keshavarz 1993) and adapted to fit each of the three levels linguistic levels. The results showed errors occurred most frequently at the grammatical level, followed by the discoursal- then lexico-semantic levels. Results were subsequently used in feedback on syllabus design and teaching. Also in Iran, Falhasiri, Tavakoli, Hasiri and Mohammadzadeh (2011) examined the effects of corrective feedback (explicit and implicit) on errors, using an EA methodology to uncover the most frequently occurring grammatical and lexical errors in L1 Farsi learners' written compositions, which were subsequently targeted for corrective feedback as a treatment. Pre- and post-treatment error frequencies indicated that most errors were intralingual, and that explicit feedback for interlingual errors was the most effective form of corrective feedback.

Moreover, in Malaysia, in a study of collocation errors in L1 Malaysian learners' written essays, Hong, Rahim, Hua and Salehuddin (2011) employed a corpus-based EA

[18] Chen's (2006) four EA stages are (a) data collection, (b) identification of errors, (c) classification of errors into error types, and (d) statement of error frequency. The 15 error categories are errors in (a) nouns, (b) articles, (c) pronouns, (d) verbs, (e) prepositions, (f) adjectives, (g) adverbs, (h) conjunction, (i) sentence fragments, (j) syntax, (k) lexicon, (l) punctuation, (m) spelling, (n) capitalization, (o) subject omission.

framework in five stages to examine the types and sources of collocation errors, specifically verb-noun collocation units.[19] Within these collocation units, results indicate that preposition-related errors are most frequent, and intralingual errors are dominant. The researchers used the findings to generate a series of recommendations for language pedagogy targeted at collocation learning for Malaysians.

Although EA remains an effective tool for ELT research, EA findings can only be as valid and effective as the analytical frameworks and systems of linguistic descriptions utilised. Hong et al.'s (2011) corpus-based approach highlights one potential area for enriching the traditional EA framework (see **Section 2.4**). One further area of development within the EA literature in recent years is the adoption of a Systemic Functional Linguistic approach in the examination and description of errors. This latter approach is outlined in the following section.

[19] Hong et al.'s five stages of EA are (a) data generation, (b) identification of errors, (c) error classification, (d) error quantification, and (e) analysis of error sources.

2.3 Systemic Functional Linguistics

Systemic Functional Linguistics (SFL) is a comprehensive theory of language as a socio-semiotic system (i.e., language as it functions in context as a resource for meaning-making; Matthiessen & Halliday, 2009, pp. 8-16). By integrating SFL descriptions into an EA analytical framework, research into errors benefits from the decades of theoretical advances made in SFL since EA was developed. In addition to furthering traditional grammatical descriptions, such an integration enables analysis along functional lines, and provides insight into the paradigmatic systems of choice underlying grammatical structures.

SFL views language as a *stratified* model across three levels of expression (i.e., phonology and graphology), lexicogrammar, and discourse semantics, as illustrated in **Figure 2.3**.

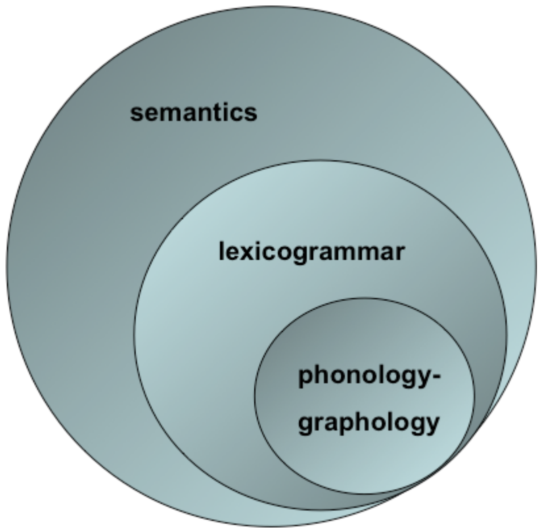

Figure 2.3 Canonical hierarchy of Systemic Functional Linguistics
Note. Adapted from Martin (2014, p. 7).

According to this model, sounds (i.e., phonology) or letters (i.e., graphology) combine to form words organised by systems into structures (i.e., lexicogrammar), which in turn construe meanings (i.e., semantics). In the internal stratum of *lexicogrammar*, grammar and lexis are viewed as a continuum (Matthiessen & Halliday, 2009, p. 10). This view enables error analysts to integrate grammatical descriptions of errors with lexical ones.

2.3.1 Rank scale and system networks.

The lexicogrammatical stratum can be further divided and analysed in SFL via the rank scale and system networks. The SFL *rank scale* orders grammatical units into a hierarchy where generally each unit consists of units from the next rank down (Matthiessen & Halliday, 2009, p. 71).[20] **Table 2.3** illustrates the rank scale within the stratum of lexicogrammar.

[20] Embedding of units from the same rank or above is also common. For example, a prepositional phrase may function within a nominal group (e.g., the ethylene gas pressure effect for this photoreaction).

Table 2.3 Rank scale at lexicogrammar stratum

Rank	Example
Clause	/Threshold voltage instability was investigated for 4H-SiC MOSFETs with phosphorus-doped and NO-annealed gate oxides/
Group/phrase	[Threshold voltage instability] [was investigated] [for 4H-SiC MOSFETs with phosphorus-doped and NO-annealed gate oxides]
Word	[{Threshold} {voltage} {instability}]
Morpheme	{<Thresh><hold>} {<volt><age>} {<in><stabil><ity>}

Note. In the denotation of rank elements, the forward slash (i.e., /) signifies the division of a clause, the square brackets (i.e., [and]) signify the division of a group or phrase, the brace (i.e., { and }) signify the division of a word, and the arrows (i.e., < and >) signify the division of a morpheme (Thompson, 1996, p. 23).

The highest-ranking unit of analysis in lexicogrammar is the clause, which consists of five classes of groups/phrases:[21] (a) nominal group, (b) verbal group, (c) adverbial group, (d) conjugation group, and (e) prepositional phrase (Halliday & Matthiessen, 2014, pp. 362-363). The group/phrase rank consists of words, and words are realised by morphemes. Each unit in the rank scale carries different types of linguistic patterns, or structural organisations, for example the clause unit is organised with participants, processes, and circumstances, while the group units is an expansion of a head word (Eggins, 2004, p. 126). By applying the rank scale in EA, analysts can examine the ways in which error patterns emerge within the constituent parts. Particularly, the nominal group is identified in SFL as a defining feature in the academic and scientific registers for

[21] In SFL terms, the distinction between a *group* and *phrase* is that a group is an expansion of a word, whereas a *phrase* is a contraction of the clause (Halliday and Matthiessen, 2014, p. 437).

its role in densely compacting meanings within the clause (e.g., Halliday, 1985, p. 73; Halliday & Martin, 1993; Martin & Veel, 1998).[22]

Another key feature of SFL is the mapping of system-structure relationships, referred to as *system networks*. System networks are a visual representation of the choices within language and their structural *realisations*. In other words, the structures, or grammatical forms, are the realisations of the choices within a grammatical system. **Figure 2.4** demonstrates a system network mapping the personal pronoun system for English.

[22] The centrality of the nominal group in academic discourse has been well documented in SFL research. For example, Fang, Schleppegrell and Cox (2006) demonstrate how nominal groups expand the clause to include more information in academic texts. McCabe and Gallagher (2008) reveal the central role of the nominal group in the ideational metafunction of meaning-expression in undergraduate academic writing, and Whittaker, Llinares, and McCabe (2011) highlight the management of nominal groups as an essential skill in developing successful L2 academic writing.

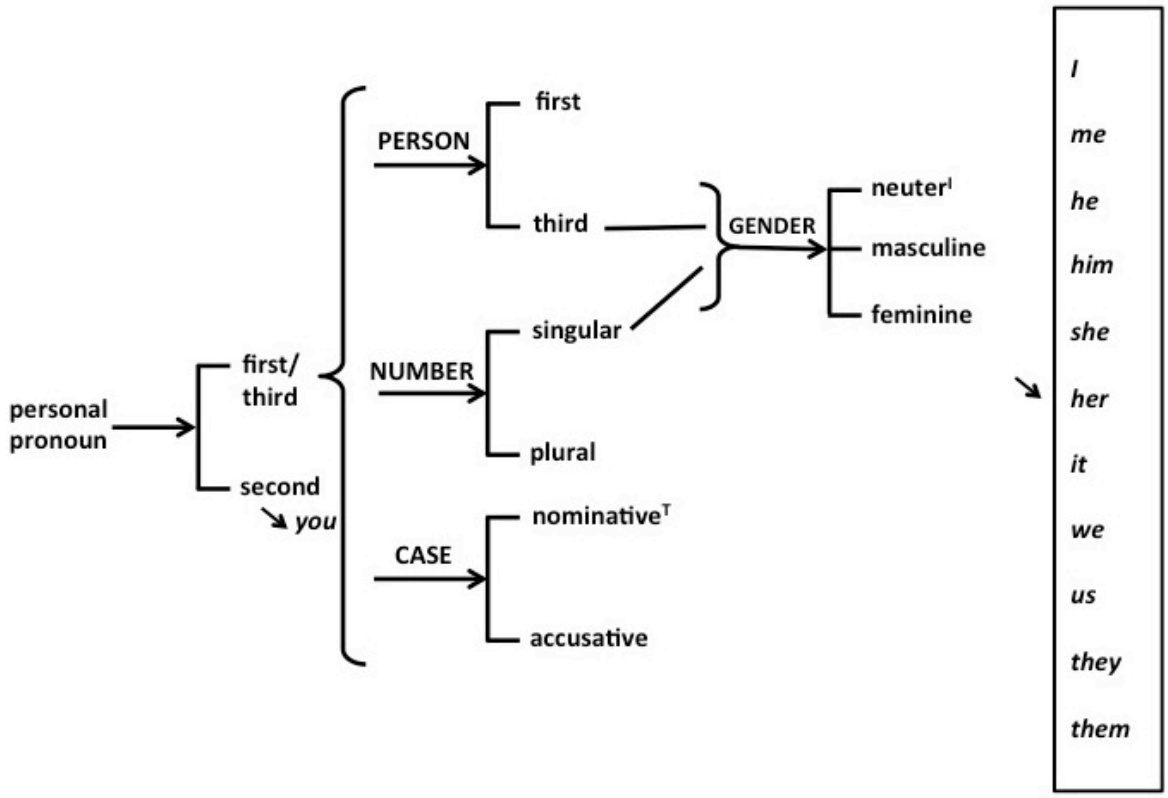

Figure 2.4 System network mapping the English personal pronoun system

Note. This system network, adapted from Martin (2013, p.23), operates at the word rank within the lexicogrammar stratum.

Following SFL conventions (e.g., Martin, 2013; pp. 13-18; Matthiessen & Halliday, 2009, pp. 98-99), the entry condition for the system network is given on the far left-hand side (i.e., personal pronoun), and the box on the far right-hand side lists the network's structural realisations (e.g., *I*, *me*, *he*, etc.). The *systems*, illustrated by right-pointing horizontal arrows, indicate the choices available within the network. The system names are given in capital letters (e.g., PERSON), and the system options, or *features*, are given in lowercase (e.g., first, third). The curly bracket, or brace, represents *simultaneous systems*, indicating that there is a concurrent choice in the systems, and the diagonal arrows point to the forms that are the realisations of the systems (e.g., *I* is the realisation of the choices of first in the PERSON system, singular in the NUMBER system, neuter in the GENDER system, and nominative in the CASE system).[23]

In this way, system networks facilitate a *paradigmatic* view of language such that grammatical forms, and subsequently the errors that manifest in those forms, can be examined through their underlying systems.

2.3.2 Systemic Functional Linguistics and Error Analysis.

In recent years, this SFL approach has been applied in the examination of errors in L2 texts. For example, Kim (2010) performed an SFL-inspired EA, examining Korean students' translations by analysing clauses according to metafunctions (e.g., within the experiential metafunction, errors where classified according to Participant, Process, or Circumstance). Errors, defined as "problematic parts in terms of accuracy and appropriateness" were marked and tallied to identify error patterns in individual students' translations from L1 Korean to L2 English and reveal areas for focussed improvement. Kim (2010) concluded that it was possible to classify errors based on meaning using an SFL approach and that this knowledge could be used to produce more systematic feedback on errors and competence in translation skills.

Similarly, Hamilton (2015) explored the potential contribution of SFL within a traditional EA framework, examining a pilot-corpus of first-year French university students' essays on two levels: the traditional morphosyntactic level, and the SFL level of metafunction.[24] Hamilton's (2015) findings indicate that within the first-level of grammatical errors, most errors are located in the nominal group, and in particular, manifest in article/determiner errors. In the second level of analysis, the results show that errors occur most frequently within participant elements in the ideational metafunction, while within the textual metafunction, almost three-quarters of the errors

[23] The superscript capitals *I* and *T* are abbreviations of *if* and *then*, respectively. They signal that if a feature with superscript I is chosen (e.g., neuter^{I}), then there is only one corresponding choice in another system (e.g., nominative^{T}) (i.e., if neuter, then nominative).

[24] Halliday's paradigmatic view of grammar led him to the discovery of three systemic clusters in language, known in SFL as *metafunctions*, which map the three distinct ways in which meaning in language is construed. The three metafunctions are: ideational, interpersonal, and textual (Hasan, 2014, p. 14).

occurred within the Rheme element of the clause.[25] Hamilton (2015) concluded that an SFL integrated approach to EA fosters fresh insight into the understanding of L2 errors, building on the traditional EA framework.

Furthermore, in a comparison of L2 English users' and native-English speakers' deployment of grammatical resources in scientific laboratory reports, Schleppegrell (2002) examined lab reports from two angles: the clause-level linguistic resources deployed in meaning-making, and the sentence-level errors. Errors are defined as clause formations departing from native-speaker norms, and found to be distracting to the instructors evaluating the lab reports. Findings indicate that errors occurred most frequently in article usage, plural and count/mass noun marking systems, and preposition choice, and highlight the need to assist L2 writers in drawing on the appropriate linguistic resources to present their intended meanings (p. 140).

This section has presented an overview of the SFL stratified model of language, its paradigmatic perspective, and notable applications in EA in recent years. Another important and often complementary approach to language analysis is Corpus Linguistics (CL).

[25] In SFL, the *Theme* is a functional term for the topic of a clause, and the *Rheme* is the remaining part (i.e., not the Theme), which is generally around 75%. Given this ratio, Hamilton's (2015) finding does not designate any special status for errors to the Rheme.

2.4 Corpus Linguistics

CL is a systematic approach to the analyses of language in collections of "real" texts or *corpora* (Meyer, 2002, pp. xi-xii). Corpora provide a basis for the empirical analyses of language; for example, frequency analyses can reveal the prevalence of linguistic items across large bodies of texts, and concordance lines can elucidate the contexts in which those items occur (Evison, 2010, pp. 123-135). This data-driven approach allows generalisations to be made across whole languages, and enables the precise characterisation of specific registers within a language (e.g., written scientific discourse; Hunston, 2002).

CL researchers can design corpora to represent a specialised text-type or genre (e.g., the scientific research article). Similarly, corpora may focus on a specific population of language users to compare their language use with others. For example, *learner corpora* can be a collection of L2 texts produced by a sample or population of language learners, and may be compared with expert or large general reference corpora.[26] In this way, the CL approach facilitates the quantitative distinction of learner constructions from expert or native ones, enabling researchers to pinpoint target areas for needs-specific, data-driven language pedagogy (Johns & King, 1991). One area where the CL approach to examining representative texts has revealed great insights is within the domain of academic discourse.

2.4.1 Corpus Linguistics and academic discourse.

CL research has contributed significantly to current understandings of academic discourse, and subsequent advances in the field of EAP. For example, the Academic Word List (AWL), now well known and applied in EAP, identifies a core group of sub-technical words used frequently and only in academic discourse (Coxhead, 2000).[27] Similarly, CL research has provided empirical evidence for the grammatical features of academic discourse, namely the extensive use of densely packed nominal groups (Biber,

---

[26] General reference corpora, such as the Corpus of Contemporary American English (COCA) and British National Corpus (BCN) comprising millions of words in thousands of language instantiation (i.e., texts) are often used in CL to represent the English language as it is generally spoken or written.

[27] The AWL is a list of 570 frequently occurring words central to academic English, as revealed through a 3.5 million-word specialised corpus (Coxhead, 2000).

1988, pp. 28-29). Furthermore, according to Biber, Johansson, Leech, Conrad, and Finegan (1999), around sixty per cent of all content words in academic discourse are nouns (p.15). Other corpus researchers have shown that this extensive use of nominal groups leads to subsequent increases in the use of determiners and prepositions (Aarts & Granger, 1998; Hunston, 2002). For example, CL analyses have revealed common *collocations*, or multi-word units, co-occurring in academic texts, such as the nominal pattern *the * of ** (e.g., *the set of rules*) (Biber et al., 1999; Biber, 2009). In addition to mapping the linguistic features most prevalent in academic English, CL provides a strong platform for investigating errors.

