4 Engaging and Empowering Scientific Writers in Different Disciplines

John Blake
University of Aizu

William R. Holden III
Japan Advanced Institute of Science and Technology

Abstract: Postgraduate students of various scientific disciplines are often required to write research articles in English. Writing for publication is an onerous task, especially when English is an additional language. This chapter describes how scientific writers from three disciplines (information, materials, and knowledge science) are engaged and empowered at a small national research institute in Japan. Based on a comprehensive needs analysis, a three-pronged approach was adopted, comprising credit-bearing courses, face-to-face writing conferences, and online support. Corpus-informed materials were developed in-house for a suite of credit-bearing courses that form the mainstay of the formal curriculum. All courses are hybrid, blending onsite instruction with online learning activities. The courses are eclectic in approach, drawing on concepts such as flipped classrooms and activity-based learning. Face-to-face writing conferences are arranged for writers who submit drafts of articles or chapters for feedback. During these meetings, tutors provide discipline-specific constructive advice. In addition, writers are introduced to online resources and in-house tailor-made tools to assist their writing. Tools harnessing string searches, such as a corpus-based error detector, are used to enable writers to receive automated feedback on their work anytime.

Keywords: scientific writing, curriculum design, needs analysis, corpus-informed materials, disciplinary variation

This chapter describes how scientific writers are engaged and empowered at a small national research institute in Japan. All the writers are studying for research degrees in materials, information, or knowledge science. Their graduation is contingent on having research articles (RAs) accepted for publication in academic journals or conference proceedings. To get published,

DOI: https://doi.org/10.37514/INT-B.2021.1220.2.04

articles need to convince reviewers of the novelty, substance, and significance of the research as well as adhere to generic expectations in terms of language and rhetoric. While breaking these generic expectations may result in rejection, simply meeting them is no guarantee of acceptance, since an excellently written paper with poor science should still be rejected by reputable venues.

The entry barrier to novice scientists is especially high. Not only do they have to deal with the intrinsic difficulties of their research field, but they also need to navigate their way into their specific discourse community to learn its forms and values (Gee, 2007). Writing for publication is an onerous task *per se*. The difficulty is exacerbated when English is an additional language (Flowerdew, 2008), and particularly so when writers may not possess the requisite vocabulary (Evans & Morrison, 2011). The dominance of English as the language of science compels researchers who want to disseminate their research widely to publish in English (Englander, 2006; Lillis & Curry, 2010). Writers need to understand the dialogic nature of RAs (Fryer, 2013) and strategies for dealing with pit bull reviewers (Walbort, 2009) and rejection (Habibie & Hyland, 2019). The journey along the cline from the periphery to expert writers at the core of the community of practice is long and arduous (Lave & Wenger, 1991; Li, 2007; O'Neill, 2001). This is evidenced by reflective accounts of the transition (Casanave & Vandrick, 2003) and numerous case studies (e.g., Canagarajah, 2015). There are many risks, notably the high rejection rates, but there are also many rewards in writing for publication (Habibie & Hyland, 2019). The primary reward, however, for doctoral candidates is the ability to graduate.

Writing conventions vary greatly among disciplines (Lillis & Turner, 2001; Trowler & Becher, 2002), and this was found to be the case for the three disciplines of materials, information, and knowledge science; in fact, even within these disciplines there is notable variation. The disciplinary variation occurs at all levels from the research paradigm, discoursal conventions, and move structure through to lexical choice. This presents a challenge to teachers of writing who, due to timetabling limitations, need to teach classes offered to students from all three disciplines. An English for Specific Purposes (ESP) approach (Dudley-Evans & St. John, 1998) was adopted to address the diverging needs of different sets of writers.

This chapter first presents a case study by describing its learning and teaching context and constraints. It then details the interdisciplinary variation discovered among the three disciplines. The next section describes the approach, needs analysis, course design, and writing lab, after which examples of the corpus-informed materials and the online resources developed are provided. The final section reflects on the program design and shares some of the evaluations given by students.

Learning and Teaching Context

Researchers have long noted the shortcomings of English language education in Japan (Fujimoto-Adamson, 2006; Koike & Tanaka, 1995). In addition, the failure to provide discipline-specific ESP instruction was initially addressed approximately two decades ago and is also well documented (Orr, 1998). A recent study by Leigh McDowell and Cassi Liardét (2019) investigated the research writing processes of Japanese materials scientists drafting manuscripts for publication in English, and discovered that materials science researchers are five times as likely to publish in English as in Japanese. Yet, few graduate programs in Japanese universities in the fields of science, technology, engineering, and mathematics offer programs to prepare graduate students to write research articles.

This case study is set in the Japan Advanced Institute of Science and Technology (JAIST), a research institute offering postgraduate degrees. The cosmopolitan campus has one of the largest percentages of non-Japanese students among Japanese universities. According to its website (JAIST, 2019), approximately half of the student body are international students. Although most research laboratories operate in Japanese, laboratories with non-Japanese speaking professors or students tend to use English as the *lingua franca*. To fulfil the institution's graduation requirements, all doctoral candidates are required to have between one and three RAs accepted for publication. The specific requirements vary by laboratory. The complexity and sophistication of RAs (Chang & Kuo, 2011; Swales, 1990) provides a challenge to which students need to rise to graduate, and which is arguably the *raison d'être* for the establishment of the new English language program.

