Engineering Registers in the 21st Century: SFL perspectives on online publications

Sheena Gardner, Xiaoyu Xu

Original citation:
Engineering Registers in the 21st Century: SFL perspectives on online publications, Sheena Gardner and Xiaoyu Xu, *Language, Context and Text: The Social Semiotics Forum*, Vol. 1, Issue 1 (2019) pp. 65-101
http://dx.doi.org/10.1075/langct.00004.gar

ISSN: 2589-7533

Publisher: John Benjamins

Engineering registers in the 21st century:
SFL perspectives on online publications

Sheena Gardner
Coventry University, ORCiD ID: 0000-0003-1924-5270
ab1248@coventry.ac.uk

Xiao-yu Xu
City University of Hong Kong, ORCiD ID: 0000-0002-7609-2912

Abstract

Following an exploration of engineering programmes in higher education, and a review of literature on engineering registers, genres and disciplines, this paper asks if there is a register for engineering. Word frequencies, n-grams and frequent n-grams in context were analysed in a 7.3 million word corpus created from four sections (Introduction, Materials & Methods, Results & Discussion, Conclusion) of over 1000 articles in civil, electrical and mechanical engineering. From systemic functional linguistics (SFL) perspectives, this reveals how engineering is construed through language that reflects the social context of high impact, open access, multi-modal, 21st century, international journal article publication, with multiple author roles, and prescribed genres, where reviewers focus on problem solving and facts, rather than persuasive claims.

Key words: academic English, engineering register, corpus linguistics, systemic functional linguistics, research article genres, English as a lingua franca, online publication contexts

1. Introduction

1.1 What is engineering?

Just as the word “history” in English suggests that it might be or have traditionally been about “his” “story”, and the word “biology” suggests that is has to do with the study (ology) of living organisms (bio), the word “engineering” suggests that it is about getting things done, about figuring out how to bring about change. This is largely accurate. Engineering is much less about interpreting events or developing theories to account for the physical or social universe, and much more about how to design and develop “things” better, about problem solving, about improving our lives through developments in the devices and physical constructs we use to interact with the world.

1.2 What do engineers do?

Engineering addresses real world problems, so the sorts of questions engineers ask are, how can we do that, or how can we do that better, cheaper, with less environmental impact. The ultimate test then, is whether something works, and how well it works as evaluated not necessarily by engineering or theoretical perspectives, but by measures that are essentially external to engineering and to the academy – usability, economic, environmental. It is this practical problem focus that unites engineering contexts.

This applies across engineering contexts. I have heard it from a professor in Hong Kong explaining the nature of doctoral theses and confirmation reports in engineering; I have heard it from an academic dean in Istanbul explaining how the different components of an undergraduate programme fit together, and I have heard it from members of an international internship team at Airbus in the Netherlands explaining the advice they were given to develop a new product using drones – start with a real world problem, don’t start with the engineering. Interestingly, the team started with the problem of air pollution in cities like Beijing and developed a drone equipped to recycle the pollution into fertiliser.

1.3 What characterises the disciplinary practices of engineering?

Three interconnected characteristics of engineering make it complicated to pin down an academic English register for engineering:

1. Engineering is problem oriented; in this sense it is an applied rather than a pure science
2. Engineering is applied in multiple sites; in this sense it is applied in many different areas of our lives: to cars, aeroplanes, ships, roads, and buildings, but also in satellites, radio, robotics and many developing areas of human exploration
3. Engineering draws on many different disciplines. Because it aims to solve real world problems it often involves teams from different backgrounds working on a problem together, and as such it is at least multidisciplinary and often interdisciplinary.

These characteristics suggest that there would be many overlapping Engineering registers, and raise one central question, which we address here:

Research question: Do the specific engineering registers overlap to the extent that we can describe a disciplinary register for engineering in English?

In SFL terms, this becomes a question of instantiation, how we view the registers of engineering in relation to both the texts that instantiate them and also the language system that is a theorisation of the meaning potential of the English language.

> In working on the grammar of a language, one tries to move freely along the instantiation cline. We are able to do this much more easily once we have a corpus. … The corpus is not a substitute for theory…but [facts and the principles behind them] can be construed much more reliably on the quantitative foundation of a modern computerised corpus. (Matthiessen & Halliday 2009: 81-2).

A corpus that aims to capture "all" of the English language will cast too wide a net; a corpus of texts from a very specialised branch of engineering equally cannot be expected to inform an account of the language of engineering more widely. The selection of texts for a corpus to inform our analysis is therefore crucial, as is an understanding of the contexts in which those texts are produced.

The approach here to such analysis is first to explore the social and cultural contexts. This will lead to an account of engineering genres and an appreciation of the role of different genres in engineering. My aim here is to identify an internationally recognised and highly valued engineering genre together with an understanding of its production context and use. The corpus construction and register analysis can then proceed, where frequent items, clusters of items and patterns of items across texts can be explained with reference to both their occurrence in individual texts and to the social contexts in which they are used. This should enable an understanding of specific engineering registers in terms of Field (and its relationship to knowledge in different branches of engineering), Tenor (addressed by and to specialists, students or laymen) and Mode (with respect to channel and degree of congruence) (following Halliday 2004: 140–141).

The paper first considers the social context of engineering in higher education – the courses, careers, associations and activities of engineering. Secondly, it reviews existing studies on the academic language of engineering. The aim here is first to get a general sense of the scope of engineering – what it is and what it is not – then to delimit a disciplinary context for engineering, and finally to select texts in their contexts and to build a corpus that can be examined to provide answers to the question about engineering registers.

2 The social context of engineering

2.1 Engineering careers

If we look at what engineers do outside the academy, we can differentiate professional accredited engineers from engineers who, at least in the UK, do not require a university degree. We are essentially interested in the former because of our focus in this special issue on disciplinary registers in higher education.

Table 1. Engineering careers that require a university degree vs those that do not

Engineering Careers that require a university degree	Engineering Careers that do not require a university degree
Aerospace engineers design, build and maintain planes, spacecraft and satellites.	**Agricultural engineers** make and maintain agricultural, horticultural and forestry machinery and equipment.
Automotive engineers design, develop, test and build cars and motorbikes.	**Broadcast engineers** make sure television, radio and online programmes are broadcast at the right times and are high quality.
Building services engineers design, install and service equipment and systems in buildings like offices and shops.	**Forklift truck engineers** service and repair lift trucks.
Civil engineers design and manage construction projects, from bridges and	**Heating and ventilation engineers** install and service heating and air conditioning in large buildings like factories, schools and hospitals.

buildings to transport links and sports stadiums.	
Electronics engineers design and develop systems for industry, from mobile communications to manufacturing and aerospace.	**Lift engineers** install, refurbish, service and repair lifts and escalators.
Energy engineers work on the research, design and construction of power generation plants, and may be involved in drilling for gas and oil.	**Measurement and control engineers** design the systems that control machinery and equipment in industry.
Manufacturing systems engineers design and install manufacturing equipment and assembly production lines.	**Nuclear engineers** are responsible for the safe running of nuclear power stations.
Marine engineers design, build, test and repair boats, ships, underwater craft, offshore platforms and drilling equipment.	**Quarry engineers** explore new sites, oversee operations and manage sites at the end of their commercial life.
Mechanical engineers develop and design the components and machinery used in manufacturing, construction, water, power, health and transport.	**Satellite engineers** install and repair telecoms equipment and networks, and satellite systems.
Motorsport engineers design, build and test racing cars and bikes.	**Sound engineers** work in studios and make recordings of music, speech and sound effects.
Structural engineers help to design and build large structures and buildings, like hospitals, sports stadiums and bridges.	**Telecoms engineers** work on satellite, digital TV and fibre optic systems, and install broadband, mobile and landline phone networks.
Chemical engineers develop ways to turn raw materials into everyday products.	**Thermal insulation engineers** install insulating materials around pipes, boilers and ductwork.
Materials engineers research the behaviour of materials used in industry to help make them stronger, lighter or more durable.	

The list in table 1 is taken from one national careers website,[1] and is not exhaustive. Similar lists from other sources would overlap significantly but would probably not be identical. The list has been subdivided here according to the information provided on the careers website, so the distinction being made here between professional engineers and technical engineers is made in this paper.

An analysis of the evidence summarised in table 1 suggests that a common expectation of engineers from relevant higher education courses in engineering is that they will design and develop engineering products. In comparison, those from apprenticeship and college programmes are more likely to install and maintain the smooth operation of engineering

[1] https://nationalcareersservice.direct.gov.uk/job-profiles/manufacturing-and-engineering.

products. The former can be described as professional engineers, who expect to command higher salaries than those who can be described as technical engineers.

From this we can see that a key capability of professional engineers is design, and that there are about a dozen different domains of engineering specialisation: aerospace, automotive, building services, civil, electronics, energy, manufacturing systems, marine, mechanical, motorsports, structural, chemical and mechanical.