2.4.1 Corpus Linguistics and Error Analysis.

The data-driven CL approach has been employed in the investigation of errors in recent years. For example, Granger (2003) developed a large-scale learner corpus, the International Corpus of Learner English (ICLE), with error-tagging enabling quantitative EA on a much larger scale than previously possible. Similarly, Gressang (2010) developed a noun phrase error tagging system to investigate errors in articles, determiners, and pronouns in a learner corpus of essays produced by L2 English learners from article-less L1 backgrounds (i.e., Korean, Chinese, and Taiwanese). Her study reveals a significant under-use of the indefinite article *a/an*. Crompton (2005) also developed a learner corpus to contrast L1 Malay speakers' use of English relativisations with those prevalent in two expert written academic English corpora, highlighting the over-use of *where*-constructions in the learner corpus and under-use of standard forms of relativisation common in written academic English. In another learner corpus examination, Flowerdew (2006) compiled a corpus of argumentative essays written by L1 Cantonese learners of English to investigate their use of signalling nouns. Based on frequency data from the corpus, he derived an error taxonomy of the various error types with signalling nouns. These corpus investigations provide a unique insight into how large collections of authentic data can reveal error patterns representative of a particular population.

2.5 Chapter Summary

This chapter has provided a background review for the investigation of L2 errors in the scientific writing of Japanese scientists. This review has mapped the predominance of English as a global language and subsequent developments in ELT, situating this research within the fields of ESP, EAP, and the emerging sub-field of ERPP. A thorough overview of the analytical and theoretical frameworks of the investigation was presented, by first focusing on the EA methodology then on SFL, and finally on CL. The following chapter will elaborate on these theoretical frameworks to define the methodology employed in this examination of Japanese scientists' L2 English scientific writing.

Chapter 3: Methodology

This study employs an elaborated Error Analysis (EA) methodology in the investigation of errors in the writing of Japanese scientists using English for Research Publication Purposes (ERPP). This corpus-assisted EA focuses on error patterns in nominal groups through the lens of Systemic Functional Linguistics (SFL). The research design and methodology for the investigation are described in three main sections. The first section (**Section 3.1**) provides an overview of the theoretical and analytical framework; the second section (**Section 3.2**) describes the corpus design, including a rich description of the Japanese scientists participating in the study and the method of data collection and preparation; the third and final section of the chapter (**Section 3.3**) outlines the elaborated EA procedure.

3.1 Integrated Framework of Analysis.

The theoretical and analytical framework integrated in this study derives from three sources: EA, SFL, and Corpus Linguistics (CL). This section provides an overview of these three bases, beginning with EA.

As discussed in the Literature Review (**Section 2.2.2**), EA research typically involves a three-stage analysis (Corder, 1981, pp. 21-25):

(1) Error recognition and reconstruction
(2) Error description
(3) Explanation of errors

The present study follows a similar trajectory towards identifying and examining errors in two stages:

(1) Error recognition and reconstruction
(2) Error classification and quantification

The first stage corresponds with proofreading, in which errors are identified and corrected (i.e., recognised and reconstructed). In the second stage, errors are classified and quantified to elucidate the most frequently occurring error patterns in the data. Classification is initially in terms of the four *error types*: omission, addition, selection, and ordering (Dulay, Burt & Krashen, 1982). This initial level of classification is further augmented with linguistic description indicating what is being omitted, added, mis-selected, or mis-ordered. For the purposes of this study, the linguistic descriptions draw from an SFL model of language.

SFL provides a lens through which errors can be better understood. As discussed in the Literature Review (**Section 2.3**), SFL views language as a stratified model along planes of expression, lexicogrammar, and semantics (Martin, 2014, pp. 6-7). Generally in EA, the focus of analysis is at the "grammatical" and "lexico-sematic" levels (Corder, 1981, pp. 36-37). In SFL, this corresponds with the stratum of *lexicogrammar*, which places grammar and lexis on a continuum; in other words, two poles on a single cline (Halliday & Matthiessen, 2014, p. 24). Within the lexicogrammar stratum, errors can be examined according to where they are located within the clause via the *rank scale*. The rank scale divides the clause into four group classes and one phrase class: nominal group, verbal group, adverbial group, conjugation group, and prepositional phrase (Halliday & Matthiessen, 2014. pp. 362-363). For purposes of scope, the present study examines error patterns within the group-phrase rank, specifically, within nominal groups.

In English, and particularly in academic and scientific registers, much of the meaning in language is packed into nominal groups. In other words, in technical English the expansion or modification of nominal groups is an important linguistic resource for extending the clause with densely packed meaning (Halliday, 1985, p. 73; Halliday & Martin, 1993; Martin & Veel, 1998). All nominal groups contain a Head.[28] In *simple nominal groups*, the Head may be realised by a pronoun (e.g., *this*), a proper noun (e.g., *Mitsubishi*), a numeral (e.g., *two*), and in certain cases an adjective (e.g., complete).[29] In *complex nominal groups*, the Head may be realised by a common noun, which is modified. Typically, complex nominal groups are pre-modified by a Deictic,[30] which specifies a subset of a referent. As more meaning is packaged into the group, additional elements may be added following the pre-determined pattern illustrated in **Table 3.4**.

[28] Following SFL convention, initial capital letters are used to distinguish functional terms from traditional grammatical classes.

[29] An Epithet may act as Head when the nominal group functions in the clause as Attribute (Halliday & Matthiessen, 2014, p. 391). For example, *complete* in the following: *After sputtering, the sol-gel process is complete*.

[30] Deictics in traditional grammar are often synonymous with determiners; for example, *each* in *each layer was prepared*, or *a* in *a layer was prepared.* Similarly in SFL, the Deictic has a "pointing out" function (Halliday & Matthiessen, 2014, p. 60).

Table 3.4 Pre-modification pattern of complex nominal groups

Functional element:	Deictic	Numerative	Epithet	Classifier	Head
Word class typically realising the function:	determiner	numeral	adjective	noun or adjective	noun
Example:	*these*	*two*	*optically-active*	*substance*	*syntheses*

Note. Adapted from Halliday and Matthiessen (2014, p. 379).

All elements before the Head (i.e., Deictic, Numerative, Epithet, and Classifier) constitute the nominal group's *Pre-modifier*. To further elaborate meaning in nominal groups, pre-modification is coupled with post-modification. Post-modification of nominal groups is typically realised by embedded prepositional phrases and rankshifted clauses. Sentences 1 and 2 demonstrate post-modification, with complex nominal groups denoted in square brackets and the *Post-modifiers* highlighted in italics.

1. [Met-heme coordination] contributes to [the stability *of the structure*] and [the ability *of electron transfer in cyt c family proteins*]. [05EST][31]
2. [Integer photonic spin *obeying Bose-Einstein statics*] confines [plural spins *with the same quantum states*]. [02ESC][31]

As evident in Sentences 1 and 2, the Post-modifiers comprise a substantial portion of the nominal group in which errors may occur. Therefore, in this framework of analysis the location of errors in the Pre-modifier, Head, and Post-modifier provides an entry point into the classification of errors occurring in nominal groups.

In addition to the error's location within the group, the error classification can be further elucidated by SFL system networks. System networks provide an efficient way to not only elucidate and visualise grammar (**Section 2.3.1**) but also to analyse the errors within those grammatical systems. **Figure 3.5** presents the system network mapped for

[31] [05EST] and [02EST] are labels for corpus extracts. See **Section 3.2** for details.

and applied in this study for the analysis of complex nominal group errors with articles and plural -*s*.

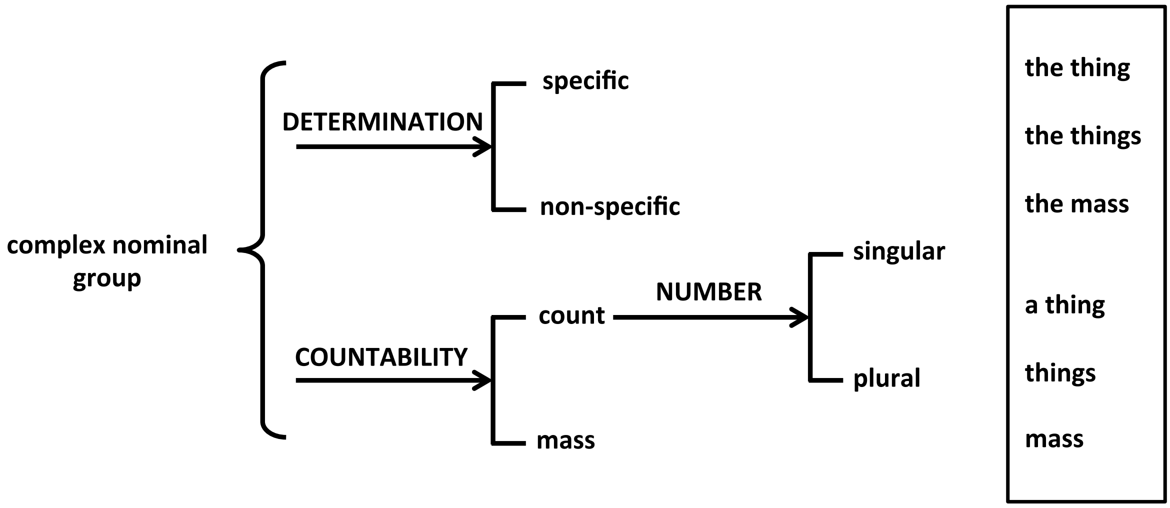

Figure 3.5 System network for articles and plural -*s*.

The complex nominal group is the entry condition for this system network; in other words, simple nominal groups do not enter the system. The network's structural realisations are given in the box on the far right-hand side. The network comprises three systems: DETERMINATION, COUNTABILITY and NUMBER. The binary choices within the two systems of DETERMINATION and COUNTABILITY—and subsequently within the NUMBER system since it is dependent on the COUNTABILITY system—are simultaneous and contingent on each other. Therefore, errors mis-construing referent specificity (i.e., *the* vs. *a* and Ø) are intricately bound with errors mis-construing countability and number (i.e., singular, plural or mass), and are therefore analysed simultaneously under this analytical framework. In this way, systems-structure relationships visualised through system networks afford a unique view into lexicogrammar and errors. To broaden that view, the third theoretical basis, CL, is employed.

CL enables the identification of error patterns across multiple texts within a sample population. Such empirical investigations may reveal error patterns distinct to certain demographics; for example, frequency analyses can reveal the most prevalent errors types and their hierarchical orders in a specialised corpus representing a target population. To this end, the present corpus is developed from a collection of scientific research articles written by Japanese scientists. From this corpus, a specialised sub-corpus of errors is compiled from the errors occurring in nominal groups. A detailed description of the corpus design follows in **Section 3.2**.

3.2 Corpus Design

The present study examines a specialised corpus of Japanese scientists' English writing for research publication purposes: the *Corpus of Japanese Scientific Discourse* (COJSD).[32] The corpus is a 46,263-word collection of research article manuscripts written by thirteen Japanese scientists working in the materials science department of a national university in Japan. The field of materials science includes work in physics, chemistry, bioscience and device engineering. To hold a faculty position in the department, participants are expected to have top-authored at least five papers published in international journals and co-authored five other papers.[33] For purposes of this study, only those faculty members who hold faculty positions were recruited, and therefore these participants are considered and herein referred to as *Established Career Scientists* (ECS).[34] Eight of the thirteen participants are Assistant Professors, four are Associate Professors, and one is a Full Professor.[35] ECS are chosen in this study over early-career scientists because their greater writing experience increases the likelihood of their errors being systematic, rather than individual mistakes.

All thirteen participants are male Japanese nationals living in Japan. According to the results of an online questionnaire administered to the participants,[36] at the time of data collection these ECS were actively writing research articles in English for publication at least once a year, and some more than 2-3 times per year (e.g., *Journal of American Chemical Society, Angewandte, PLOS One,* etc.). Throughout these processes, half of the participants had been informed that their manuscripts required proofreading

---

[32]Acronym pronounced *ko-jay-es-dee.*

[33] Within the department, international journals are considered as those published in English. This is in contrast with journals published in Japanese, which are intended for a domestic audience.

[34] In contrast with early career scientists or apprentice scholars (e.g., PhD candidates), who are typically writing their first papers.

[35] In the Japanese academic context, Full Professor is the highest academic rank; Associate Professor is one rank down from Full Professor; followed by Assistant Professors.

[36] The questionnaire and responses are presented in **Appendix C**.

for English grammar prior to being accepted for publication. All thirteen participants self-requested English proofreading from the researcher and agreed to their manuscripts being analysed for the purposes of this study.

The primary data in this study are the participants' scientific research article manuscripts. The *research article* (RA) is selected as the "pre-eminent genre" of academic and scientific discourse (Hyland, 2010, p. 117) and in this way the most representative of scientific writing. A *manuscript* is the complete RA prior to submission for publication and therefore, at this point in the revision process, is solely the voice of the authors without input from reviewers or editors. While some of the RAs were co-written with other research group members, in all cases the participant is the top-author and primary author of the manuscript. One manuscript was collected from each of the thirteen participants, and then reviewed for basic grammatical corrections, which were marked up using the Track Changes tools in Microsoft Word. One copy with mark-up was returned to the ECS, and another was further analysed using the methodology mapped in **Section 3.3** below.

For the protection of participants' anonymity, authors' names, affiliations, acknowledgments, and references are removed from the corpus.[37] Other identifying information is replaced with generic headings (e.g., *Title*, *Keywords*, etc.). Figure captions and table titles are left intact, as these contain errors forming part of the data. For labelling, each text is assigned a number from one to thirteen combined with the abbreviation for Established Career Scientists given in square brackets (e.g., [01ECS]). This labelling is applied whenever corpus excerpts are presented. An overview of the COJSD with word and sentence counts is presented in **Table 3.5**.

[37] These elements do not contribute in anyway to the data (e.g., word-counts, etc.).

Table 3.5 Overview of Corpus of Japanese Scientific Discourse (COJSD)

Text	Total words	Total sentences
01ECS	3,458	190
02ECS	4,909	245
03ECS	2,881	154
04ECS	4,497	282
05ECS	4,517	238
06ECS	1,851	96
07ECS	2,376	160
08ECS	5,123	218
09ECS	4,137	315
10ECS	2,517	90
11ECS	1,679	92
12ECS	4,144	218
13ECS	4,174	211
Total	**46,263**	**2,509**
Average	**3,559**	**193**

Note. UAM CorpusTool (O'Donnell, 2015) used for word and sentence counts.

The length of the RA manuscripts varies from 1,679 to 5,123 words, with an average word count of 3,559, and the total number of sentences ranges from 90 to 315, with an average of 193. To account for the range in text length, results of the analyses are normalised to occurrences per 1000 words or presented as relative percentages. These analyses are detailed in the following section.

3.3 Elaborated Error Analysis Procedure

This section describes the elaborated EA procedure and demonstrates how the resources of SFL and CL have been integrated to provide a more in-depth and empirical understanding of the most common nominal group errors occurring in Japanese scientists' writing. In the first stage, errors are identified and reconstructed through the initial proofreading process (**Section 3.3.1**), and in the second stage, errors are classified and quantified (**Section 3.3.2**).

3.3.1 Recognition and reconstruction of errors.

Once the manuscripts are collected and anonymised, the texts in the COJSD are proofread for grammatical errors. *Errors* are recognised as unsuccessful or idiosyncratic language interrupting reading flow and comprehension.[38] Following Corder (1981), these errors are reconstructed into well-formed sentences based on "plausible interpretation" elucidated from the context (pp. 21-23). An example of this process can be seen in Excerpts 3-4, where Excerpt 3 represents the original text prior to proofreading, and Excerpt 4 illustrates the reconstruction of identified errors (e.g., *the acid generation, which also give rise to,* etc.)

Before error recognition and reconstruction.

3. Therefore, multiple steps are involved until the acid generation, which also give rise to many decomposed fragments remained in a system for these classes of PAGs. [06ECS]

After error recognition and reconstruction.

4. Therefore, multiple steps are involved until ~~the~~ acid generation is achieved, ~~which also~~ giv~~e~~ing rise to ~~many~~ decomposed fragments remain~~ed~~ing in ~~a~~the system for these classes of PAGs. [06ECS]

During this proofreading stage, the text mark-up is annotated using the Track Changes tools in MS Word. Errors are highlighted in red, with strikethrough and underline indicating the reconstructions.

[38] This operationalisation is a synthesis of previous EA research; especially, Corder (1981) and James (1998)

Once the errors have been identified and reconstructed, the overall frequency of errors in the corpus as a whole is quantified. However, as discussed in **Section 3.1**, a complete and thorough analysis of all the errors recognised and reconstructed in the COJSD is beyond the scope of this thesis. Therefore, to refine the focus of the study, a subcorpus of nominal groups with errors is extracted. An overview of the *Nominal Group Subcorpus* (NGS) is presented in **Table 3.6.**

Table 3.6 Overview of Nominal Group Subcorpus (NGS)

Text	Approximate nominal groups per text in COJSD	Nominal groups containing errors in NGS
01ECS	562	54
02ECS	627	89
03ECS	495	78
04ECS	911	79
05ECS	731	10
06ECS	300	29
07ECS	380	20
08ECS	858	51
09ECS	831	50
10ECS	384	30
11ECS	278	32
12ECS	747	8
13ECS	776	68
Total	**7,879**	**598**
Average	**606**	**46**

Note. The labelling of texts in the NGS remains consistent with the whole-corpus labels (e.g., [01ECS]). The approximate number of nominal groups per text (i.e., second column) is based on the number of nominal groups in a 1,000-word sample from each text (see **Appendix E**).

The NGS comprises all identified nominal groups with errors (i.e. 598), which is 7.58% of the total number of nominal groups in the COJSD. These nominal groups with errors form the basis for the classification and quantification of errors, as discussed in the following sections.