Interdisciplinary Variation

To help writers, it is necessary to understand the target genre. A detailed knowledge of the target genre and disciplinary variations enables writing teachers to provide accurate actionable advice, saving novice writers valuable time and increasing their likelihood of getting published in a timely manner. The ideal scenario is one where the teacher is a specialist in both English and the specific scientific discipline. A reasonable alternative is for an English language specialist to work closely with a discipline specialist. However, given various constraints in the introductory phase of the development of the program, securing cooperation was not an option.

Pedagogic advice provided by teachers and textbooks is often rather prescriptive, and may not reflect the descriptive reality. For example, textbooks

frequently advise scientists to adopt an introduction-method-results-discussion (IMRD) model for research abstracts, yet short RAs in some engineering and information science sub-disciplines make use of a two-move result-method model (Blake, 2015). The assumption that all research articles follow the same framework is flawed and leads to such over-generalizations. Advice based on descriptive analysis may more closely reflect the type of writing that is actually published rather than an idealized envisaged form of writing. Some scholars (Gee, 1996; Wingate et al., 2011) argue that discipline-specific literacy practices are best taught by discipline teachers. Laurence Anthony (2011) states that non-specialists can teach scientific writing using a process-orientated approach rather than a product-orientated approach. This is achieved by supplementing the generic teaching materials with data-driven learning using corpora that the students compile themselves. However, tutors who are also armed with disciplinary knowledge are better placed to offer actionable advice.

To gain a clear picture of discipline-specific expectations, rhetorical organization, and lexico-grammatical patterns in each discipline, a corpus-based approach to materials development was adopted for this project. Two corpora were created: a published RA corpus comprising approximately 1,000 articles and a draft RA corpus consisting of around 200 articles. The published RA corpus included RAs co-written by JAIST faculty-students from the university repository, conference proceedings of top-tier conferences, and research articles from Institute of Electrical and Electronics Engineers (IEEE) journals.

The corpus, which was divided into knowledge, materials, and information science subcorpora, was drawn upon extensively in the creation of discipline-specific course materials. Knowledge science is an emerging discipline resulting from the demands of a knowledge-based economy to address problems in collecting, synthesizing, coordinating, and creating knowledge (Nakamori, 2011). Materials science focuses on the structure, properties, and application of materials (Nasirpouri, 2017). Information science focuses on problems in the collection, storage, retrieval, and use of information stored as bits, or binary digits (Saracevic, 2009).

Through developing corpus-based materials and investigating the corpus using standard techniques and tools, such as keyword analysis, frequency analysis, and keyword-in-context concordance line analysis, the authors became familiar with the linguistic idiosyncrasies of each of the three domains. This corpus-based knowledge combined with the insights gained from working with authors in the writing lab and the classroom led to a clearer understanding of the commonalities and differences among the three disciplines. The authors identified eight areas in which disciplinary variation impacts research writing, which are described below.

Reasoning

Knowledge science relies on arguments based on samples and uses inductive reasoning to generalize to larger populations, which is reflected in the higher incidence of hedging when making claims. Information science relies on laws and mathematical proofs, while materials science relies on the constant nature of physical elements. Falsifiability, or the principle that a proposition or theory cannot be considered "scientific" unless it is possible to empirically show it to be false, is what fundamentally separates knowledge science from materials or information science. This explains why deductive reasoning tends to dominate in information and materials science.

Document Preparation

Unlike knowledge and materials scientists, information scientists tend to prepare research documents in plain text using LaTeX rather than formatted text in word processors, such as Microsoft Word. LaTeX documents look more like HTML code than writing until they are compiled into a pdf. This means that the use of typical methods to provide feedback on Microsoft Word documents, such as track changes and insert comment features, is not possible.

Text Recycling vs. Plagiarism

Text recycling, or "language re-use" (Flowerdew & Li, 2007), is frequently used in the method and result sections in materials science, with only minor changes being made to the variables and values. The extensive use of boilerplate text as evidenced in the corpus of published articles frees up writers from having to reinvent different ways to describe very similar methods. Much research in materials science uses standard methods and produces results which vary only in the numerical quantities and names of materials. Some sub-disciplines within information science appear to permit text recycling in the introduction section as well, based on the reuse of text in the corpus of published articles. This results in widespread lift-and-switch, or patch writing (Wette, 2010) in which writers copy and paste whole sections and only change the names of variables and values. Some publications (e.g., *IEEE Transactions on Nanotechnology* and *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*) specify that 30 percent of content should be new for an article to be considered for submission, which by implication means that 70 percent of the content need not be.

Citation Practices

Citation practices differ among the disciplines (Leydesdorff et al., 2016). Knowledge science articles tend to use name-date citation styles, such as APA, while the other two disciplines invariably use numerical citation systems. Examination of hundreds of manuscripts produced by the three departments has shown that quotations are used in knowledge science, but almost never used in materials science and information science, even when exact words are copied.

