Linguistica Silesiana nr 46/1, 2025
ISSN 0208-4228
DOI: 10.24425/linsi.2025.155076

TATIANA SZCZYGŁOWSKA
University of Bielsko-Biala, Poland
Institute of Modern Languages
ORCID: 0000-0002-5051-4080
tszczyglowska@ubb.edu.pl

A CORPUS-BASED ANALYSIS OF INTRODUCTORY *IT* ACROSS RESEARCH ARTICLE PART-GENRES IN PUBLIC HEALTH

Introductory *it* is an important persuasive device in academic prose. This study analysed the variability of this structure in 200 public health research articles across the Introduction-Method-Result-Discussion part-genres. Occurrences of introductory *it* were analysed for their section-specific frequency, matrix predicate semantics, syntactic composition, and inclusion of modifying features. The results show both similarities and differences across the sections, reflecting their individual communicative purposes. For example, while all sections favour adjectives and mental and communication verbs as matrix predicates, the Methods section has a high proportion of predicates expressing dynamic and deontic modality. The findings advance our understanding of the rhetorical mechanics of academic prose, highlighting the strategic use of introductory *it* to fulfil section-specific functions. Pedagogically, the study offers practical insights for academic writers and educators who can use the structure to enhance disciplinary argumentation, rhetorical effectiveness and genre awareness in research writing.

Keywords: introductory *it*, academic writing, research article, part-genres, corpus-based

1. Introduction

As an important persuasive device that serves an evaluative function (Hunston 2011: 139), introductory *it* occurs frequently in academic writing (e.g. Biber et al. 1999; Zhang 2015), where it helps to specify and elaborate knowledge claims while enhancing their validity (Charles 2004) and de-emphasizing the writer's involvement (Hewing and Hewings 2002; Zhang 2015). However, the structure is often challenging for non-English authors (Rowley-

Jolivet and Carter-Thomas 2005). Aiming to contribute to the design of English for Academic Purposes (EAP) teaching materials, previous studies analysed introductory *it* in academic prose in terms of its functions (e.g. Herriman 2000b; Charles 2004; Larsson 2017), syntactic and lexical variability (e.g. Larsson 2016; Larsson 2019), and also the meaning of matrix predicates (e.g. Zhang 2015). These features of introductory *it* were examined across disciplines (e.g. Peacock 2011), genres (e.g. Groom 2005), discourse modes (e.g. Rowley-Jolivet and Carter-Thomas 2005), writer's linguistic backgrounds (e.g. Mur-Dueñas 2018) and academic expertise (e.g. Hewings and Hewings 2002; Larsson 2016). The findings suggest that the use of introductory *it* is both genre- and discipline-specific, varies with language proficiency and academic writing competence, but is less affected by native-speaker status.

This study draws on selected approaches to introductory *it* adopted in previous research, but it looks at the structure from an in-text perspective, analysing it by research article (RA) part-genres. It has been recognised that each RA section has different communicative goals (e.g. Swales and Feak 2012; Casal et al. 2021), which results in each using, to varying degrees, the rhetorical and linguistic resources typical of academic writing (Swales and Feak 2012: 287). This may also be the case for introductory *it*, but little is known about how the use of this structure varies across RA part-genres and contributes to their specific rhetorical goals. To gain insight into these issues, this study examines the frequency, matrix predicate semantics, syntactic composition, and the incorporation of modifying features into the structure in the Introduction-Method-Result-Discussion (IMRD) of public health RAs.

2. Theoretical background

2.1. RA part-genres

RA part-genre writing practices have attracted scholarly attention mainly in terms of highlighting the overall linguo-rhetorical construction of one specific part-genre at a time (e.g. Introduction: Samraj 2005, Del Saz Rubio 2011; Methods: Bruce 2008, Cotos et al. 2017; Results: Thompson 1993, Bruce 2009; Discussion: Parkinson 2011, Basturkmen 2012), or of consecutive sections in the whole RA (e.g. Kanoksilapatham 2005, Casal et al. 2021). Driven by the belief that each RA section is geared towards its specific communicative goals, these investigations have established, as summarised by Zhang (2022: 2), that:

> Introductions outline the context, motivation, and rationale of a study; Methods contextualize, describe, and justify the methodological choices; Results report and comment on the findings; and Discussions go beyond the factual exposition of the results by interpreting, justifying, and discussing the implications of the results.

To this end, Introductions are interpersonal and move from general to specific; Methods and Results are more factual and descriptive, although comments on findings are sometimes woven in the Results sections; and Discussions are interpretive and move from specific to general (Casal et al. 2021. Zhang 2022).

Fewer studies have examined how the use of one particular linguistic feature varies to meet the communicative goals of authors, which change as they move through the subsequent rhetorical stages of RA. Notable exceptions include a cross part-genre analysis of, for example, *we* use in biology RAs (Martínez 2005) or citation patterns in social science RAs (Zhang 2022). However, the section-specific use of introductory *it* has received less scholarly attention, although a cursory insight into this matter has been provided by Moghaddam and Ebrahimi (2019), who focused on RAs in applied linguistics.

This paper therefore examines the structure across RA part-genres in the discipline of public health, which was chosen for three reasons. First, it is concerned with the protection and improvement of human health and is a socially important field of research. However, its disciplinary writing practices have not been extensively analysed. Second, both the number of active journals and journal articles in public health increased rapidly between 2000 and 2012, but publication activity was dominated by English-speaking countries (Donner et al. 2014: 11). This suggests that there is a need to provide a wider group of authors with an insight into how a public health RA should be written. Third, the public health journals selected as the source of material in the present study offer free online access to the RAs, with each article divided into well-defined rhetorical sections that are crucial for the analysis.

2.2. Introductory *it*

The term introductory *it* (Peacock 2011) refers to the structure that begins with the impersonal pronoun *it*, which acts as the subject of the matrix clause, and continues with a matrix predicate (e.g. Herriman 2000a) or predicate lemma (Larsson 2016), followed by an extraposed subject clause in the form of a finite or non-finite *that*-clause, *wh*-clause, *to*-clause or *-ing* clause. Other labels found in the literature include subject extraposition (e.g. Quirk et al. 1985), anticipatory/preparatory/extrapositive *it* (e.g. Rodman 1991), pattern with introductory *it* (e.g. Hunston and Francis 2000), *it*-clauses (Hewings and Hewings 2002), subject *it*-extraposition (e.g. Zhang 2015), the introductory *it* pattern (e.g. Larsson 2016), or the anticipatory *it* pattern (e.g. Mur-Dueñas 2018). The pronoun *it* does not convey much information because it does not refer anaphorically to a previously mentioned referent, but it does refer cataphorically to the clausal subject (Quirk et al. 1985: 349), which carries the main

informational content of the whole message and which is commented on by the matrix predicate, which in turn "carries the most semantic content" of the structure itself and influences how the extraposed information should be interpreted (Larsson 2016: 65). This is illustrated in example (1), where the subject pronoun *it* is followed by the semantically neutral linking verb *is*, after which comes the most semantically loaded element of the matrix predicate, the deontic adjective *necessary*, followed by the extraposed infinitival subject clause *to compare our study to other ones*, which conveys the main propositional content of the message.