3.3.2 Classification and quantification of errors in nominal groups.

After the NGS is compiled and the overall frequency of errors in nominal groups quantified, the nominal group errors can be further analysed according to location (i.e., where the error occurs within the nominal group) and the error types. The following three sections outline these sub-stages in the classification of errors in the NGS.

3.3.2.1 Classification of errors by location and error type.

As outlined in the Integrated Framework of Analysis (**Section 3.1**) nominal groups can be simple or complex, with complex nominal groups containing a Head, which may be pre- and post-modified.[39] Errors in complex nominal groups can be analysed according to their location in the Head, Pre-, or Post-modifier as illustrated in Excerpts 5–7.

Error in Pre-modifier.

5. A ~~S~~similar phenomenon [01ECS]

Error in Head.

6. The ~~complete~~ completion of auto-activation [08ECS]

Error in Post-modifier.

7. the ethylene gas pressure effect ~~to~~ for this photoreaction in methylcyclohexane using pressure-resistant reaction vessel [03ECS]

These errors can be further classified according to the four error types: omission, addition, selection and ordering (see **Section 2.2.2** and **Section 3.1**). For example, in Excerpt 5, the error occurs in the Pre-modifier (i.e., *a similar phenomenon*) and can be

[39] Among all errors in nominal groups identified in this study, only one error was found in a simple nominal group. This result is presented and discussed in the Findings, **Section 4.2.1**.

classified as an omission error, where the indefinite article *a* is omitted. In Excerpt 6, the error occurs in the Head (i.e., ~~*complete*~~ *completion*) and can be classified as a selection error, where the adjective form *complete* is mis-selected for the noun form *completion*. In Excerpt 7, the error occurs in the Post-modifier (i.e., ~~*to*~~ *for this photoreaction...*) and can be classified as a selection error, where the preposition *to* is mis-selected for the preposition *for*.

Once the errors have been classified according to error location (i.e., Pre-modifier, Head, or Post-modifier) and error type (i.e., omission, addition, selection, or addition), these occurrences are examined for frequency to determine which error patterns emerge as the most prevalent across the NGS. In particular, the analysis reveals that the majority of errors in nominal groups involve pre-modifying articles and the use of plural -*s*. The second most prominent error pattern found within the COJSD involves the post-modifying -*of*. The following two sections detail how these two major error patterns are further analysed.

3.3.2.2 Errors with articles and plural -s.

Errors with articles and plural -*s* are classified by error type as outlined in **Table 3.7**.

Table 3.7 Errors with articles and plural -*s*

Excerpt	Error	Reconstruction	Error type
an increase in high order oligomers [05ECS]	Ø	*a*	Omission
both ends of the 6-system [06ECS]	Ø	*the*	Omission
the polymerization of monomers with the formation of 1c [06ECS]	Ø	*plural -s*	Omission
~~a~~ yellowish, turbid PF8T2 particles [02ECS]	*a*	Ø	Addition
~~the~~ particle scattering [02ECS]	*the*	Ø	Addition
chemical biology~~ies~~, pharmaceuticals, medicinal and agricultural areas [04ECS]	*plural -s*	Ø	Addition
the diffraction from ~~a~~ the pentacene film [01ECS]	*a*	*the*	Selection
~~the~~ a surface [01ECS]	*the*	*a*	Selection

Note. Ø indicates the absence of an article or plural -*s*.

As illustrated in **Table 3.7**, articles and plural -*s* may be erroneously omitted, added, or mis-selected. Note that a fourth error type is possible (i.e., ordering errors), however no such occurrences were identified in the COJSD.[40]

As discussed in the Integrated Framework of Analysis (**Section 3.1**), errors with articles and plural -*s* can be analysed according to the system network in which they occur (**Figure 3.5**). In order to elucidate patterning within the errors with articles and plural -*s* (i.e., exactly what is being erroneously omitted, added, or mis-selected), these

[40] This point is elaborated in the Findings (**Section 4.3.1**).

errors are analysed according to the five features of the system network (i.e., specific, non-specific, singular, plural, and mass), as illustrated in **Table 3.8**.

Table 3.8 Matrix for analysis of errors with articles and plural *-s*

	Specific	Non-specific
Singular	1. the optimization of the whole protein structure 2. the much higher catalytic activity of the mature enzyme than that of zymogen 3. the YASARA structure molecular modeling software package 4. the auto-oxidation of **the** cysteine thiol group on the enzymes	1. a very small amount 2. a combination of PAC-1 and the mature caspase-3 3. a cobalt-chelating affinity column 4. a decrease in the concentration of 2-ME
Plural	1. the largest binding energies	1. higher concentrations 2. freeze-and-thaw cycles
Mass	1. the binding of synthetic molecules 2. a docking simulation for the binding of *N*-[4-(2,3-diphenyl-3,4-dihydropyrazol-5-yl)phenyl]acetamide to an uncleavable mutant procaspase-3 reported by Clark and co-workers 3. the modulation of protein functions	

Note. Excerpts within this matrix represent all article and plural -*s* omission errors from one text: [08ECS].

The column and row headings of **Table 3.8** derive from the features of the system network (**Figure 3.5**) and classify the reconstructed referent (i.e., not the erroneous referent). Therefore, the second column (i.e. Specific) classifies errors with specific referents, and the third column (Non-specific) classifies errors with non-specific referents. The three rows enable simultaneous classification for singular, plural, and mass referents. The numbering of errors in each cell simply facilitates quantification within the given text (i.e., 08ECS). The results of this analysis are presented in **Section 4.3.1.**

3.3.2.3 Errors with preposition -of.

The second major error pattern to emerge from the data involves the preposition -*of* in either the Post- or Pre-modifiers of complex nominal groups, as illustrated in Excerpts 8 and 9.

Post-modifier error with preposition -of

8. a sudden leap ~~of~~ in reaction conversion [03ECS]

Pre-modifier error with preposition -of

9. the de peak ~~of de~~ [03ECS]

While the errors with preposition -*of* typically occur in the Post-modifier, as shown in Excerpt 8, they also involve pre-modification in the reconstruction, as illustrated in Excerpt 9. Therefore, although the error in Excerpt 9 occurs in the Post-modifier, it is classified as a Pre-modifier error because the reconstruction (i.e., *the de peak*) is within the Pre-modifier.

The errors with preposition -*of* are further analysed according to the four *error types*: omission, addition, selection, and ordering, as illustrated in Excerpts 10–13.

Omission error with preposition -of

10. knowledge and understanding of an efficient generation and switching of optically active substances [02ECS]

Addition error with preposition -of

11. the incident X-ray angle ~~of x-ray~~ [01ECS]

Selection error with preposition -of

12. One possible reason ~~of~~ for the decrease in domain size [01ECS]

Ordering error with preposition -of

13. the refractive index of the cosolvents at 589 nm (n_D) ~~of the cosolvents~~ [02ECS]

The classification of errors with preposition -*of* completes the analyses of the second major error pattern emerging from the data. The results of this analysis are presented in **Section 4.3.2**.

3.4 Chapter Summary

This chapter has presented the research design and methodology for this investigation into Japanese scientists' writing for research publication purposes. The analytical framework underpinning the study was outlined, highlighting how the three theoretical bases—EA, SFL, and CL—are combined to provide an integrated framework of analysis. This chapter has further described the development of the COJSD, a corpus of 13 Japanese Scientists' research article manuscripts submitted for proofreading (i.e., error reconstruction), and the specialised NGS. The elaborated EA procedure was outlined in two stages: error recognition and reconstruction, and error classification and quantification. Within the error classification and quantification stage, the analysis was further elaborated according to the location of the errors within the nominal groups (i.e., Head, Pre-modifier, Post-modifier), the error types (i.e., omission, addition, selection, and ordering) and the two most prominent error patterns identified in the analysis (i.e., errors with articles and plural -*s*, and errors with preposition -*of*). The results of these analyses are presented, and their findings discussed in Chapter 4.

Chapter 4: Findings

This chapter details the investigation of errors in the English writing of Japanese scientists. The chapter commences with an overview of the corpus providing the basis for analyses and a focus on errors in nominal groups by examining nominal group types, forms of nominal group modification, and location of errors within the nominal group. The chapter then elaborates the investigations of two major error patterns emerging from the data: errors with articles and plural -*s*, and errors with preposition -*of*. Throughout the chapter, the results and discussions are integrated to facilitate a comprehensive overview of the investigation.

4.1 Overview of Errors in the Corpus

This examination of error patterns in the L2 writing of Japanese scientists[41] commences with a broad overview of the key features of the Corpus of Japanese Scientific Discourse (COJSD) and the errors within the corpus. The COJSD comprises 13 texts with a total of 46,263 words in 2,509 sentences. As described in the Methodology (**Section 3.3**), the thirteen texts collated in the COJSD are proofread for grammatical errors; these reconstructed texts then serve as the data for the present analysis. **Table 4.9** identifies the number of words, sentences, and importantly the prevalence of the errors in each text in the COJSD.

[41] As noted in Chapter 1, L2 encompasses English as a foreign and additional language in this thesis.

Table 4.9 Overview of error frequencies in the Corpus of Japanese Scientific Discourse (COJSD)

Text	Words per text	Sentences per text	Sentences with errors	Percentage sentences with errors (%)	Total errors	Total errors per 1,000 words
01ECS	3,458	190	68	35.79	130	38
02ECS	4,909	245	80	32.65	179	36
03ECS	2,881	154	77	50.00	171	59
04ECS	4,497	282	96	34.04	138	31
05ECS	4,517	238	34	14.29	39	9
06ECS	1,851	96	40	41.67	63	34
07ECS	2,376	160	30	18.75	40	17
08ECS	5,123	218	76	34.86	109	21
09ECS	4,137	315	70	22.22	96	23
10ECS	2,517	90	36	40.00	89	35
11ECS	1,679	92	51	55.43	80	48
12ECS	4,144	218	27	12.39	34	8
13ECS	4,174	211	117	55.45	200	48
Total	**46,263**	**2,509**	**802**		**1,368**	
Average	**3,559**	**193**	**62**	**31.96%**	**105**	**31**

Within the COJSD, a total of 1,368 errors were recognised and reconstructed in 802 sentences, with an average of 31 errors per 1,000 words.[42] The number of errors per text spans from 8 to 59 per 1,000 words, reflecting the range in the participants' written English proficiency. In terms of sentences with errors, on average around 32% of all sentences in the data contained errors, or around one in every three sentences (**Table 4.9**). The sentences in Excerpts 1 to 4, extracted from the COJSD, illustrate the variation with which different error types occur within the various clause elements.[43]

1. For high-yielding reduction of azides, many conditions have ~~been~~ appeared, e.g. zinc metal/acetic acid, hydrogenolysis, Staudinger reactions, and other reducing agents shown in Scheme 35. [04ECS]
2. Molecular dynamics (MD) ~~simulate~~ simulations suggested three possible binding paths. [08ECS]
3. This means that the most stable conformer in the diastereoselective photoreaction of cyclohexenone can be stabilized efficiently by the clustering effect in the ground state, and may afford high~~ly~~ de values. [03ECS]
4. The weak correlation between structural dissymmetry and ~~the~~ luminescence dissymmetry factor implies a weak ~~contributes~~ contribution of the static-coupling mechanism to the circularly polarized dissymmetry of the lanthanide(III) f-f transition. [10ECS]

The first feature of Excerpts 1–4 is that while they achieve the academic register of scientific discourse with sophisticated linguistic resources, they all contain errors (highlighted in red) interrupting the reading flow and comprehension. In Excerpt 1, there is an addition error of the auxiliary verb *be* in the verbal group. In Excerpt 2, there is a selection error in the nominal group, where the verb-form *simulate* is mis-selected for the noun-form *simulation*. In Excerpt 3, two errors occur within a single nominal group: the mis-selection of the adverb-form *highly* for adjective form *high*, and the omission of plural -*s*. Finally, in Excerpt 4 there are four errors: (1) the addition of the

[42] A sample from a COJSD text demonstrating the average 31 errors per 1,000 words is presented in **Appendix D**.

[43] Extracts are labelled with text numbers (e.g., 04ECS) as described in the Corpus Design (**Section 3.2**).

definite article *the* for a non-specific referent in one nominal group (i.e., *the luminescence*), (2) the omission of both the indefinite article *a* and (3) the definite article *the* in another nominal group (i.e., *a* weak contribution, and *the* static-coupling mechanism), and (4) the mis-selection of the verb form *contributes* for the noun form *contribution*.

Another feature of Excerpts 1–4 and across the COJSD is the occurrence of multiple errors within single sentences. Excerpts 1 and 2 contain single errors, while Excerpts 3 and 4 contain multiple errors (i.e. two and four errors, respectively) typical of the COJSD. According to the totals in **Table 4.9**, among the sentences with errors there are an average of 1.7 errors per sentence.

One further feature of the sentences with errors in Excerpts 1–4 and throughout the COJSD is that while errors may occur across different clause units, (i.e., in verbal groups, nominal groups, adverbial groups, conjugation groups, and prepositional phrases) they tend to occur more frequently in nominal groups. As highlighted in the Literature Review (**Sections 2.3.1** and **2.4.1**) and Methodology (**Section 3.1**), the packaging of meaning into nominal groups is a central feature of academic and scientific discourse. As nominal groups are elaborated, and become denser, there is greater potential for errors to occur. Given this role of nominal groups and their significance for meaning-making within academic and scientific discourse,[44] the nominal group was selected as the focus for investigation in this study. The following sections examine the different patterns of errors found within nominal groups in the COJSD.

[44] See for example: Biber, 1988, pp. 28-29; Fang, Schleppegrell & Cox, 2006; Halliday, 1985, p. 73; Halliday, 2004, pp. 75-79; Halliday & Martin, 1993; Martin & Veel, 1998; McCabe and Gallagher, 2008; Whittaker, Llinares, and McCabe, 2011.

4.2 Overview of Errors in Nominal Groups

While nominal groups are a key linguistic resource for meaning-making in academic registers, they are also a major source of errors in the COJSD. This section provides an overview of errors in nominal groups, the type of nominal groups containing errors (**Section 4.2.1**), and the location of errors within the group (**Section 4.2.2**).

To begin this examination, nominal groups with errors are collated into a sub-corpus (see **Section 3.3.1**). Within the *Nominal Group Subcorpus* (NGS), a total of 654 errors are identified, comprising almost half (47.81%) of the total errors in the COJSD, as illustrated in **Table 4.10.**

Table 4.10 Overview of error frequencies in the Nominal Group Subcorpus (NGS)

Text	Nominal groups containing errors	Nominal groups containing errors as percentage of approximate nominal groups in COJSD (%)	Total number of errors in nominal groups	Nominal group errors as percentage of total errors in COJSD (%)	Nominal group errors per 1,000 words
01ECS	54	9.60	72	55.38	21
02ECS	89	14.20	98	54.75	20
03ECS	78	15.77	79	46.20	27
04ECS	79	8.67	83	60.14	18
05ECS	10	1.37	12	30.77	3
06ECS	29	9.66	29	46.03	16
07ECS	20	5.27	22	55.00	9
08ECS	51	5.95	54	49.54	11
09ECS	50	6.02	54	56.25	13
10ECS	30	7.82	40	44.94	16
11ECS	32	11.50	34	42.50	20
12ECS	8	1.07	8	23.53	2
13ECS	68	8.77	69	34.50	17
Total	**598**		**654**		
Average	**46**	**7.59%**	**50**	**47.81%**	**15**

Notably, this high proportion of errors in nominal groups is predictable given that technical meaning is frequently realised in nominal groups, and prepositional phrases also consist mostly of embedded nominal groups. Illustrating this second point, Excerpt 5 demonstrates one prepositional phrase consisting largely of one embedded nominal group (i.e., *the various mobility enhancing fabrication techniques*).

Prepositional phrase with embedded nominal group

5. Among the various mobility enhancing fabrication techniques [01ECS]

Within the NGS, 654 errors were reconstructed in 598 nominal groups, meaning that while most nominal groups with errors contain single errors, instances of multiple errors also occur. Excerpts 6–10 provide an overview of the errors in nominal groups and their features.

6. a wide range of thermoelectric SWNT-based materials [11ECS]
7. an attractive target ~~to~~ for synthetic chemists [04ECS]
8. natural marine ~~natural~~ products [05ECS]
9. ~~the~~ gate oxide formation [13ECS]
10. ~~a~~ the potential for ~~its use~~ application as ~~the~~ an elementary carrier~~s~~ [10ECS]

Typical of errors in the NGS, the nominal groups in Excerpts 6–9 contain single errors (i.e., omission, selection, ordering, and addition errors). In Excerpt 6, the indefinite article *a* is omitted, and in Excerpt 7, the preposition *to* is mis-selected for *for* (i.e., selection error). In Excerpt 8, there is an ordering error of the adjective *natural*, and in Excerpt 9 the definite article *the* is mistakenly added. Excerpt 10, is the only nominal group with multiple errors in this sample, containing three selection errors: (1) the mis-selection of the indefinite article *a* for the definite article *the*, (2) the mis-selection of *its use* for *application*, and (3) the mis-selection of the definite article *the* and plural *-s* for the singular indefinite article *an* (i.e., mis-construal of an indefinite singular referent as a definite plural referent).

Notably, all of the nominal groups in Excerpts 6-10 are complex nominal groups, as opposed to simple nominal groups, which are void of modification. This feature and the error patterns in complex and simple nominal groups are discussed in detail in following section.