Page Layout

Knowledge science articles may be written in single-column templates, whereas the other two disciplines use double-column templates (e.g., IEEE and ACM templates). Figures inserted in the double-column format tend to be inserted in portrait rather than landscape. Template instructions sometimes forbid landscape figures but, even when permitted, placement of landscape figures involves more intricate coding in LaTeX, and novice writers in information science who have not fully mastered LaTeX tend to choose the easier portrait option.

Generic Conventions

Information science and materials science manuscripts tend to display greater adherence to predictable generic conventions in their structure, organization, and development, whereas manuscripts in knowledge science tend to exhibit greater variation. There are several possible explanations for this. Generic conventions are more firmly established in the first two fields than in the latter due to the fact that the former are well-known disciplines found at universities and research institutions worldwide, and because there exists an identifiable hierarchy of research publications in these fields and their subfields. Likewise, conventions in natural and applied science articles are less flexible than those in the social sciences. In addition, research strategies in the natural sciences are anchored in "concentrated knowledge clusters," whereas those in the social sciences are more frequently adapted to numerous "small isolated knowledge clusters" (Jaffe, 2014, p. 1).

Lexical Coherence

Similarly, the information science and materials science manuscripts examined display a more coherent set of lexis and rhetorical devices than do manuscripts from knowledge science. For example, the descriptions of the materials

and methods follow similar patterns and use similar lexical items in materials science articles, while knowledge science articles do not. The simplest explanation for the discrepancy is that the taxonomies of the first two fields are clearly delineated: every scholar involved in a particular field of research shares and employs a broad overarching vocabulary and a vocabulary specific to their specialization. Because knowledge science is an emerging interdisciplinary field that comprises areas as diverse as knowledge management, perceptual information processing, media technology, data mining, and ethnography, there is little in the way of a shared vocabulary, and manuscripts tend as a result to follow the lexical conventions of a particular sub-specialization.

Research Abstracts

Research abstracts vary greatly among the three disciplines. Materials science is particularly notable since graphical abstracts are frequently used in top-tier journals (Hendges & Florek, 2019; Lane et al., 2015). The move structure of research abstracts also varies greatly, with abstracts in knowledge science tending to have lengthy introductions and less emphasis on results. Abstracts in materials science and information science are more results focused. The organization of the typical rhetorical moves of introduction (I), purpose (P), method (M), results (R), and discussion (D) (Bhatia, 1993) vary dramatically. The default order of IMRD is rarely followed in information science. Some moves may be omitted, e.g., IMR. In a corpus study of scientific research abstracts, John Blake (2015) noted that pairs of moves may be repeated, especially MRMR in wireless communication, a sub-discipline of information science. In this move pattern, the first result tends to be the new algorithm and the second result the proof that the algorithm is superior to previous algorithms. Some moves may be inverted, such as RM in both materials science and information science.

Course Design

A writing in the disciplines model (Carter et al., 2007; Wingate, 2012) was considered but was not possible at the outset. In this model, discipline specialists and language specialists work together to help novice writers acquire the requisite skill set in a timely manner and, as noted above, this was not feasible in the introductory phase of the program. An alternative approach was therefore needed to meet the needs of the novice writers. We adopted an ESP-driven eclectic approach by selecting teaching methods and materials most appropriate to achieve the aims for a particular student or cohort of students rather than rigidly adhering to a single theoretical framework. This is in line with Ken

Hyland (2019), who notes that a core strength of ESP is the ability to overcome "the theory-practice divide [and make] visible academic and professional genres to students" (p. 1). Central to the ESP framework is the importance of conducting a detailed needs analysis coupled with genre analysis (Swales, 1990). Genre analysis is a key component, arming teachers with specific knowledge of text types. Teachers and materials developers use this knowledge to make explicit the language and rhetorical features that are usually acquired through extended exposure to such texts over time. Only through investigating the genre can teachers understand the form, format, and functions that learners need to become familiar with and master. A primarily social constructivist approach was adopted in which students and teachers worked together on draft manuscripts, enabling students to move from the periphery to the core of their specific discourse community, or community of practice.

Target-context related and learning-context related needs analysis surveys (Bocanegra-Valle, 2016; Hutchinson & Waters, 1987) were used to gain a fuller picture of the perceived needs, wants, and lacks of the students (Allwright, 1982). Primary data collection methods included questionnaires, focus interviews, and observation. Secondary data sources, such as course syllabi for content subjects, lab rosters, and laboratory publications housed in the university repository, were collected and analyzed. From the surveys, we discovered that the primary difference in responses to the needs analysis surveys was not between different disciplines but between different mother tongues, with approximately half the Japanese respondents indicating that research writing in English was unnecessary, while the non-Japanese respondents universally stated that publishing in English was very important to their academic and career prospects. Analysis of the secondary data revealed that 20 percent of the labs produced 80 percent of the research output in English. Based on the extensive needs analysis, a three-pronged approach was adopted, comprising credit-bearing courses, individual conferencing in a writing lab, and provision of online resources.