2.2 Professional accrediting bodies

If we look beyond careers planning, there are many more specialisations. For example, the New York based Institute of Electronics and Electrical Engineering, the IEEE, that describes itself as the largest technical professional organisation – note the wording – technical and professional – includes 59 societies, 1800 annual conferences, and so plenty of opportunities for specialisations such as communications, circuits & systems, information theory, magnetics, photonics and reliability. Some of these areas are taught on degree courses, but most degree courses would be broader. So the IEEE represents many more specialisms than a higher education register. (It is worth knowing about, however, for those teaching engineering English for the free Massive Open Online Courses (MOOCs) and other online resources available through its website.)

The main UK-based international professional organisations seem to be the Institution of Mechanical Engineers, Institution of Civil Engineers, and Institution of Engineering and Technology. This tripartite division suggests a more manageable perspective on graduate engineers and disciplines of engineering.

2.3 University courses

If we move back from professional groupings to university courses, the QS World Ranking suggests six distinct disciplines of engineering and technology:

1. Computer Science and Information Systems
2. Chemical Engineering
3. Civil and Structural Engineering
4. Electrical and Electronic Engineering
5. Mechanical, Aeronautical and Manufacturing Engineering
6. Mineral and Mining Engineering

While these are the six main disciplines given, university courses are competing for students and so the disciplines they foreground fluctuate. They are currently listed under groupings that include biomedical engineering and environmental engineering[2] and these groupings reflect innovations in research and teaching.

2.4 University faculties, schools and departments of engineering

The integration of computer science and information systems with engineering is common, and faculties of engineering often include computer science. If we look at the top universities for engineering and technology according to the QS2018 World University Rankings[3], we see that there may be two distinct faculties (Oxford University) or there are distinct departments within a faculty of engineering (Cambridge, ETH Zurich), or computer science can be combined with electrical engineering (MIT, UC Berkeley) or the two may be combined in one school (Nanyang Technological University), or there may be parallel schools of different branches of engineering - aerospace, civil, materials, mechanical - and a school of information science & technology which includes a department of computer science and technology (Tsinghua), and my guess is that these disciplinary groupings change over time.

So while it is common to group engineering and technology courses together, and to include computer science within some engineering schools or faculties, and there is no doubt that engineering like all disciplines is increasingly dependent on computer science, for the purpose of this paper for reasons of scope, computer science registers will not be a main focus. We can also conclude that the subdisciplines and domains of engineering are variable across universities. The most stable include mechanical engineering, civil engineering and electrical engineering, and these will be the focus of the rest of this paper. If we find an engineering register that is common to these three, then further research could be conducted to explore the extent to which it also applies in other branches such as computer science, architecture, chemical engineering or environmental engineering.

3 Empirical studies on the language of engineering

3.1 Engineering compared to other disciplines

[2] https://www.topuniversities.com/courses/engineering/guide.
[3] www.topuniversities.com.

Engineering has been described as a hard, applied discipline, where accurate use of technical terms is understandably crucial for those involved in the design of vehicles, buildings or bridges. As interview data with academics and professional engineers has confirmed, engineers should be fully aware that they can be held legally accountable for their designs, reports and recommendations (Nesi & Gardner 2006; Conrad 2017).

The language of Engineering has been compared with the language of soft disciplines such as sociology, applied linguistics, marketing and philosophy. For example, in a corpus study of metadiscourse, mechanical and electronic engineering textbooks were found to include less textual metadiscourse (e.g. fewer connectives) and less interpersonal metadiscourse (e.g. fewer hedges and person markers) than textbooks in "softer" disciplines (Hyland 2000: 114). Similarly, a multidimensional analysis of university textbooks across disciplines finds the language of engineering to be the most impersonal, using combinations of conjuncts, agentless passives and past participial adverbial clauses to focus on events and circumstances rather than participants, in contrast to the more personal language of humanities and education textbooks (Biber et al. 2002: 47).

Similarly, in a multidimensional analysis of university student writing (Gardner et al. 2018), many engineering reports express compressed procedural information through long scientific nominal groups (noun pre-modifiers, common, concrete, quantity nouns) and a focus on concisely reporting experimental procedures through passive action verbs. At the other end of this dimension Gardner et al. (2018) find the "stance toward the work of others" that is typical of humanities essays, as evidenced through the absence of procedural features and the presence of 3rd person pronouns, stance nouns + that clauses, proper nouns, stance adverbials and communication verbs.

The phrases of engineering have been described as highly technical and discipline specific. In a comparison of frequent and key lexical bundles in the British Academic Written English corpus (BAWE) of student writing across three discipline-genre complex, the key engineering report bundles were found almost uniquely in engineering texts, whereas the key bundles from Business case studies occurred across multiple disciplines, with the Economics essay bundles occupying a middle position. Examples of top key bundles in student engineering reports are *moment of inertia* and *mass flow rate* (Gardner 2016: 161). A similar finding emerges in Sun (2013) which found that the hard sciences, including engineering, "recycle" strings of text more freely than the humanities and social sciences. This makes it particularly amenable to corpus analyses that identify frequent phrases.

In another study of academic phrases, Huang (2017) found that computer science bundles were more specific than their more widely used academic counterparts. For instance, computer scientists use “interesting evidence”, while the academic collocation list (Ackermann & Chen 2013) includes “compelling evidence”, or where computer scientists talk of a “significant effect” the academic collocation list includes a “profound effect”. English teachers might think that “interesting” is a rather weak word, to be avoided, but for computer scientists words such as “compelling” and “profound” may have affective or emotional connotations that they want to avoid.[4]

3.2 Engineering genres and registers compared

While engineering as a hard, applied science can be compared to Humanities or Education, it is also worthwhile exploring differences across engineering registers. In a comparison of written and spoken registers, Biber et al. (2002: 40) found that the impersonal focus of engineering textbooks did not extend to spoken registers. Indeed, engineering lectures and classroom teaching was found to be highly situation dependent and persuasive, when compared to other disciplines (*ibid*), features which may vary across cultures as research on the engineering Lecture Corpus, which compares lectures from the UK, Malaysia and New Zealand, has shown (Alsop & Nesi 2013).

Differences have also been found across genres and levels of study, for instance between student reports and professional reports in civil engineering (Conrad 2017), between research articles and PhD theses (Koutsantoni 2006), between Master’s and PhD dissertations in Hong Kong (Hyland 2004) and across levels of study at British universities both in lexical bundles (Durrant 2017) and grammatical features (Gardner et al. 2018).

Within the same registers and genres, differences have been found across disciplines: In engineering research articles, differences have been found for Swalesian steps and moves across civil, software and biomedical engineering (Kanoksilapatham 2015), and in Master’s dissertations, differences have been found in lexical bundles acrosspower, computer, control and telecommunications engineering (Rezoug & Vincent 2018).

The Introduction Methodology Results Discussion (IMRD) format emerges as a familiar structure across engineering genres. Gardner (2012) shows how it provides a macrostructure for two very different student genres: the lab report, where the task is given and the report

[4] O’Donnell, personal communication.

focuses on the methods and results, and the project report which also includes a substantial literature review that supports the development of research questions and a substantial conclusion where the contribution of the study to the literature is demonstrated. Similarly, Wang and Flowerdew (2016) demonstrate the relevance of IMRD moves and stages to successful personal statements in Hong Kong applications for U.S. PhD programmes in engineering.

A different perspective is taken on lexicogrammar from an English as a Lingua Franca perspective. Research in this area is able to track the nature of English used between non-native speakers in academic and professional engineering contexts where the focus is on having one's ideas and contributions understood, rather than on using a particular standard English. As Flowerdew (2015: 110) points out, Wood (2001) had predicted that English would become the language of international scientists rather than of the "native speaker" or English-speaking world, and this poses a very significant question for teachers of English for Academic Purposes. Do they continue to try to correct students' "errors" with "non-standard grammar"? Björkman's (2008) work analyses monologic and dialogic spoken data to identify forms that cause communication break-down, and to differentiate them from non-standard forms that do not interfere with communication. Rozycki and Johnson (2013) focus specifically on very successful written communication, where they find a range of non-standard forms such as missing articles and lack of number agreement between subject and verb. They conclude that such non-standard forms are not a problem in international engineering communication where English is used as a lingua franca.

4 Halliday's theories about scientific registers

Halliday has written extensively and eloquently about the language of science, with particular reference to Physics and Biology. His work on grammatical metaphor is captured in the four papers in Part 1 of *The Language of Science* (Halliday 2004), while his work on scientific English is captured in the four papers in Part 2 of the same volume. The influence of his ideas in these areas is evident in edited collections such as *Writing Science* (Halliday & Martin 1993), in *Reading Science* (Martin & Veel 1998) and in numerous books and papers written in different theoretical paradigms (e.g. in Hyland's (e.g. 2000: 4) work on academic discourse and Biber's (e.g. Biber & Conrad 2009: 22) work on academic registers). Nevertheless, it is worth here reminding ourselves of two of his central arguments. The first relates to the grammar of academic English and the second to its development over time.

4.1 Grammatical metaphor

The power of Halliday's work is that he not only observes patterns in text, but is also able to theorise and explain their function in their social context. Thus, he observes that the move from congruent to metaphorical language occurs alongside an increase in technical terminology.

This technical language involves not only the creation of new terms, but also in Martin's (1990) terms, the "distillation" of new meanings. Halliday gives the example of the build-up in a text of the term "glass fracture growth rate".