1. *[…], it is nevessary to compare our study to other ones.* (Discussion)

Although introductory *it* has many different realisations or variants, also known as subpatterns (Larsson 2016), its main function is to "express opinions and to comment on and evaluate propositions in a way that allows the writer to remain in the background" (Hewings and Hewings 2002: 368). This is made possible by the presence of the impersonal *it* in the subject position, so that the evaluation conveyed by the following matrix predicate, although personal and subjective, is given thematic status but is not attributed to any particular source and is therefore seen as an objective rather than a contestable viewpoint. This effect is reinforced by the fact that the newsworthy aspect of the message is extraposed, that is shifted to the end of the sentence, where it is regarded as new discourse material, albeit presented from the perspective of the thematic element. This clever way of structuring the information in a sentence ensures that the reader's full attention is focused on the postponed element, and the writer's stance on its content is explicit and thematically foregrounded, but their responsibility for that stance is hidden behind the dummy *it*. All this together gives the message an air of objectivity and authority. The result is that new knowledge is presented and the reader is imperceptibly persuaded of the validity of the claims made about it.

Even if academic writers are often advised to avoid the structure because of its "willful and nasty" character (Rodman 1991: 18), the above characteristics of introductory *it* contribute to its communicative value, which lies in the potential to claim "objective necessity or certainty for what in fact may be a matter of opinion" (Collins 1994: 20). Another reason for the popularity of introductory *it* in scientific writing (Herriman 2000b) may be the ease with which it allows long and complex discourse elements relating to the phenomena under discussion to be woven into a sentence. The complicated clause is simply placed at the end of the message, where it has space to be fully developed and where it is most likely to be remembered.

As a "grammatical feature of metadiscourse", the structure also has "four main interpersonal roles in hedging, marking the writer's attitude, emphasis, and attribution" (Hewings and Hewings 2002: 367). It thus provides a useful

means of establishing a bond with the reader and engaging them in discourse interaction, and of indicating the writer's identity and involvement in the scholarly debate taking place in the RA at hand (Hyland 2019). As Quirk et al. (1985: 1114) note, comment clauses such as *It is reported* or *It could be argued* act as hedging devices, allowing the definiteness of claims to be softened. Hewings and Hewings (2002: 370-373) add that the choice of *it* rather than a personal pronoun as the subject of the matrix clause makes it possible to depersonalise and thus objectify the evaluation of the embedded proposition encoded in the matrix predicate, for example, *It is crucial* or *It is of interest*. The structure can also be used by the writer to draw attention to a nuance, as in *It should be noted*, or to emphasise their position, as in *It is clear*, or to "lead the reader to accept the writer's judgements as being soundly based", as in *It has been proposed* (Hewings and Hewings 2002: 373). The persuasive power of introductory *it* cannot therefore be denied, and since it is "a frequent structure in the RA, where it regularly occurs as semi-formulaic 'lexical bundles'" (Rowley-Jolivet and Carter-Thomas 2005: 51), it is worth exploring how its versatile properties contribute to each rhetorical stage of RA writing.

3. Methodology

3.1. The corpus

The corpus compiled for this study includes a total of 200 RAs published in four leading public health journals: *Population Health Metrics* (PHM), *The Lancet Public Health* (LPH), *Journal of Global Health* (JGH) and *Environmental Health Perspectives* (EHP). Each journal contributed 50 RAs from 2019, with the exception of PHM (one article from 2018 and one from 2020) and LPH (four articles from 2018), which published fewer articles in the selected year. The main body of each RA was stored in plain text format, stripped of notes, long quotations, bibliographies, tables and figures, and segmented into IMRD part-genres based on section headings. In some texts, the Discussion also included the research conclusion, which was presented separately in other texts. In the latter case, the text segments marked as 'Conclusion(s)' were coded as 'Discussion' in order to ensure consistency across the corpus. Table 1 summarises the composition of the corpus.

Table 1. Summary of the corpus

Part-genres	Words	Mean	Standard Deviation
Introduction	129,197	645.98	357.62
Method	287,878	1439.39	719.99
Results	218,264	1091.32	524.17
Discussion	312,361	1561.8	523.01
Overall	947,700	4738.5	652.03

3.2. Introductory *it* identification

Samples of introductory *it* in the corpus were extracted by using the concordance feature of WordSmith Tools v. 6. (Scott 2012), which returned 1414 instances of *it* and its co-text. The concordance lines were manually inspected to exclude invalid tokens. The manual inspection allowed for the inclusion of a full range of instances, including those with *that*-deletion (e.g. *it seems this concern was accurate*), as well as those in which the matrix predicates are not adjectives (e.g. *it is a challenge to identify*, *it was assumed that smoking*). Two instances containing multiple matrix predicates, as in (2), were excluded from the study because of their semantically ambiguous nature.

(2) *It is cost prohibitive and logistically challenging to measure exposures for large populations, […]* (EHP_Introduction)

3.3. Syntactic analysis

The syntactic analysis includes a general overview of the distribution of clausal forms (i.e. types of extraposed clauses) used in the introductory *it* structure across the sections. The clausal forms found in the corpus include *that*-clauses, *to*-infinitive clauses, *wh*-clauses and *ing*-clauses. As Zhang (2015: 10) argues, there is an interaction between the meaning of the matrix predicate and the type of clausal form, so the choice of a particular clausal form may turn out to be section-specific.

Additionally, in order to provide a more comprehensive overview of the syntactic composition of the structure across the sections, this study draws on Larsson's (2016) approach to the syntactic analysis of introductory *it* in learner and expert writing. Specifically, it follows the classification system used in the COBUILD grammars (Francis et al. 1996, 1998), which describe 45 different "patterns with *it*", as shown in (3) and (4).

(3) ***it* V ADJ to-inf:** *It is timely to conduct another review […]* (JGH_Introduction)
(4) ***it* V that:** *It could be that social intimacy is multidimensional […]* (JGH_Discussion)

One change introduced concerned the link verbs, which were included in the V category, comprising individual verbs (e.g. *it remains*) and verb groups (e.g. *it was suggested*). Also, using the classification method of Hunston and Francis (2000: 32-35), some additional, rare patterns were identified in the study corpus that were not listed in the COBUILD grammars. Instances with *worth* followed by an *-ing* clause were classified as belonging to the *it V ADJ -ing* subpattern. This was motivated by Quirk et al.'s (1985: 1230) and Biber et al.'s (1999: 174) suggestion that *worth* can be considered as an adjective, Huddlestone and Pullum's (2016: 607) assertion that although *worth* "is like a preposition […] overall the case for analysing it as an adjective is strong", and Hunston and Francis' (2000: 196) claim that "*worth* has an adjective-like meaning in spite of its preposition-like behaviour".[1] The abbreviations used in this study are listed in Appendix A, together with their explanations, while the full list of patterns identified in the IMRD part-genres but not shared across them is given in Appendix B.