A suite of credit-bearing courses forms the mainstay of the formal curriculum. These courses are supplemented with online resources and a writing lab that offers individual consultations. All the credit-bearing courses are hybrid, blending onsite instruction with online learning activities. The courses are eclectic in approach, drawing on concepts such as flipped classrooms and activity-based learning. Students, thus, do the majority of their "learning" outside of the classroom. This allows class time to be devoted to activities that require students to recall and reinforce knowledge, and to develop and practice a repertoire of skills that facilitate their familiarity with and competence in research writing. Students typically watch lectures or short "how-to" videos online, or undertake reading assignments, take notes, or complete a set

of questions or tasks, and submit answers on a learning management system prior to class.

Three scientific research writing courses were developed, focusing on pre-writing, developmental writing, and research writing. The pre-writing course introduces students to the writing of scientific research documents. Students who successfully complete this course learn to analyze authentic research documents for structure, organization, language, and common features. Students on the developmental writing course produce a detailed summary or synthesis of an authentic RA that follows appropriate stylistic and linguistic conventions, and a move structure/outline which can be used as the basis for planning a future original RA. Students on the research writing course produce a manuscript for a short RA documenting original research that adheres in structure, style, and content to articles in a specific publication they have targeted. This course adopts a process-approach, engaging and empowering students to draft a manuscript of publishable quality.

The writing lab provides individualized support to writers. Students submit manuscripts for review prior to attending the writing lab. One-to-one writing conferences follow a learner-centered approach, with the learner initially identifying up to five aspects that they suggest the tutor focus on (e.g., coherence, clarity, noun phrases). Most students submit short research articles that have been vetted for content by their supervisor. The writing lab tutors provide advice based on both the learner's request and the tutor's evaluation of the research article. Constructive advice is provided on how to improve the student's ability to write in general or how to improve a particular piece of writing. During these meetings, tutors provide both generic and, where possible, discipline-specific advice. However, when the tutor is unsure about practices in a particular discipline or publication, learners are advised to consult their supervisor.

The online resources that were provided for students consisted of various tools and reference materials that could help students draft and edit their manuscripts. Some tools were proprietary, such as the plagiarism detection software, while others were open access, such as the academic writing suggestion machine (AWSuM) developed by Atushi Mizumoto (2017). Tailor-made tools were also created, including a move visualizer for research abstracts and a corpus-based error detection tool.

Materials and Online Resources

The courses make use of tailor-made materials that are corpus-informed to minimize the disjuncture between prescriptive advice and the descriptive reality of specific disciplines. Sections from authentic research articles were

chosen based on their clarity, language, and ability to provide a generic model for the analysis of authentic documents from various fields. As most incoming students are unfamiliar with the structure, organization, and language of research documents, familiarization with the prototypical genre characteristics of research documents and the way that the structure works to provide a retrospective account of the research was deemed appropriate. The concept of moves and steps (Swales, 1990) within sections is introduced, accompanied by a limited set of lexical bundles indicative of each of the functions of these various moves and steps in the development of the article. Students initially analyze a generic RA and then apply the same analytical techniques to the corresponding section of an authentic RA from their discipline.

Figure 4.1 shows one of the tasks that students complete on the generic RA. This generic research abstract follows the IMRD organization, illustrating how a research abstract can encapsulate the key sections of a research article. Figure 4.2 shows a research article in the field of information science, which was annotated by a student taking the pre-writing course. The student was able to identify the different moves within the abstract and label the functions of each of the paragraphs in the introduction. Many tasks in the writing course encourage students to analyze RAs in their specific discipline.

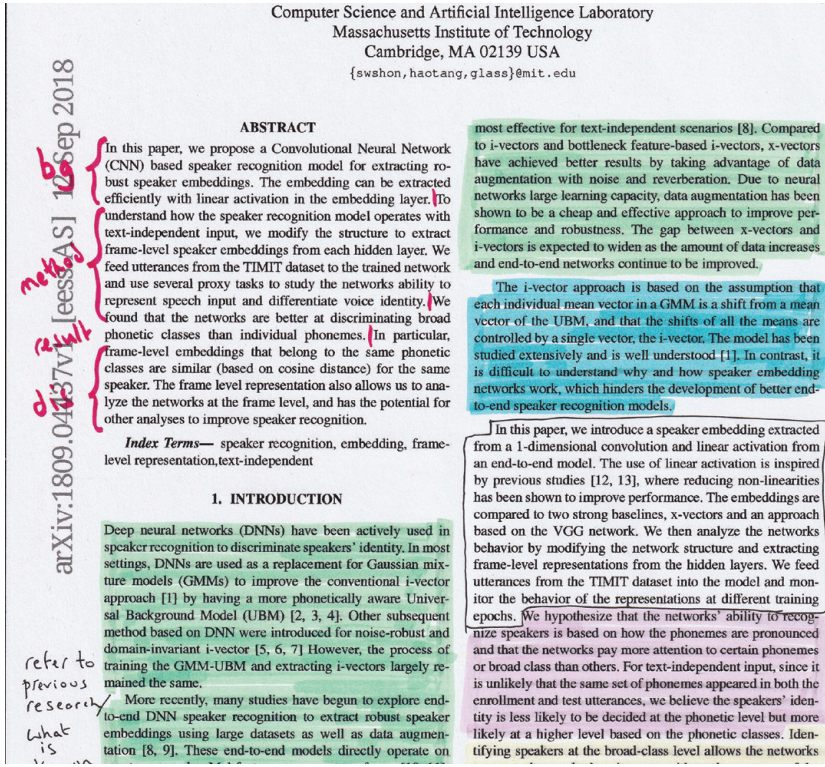
Figure 4.3 shows a task that focuses students on the need for repetition of key ideas. Students use the task to identify the sections of their selected RA that display repetition. The teacher of writing then focuses students on how sentences can be summarized as clauses, clauses as noun phrases, and noun phrases shortened even to nouns. This provides less grammatically aware writers with a systematic way to approach summarization.