> Step by step the text builds up more complex meanings: the "crack" becomes an entity which can "grow" more or less "slowly"; "grow slowly" becomes "slow growth", and "growth rate" and then at the end there is an entirely new entity called "glass fracture growth rate". (Halliday 2004: 117)

This shows how the grammar has moved from the congruent to the metaphorical over the development of the text.[5]

With grammatical metaphor and technical terminology comes the ability to construe phenomena "as if they were things" (Halliday 2004: 216). The "elaborated register of scientific knowledge reconstrues [reality] as an edifice of things" (ibid*).* Thus scientific constructs can be evaluated and can be related to each other. In this way, "grammatical metaphor … increase[s] the power that a language has for **theorizing**" (Halliday 2004: xvii, original emphasis). These understandings are founded on analyses of the type of language found in Mathematics, Physics and other sciences where the canonical forms have evolved from more clausal everyday to more nominal scientific modes, which are as a result more lexically dense (Halliday 2004: 195).

This is a theory worth exploring for engineering. We might expect engineering registers to be technical, but more to do with physical problem solving and less to do with theorizing. A specific focus of our analysis of engineering registers will be to consider whether this is evidenced in the grammar. A tentative hypothesis is that because engineering is more applied, more problem-oriented, and evaluated more by "what works", there will be technical language, but clauses will tend to select more material processes than the relational processes found in the pure sciences.

[5] The editors' note: For a schematic demonstration by Halliday of these processes see Butt Appendix 1, this volume.

4.2 The phylogenetic development of scientific language

Some of the most intriguing theories that Halliday proposes concern the development of language over time, or "semogenesis". This occurs, he argues, in three spheres where the more congruent language becomes more metaphorical: the system of a language evolves over the years (phylogenetic development), the meaning potential of a human being develops with increasing years and maturity (ontogenetic development) and the instantiated text unfolds from beginning to end (logogenic development). (Halliday 2004: 116; 2004: 220)

He traces the development of the English of science from Chaucer's instructions of use which already include technical terms, through Newton's discourse of experimentation where the passive is used, and clause complexes might be intricate with both expansion and projection, but there are also the now familiar simple clauses with long nominal groups connected by relational processes, to the last century where scientific language is lexically dense and depersonalised with simple sentence structure, where nominal groups form technical taxonomies in three fields (technological, methodological and theoretical), and where logical relations are both external (relating to the real world) and internal (relating to the argument in the text) (Halliday 2004: 153). Where clauses are nominalised, they can be treated as things in a process (Halliday 2004: 216). In this way, the language reflects and enables developments in science while at the same time leading to an increase in technicality and nominal group complexity.

Earlier developments in the nature of language resulted from the introduction of the printing press with its influence on written language, and the introduction of sound recording with its influence on spoken language. The questions for the 21st century revolve around the influences of technology and how computers and social media are changing the language of science, and engineering.

Towards the end of the 1993 paper, Halliday tries to predict how language will continue to develop. He predicts that the technical language will become too technical and people will want a simpler language. He says the language of science is "likely to back off from its present extremes of nominalization and grammatical metaphor and go back to being more preoccupied with processes and more tolerant of indeterminacy and flux" (Halliday 2004: 224). As evidence of contemporary change in use of technicality and abstraction, we might suggest the more casual, constantly updated, online news events whose language is more accessible, written in screen-size paragraphs, and less precise than that of the print news of last century. We might also suggest the anti-intellectual or anti-academic rhetoric that denies climate change, that trusts its own educational experience over the more objectively produced educational research, or where politicians aim to persuade their people through language full of affect via twitter. This paper will consider how engineering registers are influenced by online production and other cultural and situational factors that influence production in the 21st century.

5 Methodology

5.1 An online engineering corpus

Our examination of the social context of engineering in Higher Education suggests that while there are very many specialisations, some of which are taught in engineering programmes and others of which become areas of research specialisation, three broad disciplines are widely recognised – mechanical, civil and electrical – and these form the focus for data collection. Academic genres of engineering in higher education can be grouped according to their purpose into pedagogical or training genres such as essays and lab reports; professional or work-oriented genres such as design specifications and case studies; and academic or research-oriented genres, such as research reports. It was decided to focus on research reports. These are found with an IMRD macrostructure in undergraduate projects, Master's and PhD theses and published journal articles. Other related genres include confirmation reports and statements of purpose. It is anticipated therefore that while the registers of all of these genres will not be the same, the substantial similarities envisioned make this a worthwhile focus. As there are also differences between the IMRD sections of research articles, and as articles today are predominantly written and read online by engineers of many different nationalities, it was decided to investigate engineering articles from PLOS ONE.

PLOS ONE (public library of science) is a peer-reviewed open-access journal that began in 2006 and has published around 200,000 articles from across the sciences. Interestingly for our purposes, the papers may be of any length (though authors are encouraged to present and discuss findings concisely) and they are not centrally edited, which means that the language used is less influenced by the publishing process than it otherwise might be. The current guidance suggests that the articles conform to a particular macrostructure (Title, Abstract, Introduction, Materials & Methods, Results & Discussion, Conclusion, References, Appendices) and detailed guidance is given for each section[6], but the middle sections

[6] http://journals.plos.org/plosone/s/submission-guidelines.

(Materials & Methods, Results, Discussion, Conclusion) "can be renamed as needed and presented in any order".[7] As a result, and as revealed by a manual inspection, engineering papers may have an IMRD structure, with no Conclusion section.

Nevertheless, PLOS One is a highly regulated production context, and this extends to who is allowed to submit papers. All authors must have an ORCID iD; those who do not can be acknowledged, but will not be authors. Author contributions vary and should be recorded according to the CRediT Taxonomy[8], which provides a standard list of about a dozen author roles such as data curation, formal analysis and writing up.

The functional section headings of Introduction, Materials & Methods, Results and Conclusions have clear advantages for our purposes in that they are in effect macrothemes that establish expectations for the sections that follow and facilitate focused comparisons of engineering registers across disciplines. As decisions about papers focus on the originality and authenticity of experimental procedures and findings, "editorial decisions do not rely on perceived significance or impact, so authors should avoid overstating their conclusions".[9] This ensures a quick turnaround, and with a reported 50% acceptance rate, tens of thousands of papers are published each year.

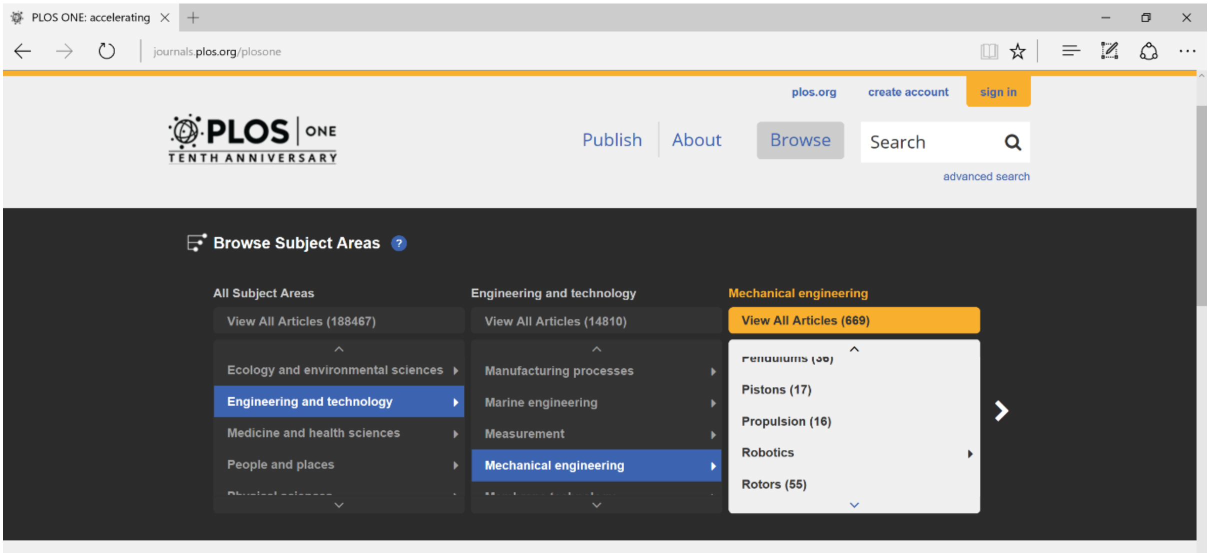

Figure 1. PLOS ONE home page browser (http://journals.plos.org/plosone/, 10/10/2017)

[7] http://journals.plos.org/plosone/s/submission-guidelines.

[8] The CRediT taxonomy was developed by a consortium of journals called the Consortia Advancing Standards in Research Administration Information (CASRAI).

[9] http://journals.plos.org/plosone/s/submission-guidelines.

Figure 1 shows the open access search engine where the c.15,000 engineering and Technology articles are subdivided into 45 branches; one of which is Mechanical engineering, which contains c.600 articles and is subdivided into fourteen categories, including pistons (17), propulsion (16) and robotics, which itself is further subdivided. The numbers of articles grows daily, but the browser gives a good overview of the contents.