3.4. Semantic analysis

Inspired by Zhang's (2022) approach to the semantic analysis of the matrix predicates in the introductory *it* structure in academic and popular writing, this study follows Biber et al.'s (1999) semantic classification of single-word (pp. 361–364) and multi-word verbs (pp. 408, 414–418) for verbal predicates. The semantic domains relevant to the present study include:

- Communication verbs involving different communication activities (e.g. *state*, *suggest*),
- Mental verbs expressing cognitive, emotional and perceptual meanings (e.g. *think*, *see*),
- Existence or relationship verbs, which report states or relationships between entities (e.g. *be*, *seem*),
- Activity verbs, which denote actions, events, static relations (e.g. *go*, *work*),
- Facilitation or causation verbs, indicating that a new state of affairs has been brought about (e.g. *cause*, *require*).

[1] Overall, the status of *worth* is unclear in the literature. For instance, Herriman (2000a: 597) classifies instances such as *worth noting* or *worth recording* as prepositional predicates expressing evaluation, while Peacock (2011: 82, 91) considers them as specific realizations of the *it v-link ADJ that* pattern.

Multi-word verb constructions, such as *keep in mind*, which "function semantically as a coherent unit that can often be replaced by a single lexical verb" (Biber et al. 1999: 427), were classified as individual verbal predicates in the semantic analysis. Since some verbs "can be used with different meanings belonging to more than one category", each matrix predicate was classified on the basis of its meaning in a given context (Biber et al. 1999: 361).

For non-verbal predicates, this study adopts Herriman's (2000a: 585–586) classification, which distinguishes four semantic domains of the matrix predicates. The (sub)categories relevant to this study are:

- Epistemic modality, concerned with the *Truth* (e.g. *it is evident/misleading*) of what is conveyed by the extraposed clause,
- Deontic modality, concerned with *Obligation*, i.e. the obligatory nature of what is conveyed by the extraposed clause (e.g. *it is essential/your duty*), and *Volition* with regard to its content (e.g. *it is desirable/my intention*),
- Dynamic modality, concerned with a participant's "ability or power to carry out a course of action" (p. 585) and subdivided into: *Potentiality* (e.g. *it is hard/a problem*), *Circumstance* (e.g. *it is cheap/early*), and *Human Attribute* (e.g. *it is prudent/vanity*),
- Evaluation, concerned with *General Evaluation* (i.e. (un)favourability; e.g. *it is nice/a pity*), *Appropriateness* (i.e. "correctness or suitability"; e.g. *it is natural/irony*), *Significance* (i.e. "degree of importance"; *it is right/an advantage*), *Frequency* (e.g. *it is rare/a custom*), or *Emotive Reaction* (e.g. *it is tempting/a relief*).

3.5. Modifying features investigated

As proposed by Larsson (2016), further insights into the use of introductory *it* in RA part-genres were obtained by considering the following modifying features[2] found in the structure: negation with *not* (e.g. *it is not possible*) and with negating prefixes (e.g. *it is unclear*), past tense (e.g. *it was proposed*), modal verbs (e.g. *it can be difficult*), optional adverbs (e.g. *it is also possible*, *it is particularly important*) or prepositional/noun phrases[3] (e.g. *it is a common misunderstanding, eg, in medical research, that*). An instance containing *not* and a negative prefix (i.e. *it is not uncommon*) was not counted as being negated because the two negating features together cancel the negation.

[2] Some of the features overlapped, as tokens such as *it would have been relevant to adjust for parental education* were considered as containing a past tense and a modal.

[3] The label 'optional prepositional/noun phrases' refers to phrases inserted between commas or dashes, but not to phrases that are an integral part of introductory *it*, as in *it is possible for a single tool to simultaneously address them*.

4. Results and discussion

4.1. Frequency

As shown in Table 2, introductory *it* is the most frequent in Discussions, followed by Introductions, Results and Methods. This trend is reflected in all three comparative measures. The clear prevalence of the structure in the Discussion part-genre is further confirmed when looking at the number of instances found in this section and those found in the other three sections (434 vs 229): the difference is highly statistically significant ($\chi^2(1)=63.38$, $p<0.0001$). This seems to be a feature of disciplinary writing in public health, as the finding contrasts with that of Moghaddam and Ebrahimi (2019) for applied linguistics RAs, where introductory *it* was the most common in Introductions.

Table 2. Frequency of introductory *it* across RA part-genres

	Introduction	**Methods**	**Results**	**Discussion**	**Total**
Raw frequency	97	61	71	434	663
Frequency per 10,000 words	7.5	2.1	3.2	13.9	7.0
%	14.6	9.2	10.7	65.5	100

The greater prominence of introductory *it* in Discussions may be explained by its potential to encode the author's knowledge claims through the means of impersonal objectivity (5). As a result, the interpretations, justifications and explanations of the findings for which the section provides discursive space are less contestable and thus more readily accepted by the reader.

(5) *Based on these results, it seems that municipalities are more suitable than parishes to analyze mortality by all cancers in mainland Portugal.* (PHM_Discussion)

4.2. Semantic analysis

The following reports on the variability of the semantic composition of introductory *it* across sections.

4.2.1. Verbal predicates

As shown in Figure 1, there are two main semantic categories of verbal predicates across sections: Communication verbs and Mental verbs, which confirms Zhang's (2015: 8) observations on academic writing. It can be seen that Communication verbs (e.g. *note*, *suggest*) have higher proportions in Results and

Discussions, where they largely dominate over the other semantic types. The clear prevalence of Communication verbs in Discussions is probably due to their potential to express the writer's stance in a way that allows it to be presented as someone else's proposition as in (6). This is particularly useful in the section where the extraposed propositions are to be negotiated with the reader, who is guided "from acceptance of the relatively uncontroversial data to acceptance of the writer's knowledge claim" (Parkinson 2011: 164). In Results, the increased presence of Communication verbs can be explained by the writer's increasing need to make it clear that s/he "has uttered the proposition before and is uttering it again in order to re-assert her/his claim" (Fetzer 2011: 261). This can be seen in (7), where the research results are reported twice: visually in a figure and in writing in the form of a sentence. Mental verbs (e.g. *assume*, *estimate*) have higher proportions in Introductions and especially in Methods, where they are used respectively to create a research context through a literature review, as in (8), and to detail the procedural steps, as in (9). The sections also differ in terms of the preferred verbal predicates belonging to the two semantic categories. The most frequent Communication verbs are: *suggest* and *note* in Introductions, *recommend* in Methods, *note* in Results and Discussions; while the most frequent Mental verbs are: *estimate* in Introductions, *assume* in Methods, *see* in Results, *estimate* and *know* in Discussions.