Section	***Function in text***
Move 1. Background	*established context of and motivation for the research*
Move 2. Purpose	*presents the aim or goal of the paper*
Move 3. Method	*information on the design, procedure, analysis, etc.*
Move 4. Results	*results, findings or product*
Move 5. Conclusion	*applications or wider implications of the results*

Task: Identify the sentences that belong to each move in this abstract.

Authorship in publications establishes accountability, responsibility, and credit. Misappropriati authorship undermines the integrity of the authorship system, but accurate data on its prevalenc limited. This paper aims to determine the prevalence of articles with gift authors and ghost auth peer-reviewed medical journals. A total of 809 corresponding authors (1179 surveyed, 69% resp rate) of articles published in 1996 in 6 peer-reviewed, general medical journals responded. A to 156 articles (19%) had evidence of gift authors; 93 articles (11%) had evidence of ghost authors 13 articles (2%) had evidence of both. In conclusion, a substantial proportion of articles in reviewed medical journals demonstrate evidence of gift authors or ghost authors.
Adapted from: Annette Flanagin, RN, et.al. Prevalence of Articles with Honorary Authors and Gl Authors in Peer-Reviewed Medical Journals. JAMA, July 15, 1998—Vol 280, No. 3

Figure 4.1. Analysis of rhetorical moves in a generic research abstract.



Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
{swshon,haotang,glass}@mit.edu

ABSTRACT

In this paper, we propose a Convolutional Neural Network (CNN) based speaker recognition model for extracting robust speaker embeddings. The embedding can be extracted efficiently with linear activation in the embedding layer. To understand how the speaker recognition model operates with text-independent input, we modify the structure to extract frame-level speaker embeddings from each hidden layer. We feed utterances from the TIMIT dataset to the trained network and use several proxy tasks to study the networks ability to represent speech input and differentiate voice identity. We found that the networks are better at discriminating broad phonetic classes than individual phonemes. In particular, frame-level embeddings that belong to the same phonetic classes are similar (based on cosine distance) for the same speaker. The frame level representation also allows us to analyze the networks at the frame level, and has the potential for other analyses to improve speaker recognition.

Index Terms— speaker recognition, embedding, frame-level representation,text-independent

1. INTRODUCTION

Deep neural networks (DNNs) have been actively used in speaker recognition to discriminate speakers' identity. In most settings, DNNs are used as a replacement for Gaussian mixture models (GMMs) to improve the conventional i-vector approach [1] by having a more phonetically aware Universal Background Model (UBM) [2, 3, 4]. Other subsequent method based on DNN were introduced for noise-robust and domain-invariant i-vector [5, 6, 7] However, the process of training the GMM-UBM and extracting i-vectors largely remained the same.

More recently, many studies have begun to explore end-to-end DNN speaker recognition to extract robust speaker embeddings using large datasets as well as data augmentation [8, 9]. These end-to-end models directly operate on most effective for text-independent scenarios [8]. Compared to i-vectors and bottleneck feature-based i-vectors, x-vectors have achieved better results by taking advantage of data augmentation with noise and reverberation. Due to neural networks large learning capacity, data augmentation has been shown to be a cheap and effective approach to improve performance and robustness. The gap between x-vectors and i-vectors is expected to widen as the amount of data increases and end-to-end networks continue to be improved.

The i-vector approach is based on the assumption that each individual mean vector in a GMM is a shift from a mean vector of the UBM, and that the shifts of all the means are controlled by a single vector, the i-vector. The model has been studied extensively and is well understood [1]. In contrast, it is difficult to understand why and how speaker embedding networks work, which hinders the development of better end-to-end speaker recognition models.

In this paper, we introduce a speaker embedding extracted from a 1-dimensional convolution and linear activation from an end-to-end model. The use of linear activation is inspired by previous studies [12, 13], where reducing non-linearities has been shown to improve performance. The embeddings are compared to two strong baselines, x-vectors and an approach based on the VGG network. We then analyze the networks behavior by modifying the network structure and extracting frame-level representations from the hidden layers. We feed utterances from the TIMIT dataset into the model and monitor the behavior of the representations at different training epochs. We hypothesize that the networks' ability to recognize speakers is based on how the phonemes are pronounced and that the networks pay more attention to certain phonemes or broad class than others. For text-independent input, since it is unlikely that the same set of phonemes appeared in both the enrollment and test utterances, we believe the speakers' identity is less likely to be decided at the phonetic level but more likely at a higher level based on the phonetic classes. Identifying speakers at the broad-class level allows the networks

Figure 4.2. Discipline-specific function analysis.