The corpus for this paper was constructed using the subject category browsing facility in the corpus generation tool AntCorGen (Anthony 2017) to select texts from four sections (Introduction, Materials & Methods, Results & Discussion, Conclusion) and from the three disciplines of civil engineering, electrical engineering and mechanical engineering. These were downloaded on 26th August 2017 in plain text format and compiled into a corpus of 7,307,279 words (table 2). Figures were automatically excluded in the process.

Table 2. The online engineering corpus with its 12 components

	Civil		Electrical		Mechanical	
	Total	Average	Total	Average	Total	Average
Research Articles (n)	425		206		672	
Introduction (words)	344,368	810	210,143	1020	617,917	920
Materials & Methods (words)	644,800	1517	303,018	1471	970,626	1444
Results & Discussion (words)	985,654	2319	747,893	3631	1,789,292	2663
Conclusions (words)	83,365	196	154,746	751	455,457	678
Total words	2,058,187	4,843	1,415,800	6,873	3,833,292	5,704

The difference in the number of texts in each discipline reflects the publication rate, with more articles being published in PLOS ONE in mechanical engineering (672) than in civil engineering (425) than in electrical engineering (205) since 2006. An explanation of this is beyond the scope of this paper, but it does reflect the greater volume of activity in Mechanical Engineering generally.

In contrast, the electrical engineering texts appear to contain the most words, with a notably larger count in the Results and Discussion sections. A manual inspection of these sections reveals that as the data from tables has not been extracted from the text, there are strings of numerical data and formulae that would increase the “word” length, as in Figure 2:

Neuronal population
θ (nA)
β (ms−1 nA−1)
RS
0.1
0.11
LTS
0.05
0.32
FS
0.28
0.35
10.1371/journal.pcbi.1002248.t002Table 2
Reference parameters for the synapses between the various types of neurons.

Figure 2. Example of text format of tables in the corpus

Similarly, a manual search of the civil conclusions shows that around one third do not have conclusions, which helps to explain the relatively small conclusion sections in civil. This is the first version of AntCorGen, which was just released in July 2017, so it is possible these anomalies can be rectified in future versions. Nevertheless, Table 2 gives us a sense of the amount of data analysed. It remains to be seen how any quantitative differences play out in the registers.

5,2 Procedures for word lists and lexical bundles

The corpus was loaded onto AntConc (Anthony 2014) to generate a word list for each discipline. This gives us a sense of shared and frequent vocabulary across engineering. Please note that counts do not take into account punctuation or differentiate by case, so, for example, *I.e.* and *i.e.* are both analysed as *i e*, which would count as two items.

Lexical bundles are used as the next entry point to register. It is expected that they will point to features worth exploring further in each component. Each of the 12 components (3 disciplines x 4 sections) was uploaded into AntConc so that they could be analysed separately or in groups. Three aspects were considered in the course of extraction: length, frequency, and dispersion/range.

Although some prior studies (Biber et al. 1999; Biber & Barbieri 2007; Hyland 2008) have suggested 4-gram as the optimum length, we decided to experiment with 3-grams, 4-grams and 5-grams so as to thoroughly capture the lexical bundles that are common and have a clear range of structures and functions.

A minimum threshold for normalised frequency (here frequency per million words) is essential to allow comparisons across the components and with prior studies. However, as

Biber and Barbieri (2007: 267) point out, “the frequency cut-off used to identify lexical bundles is somewhat arbitrary”. There are studies that set a cut-off point of 10pmw (e.g., Biber et al. 1999), 20pmw (e.g. Hyland 2008; Cortes 2004) and even stricter cut-off point of 40pmw (e.g. Biber & Barbieri 2007). In this study, we set a high cut-off point of 40pmw for 4-gram extraction, 20pmw for 5-gram extraction and 80pmw for 3-gram extraction. As AntConc only allows for minimum raw frequency set up rather than normalised frequency, we calculated the minimum raw frequency for each component in line with the word-count of each component (see table 3). For example, the 4-gram raw frequency cut-off point for civil engineering introduction is 344368 * 40/1,000,000 ≈ 14. The raw frequency cut-off points are presented in table 4.

Table 3. Minimum raw frequency for 3-, 4- and 5-grams

	Civil				**Electrical**				**Mechanical**		
Gram length	**3**	**4**	**5**		**3**	**4**	**5**		**3**	**4**	**5**
Introduction	28	14	7		16	8	4		50	25	13
Materials and methods	52	26	13		24	12	6		80	40	20
Results and discussion	80	40	20		60	30	15		144	72	36
Conclusions	6	3	2		12	6	3		36	18	9

Dispersion/range is taken account to minimise the risk of one particular author’s preferences skewing the findings (Pan et al. 2016). Different studies have set a threshold ranging from 2% to 10% of their target texts (e.g. Biber & Barbieri 2007; Hyland 2008). In this study, we set 5% for 4-grams, 2.5% for 5-grams and 10% for 3-grams. We calculated the minimum number of texts for each sub-corpus, then rounded down to ensure we captured borderline bundles. For example, civil engineering has 425 texts, of which 10% is 42.5, but we rounded down to 40 (see table 4).

Table 4. Minimum number of texts per discipline for dispersion/range

Dispersion/Range	Civil Engineering	Electrical Engineering	Mechanical Engineering
3-grams (10%)	40	20	60
4-grams (5%)	20	10	30
5-grams (2.5%)	10	5	15

Lists of frequent lexical bundles were generated following the above procedure.

In order to explore the extent of a shared engineering register, the bundles shared across all three disciplines were identified, as were those shared by two disciplines. This was done first for whole texts by discipline, and then for each of the four IMRD sections. Frequent bundles unique to one discipline were not considered here. (A better means of identifying the unique features of a specific discipline would be through key words and key lexical bundles, but this is not a focus in this paper.[10])

6 Results

6.1 Word lists

We present here the most frequent 100 words in each of the three disciplines, Civil, Electrical and mechanical engineering.

Table 5. Engineering word list

72 words shared by all three disciplines ranked by size (pmw)			16 frequent words shared by 2 disciplines	
1	the	71491	study	Civil & Mechanical
2	of	38389	during	Civil & Mechanical
3	and	27804	m	Civil & Mechanical
4	in	23636	no	Civil & Mechanical
5	to	22228	table	Civil & Mechanical
6	a	19867	they	Civil & Mechanical
7	for	11017	studies	Civil & Mechanical
8	is	10509	b	Electrical and Mech.
9	that	8607	c	Electrical and Mech.
10	with	8134	system	Electrical and Mech.
11	as	6833	d	Electrical and Mech.
12	by	6488	shown	Electrical and Mech.
13	on	6049	its	Electrical and Mech.
14	was	5843	values	Electrical and Mech.
15	are	5488	information	Electrical and Mech.
16	this	5329	network	Civil and Electrical
17	were	5005		
18	from	4918	**Remaining words unique to each discipline**	
19	we	4785	road	Civil
20	be	4620	traffic	Civil
21	at	4009	areas	Civil
22	s	3771	area	Civil
23	an	3596	species	Civil
24	or	3583	roads	Civil
25	which	3017	distance	Civil
26	not	2897	use	Civil

[10] A frequent word list (or a frequent bundle list) tells us which items are present, which is useful for characterising a register, and also for teaching it. In contrast, a key word (or key bundle) list tells us what is prominent in the data that is not prominent in the reference corpus, and so varies according to the reference corpus used.

27	between	2832
28	can	2577
29	it	2545
30	model	2535
31	time	2459
32	data	2346
33	each	2264
34	these	2261
35	have	2199
36	fig	2184
37	used	2041
38	two	2001
39	all	1931
40	using	1911
41	g	1846
42	more	1843
43	figure	1828
44	than	1805
45	e	1780
46	p	1777
47	also	1760
48	one	1728
49	our	1724
50	different	1695
51	i	1669
52	has	1644
53	when	1633
54	their	1580
55	results	1511
56	other	1502
57	such	1479
58	based	1441
59	number	1439
60	journal	1418
61	been	1387
62	both	1385
63	t	1361
64	but	1330
65	only	1309
66	pone	1271
65	high	1248
68	may	1241
69	analysis	1233
70	however	1179
71	n	1095
72	where	1088

population	Civil
within	Civil
models	Civil
most	Civil
there	Civil
vehicle	Civil
km	Civil
habitat	Civil
higher	Civil
risk	Civil
density	Civil
level	Civil
circuit	Electrical
neurons	Electrical
cells	Electrical
cell	Electrical
state	Electrical
activity	Electrical
input	Electrical
current	Electrical
circuits	Electrical
voltage	Electrical
noise	Electrical
frequency	Electrical
response	Electrical
low	Electrical
function	Electrical
synaptic	Electrical
if	Electrical
Rate	Electrical
See	Electrical
robot	Mechanical
force	Mechanical
control	Mechanical
motor	Mechanical
task	Mechanical
first	Mechanical
after	Mechanical
human	Mechanical
f	Mechanical
same	Mechanical
subjects	Mechanical
into	Mechanical
participants	Mechanical

The very high frequencies of *the* and *of* are typical of academic writing. They are also the most frequent words in general English (Mudraya 2006: 251), but the frequencies in engineering are higher. This makes sense in relation to Halliday's accounts of the important role of nominal groups and grammatical metaphor in scientific writing.