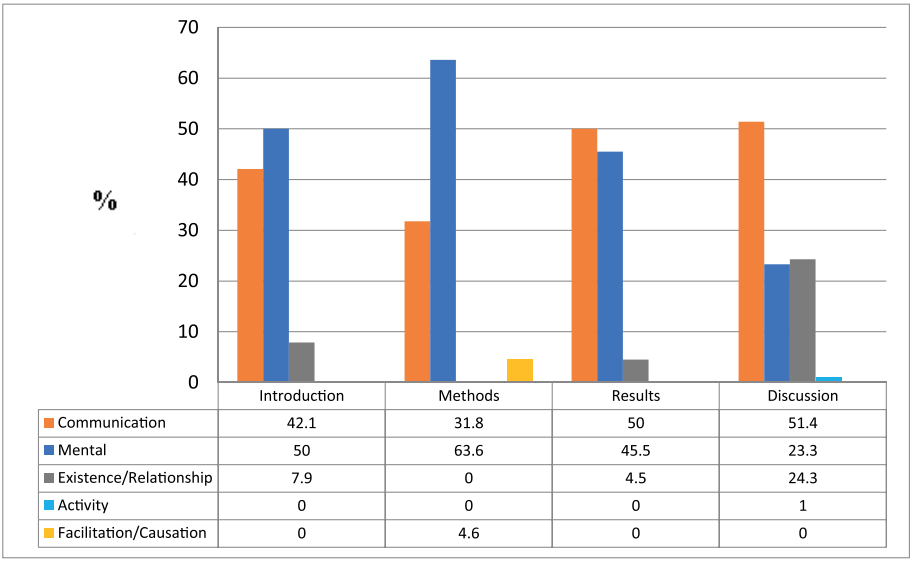
(6) *It is well documented that T3 activates TRs to transcriptionally regulate gene expression required for myelination […]* (EHP_Discussion)
(7) *In Fig. 1, it is noted that the level of ASMR by the UN estimate is obviously higher than that from the 2014 Census […]* (PHM_Results)
(8) *It is estimated that approximately 40 000 cases among inpatients are potentially underdiagnosed each year in Europe [5].* (JGH_Introduction)
(9) *For simplicity, it is assumed that every cell has a non-empty subset of members […]* (PHM_Methods)

Verbs of the other semantic types are less common and are not used in all RA sections. Single occurrences of Activity verbs were found in Discussions (10), and of Facilitation or Causation verbs, in Methods (11). Existence or Relationship verbs (12) were slightly more frequent. More than 72.4% (N=21) of these verbs were the copulas *be* and *seem*, appearing in such common stance devices as *it seems that* and *it could/may be that*, expressing probability rather than facts, which are reported in Methods, where no Existence or Relationship verbs were found. It is also worth noting that Discussions, which involve "complex causal, conditional and purposive argument" (Parkinson 2011: 164), show the strongest reliance on different types of verbal predicates, the total number of which is the highest in this section (N=103 vs 38 in Introductions, 22 each in Methods and Results), as is the number of different semantic categories (N=4).

(10) *It follows that either 1) the study physicians did not accept a key diagnosis and altered it on the basis of the clinical record or […]* (PHM_Discussion)

(11) *[…]; and second, it was required that the CHW intervention was not yet implemented in most neighborhoods.* (JGH_Methods)

(12) *However, according to our findings, it seems this concern was accurate for some countries but not generalizable to others.* (EHP_Discussion)



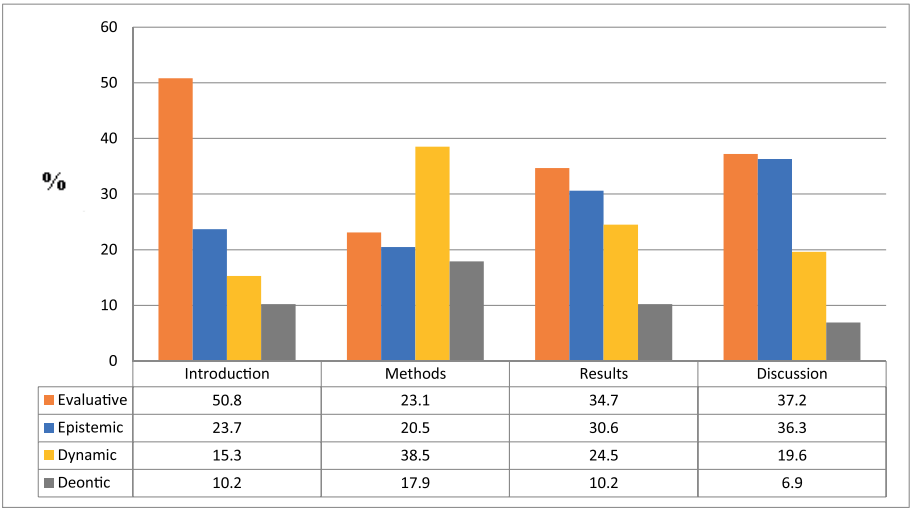
	Introduction	Methods	Results	Discussion
Communication	42.1	31.8	50	51.4
Mental	50	63.6	45.5	23.3
Existence/Relationship	7.9	0	4.5	24.3
Activity	0	0	0	1
Facilitation/Causation	0	4.6	0	0

Figure 1. The percentage distribution of the different semantic categories of the verbal predicates across sections

4.2.2. Non-verbal predicates

Figure 2 shows that, in Introductions, followed by Discussions and Results, the most common semantic category of non-verbal predicates is Evaluation, which expresses value judgements about propositions. This finding is consistent with Swales and Feak's (2012: 287) observation that the frequency of evaluative comments is high in Introductions and Discussions, variable in Results, but low in Methods. The presence of Evaluative predicates is particularly strong in Introductions, where they are often used to provide the rationale for the study as in (13). In Discussions, evaluative predicates are usually used to interpret the findings (14), while in Results, they are used to report what was found (15). In each section, the preferred Evaluation predicate is *important*, which expresses significance.