4.2.1 Errors in complex and simple nominal groups.

Nominal groups are classified as complex or simple (see **Section 3.1**). In short, complex nominal groups contrast with simple nominal groups in terms of modification. In simple nominal groups, typically realised by pronouns and proper nouns, the Head stands alone, whereas in complex groups, it takes some form of modification. Excerpts 11 and 12 illustrate this distinction.

Simple nominal group

11. Phe7 [12ECS]

Complex nominal group

12. The hydrophobic packing of Phe7, Leu27, Tyr34, Phe44, and Ile83 [12ECS]

The simple nominal group in Excerpt 11 expresses the name of a chemical compound without modification. In contrast, the complex nominal group in Excerpt 12 contains five simple nominal groups embedded in a single rankshifted prepositional phrase (i.e., *of Phe7*, *Leu27*, *Tyr34, Phe44, Ile83*). Rankshifted prepositional phrases and clauses are analysed as part of the complex nominal group in this study.

These embedded elements contribute to the elaboration that is a feature of nominal groups in academic and technical discourses, and in turn provide further possibility for error. Notably, among the 598 nominal groups with errors in the NGS, 597 were complex; in other words, only one error was identified within a simple nominal group. That simple nominal group (i.e., *PAG 1a*) is presented with its error (i.e., the addition of the definite article *the*) in Excerpt 13.

Simple nominal group error

13. The methoxy group of 1a(OMe) was deprotected to a hydroxyl group by BBr_3 and a subsequent esterification with mesyl chloride (MsCl) gave ~~the~~ PAG 1a. [06ECS]

PAG 1a is the name of a specific chemical in a group of chemicals known as *photo acid generators* (PAG). This naming makes it inherently specific and distinguishes it from other PAG, in this way, nullifying the necessity of an article.[45]

This result that 597 of the 598 nominal groups with errors are complex nominal groups (i.e., 99.83%), suggests that pedagogy aimed at improving accuracy in Japanese scientists' L2 English should focus on the forms of modification that make the complex nominal group both a central feature of scientific writing and a prime source of error. These forms of modification are further investigated in the following section.

4.2.2 Errors in Pre-modifier, Head, and Post-Modifier.

As more meaning is packed into nominal groups in technical registers, complex nominal groups are increasingly modified. This modification may precede the Head, in which case it is referred to as the Pre-modifier, or it may follow as the Post-modifier (Matthiessen, Teruya, & Lam, 2010, p. 24).[46] Additionally, the Head itself may be modified (e.g., adding plural *-s*). Pre-modification occurs in a largely fixed sequence, formalised in functional terms as follows: Deictic, Numerative, Epithet, Classifier, Head (Halliday & Matthiessen, 2014, p. 379).[47] Importantly, each element to the left of the Head modifies the one to the right, as illustrated in Excerpt 14.

Pre-modification of complex nominal group

14. a definitively-controlled absolute optically-active substance synthesis [02ECS]

In Excerpt 14, the Head (i.e., *synthesis*) is modified by the Classifier (i.e., *substance*); both are subsequently modified by three Epithets (i.e., *definitively-controlled*, *absolute*, *optically-active*). There is no Numerative in Excerpt 14; the final element on the left is the Deictic (i.e., *an*), indicating the level of specificity of the Head. As more meaning is packed into the nominal group, Classifiers, Epithets, and Deictics can be added following

---

[45] Similarly, *vitamin C* is one specific vitamin in the group of vitamins, and does not require the additional specificity of the definite article, unless it is modified (e.g., *the vitamin C on the shelves*, or *the commonly taken vitamin C*).

[46] The Post-modifier is also labelled functionally as Qualifier.

[47] Deictics are typically realised by determiners, Numeratives by numerals, Epithets by adjectives, Classifiers by nouns or adjectives, and Heads by nouns (**Table 3.4**).

this pattern of left modifying right. Notably, Excerpt 14 is free from error, but it is readily apparent how this form of modification presents challenges for Japanese scientists needing to nominalise scientific meaning in this way.

In contrast, post-modification of complex nominal groups is realised by rankshifted clauses and embedded prepositional phrases, as described in the Integrated Framework of Analysis (**Section 3.1**), and illustrated in Excerpts 15 and 16.

Post-modification of complex nominal groups

15. the E1-mechanism stabilizing the carbocation intermediate [06ECS]
16. their use from the viewpoint of acid generation [06ECS]

Excerpts 15 and 16 demonstrate post-modification through a rankshifted clause (i.e., *stabilizing the carbocation intermediate*) and an embedded prepositional phrase (i.e., *from the viewpoint of acid generation*), respectively.

Therefore, errors in complex nominal groups may occur in three locations: the Pre-modifier, Head, and Post-modifier, as shown in Excerpts 17–19, or in more than one of these locations, as shown in Excerpt 20.

Pre-modifier error

17. its amino acid sequence with that of horse cyt *c* [12ECS]

Head error

18. optically active substances [02ECS]

Post-modifier error

19. Detailed procedure ~~to~~ for calculat~~e~~ing the band-edge profile [01ECS]

Errors in Pre- and Post-modifier

20. ~~a~~ the potential for ~~its use~~ application as ~~the~~ an elementary carrier~~s~~ [10ECS]

Table 4.11 presents the prevalence of errors in these three locations, expressed as a percentage of nominal group errors in each text, and overall in the NGS.

Table 4.11 Errors in Pre-modifier, Head, and Post-modifier as percentage of nominal group errors

Text	Errors in Pre-modifier (%)	Errors in Head (%)	Errors in Post-modifier (%)
01ECS	40.28	5.56	54.17
02ECS	53.06	23.47	23.47
03ECS	78.48	13.92	7.59
04ECS	74.70	10.84	14.46
05ECS	75.00	16.67	8.33
06ECS	55.17	3.45	41.38
07ECS	40.91	18.18	40.91
08ECS	48.15	20.37	31.48
09ECS	53.70	20.37	25.93
10ECS	60.00	7.50	32.50
11ECS	64.71	23.53	11.76
12ECS	50.00	12.50	37.50
13ECS	60.87	14.49	24.64
NGS	**59.02%**	**14.98%**	**25.99%**

Notably, errors in the Pre-modifier are twice as frequent as errors in the post-modifier, and four times as frequent as errors in the Head, highlighting the significant obstacle pre-modification poses for Japanese scientists. However, further analysis is required to elucidate in detail the various errors within the three locations of the nominal group. The following section examines the two most salient error patterns occurring in nominal groups.

4.3 Error Patterns in Nominal Groups

From the analysis of errors in the Pre-modifier, Head, and Post-modifier, two major error patterns emerged: errors relating to articles and plural -*s*, and errors relating to the preposition -*of.* The following sections investigate these two major error patterns, beginning with the most salient, errors with articles and plural -*s*.

4.3.1 Errors with articles and plural -*s*.

Errors with articles *the* and *a* occur frequently in both the Pre- and Post-modifiers of complex nominal groups, as illustrated in Excerpts 21–26.[48]

Errors with articles in Pre-modifier

21. ~~a~~ the coordination geometry around Eu^{III} [10ECS]
22. ~~a~~ greater chiroptical amplitudes [02ECS]
23. a more stable conformer [03ECS]

Errors with articles in Post-modifier

24. wettability of ~~a~~ the gate insulator surface [01ECS]
25. The amount of ~~the~~ dimer produced by the treatment with ethanol [05ECS]
26. The further elimination of a proton [06ECS]

Furthermore, this pattern of errors with articles can be extended to the Head when articles are mis-selected for plural -*s* in non-specific plural referents, as shown in Excerpts 27–29.

Errors with articles in Pre-modifier and plural -s in Head

27. ~~the~~ clusters [03ECS]
28. ~~the~~ previous stud~~y~~ ies [13ECS]
29. ~~a~~ high levels of circular polarization [10ECS]

This apparent integrated nature of errors with articles and plural -*s* motivates further investigation into the systems underlying the grammar. As discussed in Chapter 3 (**Figure 3.1**), the choices in the DETERMINATION system occur simultaneously with choices in the COUNTABILITY system, and subsequently in the NUMBER system. In other words, the choice of definite or indefinite article is concurrent with and contingent on the choice in singular, plural, or mass, and vice versa, indicating that errors with articles are inextricably bound with plural -*s* errors, and thus can be analysed as one

[48] The indefinite article *a* and its phonetic equivalent *an* are treated as the same article in this investigation.

pattern. The frequency of the articles and plural -*s* error pattern is presented in **Table 4.12**, illustrating the prevalence of this error pattern throughout all thirteen texts.

Table 4.12 Errors with articles and plural -*s*

Text	Errors in nominal groups	Errors with articles and plural -*s* per text	Errors with articles and plural -*s* per 1,000 words	Errors with articles and plural -s as percentage of errors in nominal groups (%)
01ECS	72	44	13	61.11
02ECS	98	67	14	68.37
03ECS	79	55	19	69.62
04ECS	83	63	14	75.90
05ECS	12	6	1	50.00
06ECS	29	19	10	65.52
07ECS	22	16	7	72.73
08ECS	54	31	6	57.41
09ECS	54	34	8	62.96
10ECS	40	28	11	67.50
11ECS	34	26	15	76.47
12ECS	8	4	1	50.00
13ECS	69	49	12	71.01
NGS	**654**	**442**	**131**	**67.43%**
Average	**50**	**34**	**10**	

Note. NGS refers to the Nominal Group Subcorpus, which includes only those nominal groups with errors.

The errors with articles and plural *-s* pattern accounts for two in three errors in the NGS (i.e., 67.58%), and occurs on average ten times per 1,000 words. In terms of total errors, this one error pattern constitutes around 32% of all errors in the COJSD. Comparing this result with prior research requires caution given that much of that research measures percentage accuracy under controlled conditions (e.g., gap fill exercises) rather than frequency in free production,[49] however within the ELT and error analysis literature several studies are comparable. For example, with L1 Spanish university-level learners, Dotti and O'Donnell (2014) found that article errors account for around 10% of all errors. For L1 French university students, Hamilton (2015) reports that from a total of 906 errors, articles and determiner errors accounted for 11.04% of total errors. Also in Europe, Klages-Kubitzki (1995) reports that 49.39% of all errors produced by L1 German university lecturers in English departments in Germany were with articles, particularly the definite article. Schleppegrell (2002) reports "count/mass noun and article" errors account for up to 37.87% of the total surface grammar errors produced by three ESL students (p. 137).[50] In Asia, Chen (2006) reports a surprisingly lower rate of article errors of 4.23% for Taiwanese learners (i.e., 141 in 3,332; p. 89), while Hong et al. (2011) report a combined total of 19.20% for articles and plural *-s* errors in the written essays of L1 Malaysian students (i.e., 12.91% and 6.29%, respectively; p. 39).

The prevalence of this major error pattern demands closer investigation to elucidate the ways in which errors with articles and plural *-s* unfold. As previously discussed (**Sections 2.2.2**, **3.1**, and **3.3.2**), this EA classifies errors into four error types: omission, addition, selection, and ordering. Excerpts 30–40 demonstrate the four error types within the errors with articles and plural *-s* pattern.

Omission errors

30. A photograph of the MCh meta-molecule [07ECS]
31. new functional devices such as a “one-way mirror” [07ECS]

[49] For example, in a study of 60 university students in Japan, Goto-Butler (2002) reports an overall level of accuracy for "nontargetlike article use" of 34.16%.

[50] Details of the ESL students' language backgrounds are not given.

32. the function of proteins composed of several protomer units such as caspase-3 [08ECS]

Addition errors

33. ~~the~~ directional anisotropy [07ECS]
34. ~~a~~ dispersion-type features at these frequencies [07ECS]
35. the total length of the meta-molecule in millimeter~~s~~. [07ECS]

Selection errors

36. ~~a~~ the way toward the realization of an artificial gauge field [07ECS]
37. ~~the~~ a result of binding to the protomer interface cavity [08ECS]
38. the function of ~~a~~ proteins composed of several protomer units such as caspase-3[08ECS]
39. a typical reaction~~s~~ for producing a mature enzyme [08ECS]

Ordering error

40. the very weak ~~the~~ catalytic activity [i]

Note that Excerpt 40 is an invented example (i.e., [i]), created to demonstrate an ordering error. In the NGS, no ordering errors are found within this error pattern, indicating that the ordering of articles and plural -*s* presents no difficulty for Japanese scientists. This suggests that pedagogy aimed at addressing errors with articles and plural -*s* can be focused on the three other error types: omission, addition, and selection. The percentage distribution of omission, addition, and selection errors with articles and plural -*s* is given in **Table 4.13**, highlighting the dominance of omission errors.

Table 4.13 Percentage distribution of error types in errors with articles and plural *-s*

Text	Omission errors (%)	Addition errors (%)	Selection errors (%)
01ECS	63.64	18.18	18.18
02ECS	83.58	11.94	4.48
03ECS	74.55	9.09	16.36
04ECS	92.06	6.35	1.59
05ECS	83.33	16.67	0.00
06ECS	84.21	10.53	5.26
07ECS	56.25	37.50	6.25
08ECS	51.61	25.81	22.58
09ECS	70.59	23.53	5.88
10ECS	50.00	28.57	21.43
11ECS	73.08	23.08	3.85
12ECS	50.00	50.00	0.00
13ECS	81.63	12.24	6.12
Total	**74.21%**	**16.29%**	**9.50%**

Notably, omission errors are more than four times as frequent as addition errors, and eight times more frequent than selection errors, making it the dominant error type in the errors with articles and plural *-s* pattern. This result correlates with previous studies showing that L2 learners with article-less L1s (e.g., Chinese, Taiwanese, Korean) frequently omit articles in their writing (Gressang, 2010; Robertson, 2000; White, 2003). Similarly, Rozycki and Johnson (2013) found the dropping of articles to be a feature of "non-canonical" grammar in research articles written by non-native engineers, and Master (1997) has shown that even advanced learners have difficulties with articles.

This finding prompts a closer investigation into the various manifestations of errors within these three error types to identify what is being omitted, added and mis-selected. In this investigation, the 442 errors with articles and plural -*s* were analysed according to the system network in **Figure 3.5.** In short, each error is analysed according to the five features of the system network (i.e., singular, plural, mass, specific, and non-specific). The results of this analysis, presented in **Table 4.14**, illustrate the dominance of errors with singular referents, and a trend towards errors with non-specific referents.

Table 4.14 Relative frequencies of article and plural -*s* errors

	Specific	Non-specific
Singular	5. the optimization of the whole protein structure [08ECS] 6. the Pauli exclusion principle [02ECS] 7. the YASARA structure molecular modeling software package [08ECS] 8. the auto-oxidation of the cysteine thiol group on the enzymes [08ECS] 9. the gas-liquid interface [09ECS] 10. the critical point in particular [03ECS] 11. the excited state [03ECS] 12. the quenching reaction [04ECS] 13. the kealiinine family [04ECS] 14. the oxidized monomeric M61A PA cyt c_{551} [05ECS] 15. the E1-mechanism [06ECS] 16. the lower half of the bandgap [13ECS] 17. ~~a~~ the substrate surface [01ECS] 27.66%	5. a very small amount [08ECS] 6. a combination of PAC-1 and the mature caspase-3 [08ECS] 7. a cobalt-chelating affinity column [08ECS] 8. a decrease in the concentration of 2-ME [08ECS] 9. a backside alignment technique [09ECS] 10. a coupling partner [03ECS] 11. a moderate conversion [03ECS] 12. an 11-step conversion [04ECS] 13. a highly dense-functionalized cyclopentane core [04ECS] 14. An Au electrode [05ECS] 15. an increase in high order oligomers [05ECS] 16. The further elimination of a proton [06ECS] 17. ~~the~~ a large deviation [03ECS] 30.16%
Plural	2. the largest binding energies [08ECS] 3. the syntheses of bioactive molecules [04ECS] 4. the desired photoproducts [03ECS] 5. the anti-products decreased [03ECS] 9.07%	3. higher concentrations [08ECS] 4. freeze-and-thaw cycles [08ECS] 5. low spatial frequency regions [01ECS] 6. complicated structures [03ECS] 7. Highly strained cyclobutane skeletons [03ECS] 8. ~~the~~ corresponding acyl azides [04ECS] 9. ~~the~~ chemists [04ECS] 16.55%
Mass	4. the modulation of protein functions [08ECS] 5. the chlorophyll in *Chlorella* cells [09ECS] 6. the flexibility [11ECS] 7.26%	1. The amount of ~~the~~ dimer produced by the treatment with ethanol [05ECS] 2. ~~an~~ ultimate sensitivity to UV light [06ECS] 3. Most studies on ~~the~~ directional anisotropy [07ECS] 4. evidence~~s~~ of the direct observation of the MCh effects in the meta-molecule [07ECS] 9.30%

Note. Column and row headings refer to the referent *after* error reconstruction.

The samples of excerpts presented in **Table 4.14** are weighted to illustrate their relative frequencies. This illustration demonstrates that article and plural -*s* errors occur most frequently with singular referents, followed by plural and mass referents. Furthermore, a trend towards errors in non-specific referents appears. The following three sections elaborate the errors with singular, plural, and mass referents, commencing with the most prominent, singular referents.