Task: Compare your specialist paper to the table and determine whether it includes summaries in the same sections. What phrases are used to signal these moves?

Abstract	Abstract
Intro	Background
	Hypothesis or research questions
	Preview
Methods	Description of method
Results	Key data
	Key data summary
Discussion	Explanations, generalisations, comparisons and contrasts
	Summary
Conclusion	Summary
	Extension to other domains, contexts
References	List of cited works
Appendices	Place to insert details that only interest some readers

Figure 4.3. Discipline-specific summary analysis task.

Move Visualizer

To enable writers to discover move patterns prevalent in research abstracts in their discipline, a visualizer was created to automatically highlight rhetorical

moves in a corpus of 500 research abstracts collected from five journals recommended by discipline specialists. The abstracts were annotated by hand, and specialist informants were consulted to verify the accuracy. Figure 4.4 shows a screenshot of an abstract from *Transactions on Wireless Communication*, a sub-discipline of information science. The moves are color-coded to enable students to notice the patterns. Typical patterns are linear (e.g., IMRD) and cyclic (e.g., MRMR), but non-linear patterns (e.g., RM) can also be seen in this corpus.

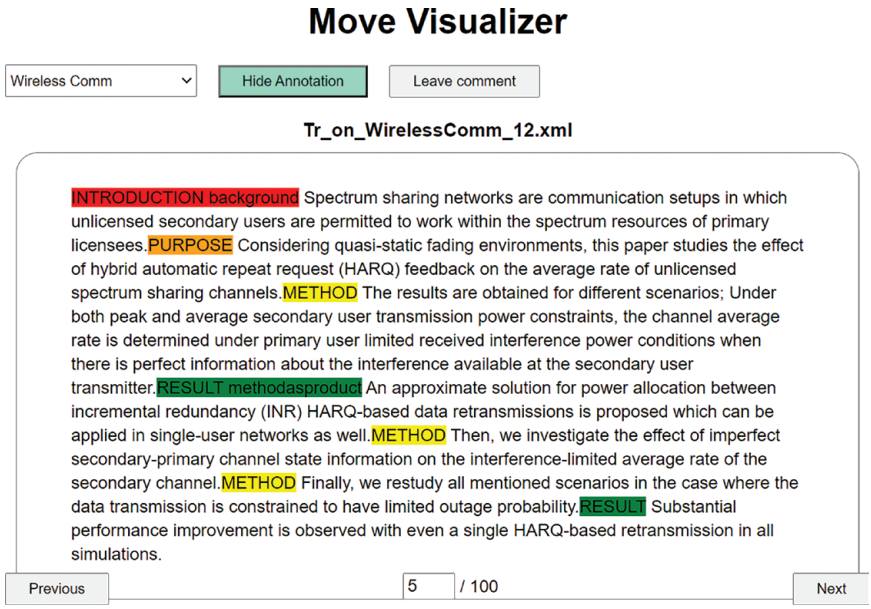


Figure 4.4. Move visualizer showing an annotated abstract.

Corpus-Based Error Detector

Error-free research articles have a higher chance of acceptance than those permeated with lexical, grammatical, or genre-related errors. A review of the pedagogic literature on scientific writing in English housed in the research institute was conducted. This survey revealed that most sources mentioned three main criteria: accuracy, brevity, and clarity, while some sources noted two additional criteria: objectivity and formality (see Table 4.1). Although these criteria are inextricably intertwined, each one can be used as a filter through which feedback on research writing can be given. Scripts to parse for these common errors were incorporated into the error detector. The common errors were identified using the corpus of 200 draft RAs submitted for internal review to the writing lab. Errors were classified manually into five catego-

ries, namely accuracy, brevity, clarity, objectivity, and formality. Feedback from ten student users of the latest version of the error detector was positive, with all students noting improvement in the accuracy of their final manuscripts. One student submitted nine drafts of a manuscript to the error detector and received a total of 227 actionable suggestions (Blake, 2020).

Tutors in the writing lab therefore do not need to deal with errors that can be automatically detected, and so can spend more time on dealing with higher level issues rather than predictable surface-level errors.