(13) *For these reasons it is important to evaluate the reliability of cause-of-death assignment and coding.* (PHM_Introduction)
(14) *It is interesting to point out that the clusters of the parishes present some similarities with the municipalities, […]* (PHM_Discussion)
(15) *It was also relevant to notice the two Low-High clusters, one joined to the North cluster and the other together with one of the other clusters.* (PHM_Results)



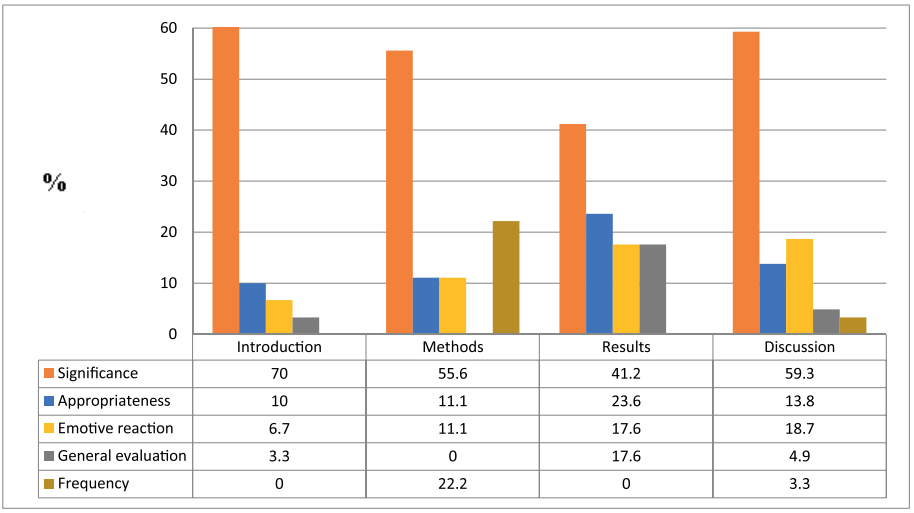
	Introduction	Methods	Results	Discussion
Evaluative	50.8	23.1	34.7	37.2
Epistemic	23.7	20.5	30.6	36.3
Dynamic	15.3	38.5	24.5	19.6
Deontic	10.2	17.9	10.2	6.9

Figure 2. The percentage distribution of the different semantic categories of the non-verbal predicates across sections

Consistent with Zhang (2015: 7), Significance is the most common Evaluation in academic writing, which is reflected in Figure 3 by the high percentages of the subcategory in all four RA sections. Appropriateness (16) and General Evaluation (17) dominate in Results, Emotive Reaction (18) is common in Discussions, and Frequency (19) dominates in Methods. The examples below show how the different subcategories of Evaluation serve the distinct communicative purposes of each section.

(16) *We think it is reasonable to consider the submission of a monthly report as a proxy for CHW activity level, […]* (JGH_Results)
(17) *To interpret the growing burden from NCDs among PLWH, it is therefore useful to compare it with projections on the population not living with HIV […]* (JGH_Results)
(18) *In future studies, it would therefore be of interest to investigate the metabolic responses to this intervention in more sedentary people […]* (TLPH_Discussion)

(19) *To increase the respondent's comfort with the interview, it was common for a community leader – [...] – to attend the introductory meeting [...]* (JGH_Methods)



	Introduction	Methods	Results	Discussion
Significance	70	55.6	41.2	59.3
Appropriateness	10	11.1	23.6	13.8
Emotive reaction	6.7	11.1	17.6	18.7
General evaluation	3.3	0	17.6	4.9
Frequency	0	22.2	0	3.3

Figure 3. The percentage distribution of the different semantic subcategories of Evaluation

Looking at Figure 2, it is also apparent that Introductions, Results and Discussions have the same percentage order of the four semantic categories of non-verbal predicates, with Epistemic modality being the second most common category, followed by Dynamic and Deontic modality. Epistemic predicates have the highest proportion in Discussions, followed by Results, Introductions and Methods. This finding only partially supports observations from other studies, which have shown that "in research articles, both the Discussion and the Introduction sections have a higher frequency of epistemic modality than the Methods and the Results sections do" (Yang et al. 2015: 2). However, this discrepancy may suggest that in public health RAs both the Discussion and Results sections share the characteristic of being heavily focused on conveying the writer's degree of confidence in the truth of propositions as in (20). Another possible explanation relates to Yang et al.'s (2015: 2) claim that "epistemic modality is especially frequent in sections analyzing phenomena or setting backgrounds", where the former would be the case in Results (21) and the latter in Discussions (22). The preferred Epistemic predicates are *unclear* in Introductions, *possible*[4] in Methods and Discussions, and *clear* in Results.

[4] Followed by a *that*-clause to mean 'probable, likely to happen' (see Herriman 2000a: 597).

(20) *As such, it remains possible that the true effects are different to what was estimated.* (TLPH_Discussion)
(21) *Because of the positive coefficients estimated, it is clear that an increase in GDP is associated with a higher diabetes rate in China.* (PHM_Results)
(22) *It is a known fact that, hospitals are amassing data at an unprecedented rate, […]* (PHM_Discussion)

As shown in Figure 2, Methods differs from the other sections in that the most common semantic category of its non-verbal predicates is Dynamic modality, and the section also has the highest proportion of Deontic predicates. The dominance of Dynamic modality can be explained by its potential to express what the thing referred to in the clause is actually able or disposed to do. This seems to fit perfectly with a factual account of the procedural steps described in the section. The majority of Dynamic predicates are concerned with Potentiality (23) and only one with Circumstance, and the preferred one is *possible*[5], similar to the other sections. The increased presence of Deontic modality in Methods may be due to the clockwork precision with which the choices made in the research process have to be described, as in (24). What Methods has in common with the other sections is the choice of *necessary* as the preferred Deontic predicate.

(23) *Therefore, it is possible to compute life expectancy at 60, but we have to truncate life expectancy at age 86.* (PHM_Methods)
(24) *[…] when calculating LE at a small-area level, it is necessary to consider sampling variation according to the occurrence of stochastic variation over time.* (PHM_Methods)

4.3. Syntactic analysis

The following reports on the variability of the syntactic composition of introductory *it* across sections.

4.3.1. Types of clausal forms

Figure 4 shows that all sections contain high percentages of *to*-clauses and *that*-clauses, but low or even zero percentages of the other clausal forms, which is in line with trends previously observed in academic writing (e.g. Biber et al. 1999; Zhang 2015). *To*-clauses are more common in Methods and Results, whereas *that*-clauses are more common in Introductions and Discussions. A possible explanation for this finding may be related to Charles' (2000) observation that *to*-clauses evaluate processes, whereas *that*-clauses evaluate

[5] Followed by a *to*-clause to mean 'practicable, can be done' (see Herriman 2000a: 597).

propositions and "often report static information in an impersonal manner" (Biber et al. 1999: 675). The preference for one clausal form or the other may therefore be due to the different rhetorical purposes of the sections. Methods and Results tend to focus on presenting the mechanisms of data collection and processing, as in (25) and (26), whereas Introductions and Discussions tend to focus on arguing and interpreting claims, as in (27) and (28). It should be noted, however, that the above preferences appear to be typical of RAs in public health, as it was reported by Moghaddam and Ebrahimi (2019), for example, that in applied linguistics all sections were dominated by *that*-clauses.