Post-modification of nouns with prepositional phrases is common. As described in detail in Biber et al. (1999: 365), six prepositions account for most of the post-modification, with *of* the most frequent accounting for 60-65% of all post-modification across registers, followed by *in* (8-10%), and *for, on, with* and *to* (3-5% each). This order of frequency is reflected in our word list, with *in, to, for* and *with* all in the top ten, though *on* is 13th and preceded by *as* and *by,* as it also is in Mudraya's engineering corpus, unlike the general English lists (Mudraya 2006: 251). An exploration of how these occur in lexical bundles will shed more light in this area.

Frequent words in the top 50 general word lists (e.g. Mudraya 2006: 235) that are absent from our list include most pronouns and possessive adjectives (I, you, he, she, her, his), certain modals (would, could, should) and communication verbs (said). Such items tend to be frequent in spoken conversation and in written fiction, and help us identify the distinctive nature of academic registers.

Also of interest in table 5 are the groups of words from the top 100s that are specific to one discipline. From these we learn that articles in civil engineering may relate to roads, traffic, population and density; that articles in electrical engineering may relate to neurons, cells, currents, voltage and circuits; while articles in mechanical engineering may relate to robots, motors, control, humans and subjects. This provides insights into the activity focus, or Field, of the three disciplines in our study.

6.2 N-grams

AntConc yielded around 40 4-grams for each discipline, which includes items such as "journal pone t table" referring to a missing table in the journal PLOS ONE. Although this is important information, we have omitted it here to focus on the wording of the original text.

Table 6. Frequencies of the ten 4-grams shared by all three disciplines

4 grams	Civil pmw	Civil n	Electrical pmw	Electrical n	Mechanical pmw	Mechanical n	Total pmw	Total N
1 as a function of	116	239	208	295	172	660	163	1194

2 on the other hand	92	190	116	165	121	467	112	822
3 in the case of	58	120	125	177	102	393	94	690
4 as well as the	71	147	99	141	83	318	83	606
5 in the present study	56	117	42	59	90	345	71	521
6 a function of the	56	117	101	143	64	247	69	507
7 can be used to	47	98	64	91	63	243	59	432
8 the total number of	84	174	43	61	46	178	57	413
9 in this study we	49	101	50	72	52	198	51	371
10 it is important to	52	108	49	70	48	183	49	361

The bundles in table 6 are frequent and shared across the engineering disciplines. They suggest that engineering has to do with “function” and “use” “in the case of” and that authors refer to themselves as “we”, which reflects project teams and multiple authorship. Six of the ten, including the “function” bundles, are particularly frequent in electrical engineering.

13 4-grams are shared by electrical and mechanical (table 7), three are shared by civil and mechanical (table 8) and none are shared by civil and electrical. Only one of the remaining discipline specific 4-grams indicates Field – *based liquid metal ink* in electrical engineering.

Table 7. Thirteen 4-grams shared by electrical and mechanical engineering

	Electrical pmw	Electrical n	Mechanical pmw	Mechanical n	Total E&M pmw	Total E&M n
1. with respect to the	65	93	102	394	93	487
2. in the absence of	110	156	73	281	83	437
3. as shown in fig	78	111	73	280	74	391

4. in the presence of	130	185	53	202	74	387
5. in the context of	62	88	53	203	55	291
6. in this paper we	62	88	50	191	53	279
7. the performance of the	46	66	51	195	50	261
8. the size of the	53	76	48	184	50	260
9. the fact that the	50	72	48	184	49	256
10. as shown in figure	61	87	40	153	46	240
11. it is possible to	57	82	41	157	46	239
12. has been shown to	43	61	43	164	43	225
13. is shown in fig	50	71	40	154	43	225

Three of these refer to figures (*as shown in fig, as shown in figure, is shown in fig*), which is interesting for several reasons. First, this seems to be a typical feature of electrical and mechanical engineering not shared by civil engineering. This does not necessarily mean that civil texts do not have figures, but rather that either they do not have as many figures or they do not refer to them as regularly in this way. Moreover, the use of the abbreviation "fig" seems to be widespread and acceptable in both disciplines. A manual inspection of papers in civil engineering indicates that they do include figures, referred to in different ways. For instance, in some papers there is an effort to not be formulaic by using the same expression every time, but to use "figure 1 shows", "in figure 2 we can see" as well as "as shown in figure 3". In other papers the references to figures are entirely non-integral, occurring in brackets as references.

Another bundle that has a textual function is *in this paper we* and again this seems to be more used in electrical and mechanical than in civil engineering.

Four bundles refer to context (*in the absence of, in the presence of, with respect to, in the context of*), where three follow the "in the x of" pattern so typical of academic writing.

Finally, it is worth noting that this group includes another quantitative bundle, *the size of the* in addition to *the total number of* in table 6. The relative lack of such bundles suggests that numbers and formulae may be preferred to wording.

Table 8. Three 4-grams shared by civil and mechanical engineering

	Civil pmw	Civil n	Mechanical pmw	Mechanical n	Total C&M wpm	Total C&M n
1. the results of the	48	99	53	203	51	302
2. as a result of	62	129	40	153	48	282
3. for each of the	43	89	44	169	44	258

The striking feature here is that "results" bundles are frequent in civil and mechanical engineering text, unlike in electrical engineering. It is also worth noting that there have been no bundles with *introduce/introduction*, *methods* or *conclude/conclusion* despite these also being variants of prescribed section headings.

Before we explore in more detail how these bundles are used and what they tell us about engineering registers, we present the results of the 4, 3 and 5-gram analysis from all of the disciplines across sections. The entries in bold appear in all three sections, albeit in varying frequencies. The normalised frequency, raw frequency and range data is available in the Appendix.

Table 9. Frequent lexical bundles by section

	Introduction	Materials and methods	Results and discussion
4-grams	one of the most in this study we in this paper we as well as the **on the other hand** is one of the is organized as follows paper is organized as in the context of studies have shown that	as a function of the total number of is the number of **on the other hand** as well as the at the end of	as a function of **on the other hand** in the case of a function of the in the present study it is important to as shown in fig the size of the
3-grams	**as well as** **the number of** **one of the** **in order to**	**the number of** **in order to** was used to **as well as**	**in order to** **the number of** **based on the** according to the

	the use of the effects of **due to the** **based on the** the effect of in this study in terms of in this paper such as the a number of according to the in addition to the development of the presence of a variety of there is a it has been	**based on the** according to the the effect of in terms of in this study each of the i e the **due to the** the sum of a set of the presence of used in the a function of approved by the to determine the were used to **one of the** part of the in the same used in this defined as the we used the in which the end of the the end of is based on the use of	**as well as** each of the in this study with respect to **due to the** shown in fig of the two **one of the** i e the as shown in
5-grams	of this study was to paper is organized as follows of this study is to this paper is organized as is one of the most is organized as follows section of this paper is organized of this paper is to the paper is organized as of the present study was	at the end of the study was approved by the the study was approved by as a function of the by the ethics committee of the ethics committee of the animal care and use committee approved by the institutional review	N/A

The method of development of the engineering research article can be inferred from the items in the three main sections shown on table 9.

6.2.1 N-grams in Introductions

The Introduction sections contain moves that are familiar to research article introductions across disciplines. They introduce the study (*in this study/paper we)* and indicate why it is worth investigating (*is one of the most*) or where there is an unresolved issue *(on the other hand*). They may also refer to previous studies (*studies have shown that*) and preview the organisation of the paper (*the paper is organized as follows*).

6.2.2 N-grams in Materials and Methods

The Materials and Methods sections explain the procedures used in the study, with the ethical approval particularly notable in the 5-grams. The other items in table 9 include expressions of use (*the use of, used in the, we used the, was/were used to*), expressions of number and size (*is the number of, the total number of, a set of, each of the, the sum of, part of the*), and expressions of function and relationship (*is based on the, as a function of the, in order to, according to, to determine the*). In her study of engineering textbooks, Mudraya (2006: 234) found that *use* was the most frequent word family with *used* being the most frequent form, followed by *use* and *using*. In relation to the expressions of number and size, Biber et al. (1999: 636) observe that many of the most common bundles in academic writing include an *of* phrase, and that many of these relate to size and amount (ibid.: 1015).

6.2.3 N-grams in Results and Discussion

The items in the Results and Discussion sections overlap significantly with those in the Materials and Methods, with the notable exception of *as shown in fig,* which reminds us that the results are generally presented in tables and figures. Mudraya (2006: 234) finds that *figure* is the most frequent content word form in her engineering corpus (compared with general English corpora). Biber et al. (1999: 1020) observe that few academic bundles include verbs, and that those with *shown* are notable exceptions. Others of note in table 9 include *with respect to, according to* and *it is important to,* which anticipate more specific content to complete the group or clause.

Before we look at further individual sections, it is worth noting that many of the bundles form part of long nominal groups, with the structure: *the x of the y,* as in these examples:

The case/ context/ development/ effect(s)/end/ number/ presence/ size/ sum / total number/ use of (the).