4.3.1.1 Singular referent errors with articles and plural -s.

Errors with singular referents account for a combined 57.82% of the errors with articles and plural *-s* error pattern (**Table 4.14**). Generally in English, singular nouns occur three times more frequently than plural nouns (Taylor, 2012, pp. 154-155). Moreover, English has many more count nouns than mass nouns (Fieder, Nickels, & Biedermann, 2014, p. 14). Therefore, the error frequencies presented in **Table 4.14** may simply be a reflection of the naturally occurring frequencies of nouns in English, with singular nouns the majority.

Within the DETERMINATION system (i.e., specific and non-specific referents), errors with singular referents are distributed evenly between specific and non-specific referents (i.e., 27.66% and 30.16%, respectively), with just a slight trend towards errors with non-specific referents. However, when considering the naturally occurring frequencies of specific and non-specific referents in English, the trend towards errors with non-specific referents is more pronounced. In large corpora representing general English (e.g., the Brown Corpus and the British National Corpus), the definite article occurs around 2.6 times more frequently than the indefinite article.[51] In contrast to this trend in larger corpora and the greater potential for error in the more frequently occurring definite article, errors occur more frequently with the indefinite article in the NGS. This finding suggests the indefinite article presents greater difficulties for Japanese scientists than the definite article and should be considered carefully in pedagogy aimed at addressing this error pattern.

The errors with singular referents, both specific and non-specific, occur in the following three ways: the article is (1) omitted, (2) mis-selected or (3) inappropriately added (i.e., omission, selection, and addition errors, respectively). Within the errors with singular referents, omission errors are the majority (85.16%), followed by selection, then addition errors, as outlined in **Table 4.15**.

---

[51] In the one-million word Brown Corpus of written American English, the definite article occurs 69,971 times, while the indefinite article occurs 23,225 times (Francis & Kucera, 1961). Similarly, in the British National Corpus the definite article occurs 5,973,437 times, compared with the indefinite article occurring 2,136,923 times (BNC, 2007). In both corpora, this is a 2.6 times increase in the definite article.

Table 4.15 Error types within article and plural *-s* errors with singular referents

Error type	Total frequency in NGS	Proportion (%)	Average frequency per text	Average frequency per 1,000 words
Omission	218	85.16	16.77	4.71
Selection	35	13.67	2.69	0.76
Addition	3	1.17	0.23	0.06
Total	**256**	**100.00%**	**19.69**	**5.53**

These results indicate that the omission of articles with singular referents is a persistent error in the L2 writing of Japanese scientists, occurring around 4–5 times per 1,000 words. To illustrate the effects of these errors on reading flow and comprehension, Excerpts 41–44 demonstrate omission errors with specific and non-specific singular referents, before and after error reconstruction.

Omission of article errors with singular referents before error reconstruction

41. As shown in Fig. 1, the gap electrodes were fabricated on thermally oxidized Si chip. [ECS09]
42. To elucidate this mechanism, we investigated relationship between conductivity and thermoelectric properties of tpp-doped SWNT films at various dopant concentrations. [11ECS]

Omission of article errors with singular referents after error reconstruction

43. As shown in Fig. 1, the gap electrodes were fabricated on a thermally oxidized Si chip. [ECS09]
44. To elucidate this mechanism, we investigated the relationship between conductivity and thermoelectric properties of tpp-doped SWNT films at various dopant concentrations. [11ECS]

While the impact of each article omission error may not be critical to intelligibility, the prevalence and subsequent accumulation of these errors over a whole text distracts from the content and reflects negatively on the accuracy of the writing (and subsequently, the research). Given that these omission errors with singular referents occur on average 16-17 times per text (**Table 4.15**), the frequent disruptions to the reading flow and comprehension that these errors cause may be consequential.

The negative impact of the omission errors is further compounded by the selection errors. The selection errors in singular referents account for 13.67% of the errors with singular referents (**Table 4.15**) and combine with the omission errors to interrupt the reading flow and comprehension, as demonstrated in Excerpts 45-48, before and after error reconstruction.

Selection of article errors with singular referents before error reconstruction

45. Power factors are the useful measure of thermoelectric properties. [11ECS]
46. Since the splitting of narrow f-f emission lines is highly sensitive to a coordination geometry around Eu^{III}, the Stark splitting pattern of the Eu^{III} complexes is a good optical fingerprint for identification of their coordination structures in solution. [10ECS]

Selection of article errors with singular referents after error reconstruction

47. Power factors are ~~the~~ a useful measure of thermoelectric properties. [11ECS]
48. Since the splitting of narrow f-f emission lines is highly sensitive to the ~~a~~ the coordination geometry around Eu^{III}, the Stark splitting pattern of the Eu^{III} complexes is a good optical fingerprint for identification of their coordination structures in solution. [10ECS]

The third error type within singular referents, addition errors, occurs only three times throughout the NGS. Those three occurrences are presented in Excerpts 49-51.

Addition of article errors with singular referents

49. For the purpose of clarity, Figure 1c, 1g, 1k, and 1p show ~~the~~ either one of the crystal structures. [10ECS]
50. The weak correlation between the structural dissymmetry and ~~the~~ luminescence dissymmetry factor implies a weak contribution of the static-

coupling mechanism to the circularly polarized dissymmetry of the lanthanide(III) f-f transition. [10ECS]

51. The microplasma sources, which supply reactive species, are located on the back of ~~the~~ each microwell. [09ECS]

Interestingly, all three occurrences involve the addition of the definite article, rather than the indefinite article; however, the very low frequency and irregularity of these addition errors with singular referents suggests that they are more likely to be random mistakes than systematic errors.

4.3.1.2 Plural referent errors with articles and plural -s.

After the singular referent errors, article and plural -*s* errors with plural referents constitute the second most frequently occurring errors in this error pattern, accounting for 25.62% of the articles and plural -*s* errors (**Table 4.14**). In plural referent errors, non-specific referent errors are almost twice as frequent as specific referent errors in the data (i.e., 16.55% and 9.07%, respectively; **Table 4.14**). As previously discussed (**Section 4.3.1.1**), this trend is more pronounced when considering that non-specific referents actually occur less frequently than specific referents in English.

Similar to singular referent errors, plural referent errors are dominated by omission errors, as illustrated in **Table 4.16**.

Table 4.16 Error types within article and plural -*s* errors with plural referents

Error type	Frequency	Proportion (%)	Average frequency per text	Average frequency per 1,000 words
Omission	87	76.99	6.69	1.88
Selection	3	2.65	0.23	0.06
Addition	23	20.35	1.77	0.50
Total	**113**	**100.00%**	**8.69**	**2.44**

Since errors occur almost twice as frequently with non-specific referents compared with specific referents (i.e., 16.55% and 9.07%, respectively; **Table 4.14**), the tendency of the Japanese scientists in this study is to omit plural -*s*, as demonstrated in Excerpt 52.

Omission of plural -s error in non-specific plural referent

52. Different effects from surface treatment [01ECS]

In contrast, omission errors in specific plural referents may occur with either the definite article or the plural -*s* as demonstrated in Excerpt 53 and 54, respectively.

Omission of definite article and plural -s errors in specific plural referents

53. an apparent weakening of the bright spots in the FOM image of *h*-PF8T2 after the *r*-CP irradiation [02ECS]
54. the mechanisms underlying plasma medicine [09ECS]

Combined together, these omission errors with specific and non-specific referents occur on average two times per 1,000 words in the COJSD.

After the omission errors, addition errors account for 20.35% of the errors in plural referents (**Table 4.16**). In all 23 of these addition errors in the data, the errors occur with non-specific plural referents, as illustrated in Excerpts 55 and 56.

Addition of article errors in plural referents

55. all kinds of ~~the~~ substrates used in this work [01ECS]

56. ~~a~~ greater chiroptical amplitudes [02ECS]

Notably, the definite article is erroneously added 19 times, while the indefinite article is added only 4 times, further highlighting the Japanese scientists' infrequent use of the indefinite article.

The third type of plural referents error, selection, is rare, only occurring three times in the NGS (**Table 4.16**). These three selection errors are presented in Excerpts 57–59.

Selection of article and plural -s errors in plural referents

57. ~~a~~ reducing reagents and/or a stabilizer [08ECS]
58. the function of ~~a~~ proteins composed of several protomer units such as caspase-3 [08ECS]
59. ~~the~~ previous stud~~y~~ies studies [13ECS]

In all three occurrences, the error is the mis-selection of a singular referent for a plural one. Given that this error occurs only three times in the total 442 errors with articles and plural -*s*, and given two of those occurrences are within the one text (i.e., [08ECS]), this error is not considered an endemic error pattern for Japanese scientists.

Compared with the singular referent errors, errors with plural referents tend to produce different effects on reading flow and comprehension. To illustrate this point, Excerpt 60 demonstrates the most frequent, omission error with a non-specific referent, before and after error reconstruction.

Errors with articles and plural -s in plural referents before error reconstruction

60. Highly strained cyclobutane skeleton shows important potential as the precursor of not only natural products but also various unique compounds, and much effort has been dedicated to this reaction in various fields. [03ECS]

Errors with articles and plural -s in plural referents after error reconstruction

61. Highly strained cyclobutane skeletons show~~s~~ important potential as the precursor of not only natural products but also various unique compounds,

> and much effort has been dedicated to this reaction in various fields. [03ECS]

The impact of these errors on reading comprehension and the accuracy of the content cannot be considered minimal. With an average frequency of 8–9 times per text, these errors disrupt the reading flow and potentially detract from the quality of the Japanese scientists' work, especially when combined with the singular referent errors discussed above, and the mass referent errors examined in the following section.

4.3.1.3 Mass referent errors with articles and plural -s.

The final descriptor in the investigation of errors with articles and plural -*s* error pattern, errors with mass referents, accounts for 16.56% of the errors with articles and plural -*s* (**Table 4.14**). As previously discussed (**Section 4.3.1.1**), this relatively low frequency of errors with mass referents reflects the relatively low frequency of mass nouns in English compared with singular and plural nouns. Furthermore, this result may be an artefact of the lack of indefinite article and plural -*s* with mass nouns. In other words, given there is no indefinite article or plural -*s* with mass referents, there is less potential for the most common error type, omission, to occur. Indeed, **Table 4.17** shows that, compared with omission errors, addition errors account for twice the number of errors in mass referents.

Table 4.17 Error types within article and plural -*s* errors with mass referents

Error type	Frequency	Proportion (%)	Average frequency per text	Average frequency per 1,000 words
Omission	23	31.51	1.77	0.50
Selection	4	5.48	0.31	0.09
Addition	46	63.01	3.54	0.99
Total	**73**	**100.00%**	**5.62**	**1.58**

Errors with mass referents occur on average 1.58 times for every 1,000 words in the COJSD. Excerpts 62 and 63 demonstrate the most prominent error type with mass referents, addition, highlighting the nuanced choice in the COUNTABILITY system (i.e., the choice to construe a noun as a count or mass noun).

Addition errors with articles and plural -s in mass referents

62. chemical biology~~ies~~, pharmaceuticals, medicinal and agricultural areas [04ECS]
63. ~~an~~ evidence of an optofluidic effect [02ECS]

While it may be possible to construe the mass nouns in Excerpts 62 and 63 as count nouns (e.g., *biologies*, *an evidence*), that choice is atypical[52] and rather than enriching the meaning in context, functions only as a disruption to reading flow and comprehension. Notably, nouns that can be construed as count or mass (e.g., *difficulty*, *influence, temperature, condition,* etc.) are prevalent in the COJSD, but in these cases the tendency of the Japanese scientists in this study is to omit the definite article, as illustrated in Excerpts 64 and 65, rather than mis-construe them as count nouns.

Omission of definite article errors in mass referents

64. the difficulty of reaction control [04EST_43]

[52] In the British National Corpus, *biology* occurs 1,074 times, while *biologies* occurs only once. Similarly, *evidence* occurs 20,995 times while *an evidence* occurs only 11 times (BNC, 2007).

65. the influence of the surface treatments correctly [01EST_22]

The second most common mass referent error, omission errors, account for 31.51% of the mass referent errors (**Table 4.17**). Notably, the omission of the definite article, shown in Excerpts 64 and 65, is the only form of omission that may occur with mass referents, as there is no article in non-specific mass referents. The omission of the definite article with specific mass referents occurs in 9 of the 13 texts in the COJSD for a total of 23 times (**Table 4.17**). Excerpts 66 and 67 illustrate the effects of both an omission error and the more prevalent plural -*s* addition error for mass referents, before and after error reconstruction.

Addition of plural -s error and omission of definite article error in mass referents before error reconstruction

66. This finding suggests that we should pay attentions to the effect of concomitant reagents in reaction media evaluating the abilities of small organic molecules in modulation of protein functions. [08ECS]

Addition of plural -s error and omission of definite article error in mass referents after error reconstruction

67. This finding suggests that we should pay attention~~s~~ to the effect of concomitant reagents in reaction media evaluating the abilities of small organic molecules in the modulation of protein functions. [08ECS]

The overall effect of these errors is once again not critical for intelligibility, but interrupts reading flow and comprehension, negatively impacting the level of sophistication and accuracy central to academic and scientific registers.

4.3.1.4 Summary of errors with articles and plural -s.

This investigation of errors with articles and plural -*s* indicates these errors comprise a major error pattern accounting for two in three of the errors in the NGS, and occurring on average 34 times per text, or 10 times per 1,000 words (**Table 4.12**). The prevalence of this error pattern can be expected given the absence of an article system in the participants' L1 (Butler, 2002, p. 453), the abstract nature of the DETERMINATION and COUNTABILITY systems (Barner, Inagaki, & Li, 2009), and the relatively high occurrence of articles in English (Berry, 1991). These errors were classified according to the four error types within the EA analytical framework (i.e., omission, addition, selection, ordering), with findings revealing omission errors as the predominant error type, followed by addition and selection errors (i.e., 74.21%, 16.29%, and 9.50%, respectively; **Table 4.13**). These error types were further analysed according to the DETERMINATION, COUNTABILITY and NUMBER systems. In other words, errors were further classified simultaneously according to whether they occur in specific or non-specific referents, and singular, plural or mass referents. Findings indicate that the distribution of errors in the DETERMINATION system tends towards errors with non-specific referents. Furthermore, in the COUNTABILITY and NUMBER systems, errors occur twice as frequently in singular referents than plural (i.e., 57.82%, 25.62%, respectively), and more than three times more frequently than in mass referents (i.e., 16.56%; **Table 4.14**). Therefore, the Japanese scientists in this study primarily omit articles in singular non-specific referents, as illustrated in Excerpt 68.

Predominant article error with indefinite singular referent

68. As shown in Fig. 1, the gap electrodes were fabricated on a thermally oxidized Si chip. [09ECS]

The type of error demonstrated in Excerpt 68 occurs on average 16–17 times per text in the COJSD, and while it may not be critical to intelligibility, the regular disruption to the reading flow and comprehension may negatively impact the accuracy of the writing and subsequent accuracy of the content. The integration of the three systems (i.e., DETERMINATION, COUNTABILITY, and NUMBER) and the errors occurring within those systems suggests the need for Japanese scientists to understand the whole grammatical system. In other words, given that the choices in each of the three systems are concurrent with and contingent on each other, an understanding of the whole grammatical system network is necessary to reduce any of the errors in this major error pattern. Needs-specific pedagogy aimed at improving accuracy in the L2 English of Japanese scientists must take account of this integration.

4.3.2 Errors with preposition *-of*

This section examines the second major error pattern to emerge from the analysis of errors in the NGS, namely errors involving the post-modifying -*of*. Throughout the NGS, errors with -*of* appear predominantly in the Post-modifiers of complex nominal groups, but also may involve pre-modification. In other words, although the error originates in the post-modifier, the reconstruction of the error may relocate it to the pre-modifier. **Table 4.18** outlines the frequency of the errors with preposition -*of* in relation to the total number of errors in the NGS, highlighting the prevalence of these errors throughout all 13 texts in the COJSD.

Table 4.18 Errors with preposition -*of*

Text	Errors in nominal groups	Errors with -*of* per text	Errors with -*of* per 1,000 words	Errors with -*of* as percentage of errors in nominal groups (%)
01ECS	72	18	5	25.00
02ECS	98	11	2	11.22
03ECS	79	3	1	3.80
04ECS	83	5	1	6.02
05ECS	12	4	1	33.33
06ECS	29	4	2	13.79
07ECS	22	3	1	13.64
08ECS	54	8	2	14.81
09ECS	54	9	4	16.67
10ECS	40	4	1	10.00
11ECS	34	3	1	8.82
12ECS	8	1	0	12.50
13ECS	69	5	3	7.25
NGS	**654**	**78**	**24**	**11.93%**
Average	**50**	**6**	**2**	

Note. NGS refers to the Nominal Group Subcorpus.

As illustrated in **Table 4.18**, the errors with -*of* pattern comprises around 12% of all the errors in the NGS, and occurs on average twice per 1,000 words in COJSD. While these errors are less frequent than those involving articles and plural -*s*, they nonetheless constitute the second most frequently occurring error pattern in the NGS, requiring a closer examination to identify the specific obstacles they present to Japanese scientists striving to produce sophisticated and accurate grammatical English for research and publication purposes.

As outlined in the Methodology (**Section 3.3.2.3**), the errors with -*of* pattern can be analysed according to the four error types in the EA analytical framework (i.e. omission, addition, selection, and ordering). Excerpts 69-72 illustrate each of the four error types involving the post-modifying -*of*.