(25) *Once the model structure is defined, it is necessary to identify the transition probabilities by age among the different states being considered.* (PHM_Methods)
(26) *Based on these predicted biomarkers, it was possible to characterize highly exposed subpopulations.* (EHP_Results)
(27) *It is well established that HPV is the causative factor in most cases of cervical cancer [2, 3].* (PHM_Introduction)
(28) *It is possible that these uncertainties can be reduced based on updated toxicity data and dose-response analyses.* (EHP_Discussion)

All sections also contain *wh*-clauses, as in (29), introduced by *whether*, *if*, and *to what extent* in Introductions; *whether* in Methods; *whether*, *if*[6], *to what extent*, and *what* in Results; and *whether*, *if*, *to what extent/degree*, *how*, *why*, and *how many* in Discussions. In Methods, Results, and Discussions, there are some clausal forms introduced by *–ing* as in (30), which are "uncommon outside informal speech" (Quirk et al. 1985: 1393), while in Methods and Discussions there are some zero *that*-clauses, as in (31), which are generally rare in academic prose, where the retention of *that* is the norm (Biber et al. 1999: 680).

(29) *Pictorial information presentation featured in 20% of identified priorities, but it is unclear if these related to digital systems.* (JGH_Results)
(30) *It is worth highlighting here the short timeframe of the study period […]* (JGH_Discussion)
(31) *Finally, it could be argued there are other means of measuring the quality of birth registration data; […]* (PHM_Discussion)

4.3.2. Subpatterns of introductory *it*

In the corpus, there were 26 different subpatterns of introductory *it* among a total of 663 valid tokens. As many as 20 different subpatterns were found in Discussions, followed by 12 in Methods, 9 in Introductions, and 8 in Results. The

[6] Meaning 'whether'.

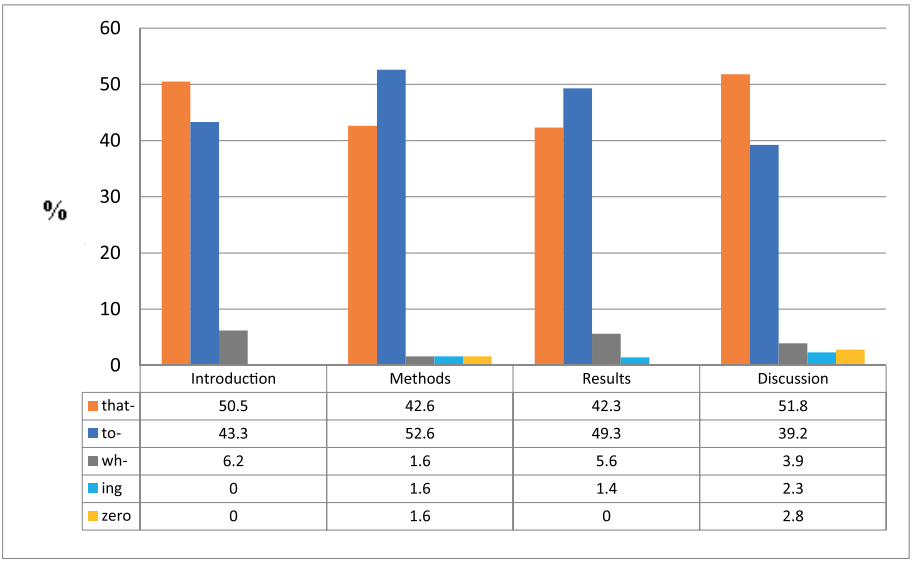


Figure 4. The percentage distribution of the different types of clausal forms across sections[7]

full list of the subpatterns identified in the corpus with their raw frequencies in the different sections can be found in Appendix B.

Table 3 shows the distribution of shared and section-specific subpatterns. It can be seen that section-specific subpatterns have the highest proportion, particularly in the Discussion, which is the section with the highest number of

Table 3. Distribution information on shared and section-specific subpatterns

Subpatterns found in # sections											
four	**No**	**%**	**three**	**No**	**%**	**two**	**No**	**%**	**one**	**No**	**%**
I, M, R, D	5	19.2	I, M, R	–	–	I, D	2	7.7	I	1	3.85
			I, M, D	1	3.85	I, M	–	–	M	3	11.5
			I, R, D	–	–	I, R	–	–	R	1	3.85
			M, R, D	3	11.5	M, R	–	–	D	10	38.5
						M, D	–	–			
						R, D	–	–			
Totals	5	19.2		4	15.4		2	7.7		15	57.7

[7] One clausal form introduced by *like* is not included in the figure: *Based on these results, it seems like the typical X59 death occurred in an elderly woman with an injury (fracture) in the hip or thigh region, dying in a nursing home.* (PHM_Results).

unique subpatterns. This finding supports Parkinson's (2011: 174) observation that the "section draws on a wide range of lexico-grammatical resources to achieve" its meanings. In terms of shared subpatterns, the majority are found in all four sections, followed by those shared by three sections, with the lowest proportion of shared subpatterns found in only two sections.

Pairwise comparisons revealed that the closest affinity was between Methods and Discussions, which shared nine subpatterns, although only *it V ADJ –ing* (32), *it be V-ed to-inf* (33) and *it be V-ed prep that* (34) were exclusive to these two sections. Methods also shared eight subpatterns with Results and six with Introductions, but all of them also appeared in at least one of the other sections. Results and Discussions shared eight subpatterns, none of which were unique to these two sections, while Introductions and Discussions also shared eight subpatterns, two of which were not found in the other sections: *it V that* (35) and *it V prep that* (36).

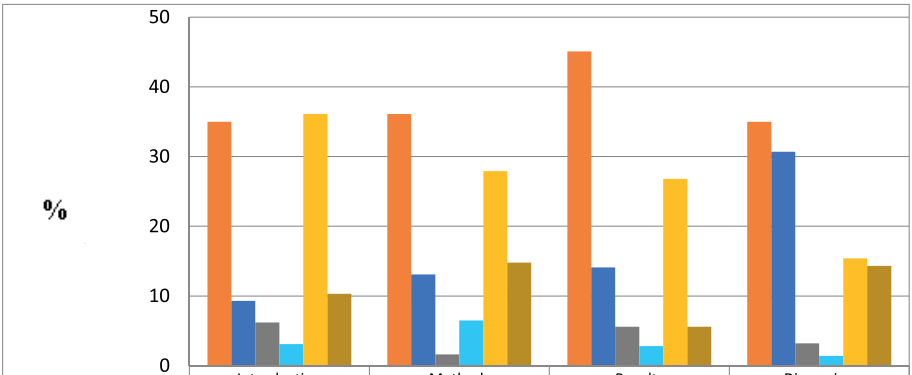
(32) *It is worth noting that the list of garbage codes in the GBD is very large, […]* (PHM_Discussion)
(33) *It is recommended to triangulate credible prior information to develop a single, final prior estimate […]* (PHM_Methods)
(34) *[…], it would stand to reason that the effect of MeHg in these cells is associated with protein degradation.* (EHP_Discussion)
(35) *Based on these results, it seems that municipalities are more suitable than parishes to analyze mortality […].* (PHM_Discussion)
(36) *[…] it is of note that HPV types 16 and 18 are implicated in more than 70% of cervical cancers in Australia.* (TLPH_Introduction)