Such bundles can be tricky to use because of the rank shifting and nominalisation involved. As examples (1)-(3) suggest, a series of prepositional phrases (shown in italics) can make it difficult to identify the function of each phrase:

(1) The only indirect effect *of significant magnitude* was the effect *of the density of forest reserves on recent deforestation through deforestation before 2000*, which was strongly negative (-0.49). (Civil)
(2) We have further shown *in idealised 2D models* that the metric is robust *against the spatial resolution of the clinical recordings* and the size *of the search radius* used *in its calculation*, whilst *in realistic whole ventricular models* we have demonstrated it works well *in the case of highly complex and intramural scar anatomies*. (Electrical)
(3) This study aims to assess the extent to which accelerometers can be used to determine the effect *of robot-supported task-oriented arm-hand training*, relative to task-oriented arm-hand training alone, *on the actual amount of arm-hand use of chronic stroke patients in their home situation*. (Mechanical)

On the other hand, it is *of* structures such as *the effect of* in (1) that enable clauses to be construed as the Thing being assessed.

This intelligent guesswork of course is not enough, but it will guide our investigation of specific instances from corpus texts.

7. N-grams in context

7.1 Introductions in engineering

One of the most is typically found in the first sentence of the Introduction section, as in these examples (4)-(8):

(4) The Scandinavian moose population has been *one of the most* productive and most extensively harvested in the world since the 1960s (Civil)
(5) *One of the most* fundamental challenges in network science is to understand the impact of structural properties on network functionality. (Electrical)
(6) The blade is *one of the most* critical parts of an aviation engine, and a small change in the blade geometry may significantly affect the dynamics performance of the aviation engine. (Mechanical)
(7) Bottom trawling is *one of the most* efficient fishing activities, but serious and persistent ecological issues have been observed by fishers, scientists and fishery managers. (Mechanical)
(8) Sleep is *one of the most* important sources for regeneration of the body and “needs its integrity to allow the living organism to recuperate normally” [1]. Disturbed sleep can therefore be of consequence for immediate and long term health [2]. In today’s society a multitude of sources exist with the potential to disturb sleep, *one of the most* prevalent being environmental exposure from transportation. (Mechanical)

These first few sentences in the Introduction sections not only present the value of the area of study as “important”, “critical”, fundamental” et al., but also straight away state the problem to be addressed. This immediately responds to the journal’s instructions to focus on the science and the results, rather than spending too much time discussing the value of the project. Nevertheless, some attention may be given to demonstrating the value of the project.

An examination of a complete Introduction section reveals how this can be done, as well as further typical features, as in this instance from electrical engineering. The Introduction starts with background details about the topic, reports what previous studies have shown, then outlines the scope of the article. Throughout there is technical language (shown in bold), explicit mention of function and use (underlined), evaluative language with positive appraisal (shown in italics), and non-standard English (exemplified below).

(9) Introduction:

The **metal-semiconductor (MS) contact** and the **metal-oxide-semiconductor (MOS) capacitor** are the *most* useful device in the study of semiconductor surfaces and *essential* component in semiconductor device. **MS contact with rectifying characteristic** is *widely* used in **MESFETs, HEMTs,** optical sensors, and gas sensors. **MOS capacitor with voltage-controlled variable** is used in **MOSFETs** for **forefront high-density integrated circuits** [1]–[4]. Recently, Hydrogen has been *widely* used in hydrogen-fueled vehicles, medical treatment, chemical industry, and semiconductor fabrication. However, hydrogen-containing gases have the risk to cause explosion. Therefore, **the development of hydrogen sensors for real-time in situ detection** is *highly required*. A number of **palladium and platinum-based hydrogen sensors** have been demonstrated [5]–[22]. Among them, **MS diodes** [5]–[13] have been addressed to be *one of the most promising* devices. Hydrogen sensors employing **MOS diodes** have also been *extensively studied* [14]–[18]. In addition, Chiu et al. [19]–[22] reported a *new* **MSM hydrogen sensor with two multifinger Schottky contacts**. Unlike conventional **MS and MOS diodes**, a mixture of **palladium and silicon dioxide** (SiO2) is deposited upon the semiconductor layer. Compared to *commonly* used **MS and MOS diodes, M-MSM diodes** obtained *excellent* performance of high sensitivity. However, the current–voltage (I–V) curve represents the **diode current operated as sensor in N2**. **I-V curve for M-MSM diodes** differ from one for **MS diodes** in that the former exhibit the **multiple-step phenomenon**, while the latter are not. The reason of causing the **multiple-step phenomenon** is *very interesting* but there are no descriptions in Chiu et al. reported [22]. In this paper, characterization and modeling of **M-MSM GaAs diodes** were reported. The **φb and the A*** were determined by a deduced equations from the **I-V curve** that operated at various temperature. The carrier over both the metal-semiconductor barrier and the insulator-semiconductor barrier are considered simultaneously on the **thermionic emission process** that can be used to describe *well* the current **transport for M-MSM diodes**. With increasing the applied voltage, the number of **minority carrier** at the semiconductor surface is larger than of the **majority carrier**. The carrier recombination will be taken into consideration. Furthermore, a **composite current (CC) model** is developed to evidence the concepts.

The calculated results are in *good* agreement with the experimental ones. Finally, conclusions were made.

The language used to introduce the topic is technical from the outset, and makes no concession to non-specialists. If the reader does not know what *the metal-oxide-semiconductor (MOS) capacitor* or *metal-oxide-semiconductor (MOS) capacitor* is, they need to find out elsewhere, and are unlikely to understand why the *development of hydrogen sensors for real-time in situ detection* is so highly *required*. In addition to the noun+noun compounds such as *metal-oxide semiconductor capacitor* and grammatical metaphor such as *development* and *detection*, there are abbreviations (*M-MSM*), scientific symbols (*The φb and the A**), and technical vocabulary (*voltage*).. Even where the individual lexical items are familiar, the meanings construed can be technical, as in *multiple-step phenomenon,* or *majority carrier.* This is therefore not only a highly elaborated academic code, written for the educated reader with its long subjects and grammatical metaphor, but it is also a highly technical code, written by specialists for specialists. Hyland quotes a Mechanical Engineer he interviewed as saying "There are certain things one expects one's readers to know. It would be insulting to spell everything out for them" (Hyland 2000: 71).

The middle of the paragraph (9) reports previous research with references to other studies, shown in the square brackets typically using the agentless passive as in *have been demonstrated [5]-[22]* and *have also been extensively studied [14]-[18]* with multiple sources acknowledged numerically. Thus the citation is predominantly non-integral, with hidden authors, although *Chui* et al. are mentioned twice and do appear as active participants, *e.g. Chui et al. [19]-[22] reported'.* (9) finishes with a preview of the rest of the paper, introduced with *In this paper*.

Evaluative language is used throughout to indicate the merits of the study. As shown in italics in (9), it is all positive: *the most useful, essential component, highly required, one of the most promising, extensively studied, new, commonly used, excellent performance, very interesting, well, good agreement.* It is worth noting that there is seldom any justification given for these evaluations. They are mostly stated as facts. The overall effect of this positive appraisal is to create an impression of the great worth of the project, at least from the writers' perspective.

Finally, the non-standard English forms reflect the ELF context. We assume this text was written by non-native speakers of English, and that proof reading for accuracy in terms of standard English has not been deemed necessary. For example, there is omission of plural "-s" in “*are the most useful device*”, omission (^) of articles with singular countable nouns, e.g. “*and ^ essential component in ^ semiconductor device”*. While the text is largely comprehensible, and perhaps fully comprehensible to those who know the theory, there may also be ambiguous clauses. In *MS contact with rectifying characteristic is widely used*, it is not clear if *characteristic* should be singular or plural. Similarly, with unusual word choices, such as *addressed* in *MS diodes have been addressed to be one of the most promising devices* the meaning can be inferred, but an element of uncertainty remains which extends beyond that already there with the use of the passive (where? by whom?) to a lack of certainty about “how?” (is this a verbal or material process?).

Generally, however, introductions provide information about the topic, aims and practical motivation for the paper, as illustrated in example (10) from mechanical engineering.

(10) Our long-term goal is to enable a robot to engage in partner dance for use in rehabilitation therapy, assessment, diagnosis, and scientific investigations of two-person whole-body motor coordination. Partner dance has been shown to improve balance and gait in people with Parkinson's disease and in older adults, which motivates our work. During partner dance, dance couples rely heavily on haptic interaction to convey motor intent such as speed and direction. (Mechanical)

7.2 Materials and Methods in engineering

The Materials and Methods sections describe the experimental procedures used with agentless passives (e.g., *are considered*) and long nominal groups (e.g., *an analytic formula for the optimal current density as a function of radius*) as in this extract.

(11) Two methods for reactive power optimization are considered. The first method computes the optimal current distribution of a flat disc coil, resulting in an analytic formula for the optimal current density as function of radius. In the second method the locations of the current loops within the transmitter coil are predefined, and the objective is to calculate the optimal amplitudes. (Methods)

Here the relationship between components is described using bundles such as is/*as a function of* and *with respect to* as in italics in examples (12)-(22) taken from the Methods and Results sections of papers.