Table 4.1. Criteria for scientific research writing

Criteria	Typical errors
Accuracy	Factual, numerical, and language errors
Brevity	Verbosity
Clarity	Vagueness and ambiguity
Objectivity	Overly personal and emotional
Formality	Unexplained abbreviations, contractions, and informal terms

Source: In-house writing course

Conclusion

Response to the materials and courses collected on student evaluation forms has been very positive. Students frequently commented on the usefulness and practicality of these courses, and stated that they feel they can write more fluently and coherently. The real success, however, is not related to the student feedback questionnaires, but in enabling students to get published. Numerous students have attributed their success in this respect to the research writing program and the tutors in the writing lab. In addition to sending emails thanking tutors for their help, students leave comments on the writing lab record. Some of their comments are reproduced below verbatim:

> Student 1: Thank you for your reviewing of my conference paper in this October. The result comes today, and my paper has been accepted! I know that my paper would not have been accepted without your help.
>
> Student 2: Thank you for help me to improve the quality of my paper. I really liked the quick response from the professor and of course I really appreciate all his comments.
>
> Student 3: I have learnt a lot of things that I never knew before. One important thing I have learnt . . . is that using

> short and simple terms can make your writing become more powerful than using some difficult terms or vocabularies.
>
> Student 4: Comments and suggestions were so practical and specific that I could use them directly.

Numerous students who have taken writing courses and participated in the writing lab secured their required publications, gained their doctoral degrees, and started their professional or academic careers. This itself is the main driver of satisfaction for the teachers of writing in this program.

Thanks to the positive word-of-mouth feedback from students completing writing courses and participating in one-to-one writing conferences, some discipline specialists now work directly with writing lab tutors. Students enrolled in the research writing courses benefit from language advice from the writing tutor in tandem with advice from their research supervisor. The initial ESP approach has slowly started to transform into a writing-in-the-disciplines approach.

The corpus-informed tailor-made materials provide authenticity and minimize the disjuncture between prescriptive advice and the descriptive reality of specific disciplines. The focus of the pre-writing course is on enabling learners to understand the generic characteristics and language features of scientific research articles. For the more advanced courses, the focus is on engaging and empowering students to develop knowledge and skills that will better enable them to write their own research articles. As shown in the materials section, students were able to apply the knowledge of structure, language, and organization learned in the classroom on generic materials to authentic research documents from their own field.

In this context, engaging and empowering students means focusing on enabling them to develop, both under supervision and on their own, in order to meet the external needs and demands of their academic and professional communities of practice. Students' efforts are guided, encouraged, and supported by writing center faculty and their disciplinary supervisor, but students remain in charge of both the development of their skills repertoire and the documents needed to satisfy the requirements of the course and an external audience. Finally, the focus of the writing courses is on initiating and developing students as members of their specific discourse community, so that they can participate knowledgeably, competently, and confidently.

Novice writers are engaged in reading, analyzing, and understanding the form, format, and function of each of the sections of RAs in their respective disciplines. The credit-bearing courses provide the foundation on which writers build. The individual one-to-one writing consultations help writers

improve their draft research articles, while the tailor-made online tools help individualize the learning to each specific discipline. By focusing on the five filters of accuracy, brevity, clarity, objectivity, and formality, writers have a tangible framework through which to assess the language of their draft RAs. Following this approach, writers are empowered to draft RAs that adhere to the generic integrity, expectations, and conventions of their community of practice.

References

Allwright, R. (1982). Perceiving and pursuing learner's needs. In M. Geddes & G. Sturtridge (Eds.), *Individualisation (pp. 24-31).* Modern English Publications.

Anthony, L. (2011). Products, processes and practitioners: A critical look at the importance of specificity in ESP. *Taiwan International ESP Journal*, 3(2), 19-50.

Bhatia, V. K. (1993). *Analyzing genre: Language use in professional settings*. Longman Publishing.

Blake, J. (2015). Prescriptive-descriptive disjuncture: Rhetorical organisation of research abstracts in information science. In F. Formato & A. Hardie (Eds.), *8th International Corpus Linguistics Conference* (pp. 377-378). Lancaster University.

Blake, J. (2020). Genre-specific error detection with multimodal feedback. *RELC Journal*, *51*(1), 179-187. https://doi.org/10.1177/0033688219898282

Bocanegra-Valle, A. (2016). Needs analysis for curriculum design. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 560-576). Routledge.

Canagarajah, A. S. (2015). "Blessed in my own way:" Pedagogical affordances for dialogical voice construction in multilingual student writing. *Journal of Second Language Writing*, *27*, 122-139. https://doi.org/10.1016/j.jslw.2014.09.001

Carter, C., Ferzli, M., & Wiebe, E.N. (2007). Writing to learn by learning to write in the disciplines. *Journal of Business and Technical Communication*, *21*(3), 278-302. https://doi.org/10.1177/1050651907300466

Casanave, C. P., & Vandrick, S. (Eds.). (2003). *Writing for scholarly publication: Behind the scenes in language education*. Erlbaum.

Chang, C.F., & Kuo, C.H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, *30*(3), 222-234. https://doi.org/10.1016/j.esp.2011.04.001

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

Englander, K. (2006). Revision of scientific manuscripts by non-native-English-speaking scientists in response to journal editors' criticism of the language. *Journal of Applied Linguistics*, *3*(2), 129-161. https://doi.org/10.1558/japl.v3i2.129

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, *30*(3), 198–208. https://doi.org/10.1016/j.esp.2011.01.001

Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "stigma" tell us? *Journal of English for Academic Purposes*, *7*(2),77-86. https://doi.org/10.1016/j.jeap.2008.03.002

Flowerdew, J., & Li, Y.Y. (2007). Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics*, *28*(3), 440-465. https://doi.org/10.1093/applin/amm031

Fryer, D. L. (2013). Exploring the dialogism of academic discourse: Heteroglossic engagement in medical research articles. In G. Andersen & K. Bech (Eds.), *English corpus linguistics: Variation in time, space and genre* (pp. 183-207). Rodopi.