Omission error with preposition -of

69. knowledge and understanding of an efficient generation and switching of optically active substances [02ECS]

Addition error with preposition -of

70. the incident X-ray angle ~~of x-ray~~ [01ECS]

Selection error with preposition -of

71. One possible reason ~~of~~ for the decrease in domain size [01ECS]

Ordering error with preposition -of

72. the refractive index of the cosolvents at 589 nm (n_D) ~~of the cosolvents~~ [02ECS]

Excerpt 69 demonstrates an omission error with -*of* in the Post-modifier, where the omission mis-construes the noun *understanding* as the verb-form, causing interruption to reading flow and comprehension. In other words, in the case of nominalisations such as *understanding*, the nominal form is morphologically identical to its process form (e.g., *understanding something is important* vs. *the understanding of something is important*) and is only distinguishable through pre- and post-modifying elements (i.e., a pre-modifying *the*- and post-modifying -*of*).[53] Therefore, in these circumstances, the inclusion of the post-modifying -*of* is particularly important to designate the nominal

[53] This particular type of nominalisation is known as *gerundive nominalisation* (Heyvaert, 2008; Lees, 1960).

grammatical function. Excerpt 70 contains an addition error in which -*of* is inappropriately used to embed the classifying meaning of x-ray in the Post-modifier (i.e., *angle of x-ray*) rather than in pre-modifying position (i.e., *x-ray angle*; detailed in **Section 4.3.2.3**). In the selection error in Excerpt 71, -*of* is mis-selected for *for* (i.e., *reason of the decrease* rather than *reason for the decrease*), and in Excerpt 72, the -*of* construction is mis-ordered (i.e., in its original error form, it is modifying the circumstance, *at 589,* rather than the Head of the nominal group, *index*), again interrupting the reading flow and comprehension.

A closer look at the frequencies of these errors involving -*of* reveals that addition and selection errors are the most common error types, as illustrated in **Table 4.19** (i.e., a combined 82.05% of all -*of* errors).

Table 4.19 Proportion of error types in errors with -*of* constructions

Text	Omission errors (%)	Addition errors (%)	Selection errors (%)	Ordering errors (%)
01ECS	5.56	22.22	72.22	0.00
02ECS	54.55	0.00	36.36	9.09
03ECS	33.33	33.33	33.33	0.00
04ECS	20.00	80.00	0.00	0.00
05ECS	0.00	100.00	0.00	0.00
06ECS	0.00	100.00	0.00	0.00
07ECS	33.33	0.00	66.67	0.00
08ECS	12.50	50.00	37.50	0.00
09ECS	11.11	66.67	22.22	0.00
10ECS	0.00	100.00	0.00	0.00
11ECS	0.00	0.00	66.67	33.33
12ECS	0.00	100.00	0.00	0.00
13ECS	0.00	60.00	40.00	0.00
Total	**15.38**	**44.87**	**37.18**	**2.56**

To further understand the various errors involving -*of*, the following sections examine each of the four error types in detail (i.e., omission, addition, selection, and ordering), beginning with the most frequently occurring addition errors.

4.3.2.1 Addition errors with preposition -of

Addition errors account for almost half of the errors with *-of* pattern (i.e., 44.87%; **Table 4.19**). Within these addition errors, there are two ways in which *-of* is inappropriately added, as illustrated in Excerpts 73 and 74.

Addition errors with -of

73. affection of the whole ~~of~~ protein structure [08ECS]
74. One 2x2 cm^2 plasma-on-chip ~~of 2x2 cm^2~~ [09ECS]

Excerpt 73 demonstrates a straightforward addition error simply involving the unnecessary addition of *-of*. This error occurs a total of six times in five of the COJSD texts. In contrast, the addition error illustrated in Excerpt 74 is more common throughout the corpus, occurring 27 times in 10 of the 13 texts, and involving the more complex addition of an embedded *of*-prepositional phrase in the Post-modifier. While the addition is in the Post-modifier, this error is classified as a Pre-modifier error because the reconstruction requires the re-configuration of the meaning into the Pre-modifier as an Epithet.[54] The prevalence and more complex nature of this error warrant further explanation.

In Excerpt 74, while the structuring of the meaning as a Qualifier in the Post-modifier is grammatically possible and acceptable (i.e., *one plasma-on-chip of 2x2 cm^2*), there is an important nuanced choice in the textual meaning to consider. Notably, the unmarked focus of a nominal group falls on the last element of the group, which is often the Head but moves to the Qualifier when post-modification is present (Halliday & Matthiessen, 2014, p. 387; Halliday & Webster, 2009, p. 231). Therefore, compared with packaging meaning in the Pre-modifier as a Classifier or Epithet, locating the meaning in the post-modifying Qualifier has the "greater potential for news" (Halliday & Matthiessen, 2014, p. 388). This focus on the last element of the group has important implications in the *given-new* textual flow of meaning in English.[55] In given-new prosodies, the Post-modifier conveys information as *new*, whereas the same information

[54] The convention throughout the thesis is to classify errors based on their location in the nominal group *after* error reconstruction.

[55] See Halliday and Webster (2009) for an overview of the given-new prosodies of English.

located in the Pre-modifier construes it as *given*. As Halliday and Matthiessen (2014) summarise:

> If a property or class relating to a thing has been introduced in the discourse preceding a nominal group, it is likely to be presented as an Epithet or Classifier of the Thing rather than as (part of) a Qualifier. (p. 388)

Excerpt 75 further illustrates this prosody.

Addition error with -of reconfigured as Classifier

75. Conversely, the g_{lum} values of 1-5^{iPr} are rather sensitive to solvent since their 1-5^{iPr} coordination geometries ~~of 1-5^{iPr}~~ rearrange upon dissolution in solvent. [10ECS]

In Excerpt 75, the embedded *-of* prepositional phrase in the Post-modifier (i.e., *of 1-5^{iPr}*) is more appropriately reconstructed as a pre-modifying Classifier (i.e., *1-5^{iPr} coordination geometries*) to become the 'given' information the second time it is used.

The prevalence of this addition error with *-of* prepositional phrases in the Post-modifier once again highlights the difficulties Japanese scientists face with complex nominal group modification, specifically with pre-modification. It further reveals the Japanese scientists' tendency to overuse *-of* constructions in the Post-modifier. This over-reliance on the preposition *-of* is further compounded by the selection errors described in the following section.

4.3.2.2 Selection errors with preposition -of

Selection errors with preposition *-of* are the second most frequent error type in this error pattern, comprising 37.18% of the total errors with *-of* pattern (**Table 4.19**). Excerpts 76 and 77 illustrate two selection errors with *-of*.

Selection errors with preposition -of

76. active components ~~of~~ for fully flexible thermoelectric modules [11ECS]
77. half of ~~as~~ that in a vacuum of 3.0 x 10^8 m sec^{-1} [02ECS]

In Excerpt 76, *of* is mis-selected for *for*, and in Excerpt 77, *as* is mis-selected for *of*. The tendency of the Japanese scientists in this study is to over-select *of*, as in Excerpt 76, rather than under-select it, as in Excerpt 77. Notably, *of* is mis-selected for another preposition (i.e., *of* is used when another preposition is more appropriate) in 26 of the 28 selection errors, while it is under-selected only twice. **Table 4.20** presents the frequencies of the various selection errors with -*of* in the COJSD, highlighting the mis-selection of *of* for *in* as the most frequent.

Table 4.20 Selection errors with preposition -*of*

Excerpt		Mis-selection	Frequency across the NGS
78.	the decrease ~~of~~ in the number of energy barriers at domain boundaries in the TFT channel [01ECS]	*of* for *in*	12
79.	The reason ~~of~~ for this investigation [08ECS]	*of* for *for*	9
80.	temperature dependence ~~of~~ with V_{th} instability [13ECS]	*of* for *with*	2
81.	Such a drastic influence ~~of~~ from the OTS treatment [01ECS]	*of* for *from*	1
82.	the auto-oxidation of the cysteine thiol group ~~of~~ on the enzymes [08ECS]	*of* for *on*	1
83.	the on-state V_{gs} ~~of~~ at 10 V [13ECS]	*of* for *at*	1
84.	half of ~~as~~ that in a vacuum of 3.0 x 10^8 m sec^{-1} [02ECS]	*as* for *of*	1
85.	a swapping of signs in chiroptical polarization ~~to~~ of the particles [02ECS]	*to* for *of*	1
TOTAL			28

Note. NGS refers to the Nominal Group Subcorpus.

In short, the results for selection errors further highlight the Japanese scientists' over-reliance on the preposition -*of*. Corroborating this finding, a word frequency analysis[56] reveals that *of* occurs 1942 times in the 46,263-word COJSD, indicating a coverage of 4.20%. Compared with the coverage in the Brown and BNC corpora (i.e., 3.09%), this is an increase of more than 1%.[57] Given the average length of a text in the COJSD is 3,559 words, a 1% increase in the use of *of* results in an additional 35 occurrences per text, thus greatly increasing the potential for error. The identification of this over-use could be a first step towards both consciousness-raising and instructional intervention to address this error pattern. In other words, simply highlighting this overuse to Japanese scientists may motivate and help them understand the need to explore alternative forms of complex nominal group modification, especially pre-modification.

4.3.2.3 Omission errors with preposition -of

Omission errors with -*of* occur in 7 of the 13 COJSD texts and comprise 15.38% of all -*of* errors (**Table 4.19**). There are straightforward cases of omission in which the -*of* preposition is simply omitted, such as that illustrated in Excerpt 86.

Straightforward -of omission error

86. the participation of Glul24 and/or Arg164 in the fixation of PAC-1 [08ECS]

However, the omission errors with -*of* are typically more complex. Excerpt 87 illustrates a more common -*of* omission error.

Complex -of omission error

87. the equilibrium among the conformation of various pro-chiral substrates~~'~~ ~~conformation~~ [03ECS]

In omitting -*of*, the error in Excerpt 87 inappropriately deploys the *various pro-chiral substrates* as the possessive of the Head *conformation* (i.e., *the various pro-chiral substrates' conformation*). While this possessive form is a grammatical possibility, with an understanding of the context, it is apparent that the focus of the *equilibrium* is in fact

56 Using AntConc 3.4.3 (Anthony, 2014)

57 In the Brown Corpus (written), *of* occurs 30,971 times per million words (i.e., coverage of 3.09%) (Francis & Kucera, 1961). Similarly, in the 100 million word spoken and written British National Corpus, *of* occurs 3,009,801 times (British National Corpus, 2007).

on *the conformation*, and not the *various prochiral substrates*. The complex nature of this error prompts further elaboration.

As discussed in **Section 4.3.2.1**, the "unmarked" focus of the nominal group falls on the final element of the group.[58] However, there is an important exception involving -*of*, known functionally as the Focus (Martin, Matthiessen & Painter, 2010, pp. 169-171; originally termed the Facet function by Halliday and Matthiessen, 2004, pp. 333-335). The Focus function, realised by an embedded nominal group with -*of* as Linker, marks the focus of meaning at the front of the nominal group, as illustrated in Excerpt 88.[59]

	Focus		Linker	Deictic	Classifier	Thing	
88.	the	conformation	of	various	pro-chiral	substrates	[03ECS]

The deployment of this linguistic resource has a distinct function in the textual flow of meaning in discourse. In other words, the positioning of elements in the nominal group can change the intended meaning. Without an understanding of the Focus function, Japanese scientists are more likely to misplace the focus of nominal groups. Therefore, pedagogy aimed at improving accuracy in nominal group modification for Japanese scientists must take account of this function. Given the related nature of the -*of* addition errors (**Section 4.3.2.1**), this pedagogy may contrast the Focus function with the given-new textual prosodies of nominal groups.

4.3.2.4 Ordering errors with preposition -of

Only two ordering errors with -*of* are identified in the NGS, making this the least frequent error type in the errors with -*of* pattern. Notably, the issue is once again with the choice in location of the qualifying meaning, as illustrated in Excerpt 89.

---

58 For example, the meaning in (a) and (b) is experientially the same: (a) *the room temperature samples*, and (b) *the samples at room temperature*. However textually, in (b) the post-modifying element *at room temperature* is likely to be carried forth as the "newsworthy" element in the given-new textual flow of the discourse.

59 Adapted from Martin (2013, p. 65), who lists the following as examples: *the front of stage*, *a collection of songs*, *a chapter of the book*, etc.

Ordering error with -of

89. the refractive index of the cosolvents at 589 nm (n_D) ~~of the cosolvents~~ [02ECS]

In Excerpt 89, the erroneous embedded -*of* preposition phrase (i.e., *of the cosolvents*) at the end of the group causes an interruption to reading flow and is thus more appropriately relocated directly following the Head, *index*. While this ordering error type is infrequent in the data, for pedagogical purposes it can be treated as part of the errors with -*of* pattern, and in particular as part of the larger issue of locating meaning within the nominal group to most appropriately maintain the given-new prosody, as discussed in **Section 4.3.2.1**.

4.3.2.5 Summary of errors with -of

Due to its common colligation with nouns and nominalisations (Hunston, 2002), the preposition -*of* is one of the most frequently occurring words in English, accounting for around 3% of all written and spoken discourse (BNC, 2007). As previously demonstrated with the errors with articles and plural -*s* (**Section 4.3.1**), this high occurrence affords greater potential for error. Errors with -*of* constitute the second most prevalent error pattern to emerge from the NGS, accounting for one in every ten errors in the NGS (**Table 4.18**). Errors with -*of* are primarily located in the Post-modifier of the nominal group, though they also involve pre-modification. Addition and selection errors are the most common error types occurring within this pattern, accounting for 82.05% of the errors (i.e., 44.87% and 37.18%, respectively; **Table 4.19**). These addition and selection errors highlight the participants' over-reliance on *of*; particularly, their tendency to employ embedded *of*-prepositional phrases in the Post-modifier when the meaning would be more appropriately configured as Epithets or Classifiers in the Pre-modifier.

The findings from this investigation suggest that pedagogy aimed at improving accuracy in the L2 English writing of Japanese scientists should consider this over-reliance on *of* and the various ways it manifests as errors. Instruction may highlight the subtle but important difference between placing meanings in the Pre- or Post-modifier; namely, the last part of the nominal group has the greater potential for news, but -*of* constructions can shift the focus of the nominal group to the front.

4.4 Chapter Summary

This chapter has outlined the analysis of errors in the 46,263-word Corpus of Japanese Scientific Discourse (COJSD). A total of 1,368 errors were recognised and reconstructed, corresponding with an average of 31 errors per 1,000 words, or one in every three sentences. Around half of the errors were found to occur within nominal groups (i.e., 598 errors) and these errors were examined closely as the focus of this investigation (i.e., the NGS). Analyses of the errors in nominal groups reveals that complex nominal groups, and the various forms of modification that are a feature of complex nominal groups, account for 99.83% of the errors in the NGS. In contrast, simple nominal groups, which have no modification, present no difficulties for the Japanese scientists in this study. The various errors within complex nominal groups were classified according to modification and location within the group (i.e., their location in the Pre-modifier, Head, or Post-modifier). Results indicate that errors in the Pre-modifier occur twice as frequently as errors in the Post-modifier, and four times more frequently than errors in the Head, highlighting the difficulties of nominal group pre-modification for Japanese scientists. From the analysis of errors in the modifying elements of complex nominal groups, two major error patterns emerged from the data: errors with articles and plural *-s*, and errors with preposition *-of*.

Errors with articles occur in the Deictic in both Pre- and Post-modifiers. This error pattern also extends to the Head to include errors with plural *-s*. Errors with articles and plural *-s* were classified according to the four error types: omission, addition, selection, and ordering. This analysis reveals that the Japanese scientists in this study primarily omit articles and plural *-s*. Further analysis employing the system network mapping the grammatical choices of articles and plural *-s* demonstrates that the Japanese scientists in the study make errors most frequently with singular referents, and in particular with non-specific singular referents.

This investigation of errors with articles and plural *-s* highlights the integrated nature of the choices in the underlying grammatical systems and the errors that occur within those systems (e.g., ~~a~~ high levels of circular polarization [10ECS]). In other words, the choice of definite or indefinite article is concurrent with and contingent on the choice in singular, plural, or mass. It is therefore necessary for needs-based pedagogy aimed at addressing these errors to account for this integration. For example, the system network presented in **Figure 3.5** in this study could be utilised to illustrate those grammatical choices to Japanese scientists, and samples from this study could be presented to illustrate the ways the errors in the systems unfold.

The second major error pattern that emerged from the NGS data was errors involving the preposition *-of*. Errors with *-of* are prevalent throughout the COJSD, occurring in all 13 texts and accounting for around 12% of the total errors in the NGS. Analysis of these errors according to the four error types (i.e., omission, addition, selection, and ordering) reveals addition and selection errors occur most frequently in the NGS. Overall, the Japanese scientists tend to over-rely on *-of* constructions to construe meaning in nominal groups, particularly through the use of embedded *-of* prepositional phrases to package qualifying meaning in the Post-modifier, when that meaning is more appropriately configured as an Epithet or Classifier in the Pre-modifier. The implications of these findings highlight the need to preserve the given-new prosodies in nominal groups to effectively focus on the appropriate nominal group element. Furthermore, the results indicate that the Focus function is an area requiring consideration (e.g. *the results of this study*, in comparison with *this study's results*). Needs-based pedagogy aimed at addressing the use of *-of* may equip Japanese scientists with a better understanding of nominal group pre-modification and the various linguistic resources available to more accurately and coherently package meaning in nominal groups.