(12) Profiles *as a function of* distance,
(13) Spring constant *as a function of* the moving mass
(14) the optimal current density *as a function of* radius

(15) Instantaneous synaptic output *as a function of* net input when the head sweep oscillation is present
(16) As in the model, robots decide to stay at or leave a resource *as a function of* the density of robots already present.
(17) The test score *is a function of* the expression levels of 23 genes which are grouped into highly correlated terms reflecting biological processes or cell types [8]
(18) Due to the mechanical properties of connective tissue structures, we hypothesized that the forces applied by the non-instructed fingers will increase *as a function of* the degree of flexion of the instructed finger, and that the enslaved force will be related to the magnitude of the applied force by that finger.
(19) In addition, the values of the normalized critical depths zc/R and zs/R are provided *as a function of* the reduced acceleration γ/g.

(20) The data from Figure 1B were normalized *with respect to* their maximal value.
(21) The value of the unperturbed model *with respect to* this measurement is indicated by markers along the abscissa, at the top of each panel.
(22) Here we have shown how biochemical noise modifies the location of bifurcation points of the epigenetic landscape *with respect to* a noise-free system and the impact of this phenomenon for promoting the stability of phenotypic states.

Has been shown to in the Methods section (23)-(24) is used to justify the use of a particular piece of equipment or test, as in these examples. In the Introduction it generally refers to previous research, and in the Results generally refers to the findings. It is therefore not surprising that it is used frequently in electrical and mechanical engineering (table 7).

(23) Here we chose the TUG test in part because an interaction of breathing with locomotion would be clinically relevant to chronic respiratory disorders [28, 29] and because the TUG test *has been shown to* be sensitive to dual tasking. (Methods, Mechanics)
(24) Our approach was to bypass this experimental difficulty by using a simulated mobile robot (Madame). Of all possible implementations of visual sensors, in Madame we implemented variants of retinal morphologies with a "log-polar" distribution of photoreceptors (here, only cones) [21–23]. The log-polar geometry *has been shown to* model accurately the topographical (retino–cortical) mapping of retinal cells (cones or ganglion cells) to the geniculate body and the striate cortex (area V1) [16,24]. (Methods, Mechanical)

7.3 Results and Discussions in engineering

As mentioned earlier, the frequent bundles in the Results sections are similar to those in the Methods section, with the addition of references to (results in) figures, as in example (25) from electrical engineering.

(25) We first compared the time-dependent output of the standard energy model (i.e., lacking the integrator) against the output of the extended model. The stimulus contained directional motion for the first 120 seconds, followed by a stationary test pattern for the remaining stimulus period. The output of the models *is shown in Figure* 4; combining the four filters in the way described above results in a net energy value that, when averaged over the spatial dimension, can be visualised *as a function of* time. *As can be seen in the Figure*, the standard model maintains a constant rightward output during adaptation, and a constant non-directional output during the test interval. The extended model, *on the other hand*, shows an initial drop in directional energy at the start of the adapting period, and an "after-effect" lasting approximately 20 seconds at the cessation of adaptation; net energy signalled by the sensors is in the opposite direction to the adapting stimulus.

Here we also see *on the other hand* used to compare results.

7.4 Conclusions in engineering

There have been problems locating the Conclusions sections, but example (26) from civil engineering suggests that where they do occur they are short and include a summary of the findings (introduced here by *This study indicates*) and suggestions for further research (indicated here by *Further research should aim to*):

(26) *This study indicates* that in a rural high coverage context a 40% reduction of delivery sites will lead to a 7% loss of geographical access. Such careful reduction of delivery sites using GIS modelling methods has the potential to assist decision makers on where to concentrate scarce resources by creating higher volume settings. Although a small percentage of the population will suffer an increase in distance to health facilities, a policy change in the organization of obstetric services might provide overall improved childbirth care, particularly for the rural poor who preferentially use first-line facilities. *Further research should aim* to investigate the effects of the proposed policy adjustment in a limited geographical area. In particular, the effect of fewer strengthened delivery sites on maternal and newborn mortality should be assessed, and whether the loss of proximity affects institutional delivery coverage. (Civil Conclusion)

8 Conclusions

In order to explore the central question of whether a core academic engineering register exists, we focused on one genre in one specific context: peer reviewed English research articles in engineering in the PLOS ONE open access journal. A review of the literature suggests that other engineering genres, such as spoken lectures or written student assignments, would include features that reflect their specific cultural and situational contexts. Further research could explore the extent to which they overlap with the registers described here for PLOS ONE articles.

We explored whether engineering registers, related to a hard applied science, would reflect the features described by Halliday for the hard pure sciences of Physics and Biology, tempered with a focus on problem solving, and that this might be seen in the use of more material processes.

From the perspective of lists of frequent words and lexical bundles shared across three engineering disciplines, there is evidence that engineering is concerned with *function* and *use*. *As a function of* was found to construe the interaction between variables in the real world and in scientific calculations. *Can be used to* is one of the most frequent bundles (table 6), suggesting that this is not the only way of doing things. In other words, unlike in the pure sciences where authors are more concerned with developing theories, in the applied sciences there is a greater concern with how parts interact and what works. Such bundles occur particularly in the Materials and Methods sections as *was/were used to*, *used in the/this*, *we used the* and *the use of*.

In contrast, the process that was foregrounded in the Results section was *shown*, particularly in Electrical and Mechanical engineering, as in *as/is shown in fig/figure* and *has been shown to* (table 7) where the former is generally used with an internal sense as part of the argument of the paper, and the latter is found with an external sense when it refers to previous uses of equipment, for instance, although the participants might be abstract. Halliday discusses the use of *show* as a verb expressing an internal process (Halliday 2004:154-5) that increased in use as scientific writing became less personal, as, for instance, in impersonal projections such as *our results show that*… (155). In our frequent bundles it is invariably in its past participle form *shown* suggesting a further step in impersonalisation, but also a greater remove from any Agent or Goal. In our contexts, the Subject is generally a figure or a study, but it is not really the figure *per se* that shows, but rather the evidence in the figure. In these contexts, *show* includes its meaning of “displays” but goes beyond this to include an element of “demonstrates”. People who read the figures or studies are expected to understand the implicit conclusions. This gives *show* a mental sense, but the layers of Participants and participant roles are tricky to unpack.

The frequent references to figures indicates that the papers are multimodal, where the written text refers to figures and also, but less frequently, to information on tables. This aspect of scientific writing is now widespread (e.g. Kress 2000: “multimodality”; Canagarajah 2018:

"strategic alignment of semiotic resources"), extending the nature of scientific literacies beyond those described in Halliday's earlier work.

In terms of Tenor, the texts are also highly specialised. There is no evidence of the explanatory text that would be expected in engineering textbooks or lectures, and very little of the display of accepted knowledge that might be expected in some student writing. This is reflected in the technical terms used. The authors are experts writing for fellow experts with the aim of publishing their results for other Engineers across the world to see.

In terms of Field, this paper has focused on a shared language of engineering, and has explored this through articles from three engineering disciplines: civil, electrical and mechanical. The evidence suggests that while all three do share some features (table 7), the language of electrical is closer to mechanical than to civil (tables 8 and 9). Within each of these disciplines, there will be subfields of activity that share more language, such as Pistons or Robotics within mechanical engineering. Such registers would be closer to the language of individual texts along the cline of instantiation than those in the mid-level registers of civil, electrical and mechanical, which in turn are more diversified (as suggested by the top 100 words frequent in only one discipline (table 3) than a register of engineering more broadly. This means that an engineering register has to be construed as more abstract and general than the registers of specific disciplines or texts. A focus on lexical bundles in a multidisciplinary corpus has provided a useful tool to uncover such patterns across engineering texts.

Finally, perhaps the most striking features of the texts chosen to represent engineering is the context in which they were produced. From interviews (Canagarajah 2018) with Chinese academics in the US, including Engineers, it appears that being an effective communicator in academic English may have less to do with accurate grammar and more to do with multimodality, useful lexical bundles and genre awareness. All three are evident in the texts examined here that were produced, using English as a lingua franca, for a 21st century social context: open access, multiple author roles, genre prescribed, multimodal, focus on the facts rather than their value/impact.

Acknowledgment

We acknowledge three phrases of support: The initial research was conducted by both authors at Coventry University in summer 2017. A version of this paper was presented at the Second Halliday-Hasan Forum on Language: Registerial Perspectives on Disciplinary English Study, at the Guangdong University of Foreign Studies in December 2017. The published version of this paper has benefited from advice from reviewers and journal editors, Geoff Williams and Alex Xuanwei Peng.

References

Ackermann, Kirsten & Yu-hua Chen. 2013. Developing the Academic Collocation List (ACL)–A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes* 12. 235–247.

Alsop, Siân & Hilary Nesi. 2013. The summarising function of university Engineering lectures: A cross-cultural perspective. In Alasdair N. Archibald (ed.), *Multilingual theory and practice in applied linguistics*, 11–14. London: Scitsiugnil Press.

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

Anthony, Laurence. 2017. AntCorGen (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Biber, Douglas & Frederica Barbieri. 2007. Lexical bundles in university spoken and written register. *English for Specific Purposes* 26. 263–286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, Douglas, Susan Conrad, Randi Reppen, Pat Byrd, & Marie Helt. 2002. Speaking and writing in the university: a multidimensional comparison. *TESOL Quarterly* 36. 9–48.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.