The percentage information of the five subpatterns shared by all sections is shown in Figure 5. It can be seen that in each part-genre the subpatterns that are common to all four sections greatly outnumber the subpatterns that are either section-specific or found in only two or three sections. The five shared subpatterns account for 89.7% (87/97) of the total number of tokens in Introductions, for 85.2% (52/61) in Methods, for 94.4% (67/71) in Results, and for 85.7% (372/434) in Discussions. This suggests that the RA is a coherent whole, in which meanings are expressed by a limited set of lexico-grammatical features. However, its different part-genres draw on these resources to varying degrees to achieve their individual communicative purposes, as reflected in the section-specific distribution of the five shared subpatterns of introductory *it*.

Another interesting finding is that all four sections show a strong preference for those subpatterns in which the matrix predicates are adjectives, as in (37). This trend is reflected in Figure 5, where four of the shared subppaterns contain matrix adjective predicates, and in the fact that the subpatterns with adjectives generally stand out in each section. In Introductions, they account for 53.6% (52/97) of the total number of tokens, in Methods for 60.6% (37/61), in Results for

69% (49/71), and in Discussions for 73.5% (319/434). This confirms previous findings showing the dominance of adjectival predicates in academic prose (e.g. Collins 1994: 11; Zhang 2015: 6; Fišerová 2016: 42, Larsson 2019: 325). This may be due to their potential to "allow the writer to encode an evaluation" that influences how the following clause is to be interpreted (Hewings and Hewings 2002: 370).

(37) *In that context, it was not essential to work at a high mass resolution.* (EHP_Results)



	Introduction	Methods	Results	Discussion
it V ADJ to-inf	35	36.1	45.1	35
it V ADJ that	9.3	13.1	14.1	30.7
it V ADJ wh	6.2	1.6	5.6	3.2
it V ADJ for n to-inf	3.1	6.5	2.8	1.4
it be V-ed that	36.1	27.9	26.8	15.4
others	10.3	14.8	5.6	14.3

Figure 5. The percentage distribution of the five subpatterns shared across sections

The remainder of this section examines variation in the use of the five shared subpatterns to see how they meet the different rhetorical demands of each part-genre. As shown in Figure 5, the subpatterns *it V ADJ to-inf*, *it V ADJ that* and *it be V-ed that* stand out in each section, together accounting for 80.4% (78/97) of the total number of tokens in Introductions, 77% (47/61) in Methods, 85.9% (61/71) in Results, and 81.1% (352/434) in Discussions. This reflects the trends observed by Larsson (2016) in academic writing.

The *it V ADJ to-inf* subpattern, as in (38), has the highest proportion in Results, followed by Methods, Discussions, and Introductions. The four most frequent predicate adjectives used in the subpattern in each section, listed in order of frequency, account for 67.6% (23/34) of its tokens in Introductions: *important*, *possible*, *necessary*, *difficult*; for 81.8% (18/22) in Methods: *possible*, *necessary*, *important*, *difficult*; for 56.3% (18/32) in Results: *possible*, *important*, *useful*, *interesting*; and for 65.4% (100/153) in Discussions: *important*, *possible*,

difficult, *necessary*. While Introductions and Discussions show subtle variations in the proportion of preferred predicates and in the choice of specific adjectives, Methods and Results differ considerably in both respects, except for the prominence each gives to the adjective *possible*.

(38) *It is important to highlight the decrease in heterogeneity in life expectancy among states in 2015 compared to 1990.* (PHM_Results)

Another common subpattern is *it be V-ed that*, which dominates in Introductions, has high proportions in Methods and Results, but is relatively rare in Discussions. It is often used to formulate comment clauses that act as hedging devices, expressing "the speaker's tentativeness over the truth of the matrix clause" (Quirk et al. 1985: 1114), as in (39), so its low proportion in the Discussion section is somewhat counterintuitive. The top predicate verbs, with at least three occurrences in the subpattern in a given section, listed in order of frequency, are: *estimated*, *suggested*, *noted* and *shown* in Introductions; *assumed* in Methods; *noted* in Results; *shown*, *noted*, *reported*, *suggested*, *estimated* and *known* in Discussions.

(39) *It is suggested that diet-related health outcomes, such as obesity and diabetes, cannot fully be examined on the individual level [13, 14].* (PHM_Introduction)

The third relatively common subpattern in all sections is *it V ADJ that*, shown in (40), which has the highest proportion in Discussions, and is definitely less frequent in the other sections, especially in Introductions. The two most frequent predicate adjectives used in the subpattern in each section, listed in order of frequency, account for 55.6% (5/9) of its tokens in Introductions: *possible*, *plausible*; for 62.5% (5/8) in Methods: *possible*, *unlikely*; for 50% (5/10) in Results: *clear*, *evident*; and for 51.9% (69/133) in Discussions: *possible*, *likely*. It is interesting to note that *possible* was used in the subpattern only once in Results, despite being the preferred adjective in the other sections.

(40) *Generally, it seems plausible that the proposed methodology could be applied to different forms of medical research, e.g. clinical studies.* (PHM_Discussion)

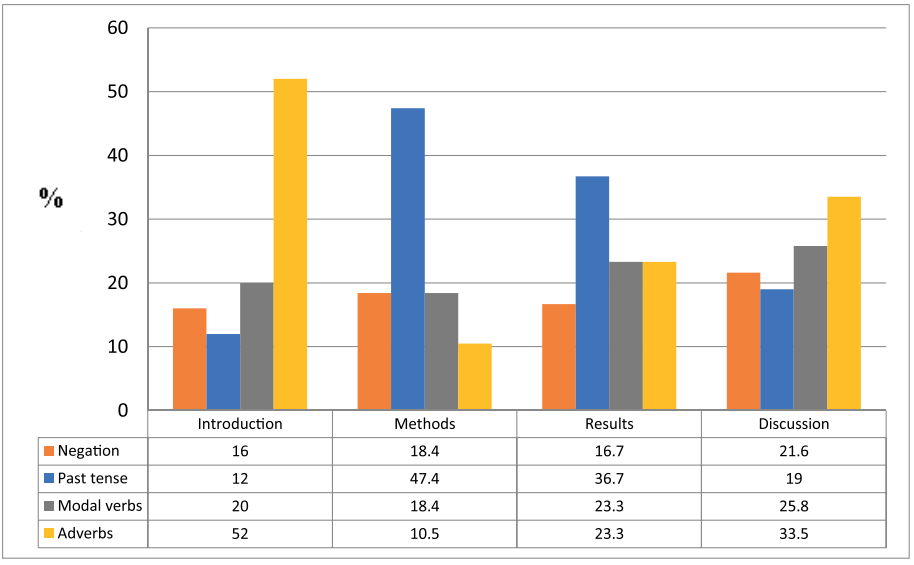
The subpatterns *it V ADJ wh* (41) and *it V ADJ for n to-inf* (42), although attested in all four sections, are generally rare. However, a trend in the use of the first subpattern was observed in the data: the preferred adjectival predicate in the sections is *unclear*. The typical ways in which the five shared subpatterns are used in each section are shown in examples (38) to (42). As can be seen, in Introductions and Discussions the subpatterns help to present arguments that make up the claim of the article, in Methods they serve to provide details of the procedural steps, while in Results they support the presentation of data.