The two major errors patterns combined (i.e., errors with articles and plural *-s*, and errors with preposition *-of*) account for 79.36% (**Table 4.12** and **Table 4.18**) of all errors in the NGS. Given that the NGS comprises 47.81% of the total number of errors in the COJSD, two in every five errors occurring in the COJSD have been elaborated in this investigation, revealing the most frequent and prevalent patterns requiring instructional intervention. Further investigation is necessary, however, to provide a fuller account of the obstacles Japanese scientists face when preparing manuscripts for publication.

Chapter 5: Conclusion

This study has provided an introductory examination of the second language (L2) English errors in the scientific writing of Japanese scientists, focussing on errors in nominal groups. Significantly, two major error patterns are identified: errors with articles and plural -*s*, and errors with preposition -*of*. Together these two error patterns account for two in every five errors in the 46,263-word corpus analysed. This concluding chapter summarises the major achievements of the study along with its limitations, and outlines recommendations for future research.

This study contributes to the wider understanding of L2 errors by providing an elaborated framework for Error Analysis (EA). This elaborated EA framework integrates the delicate grammatical descriptions of Systemic Functional Linguistics (SFL) and the empirical basis of Corpus Linguistics (CL) within the traditional EA framework of analysis. By integrating these approaches, the present investigation and future research into L2 errors benefits from the decades of theoretical and analytical advances made in the fields of SFL and CL since EA was developed in the 1960s-70s.

The elaborated EA framework resulted in the identification of two major error patterns in the Corpus of Japanese Scientific Discourse (COJSD): errors with articles and plural -*s*, and errors with preposition -*of*. These findings inform a more focused pedagogy supporting the needs of Japanese scientists striving to achieve the levels of accuracy demanded in scientific publications. Moreover, the excerpts and error samples provided in this investigation serve as field-specific and data-driven resources to inform that specialised pedagogy.

While this study has established an elaborated EA framework and identified two major L2 error patterns in the writing of Japanese scientists, due to its limited scope, the current thesis serves as an introductory investigation of errors in the L2 English of Japanese scientists. Among the limitations of this study is the sample size of thirteen texts. This sample size limits potential to infer the results beyond the study's participants towards the wider population of Japanese scientists. Similarly, the sample of Japanese scientists participating in this study is limited to established-career materials scientists. Consequently, the findings cannot be corroborated with errors in the L2 writing of other Japanese scientists, especially early-career scientists such as PhD students who are typically writing their first papers in English and may be prone to different error patterns.

Further among the limitations of this study is the restricted focus on errors in nominal groups. While nominal groups are a central feature of scientific writing and a major source of the errors in this study, a full analysis of errors both within and beyond the clause remains; particularly, the remaining groups and phrase units (i.e., verbal groups, adverbial groups, conjunction groups, and prepositional phrases). Furthermore, for purposes of scope, this examination of errors in nominal groups is restricted to the two most prevalent error patterns, leaving additional patterns to investigate further. Finally, the analysis of errors is only as effective as the initial identification and classification. The examination of errors in this study is limited to one analyst. While this approach facilitates consistency, it subsequently limits the reliability of the findings. Due to these limitations, this investigation may best serve as a pilot study for a larger examination of L2 errors in the scientific writing of Japanese scientists.

Building from this introductory study, a complete understanding of L2 errors in the scientific writing of Japanese scientists requires further investigation. Specifically, in order to infer results to the wider population of Japanese scientists, future research will require a larger sample size approximating a normal population. Furthermore, in addition to established-career Japanese scientists, such a sample should include early-career scientists. The examination of a larger and more diverse range of participant texts may establish whether the error patterns identified in this study are consistent across and within the population. Subsequently, this increase in data requires more efficient analyses, particularly, corpus software tools for electronic annotation and automated analyses.

Additionally, future work investigating L2 errors in the scientific writing of Japanese scientists requires an examination of errors beyond nominal groups to include the whole clause. Such an examination includes not only the remaining group and phrase units (i.e., verbal groups, prepositional phrases, conjunction groups, adverbial groups), but also potentially functional elements within and above the clause (e.g., errors in Participants, Processes, Circumstances; and errors in Theme and Rheme). These analyses may also benefit from intelligibility studies measuring the impact of errors on comprehension, and obligatory occasions analyses to corroborate frequency data.

In conclusion, the present investigation of L2 errors has contributed significantly to the analysis of errors in Japanese Scientists' research writing through the development of an elaborated EA framework and the identification of the two most frequent error patterns in nominal groups: errors with articles and plural -*s*, and errors with preposition -*of*. This study serves as an initial step towards a larger end goal, which is the development and implementation of needs-based, field-specific, and data-driven pedagogy to help Japanese scientists achieve the levels of sophistication and accuracy required for successful international communication and publication in scientific English. Towards this end, this investigation of L2 errors in the scientific writing of Japanese scientists will be followed by a more in-depth and extensive examination of a larger population of Japanese scientists' research articles.

References

Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora. In S. Granger (Ed.), *Learner English on computer* (pp. pp. 132–141). London: Longman.

Abbott, G. (1980). Towards a more rigorous analysis of foreign language errors. *International Review of Applied Linguistics in Language Teaching, 18*(1-4), 121-134.

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. (2003). The international standing of the German language. In J. Maurais & M. Morris (Eds.), *Languages in a Globalising World* (pp. 231-249). Cambridge: Cambridge University Press.

Ammon, U. (2006). Language conflicts in the European Union. *International Journal of Applied Linguistics, 16*(3), 319-338.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo: Waseda University. Retrieved from http://www.laurenceanthony.net/

Austin, J. L. (1962). *How to do things with words*. Fair Lawn, NJ: Oxford University Press.

Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition, 111*(3), 329-344.

Basturkmen, H. (2006). *Ideas and options in English for specific purposes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Belcher, D. D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly, 40*(1), 133-156.

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

Berry, R. (1991). Re-articulating the articles. *ELT journal, 45*(3), 252-259.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275-311.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: Longman.

Blake, N. F., Lass, R., & Romaine, S. (1992). *The Cambridge history of the English language* (Vol. 6). Cambridge: Cambridge University Press.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*(1), 1-17.

BNC. (2007). The British National Corpus, version 3 (BNC XML edition). *Oxford University Computing Services on behalf of the BNC Consortium.* Retrieved from http://www.natcorp.ox.ac.uk/

Burt, M. K., & Kiparsky, C. (1972). *The gooficon: A repair manual for English*. Rowley, MA: Newbury House Publishers, Inc.

Butler, C. S. (1999). Some possible contributions of corpus linguistics to the functional-lexematic model. In M. J. Feu Guijarro & S. Molina Plaza (Eds.), *Estudios funcionales sobre léxico, sintaxis y traducción: Un homenaje a Leocadio Martín Mingorance (Functional studies on lexis, syntax and translation: A tribute to Leacadio Martin Mingorance* (pp. 19-38). Cuenca: Universidad de Castilla-La Mancha.

Butler, Y. G. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition, 24*(03), 451-480.

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

Cargill, M., & Burgess, S. (2008). Introduction to the special issue: English for research publication purposes. *Journal of English for Academic Purposes, 7*(2), 75-76.

Chen, L. L. (2006). The effect of the use of L1 in a multimedia tutorial on grammar learning: An error analysis of Taiwanese beginning EFL learners' English essays. *Asian EFL Journal, 8*(2), 76-110.

Cook, V. J. (1993). *Linguistics and second language acquisition*. London: Macmillan.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching, 5*(1-4), 161-170.

Corder, S. P. (1971). Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics in Language Teaching, 9*(2), 147-160.

Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Crompton, P. (2005). 'Where','In Which', and 'In That': A Corpus-Based Approach to Error Analysis. *RELC Journal, 36*(2), 157-176.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

Curry, M. J., & Lillis, T. (2004). Multilingual scholars and the imperative to publish in English: Negotiating interests, demands, and rewards. *TESOL Quarterly, 38*(4), 663-688.

Dotti, F., & O'Donnell, M. (2014). *Addressing article errors in Spanish learners of English using a learner corpus*. Paper presented at the 2014 TSLL Conference, Iowa State University, September 2014. http://www.uam.es/proyectosinv/treacle/Publications/Dotti-ODonnell-ArticlePaper.pdf

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York, NY: Oxford University Press.

Eggins, S. (2004). *Introduction to systemic functional linguistics* (2nd ed.). London: Continuum.

Evison, J. (2010). What are the basics of analysing a corpus? In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 122-135). London: Routledge.

Falhasiri, M., Tavakoli, M., Hasiri, F., & Mohammadzadeh, A. (2011). The effectiveness of explicit and implicit corrective feedback on interlingual and intralingual errors: A case of error analysis of students' compositions. *English Language Teaching, 4*(3), 251-264.

Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research, 38*(3), 247-273.

Ferguson, G. (2007). The global spread of English, scientific communication and ESP: Questions of equity, access and domain loss. *Ibérica, 13*, 7-38.

Fieder, N., Nickels, L., & Biedermann, B. (2014). Representation and processing of mass and count nouns: a review. *Frontiers in Psychology, 5*.

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

Flowerdew, J. (2013a). English for research publication purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 301-321). Chichester: John Wiley & Sons, Ltd.

Flowerdew, J. (2013b). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching*, 1-13.

Francis, W. N., & Kucera, H. (1961). Brown Corpus. Retrieved from https://the.sketchengine.co.uk/open/

Glover, A., Johnson, E., Strevens, P., & Weeks, F. (1988). *SeaSpeak training manual*. Oxford: Pergamon Press.

Gordin, M. D. (2015). *Scientific babel: How science was done before and after global English*. Chicago: University of Chicago Press.

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.

References

Granger, S. (2003). The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly, 37*(3), 538-546.

Gressang, J. E. (2010). *A frequency and error analysis of the use of determiners, the relationships between noun phrases, and the structure of discourse in English essays by native English writers and native Chinese, Taiwanese, and Korean learners of English as a Second language.* (Doctor of Philosophy), University of Iowa. Retrieved from http://ir.uiowa.edu/etd/507

Hagen, A. S. (2016). English spreading worldwide. Retrieved from http://ndla.no/en/node/9166

Halliday, M. A. K. (1985). *Spoken and written language*. Geelong, Victoria: Deakin University Press (republished by Oxford University Press 1989).

Halliday, M. A. K. (2004). *The language of science*. London: Continuum.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: The Falmer Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (revised 3rd ed.). London: Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. Oxford: Routledge.

Halliday, M. A. K., & Webster, J. J. (2009). *Studies in English language* (Vol. 7). London: Continuum.

Hamilton, C. E. (2015). The contribution of systemic functional grammar to the error analysis framework. *TESOL International Journal, 10*(1), 11-28.

Hammarberg, B. (1974). The insufficiency of error analysis. *International Review of Applied Linguistics in Language Teaching, 12*(1-4), 185-192.

Hasan, R. (2014). Towards a paradigmatic description of context: systems, metafunctions, and semantics. *Functional Linguistics, 1*(1), 1-54.

Heyvaert, L. (2008). On the constructional semantics of gerundive nominalizations. *Folia Linguistica, 42*(1-2), 39-82.

Hong, A. L., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). Collocations in Malaysian English Learners' Writing: A Corpus-based Error Analysis. *3L: The Southeast Asian Journal of English Language Studies, 17*(Special Issue), 31-44.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge: Cambridge University Press.

Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum International Publishing Group.

Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes, 9*(2), 116-127.

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

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. Harlow: Longman.

Johns, T., & King, P. (1991). Classroom concordancing. *English Language Research Journal, 4*.

Jordan, R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.

Kachru, B. B. (Ed.) (1992). *The other tongue: English across cultures* (2nd ed.). Urbana. IL: University of Illinois Press.

Keshavarz, M. H. (1999). *Contrastive analysis and error analysis*. Tehran: Rahnama.

Khodabandeh, F. (2007). Analysis of students' errors: The case of headlines. *The Asian ESP Journal, 3*(1), 6-21.

Kim, M. (2010). Translation error analysis: A systemic functional grammar approach. In C. Coffin, T. Lillis, & K. O'Halloran (Eds.), *Applied Linguistics Methods: A Reader* (pp. 84-94). London: Routledge.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Klages-Kubitzki, M. (1995). *Article usage in English: a computer-based self-teaching programme on the basis of a functional theory of reference*. Frankfurt am Main: Peter Lang.

Klein, W. (1998). The contribution of second language acquisition research. *Language Learning, 48*(4), 527-549.

Lees, R. B. (1960). *The grammar of English nominalisation* (Vol. 12). Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Levelt, W. J. M. (1978). Skill theory and language teaching. *Studies in Second Language Acquisition, 1*(1), 53-70.

Lillis, T. M., & Curry, M. J. (2010). *Academic writing in global context: The politics and practices of publishing in English*. London: Routledge.

Long, M. H. (Ed.) (2006). *Second language needs analysis*. Cambridge: Cambridge University Press.

Mabe, M. (2003). The growth and number of journals. *Serials, 16*(2), 191-197.

Martin, J. R. (2013). *Systemic functional grammar: A next step into the theory: Axial relations*. Beijing: Higher Education Press.

Martin, J. R. (2014). Evolving systemic functional linguistics: Beyond the clause. *Functional Linguistics, 1*(1), 1-24.

Martin, J. R., Matthiessen, C. M., & Painter, C. (2010). *Deploying functional grammar* (2nd ed.). Beijing: Commercial Press.

Martin, J. R., & Veel, R. (Eds.). (1998). *Reading science: Critical and functional perspectives on discourses of science*. London: Routledge.

Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System, 25*(2), 215-232.

Matthiessen, C. M., & Halliday, M. A. (2009). Systemic functional grammar: a first step into the theory. Beijing: Higher Education Press.

Matthiessen, C. M. I. M., Teruya, K., & Lam, M. (2010). *Key terms in systemic functional linguistics*. London: Continuum.

McCabe, A., & Gallagher, C. (2008). The role of the nominal group in undergraduate academic writing. In C. Jones & E. Ventola (Eds.), *New developments in the study of ideational meaning: From language to multimodality* (pp. 189-208). London: Equinox.

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Neeley, T. (2012). Global business speaks English: Why you need a language strategy now. *Harvard business review, 90,* 116-124.

O'Donnell, M. (2015). UAM CorpusTool 3.2 [Computer software]. Retrieved from http://www.wagsoft.com/CorpusTool/download.html

OECD. (2014). *Education at a Glance 2014: OECD Indicators*. Retrieved from http://www.oecd.org/education/eag.htm.

Okamura, A. (2006). How do Japanese researchers cope with language difficulties and succeed in scientific discourse in English?: Interviews with Japanese research article writers. *The Economic Journal of Takasaki City University of Economics, 48*(3), 61-78.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning, 63*(s1), 1-24.

Paltridge, B., & Starfield, S. (2013). *The handbook of English for specific purposes.* Chichester: John Wiley & Sons.

Park, K. (2012). Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics, 33*(4), 361-385.

Pennycook, A. (2014). *The cultural politics of English as an international language*. Oxford: Routledge.

Phillipson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.

Richards, J. C. (1971). A non-contrastive approach to error analysis. *English Language Teaching, 25*, 204-219.

Richards, J. C. (1974). *Error analysis: Perspectives on second language acquisition*. London: Longman.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research, 16*(2), 135-172.

Robertson, F. A. (1988). *Airspeak: Radiotelephony communication for pilots*. Hemel Hempstead: Prentice Hall.

Ronen, S., Gonçalves, B., Hu, K. Z., Vespignani, A., Pinker, S., & Hidalgo, C. A. (2014). Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences, 111*(52), E5616-E5622.

Rozycki, W., & Johnson, N. H. (2013). Non-canonical grammar in Best Paper award winners in engineering. *English for Specific Purposes, 32*(3), 157-169.

Schachter, J. (1974). An error in error analysis. *Language Learning, 24*(2), 205-214.

Schachter, J., & Celce-Murcia, M. (1977). Some reservations concerning error analysis. *TESOL Quarterly, 11*(4), 441-451.

Schleppegrell, M. J. (2002). Challenges of the science register for ESL students: Errors and meaning-making. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 119-142). Mahwah, N.J.: Lawrence Erlbaum Associates

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*, 209-231.

Selinker, L. (1992). *Rediscovering interlanguage*. London: Longman.

Strevens, P. (1969). *Two ways of looking at error-analysis*. Paper presented at the GAL Meeting, Stuttgart.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.

Thompson, G. (1996). *Introducing functional grammar*. London: Arnold.

Thomson-Reuters. (2009). Science Watch. Retrieved from http://archive.sciencewatch.com/dr/cou/2009/09decALL/

W3Techs. (2016) Usage of content languages for websites. Retrieved from http://w3techs.com/technologies/overview/content_language/all

Weber, G. (1997). The world's 10 most influential languages. *Language Today, 2*(3), 12-18.

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and cognition, 6*(02), 129-141.

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343-362.

Willey, I., & Tanimoto, K. (2012). “Convenience Editing” in action: Comparing English teachers’ and medical professionals’ revisions of a medical abstract. *English for Specific Purposes, 31*(4), 249-260.

Willey, I., & Tanimoto, K. (2013). “Convenience editors” as legitimate participants in the practice of scientific editing: An interview study. *Journal of English for Academic Purposes, 12*(1), 23-32.

WIPO. (2014). World intellectual property indicators: Total patent grants by country of origin: 2000-2014. Retrieved from http://www.wipo.int/ipstats/en/statistics/country_profile/profile.jsp?code=JP

Yamazaki, S. (1995). Refereeing system of 29 life science journals preferred by Japanese scientists. *Scientometrics, 33*(1), 123-129.