Björkman, Beyza. 2008. English as a lingua franca of engineering: The morphosyntax of academic speech events. *Nordic Journal of English Studies* 7. 103–122.

Canagarajah Suresh. 2018. English as a spatial resource and the claimed competence of Chinese STEM professionals. *World Englishes* 37. 34–50. https://doi.org/10.1111/weng.12301

Conrad, Susan. 2017. From a plate of spaghetti to a cable-stayed bridge: increasing the impact of corpus linguistics in disciplinary education. Plenary presentation, Corpus Linguistics 2017, Birmingham: University of Birmingham. Available at https://www.birmingham.ac.uk/research/activity/corpus/events/2017/cl2017/index.aspx.

Cortes, Vivian. 2004. Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes* 23. 397–423.

Durrant, Philip. 2017. Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics* 38. 165–193. Doi.org/10.1093/applin/amv011

Flowerdew, John. 2015. John Swales's approach to pedagogy in Genre Analysis: A perspective from 25 years on. *Journal of English for Academic Purposes* 19. 102–112. Doi.org/10.1016/j.jeap.2015.02.003

Gardner, Sheena. 2012. Genres and registers of student report writing: An SFL perspective on texts and practices. *Journal of English for Academic Purposes* 11. 52–63. Doi.org/10.1016/j.jeap.2011.11.002.

Gardner, Sheena. 2016. A genre-instantiation approach to teaching English for specific academic purposes: Student writing in business, economics and engineering. *Writing and Pedagogy* 8. 117–144. DOI: 10.1558/wap.v8i1.27934.

Gardner, Sheena, Hilary Nesi & Douglas Biber. 2018. Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics* DOI – 10.1093/applin/amy005

Halliday, Michael A.K. 2004. *The language of science*. (Volume 5 in the Collected Works of Michael Halliday edited by Jonathan J. Webster). London/New York: Continuum.

Halliday, Michael A.K. & James R. Martin. (eds.) 1993. *Writing science: Literacy and discursive power.* Bristol & London: Falmer.

Huang, Ping-yu. 2017. Academic collocations in Computer Science research articles: A corpus-based study. Paper presented at Corpus Linguistics 2017, University of Birmingham.

Hyland, Ken. 2000. *Disciplinary discourses: Social interaction in academic writing*. Harlow, Essex: Pearson Education.

Hyland, Ken. 2004. Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing* 13. 133–151.

Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27. 4–21.

Kanoksilapatham, Budsaba. 2015. Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes* 37. 74–86.

Koutsantoni, Dimitra. 2006. Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes* 5. 19–36.

Kress, Gunter. 2000. Multimodality: Challenges to thinking about language. *TESOL Quarterly* 34. 337–340.

Martin, James R. 1990. Literacy in science: learning to handle text as technology. In Frances Christie (ed.), *Literacy for a changing world*. Hawthorn, Australia: Australian Council for Educational Research. Reprinted in Halliday & Martin 1993.

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

Matthiessen, Christian M.I.M. & Halliday, Michael A.K. 2009. *Systemic functional grammar: A first step into the theory.* Beijing: Higher Education Press.

Mudraya, Olga. 2006. Engineering English: A lexical frequency instructional model. *English for Specific Purposes* 25. 235–256.

Nesi, Hilary & Sheena Gardner. 2006. Variation in disciplinary culture: University tutors' views on assessed writing tasks. *British Studies in Applied Linguistics* 21. 99–117.

Pan, Fan, Randi Reppen & Douglas Biber. 2016. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes* 21. 60–71.

Rezoug, Fares & Benet Vincent. 2018. Exploring lexical bundles in the Algerian Corpus of Engineering (ACE). *Arab Journal of Applied Linguistics* 3. 47–77.

Rozycki, William & Neil H. Johnson. 2013. Non-canonical grammar in best paper award winners in engineering. *English for Specific Purposes* 32. 157–169.

Sun, Yu-chih. 2013. Do journal authors plagiarize? Using plagiarism detecting software to uncover matching text across disciplines. *Journal of English for Academic Purposes* 12. 264–272. http://dx.doi.org/10.1016/j.jeap.2013.07.002

Wang, Simon & John Flowerdew. 2016. Participatory genre analysis of statements of purpose: an identity focused study. *Writing and Pedagogy.* 8. 33–57. doi: 10.1558/wap.v8i1.29699

Wood, Alistair. 2001. International scientific English. In John Flowerdew & Matthew Peacock (eds.), *Research perspectives on English for academic purposes*, 71–83. Cambridge: Cambridge University Press.

Appendix: Supplementary Data

	Introduction							Materials and methods							Results and discussion						
	C		E		M			C		E		M			C		E		M		
	Raw freq.	Range	Raw freq.	Range	Raw freq.	Range		Raw freq.	Range	Raw freq.	Range	Raw freq.	Range		Raw freq.	Range	Raw freq.	Range	Raw freq.	Range	
4-gramS	50	45	19	18	45	40	one of the most	50	34	32	23	98	67	as a function of	167	51	227	79	478	147	as a function of
	44	39	23	20	55	47	in this study we	88	58	29	19	95	57	the total number of	119	84	87	51	247	150	on the other hand
	42	33	29	25	77	64	in this paper we	46	27	26	17	77	47	is the number of	84	50	120	51	207	113	in the case of
	38	32	29	22	62	53	as well as the	27	20	24	19	41	36	on the other hand	82	28	112	51	157	76	a function of the
	38	31	34	25	105	89	on the other hand	48	36	23	19	80	62	as well as the	74	37	39	19	228	103	in the present study
	36	34	20	19	44	42	is one of the	46	38	19	17	109	84	at the end of	61	49	40	31	94	80	it is important to
	27	27	17	17	35	35	is organized as follows								55	23	66	38	190	83	as shown in fig
	26	26	16	16	34	34	paper is organized as								49	30	41	27	98	68	the size of the
	22	22	20	17	44	43	in the context of														
	21	21	15	13	33	31	studies have shown that														
3-grams	184	135	104	64	287	196	as well as	545	201	191	68	532	195	the number of	215	111	246	117	607	244	in order to
	165	95	79	40	194	97	the number of	300	134	155	72	607	244	in order to	781	213	416	96	532	195	the number of
	157	124	80	59	178	142	one of the	209	131	95	51	326	182	was used to	212	132	169	94	360	199	based on the
	130	85	111	64	351	218	in order to	191	127	84	46	247	137	as well as	148	90	67	39	324	191	according to the
	103	80	63	42	168	130	the use of	314	176	76	49	360	199	based on the	272	151	247	117	247	147	as well as
	100	74	37	26	126	96	the effects of	167	115	72	46	324	191	according to the	89	59	84	46	232	143	each of the
	99	80	68	51	180	137	due to the	121	79	69	38	124	93	the effect of	283	129	90	55	219	133	in this study
	92	64	71	48	253	166	based on the	111	75	59	27	111	77	in terms of	103	61	126	60	201	110	with respect to
	91	62	37	26	141	91	the effect of	177	110	59	35	219	133	in this study	281	158	252	116	174	113	due to the
	82	70	45	36	112	87	in this study	181	104	55	29	232	143	each of the	117	46	212	58	170	75	shown in fig
	81	59	44	29	146	105	in terms of	92	63	54	33	150	101	i e the	62	46	99	67	155	88	of the two
	76	57	70	56	160	122	in this paper	140	99	52	40	174	113	due to the	203	117	141	84	153	112	one of the

	69	57	43	37	146	122	such as the	63	40	46	27	89	65	the sum of	85	55	71	40	150	101	i e the
	68	49	47	36	121	87	a number of	63	47	45	24	119	73	a set of	110	51	183	87	144	79	as shown in
	56	44	33	20	75	66	according to the	95	66	44	29	83	60	the presence of							
	56	47	30	26	76	68	in addition to	99	69	44	34	185	135	used in the							
	56	47	38	35	176	113	the development of	61	40	41	27	118	74	a function of							
	56	45	52	33	107	77	the presence of	112	105	41	38	191	173	approved by the							
	52	47	37	34	101	89	a variety of	80	62	41	28	124	89	to determine the							
	51	45	23	21	78	70	there is a	152	111	41	30	183	133	were used to							
	45	44	46	39	176	125	it has been	123	95	39	26	153	112	one of the							
								92	71	39	28	122	89	part of the							
								63	51	38	27	126	93	in the same							
								62	49	38	30	82	64	used in this							
								66	48	35	21	145	98	defined as the							
								156	94	35	22	102	78	we used the							
								60	44	33	24	134	92	in which the							
								61	45	29	23	150	107	end of the							
								79	59	27	21	171	113	the end of							
								60	44	26	22	91	64	is based on							
								77	53	24	21	108	78	the use of							
5-grams	26	26	16	16	34	34	paper is organized as follows	27	22	8	6	61	53	at the end of the							
	15	15	12	12	23	23	this paper is organized as	35	35	6	5	48	46	study was approved by the							
	14	14	7	7	21	20	is one of the most	13	12	13	11	28	20	as a function of the							
								31	30	10	9	21		animal care and use committee							

Address for correspondence

Professor Sheena Gardner

School of Humanities, FAH, Coventry University

Priory Street, Coventry CV1 5FB, UK

Sheena.gardner@coventry.ac.uk