Fujimoto-Adamson, N. (2006). Globalization and history of English education in Japan. *Asian EFL Journal*, *8*(3), 259-282.

Gee, J. P. (1996). *Social linguistics and literacies: Ideology in discourses, critical perspectives on literacy and education* (2nd ed.). Routledge.

Gee, J. P. (2007). *Social linguistics and literacies: Ideology in discourses* (3rd ed.). Taylor & Francis.

Habibie, P., & Hyland, K. (2019). Introduction: The risks and rewards of scholarly publishing. In K. Hyland & P. Habibie (Eds.), *Novice writers and scholarly publication: Authors, mentors, gatekeepers* (pp. 1-10). Palgrave Macmillian. https://doi.org/10.1007/978-3-319-95333-5

Hendges, G. R., & Florek, C. S. (2019). The graphical abstract as a new genre in the promotion of science. In M.-J. Luzón & C. Pérez-Llantada (Eds.), *Science communication on the internet: Old genres meet new genres* (pp. 59-79). John Benjamins. https://doi.org/10.1075/pbns.308.04hen

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

Hyland K. (2019) English for specific purposes: Some influences and impacts. In X. Gao (Ed.), *Second handbook of English language teaching*. Springer. https://doi.org/10.1007/978-3-030-02899-2_19

Jaffe, K. (2014). Social and natural sciences differ in their research strategies, adapted to work for different knowledge landscapes. *PloS One. 9*(11), e113901. https://doi.org/10.1371/journal.pone.0113901

Japan Advanced Institute of Science and Technology (JAIST). (2019). *About JAIST*. https://www.jaist.ac.jp/english/about/outline/foreigner-student.html

Koike, I., & Tanaka, H. (1995). English in foreign language education policy in Japan: Toward the twenty-first century. *World Englishes*, *14*(1), 13-25.

Lane, S., Karatsolis, A., & Bui, L. (2015, July). Graphical abstracts: A taxonomy and critique of an emerging genre. In *Proceedings of the 33rd Annual International Conference on the Design of Communication* (pp. 1-9). ACM. https://doi.org/10.1145/2775441.2775465

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press.

Leydesdorff, L., Bornmann, L., Comins, J. A., & Milojević, S. (2016). Citations: Indicators of quality? The impact fallacy. *Frontiers in Research Metrics and Analytics*, *1*(1). https://doi.org/10.3389/frma.2016.00001

Li, Y. (2007). Apprentice scholarly writing in a community of practice: An intra-view of an NNES graduate student writing a research article. *TESOL Quarterly*, *41*(1), 55-79. https://doi.org/10.1002/j.1545-7249.2007.tb00040.x

Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context.* Routledge.

Lillis, T., & Turner, J. (2001). Student writing in higher education: Contemporary confusion, traditional concerns. *Teaching in Higher Education*, *6*(1), 57-68. https://doi.org/10.1080/13562510020029608

McDowell, L., & Liardét, C. L. (2019). Japanese materials scientists' experiences with English for research publication purposes. *Journal of English for Academic Purposes*, 37, 141-153. https://doi.org/10.1016/j.jeap.2018.11.011

Mizumoto, A. (2017). Initial evaluation of AWSuM: A pilot study. *Vocabulary Learning and Instruction*, *6 (2)*, 46-51.

Nakamori Y. (2011) Introduction. In Y. Nakamori (Ed.), *Knowledge science – Modeling the knowledge creation process* (pp. 1-10). CRC Press.

Nasirpouri, F. (2017). *Electrodeposition of nanostructured materials*. Springer.

O'Neill, D. K. (2001). Knowing when you've brought them in: Scientific genre knowledge and communities of practice. *The Journal of the Learning Sciences*, *10*(3), 223-264.

Orr, T. (1998). ESP for Japanese universities: A guide for intelligent reform. *The Language Teacher*, *22*(11). https://jalt-publications.org/tlt/articles/2428-esp-japanese-universities-guide-intelligent-reform

Saracevic, T. (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2570-2585). Taylor and Francis.

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

Trowler, P., & Becher, T. (2002). Academic tribes and territories: Intellectual enquiry and the cultures of discipline (2nd ed.). Open University Press.

Walbort, V. (2009). Are we training pit bulls to review our manuscripts? *Journal of Biology*, *8*(3), 24.

Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, *19*(3), 158-177. https://doi.org/10.1016/j.jslw.2010.06.002

Wingate, U. (2012). 'Argument!' Helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*(2), 145-154. https://doi.org/10.1016/j.jeap.2011.11.001

Wingate, U., Andon, N., & Cogo, A. (2011). Embedding academic writing instruction into subject teaching: A case study. *Active Learning in Higher Education*, *12*(1), 69-81. https://doi.org/10.1177/1469787410387814