Appendices

Appendix A: Ethics Approval Letter

Office of the Deputy Vice-Chancellor
(Research)

Research Office
Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 4459
http://www.research.mq.edu.au/
ABN 90 952 801 237

MACQUARIE University
SYDNEY·AUSTRALIA

18 May 2015

Dr Cassi Liardet
Macquarie University
NSW 2109

Dear Dr Liardet,

Reference No: 5201500154

Title: *Japanese scientists' research articles: An error analysis*

Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Human Sciences & Humanities)) at its meeting on 27 March 2015 at which further information was requested to be reviewed by the Ethics Secretariat.

The requested information was received with correspondence on 8 and 14 May 2015.

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted at:

- Macquarie University

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007 – Updated March 2014) (the *National Statement*).

This letter constitutes ethical and scientific approval only.

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the *National Statement*, which is available at the following website:

http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.

3. All adverse events, including events which might affect the continued ethical and scientific acceptability of the project, must be reported to the HREC within 72 hours.

4. Proposed changes to the protocol must be submitted to the Committee for approval before implementation.

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 9850 4194 or by email ethics.secretariat@mq.edu.au

The HREC (Human Sciences and Humanities) Terms of Reference and Standard Operating Procedures are available from the Research Office website at:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics

The HREC (Human Sciences and Humanities) wishes you every success in your research.

Yours sincerely

Dr Karolyn White
Director, Research Ethics & Integrity,
Chair, Human Research Ethics Committee (Human Sciences and Humanities)

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007) and the *CPMP/ICH Note for Guidance on Good Clinical Practice*.

Appendix B: Participant Consent Form

DEPARTMENT OF LINGUISTICS
Faculty of Human Sciences

Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 6701
F: +61 (2) 9850 9199

Cassi.Liardet@mq.edu.au

CRICOS Provider No 00002J

MACQUARIE University

Chief Investigator: Dr. Cassi Liardét, Lecturer, Linguistics
Co-Investigator: Leigh McDowell, Student, Macquarie University

Participant Information and Consent Form

Name of Project: Japanese scientists' research articles: A needs analysis

Dear Japanese scientist,

You are invited to participate in a study of Japanese scientists' research articles. The purpose of the study is to collect English manuscripts to build a corpus titled the *Corpus of Japanese scientific discourse*. This corpus will be analysed for patterns of lexico-gramatical errors (e.g. the misuse of "consisted with" for "consisted of"). By understanding these error patters better, we hope to provide better education and support for Japanese scientists who need to publish their research in international journals.

The study is being conducted by Cassi Liardét (Chief Investigator) of the Department of Linguistics, Faculty of Human Sciences, Macquarie University, and Leigh McDowell, a Higher Degree Research student at Macquarie University. This project is designed for Leigh McDowell's master's degree research. Contact details of the investigators are as follows.

Dr. Cassi Liardét (Chief Investigator)
cassi.liardet@mq.edu.au
Bldg C5A, Rm. 537
Macquarie University, NSW 2109
Ph: +61 (0)2 9850 6704

Leigh McDowell (Co Investigator)
leigh.mcdowell@students.mq.edu.au
Department of Linguistics, Faculty of Human Sciences
Macquarie University, NSW 2109

If you decide to participate, you will be asked to send by email an electronic copy of the English manuscript of one of your recent publication drafts. You will also be asked to complete a brief questionnaire to provide a few details (e.g. your position at NAIST, name of the journal to which you submitted the manuscript, etc.). This is a one-off process which should not take any more than 20 minutes in total.

Local contact details if you have any questions:
ご質問があればお気軽にお問い合わせください。
物質創成科学研究科　研究科長室　〒630-0192 奈良県生駒市高山町 8916-5　電話 0743-72-6172
a-tozawa＠ad.naist.jp

DEPARTMENT OF LINGUISTICS
Faculty of Human Sciences

Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 6701
F: +61 (2) 9850 9199

Cassi.Liardet@mq.edu.au

CRICOS Provider No 00002J

MACQUARIE University

Any information or personal details gathered in the course of the study are confidential, except as required by law. No individual will be identified in any publication of the results. Only the chief investigator (Cassi Liardét) and her co investigator (Leigh McDowell) will have access to your name. If you choose to participate, you will be assigned a number to protect your anonymity both in the collected corpus and in any related publications. A short report of the results of the study will be made available to you via email on an annual basis.

Participation in this study is entirely voluntary: you are not obliged to participate and if you decide to participate, you are free to withdraw at any time without having to give a reason and without consequence. Note: If you decide not to participate in the current study, your manuscript will still be reviewed by the researchers but it will not be included in the corpus for any research analysis.

I, _________________________ *(participant's name)* have read and understand the information above and any questions I have asked have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw from further participation in the research at any time without consequence. I have been given a copy of this form to keep.

Participant's Name:____
(Block letters)

Participant's Signature:_____________________________Date:_____________________

Investigator's Name:____
(Block letters)

Investigator's Signature:________________________ ___Date:_____________________

The ethical aspects of this study have been approved by the Macquarie University Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Committee through the Director, Research Ethics (telephone (02) 9850 7854; email ethics@mq.edu.au). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

PARTICIPANT'S COPY

Local contact details if you have any questions:
ご質問があればお気軽にお問い合わせください。
物質創成科学研究科　研究科長室　〒630-0192 奈良県生駒市高山町 8916-5　電話 0743-72-6172
a-tozawa@ad.naist.jp

Appendix C: Participant Questionnaire and Responses

Q1. I need to publish my research in international journals in English.

#	Answer		Response	%
1	Strongly Agree		11	100%
2	Agree		0	0%
3	Neither Agree nor Disagree		0	0%
4	Disagree		0	0%
5	Strongly Disagree		0	0%
	Total		11	100%

Q2. I need to publish my research in Japanese journals in Japanese.

#	Answer		Response	%
1	Strongly Agree		0	0%
2	Agree		7	64%
3	Neither Agree nor Disagree		3	27%
4	Disagree		1	9%
5	Strongly Disagree		0	0%
	Total		11	100%

Q3. How often do you write research articles in English for publication?

#	Answer		Response	%
1	More than 2-3 Times a Year		4	36%
2	2-3 Times a Year		5	45%
3	Once a Year		2	18%
4	Less than Once a Year		0	0%
5	Never		0	0%
	Total		11	100%

Q4. How often do you write research articles in Japanese for publication?

#	Answer		Response	%
1	More than 2-3 Times a Year		0	0%
2	2-3 Times a Year		1	9%
3	Once a Year		5	45%
4	Less than Once a Year		5	45%
5	Never		0	0%
	Total		11	100%

Q5. Writing research articles in English is difficult for me.

#	Answer		Response	%
1	Strongly Agree		2	18%
2	Agree		9	82%
3	Neither Agree nor Disagree		0	0%
4	Disagree		0	0%
5	Strongly Disagree		0	0%
	Total		11	100%

Q6. My English manuscripts need proofreading before submission to journals.

#	Answer		Response	%
1	Strongly Agree		8	73%
2	Agree		3	27%
3	Neither Agree nor Disagree		0	0%
4	Disagree		0	0%
5	Strongly Disagree		0	0%
	Total		11	100%

Q7. What is your current position?

#	Answer		Response	%
1	Professor		0	0%
2	Associate Professor		3	27%
3	Assistant Profosser		8	73%
4	PhD Candidate		0	0%
5	Other		0	0%
	Total		11	100%

Q8. Which sections are the most difficult for you to write in English? You can choose more than one if needed.

#	Answer		Response	%
1	Title		3	27%
2	Abstract		0	0%
3	Introduction		9	82%
4	Method		1	9%
5	Results		0	0%
6	Discussion		5	45%
7	Conclusion		2	18%
8	References		0	0%
9	Supporting information		0	0%
10	Other		0	0%

Q9. Have you ever been told by a journal editor or reviewer to have your English manuscript proofread?

#	Answer		Response	%
1	Yes		6	55%
2	No		5	45%
	Total		11	100%

Appendix D: 1,000-word Excerpt

In organic and medicinal chemistry, ~~Organic~~ organic ~~Azides~~ azides ($R\text{-}N_3$) have been well studied ~~among organic chemistry and medicinal chemistry~~ since Peter Grieß synthesized the first organic azide, phenyl azide. After the development of Curtius rearrangement using hydrogen azide (HN_3), which produces isocyanates from ~~the~~ corresponding acyl azides, many organic azide-utilizing reactions were produced. Despite their explosive and toxic properties, organic azides are attractive not only industrially, but also agriculturally, and pharmaceutically. For these reasons, organic azide chemistry has developed extensively. Interestingly, natural products possessing azide groups have not been isolated to date, while those having 1,2,3-triazine and 1,2,3-triazene structures ~~were~~ have been found in nature.

Structurally, azides consist of three nitrogen atoms in linear form (not straight, but slightly bent; ~~a~~ calculated angle~~s~~ of N1-N2-N3 is 172.7° and ~~of~~ R-N1-N2N3 is 115.2° in methyl azide) (Scheme x). Organic azides show different chemical reactivities: the N1 atom can work as a nucleophile, and the N3 position nitrogen atom shows electrophilic reactivity (Scheme x). The specific efficiency of organic azides is its character as 1,3-dipolar and this provides [3+2] cycloadditions with unsaturated bonds to give triazolines, triazoles and tetrazoles. Recently, the cyclization reactions of organic azides with alkynes (Huisgen reaction) have been a focus~~ed~~ in the area of chemical biology ~~area~~ and extensive reports have been published (Meldal-Sharpless click reaction or Copper-Catalyzed Azide-Alkyne Cycloaddition—CuAAC).

Organic azides consist ~~with~~ of amines (N1) and the excellent leaving group diazonium cation (N2N3) producing nitrogen gas. Actually, the bond length of N1–N2 (1.237 Å) is computationally longer than that of N2–N3 (1.156 Å) in methyl azide. Thus, azides can easily evolve nitrogen gas in many reactions (Scheme x). Especially, heating conditions or photoirradiations produce nitrenes from organic azides, which are highly reactive and give aziridinations and C-H aminations (eq. x).

In this review, we describe the more recent applications of organic azides in the synthesis of natural products. The reaction steps involving the use of this functional group as well as ~~the~~ a description of the methods of decomposition/reduction are listed ~~according to the~~ as follow~~sing order~~: (1) preparation of organic azides; (2) C-H insertion reaction by nitrenes; (3) Curtius rearrangement; (4) Schmidt reaction; (5) [3+n] cycloaddition reaction; (6) Staudinger/aza-Wittig reaction and (7) organic azides as masked amino functional group.

Before ~~starting~~ reviewing the more recent synthetic applications of organic azides, we show their general preparation methods. General procedures are summarized in equations 4–10 in Scheme x. To prepare organic azides, use of volatile, toxic and highly explosive hydrogen azide should ~~strongly~~ be completely avoided. S_N2 azidation of alkyl halides and acyl halides using nucleophilic azides, mostly sodium azide (NaN_3), is the most general method (Scheme x). For transformation of alcohols and carboxylic acids, Shioiri reagent (DPPA—diphenylphosphoryl azide) is often used. To replace hydroxy groups, Mitsunobu reaction conditions (PPh_3 and azodicarboxylates DEAD/DIAD) can introduce azide groups with DPPA or zinc azide, and later direct conversion without Mitsunobu conditions. Recently, Kitamura and co-workers demonstrated that the safer and more stable azido compound ADMP (2-azido-1,3-dimethylimidazolinium hexafluorophosphate) was a good azidation reagent. Azides can be prepared from primary amines by way of diazoniumion formation followed by nucleophilic azidations. Aryl azides are usually synthesized with these procedures. From primary amines, diazotransfer reaction can also deliver organic azides. However, use of unstable and explosive trifluoromethanesulfonyl azide (TfN_3) was a problem. Recently, safer diazotransfer reagents, ADMP and Goddart-Borger reagent (1), ~~were~~ have appeared and can be used instead of TfN_3.

1,4-Addition of azides ~~are~~ is problematic because a formal [3+2] reaction could occur as a side reaction. Miller et al. reported successful reaction conditions to perform preferential 1,4-addition of azides to α,β-unsaturated carbonyl compounds, which was recently demonstrated in a model study in the total synthesis of cortistatin A 4 and J 5 by Yamashita and Hirama's group.

1,4-Azidation ~~was~~ has also been reported in the total syntheses of marine bisindole alkaloids hamacanthin A (11), B (12) and the antipodode of *cis*-dihydrohamacanthin B (13) by

Kawasaki et al. (Scheme x). Treatment of 2-methoxyindoline derivative 6 with methanesulfonic acid generated eniminium intermediate 7, and then 1,4-azidation ~~was~~ proceeded with $TMSN_3$ to give a desired azide compound 9 in 56% along with its epimer 8. The azide group in 9 was reduced by the Staudinger reaction followed by protection to afford 10, which was converted to natural products 11, 12 and antipode 13.

Not only nucleophilic azidation, but also the direct introduction of azides by electrophilic azidations of carbanions have been reported, which are performed with sulfonyl azides, usually $TrisN_3$ (2,4,6-trimethylbenzenesulfonylazide). In the reaction with enolates, the proton source for the quenching reaction is critical and Evans and co-workers revealed that acetic acid was the best. Kozmin et al. reported total syntheses and biological activities of streptolydigin 17 and its analogues using this stereoselective electrophilic azidation reaction (Scheme x).

Lovely and co-workers reported the total synthesis of some components of the kealiinine family (Scheme x). In their synthesis, the 2-aminoimidazole moieties were successfully constructed in the late stages of syntheses by way of lithiation-electrophilic azidation of the corresponding imidazoles 18–20. The following hydrogenolysis gave kealiinines A 24, B 25 and C 26. His group also achieved concise total synthesis of related naamine G 27 and naamidine H 28 with the same strategy.

Very recently, Fukuyama and co-workers reported total synthesis of pentacyclic alkaloid lyconadin A 34 and B 35 (Scheme x). After some investigation~~s~~ in modified syntheses, they found that nucleophilic azidation was the best way to construct the enone 33, which was essential for pyridone core synthesis. Lithiation of vinyl bromide 29 followed by trapping with Tris-N_3–acetic acid quenching gave vinyl azide 30. Transformation of 2-methylvinyl azide moiety of 30 into enone was performed by heating in the presence of acid. The protonation of the azide triggered nitrogen evolution (Schmidt reaction) to give unsaturated imine 32. The following hydrolysis afforded desired enone 33.

In 1992, Magnus et al. discovered allylic azidation of triisopropylsilyl enol ethers with $TMSN_3$ in the presence of hypervalent iodines. Recently, White et al. utilized this reaction to introduce amino groups in the total synthesis of huperzine A 40 (Scheme x).

Figure D 1,000-word excerpt illustrating error recognition and reconstruction

Note. Excerpt taken from Introduction section of [04ECS].

Appendix E: Nominal Group Denotation and Quantification

[Met-heme coordination] contributes to [the stability *of the structure*] and [the ability *of electron transfer in cyt c family proteins*]. Although [the optical absorption spectra] and [redox potentials] were similar between [monomeric and dimeric WT PA cyt c_{551}], [heme-ligating His and Met] originated from [different protomers *in the dimer*], similar to [the case *of dimeric HT cyt* c_{55}]. In [the case *of dimeric horse cyt c*], [Met-heme coordination] was perturbed and [a hydroxide ion] was coordinated to [the heme iron]. [The difference *in the heme coordination structure between dimeric PA cyt* c_{551} *and dimeric horse cyt c*] may be due to [the differences *in the stability of the Met-heme coordination bond*] and [the rigidity *of the loop containing the heme-ligating Met*]. According to [DSC measurements], [ΔH *for the dissociation of dimeric horse cyt c to monomers*] exhibited [a large, negative value], whereas [the ΔH values *for the dissociation of dimeric PA cyt* c_{551} *and dimeric HT cyt* c_{552}] were [~0] and [+14 kcal/mol], respectively. [These results] show that [the coordination *of Met to the heme*] contributes to [stabilization *of the dimer*] enthalpically.

Note. This excerpt is one paragraph taken from the Discussion section of [05EST]. Nominal groups are denoted in square brackets highlighted in red with Postmodifiers in italics. Total words: 182; Words in nominal groups: 148 (81%); Total nominal groups: 25 (⇒ 137 nominal groups per 1,000 words).

Table E Approximate number of nominal groups per 1,000 words and per text in the Corpus of Japanese Scientific Discourse (COJSD)

Text	Approximate nominal groups per 1,000 words	Approximate nominal groups per text
01ECS	157	562
02ECS	118	627
03ECS	160	495
04ECS	183	911
05ECS	154	731
06ECS	158	300
07ECS	150	380
08ECS	161	858
09ECS	192	831
10ECS	151	384
11ECS	161	278
12ECS	173	747
13ECS	176	776
Total	**2,094**	**7,879**
Average	**161**	**606**

Note. The approximate number of nominal groups per 1,000 words (i.e., second column) is based on the manual analysis of 1,000 words samples from the Discussion and Conclusion sections of each text in the COJSD. The approximate number of nominal groups per text (i.e., third column) is an extrapolation from the 1,000 word samples.

Appendix F: Thesis Word Count

Table F Words per chapter

	Words
Introduction	954
Literature Review	5,600
Method	3,560
Findings	9,869
Conclusion	870
Total	**20,853**

Note. Word counts include tables and figures. Footnotes are not included.