(41) *It is unclear whether MeHg-induced toxicity is associated with the expression of HIF-1?, […]* (EHP_Introduction)
(42) *Indeed, it is possible for improvements overall to be accompanied by increasing rather than decreasing inequalities [21].* (PHM_Discussion)

4.4. Modifying features

Figure 6 shows that the largest number of features examined was found in Discussions (N=263), followed by Results (N=60), Introductions (N=50), and Methods (N=38). However, the percentage of tokens with at least one modifying feature is highest in Results (43/71, 60.6%), followed by Methods (30/61, 49.2%), Discussions (193/434, 44.5%) and Introductions (42/97, 43.3%). Thus, Results seems to be the section in which introductory *it* is most varied with additional elements. This is probably because, in addition to objective reporting of findings, this section tends to be used to justify methodological choices, to evaluate data, to explain, interpret and compare findings with previous work, or to highlight surprising findings (Thompson 1993). The advanced nature of this type of scientific argumentation often requires a variety of additional features.

Another distinct finding is the relatively high proportion of past tense tokens in Methods and Results and of adverbs in Introductions and Discussions. The



	Introduction	Methods	Results	Discussion
Negation	16	18.4	16.7	21.6
Past tense	12	47.4	36.7	19
Modal verbs	20	18.4	23.3	25.8
Adverbs	52	10.5	23.3	33.5

Figure 6. The percentage distribution of modifying features across sections[8]

[8] Three modifying features (i.e. optional prepositional/noun phrases) are not included in the figure: two in Methods and one in Discussions.

former two sections are oriented towards reporting what was done and how, and what results were obtained, as in (43) and (44), which may explain the pronounced presence of the past tense, also reported by Swales and Feak (2012: 287). An alternative explanation is that the past tense makes it possible to identify the responsibility for the details of the experiment "as that of the writer exclusively", who then becomes the main actor of the described study (Charles 2004: 82). The latter two sections are "interpretive and somewhat more abstract", and serve to "connect findings to a broader research community" (Casal et al. 2021: 8). Adverbs seem to support such meanings because of their potential to modify a wide range of expressions and their ability to provide a variety of additional information in the sentence, as in (45).

(43) *It was estimated that 46 villages were required to find 96 SAM cases.* (PHM_Methods)
(44) *Based on these predicted biomarkers, it was possible to characterize highly exposed subpopulations.* (EHP_Results)
(45) *It is also increasingly of interest to measure the change in key injuries over time and to identify priority areas for injury prevention.* (TLPH_Introduction)

A slightly stronger presence of negation in Discussions and Methods may be related to an increased need to indicate problems and limitations of the study without undermining the author's credibility, as in (46) and (47). The higher proportion of modal verbs in Results and Discussions may in turn be due to the fact that, as both sections revolve around the crucial aspect of the study reported in the RA – its findings – the writer may be particularly interested in intensifying the meanings they convey by using introductory *it*, as shown in (48).

(46) *Thirdly, MBDS data are anonymous, and it is impossible to identify whether a patient has been hospitalized more than once in different hospitals.* (PHM_Discussion)
(47) *Given the deficiencies in the CRVS system, it is not possible to obtain statistically meaningful and reliable mortality indicators from registered deaths.* (PHM_Methods)
(48) *When interpreting the coefficients it should be kept in mind that the model was fitted to optimize predictive precision and not to investigate the associations.* (JGH_Results)

Also, Discussions exhibit the most balanced frequencies of the modifying features. This may be due to the complexity of rhetorical goals that the section has to achieve on its "journey from data to claim", including "explanations about the cause of elements of the data, about the purpose of performing the experimental work in the way that it was performed, and about the conditions needed for the experiment to function" (Parkinson 2011: 174).

5. Conclusion

Using a corpus of public health RAs, this study examined the use of introductory *it* across IMRD sections. The results indicate that the use of the structure is not uniform within RAs: it occurs most frequently in the Discussion section and least frequently in Methods. This pattern appears to be discipline-specific, as it contrasts with findings from applied linguistics, where introductory *it* is most common in Introductions (Moghaddam and Ebrahimi 2019). Semantically, there is a general preference for mental and communication verbs, although each section demonstrates different lexical choices and degrees of usage. Non-verbal predicates also show variation, with Methods favouring Dynamic and Deontic predicates, and Introductions, Results and Discussions featuring a high occurrence of Evaluative predicates. However, evaluative meanings indicating Significance, usually through the adjective *important*, are common in all sections. Syntactic analysis shows that *to*-infinitive clauses dominate in Methods and Results, whereas *that*-clauses are more frequent in Introductions and Discussions. The latter appears to be typical of RAs in public health, as sections in applied linguistics, for example, tend to be dominated by *that*-clauses (Moghaddam and Ebrahimi 2019). Despite these differences, five recurring subpatterns of introductory *it* are present in all sections, mainly with adjectival predicates. Each section, however, adapts these patterns to its individual rhetorical purposes. Furthermore, Results shows the greatest variation in terms of the inclusion of different modifying features within introductory *it*, while Methods shows the least variation in this respect. Also, while Methods and Results show a high frequency of past tense tokens, Introductions and Discussions show a clear preference for tokens containing adverbs.

Overall, the analysis provides insights into discipline-specific writing practices in public health and enriches the general understanding of RA writing. The reported findings can therefore inform the design of materials for experienced and novice users of EAP, raising their awareness of the inter-dependence of the distributional and lexico-syntactic variability of the structure and the communicative purposes of RA part-genres. Because the study has covered only one discipline, future research can extend the proposed part-genre comparison to include RAs in other disciplines or to include other part-genre formats, such as combined Results-Discussion section.

References:

Basturkmen, H. 2012. A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes* 11: 134-144.

Biber, D., and S. Johansson, G. Leech, S. Conrad, E. Finegan 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Bruce, I. 2008. Cognitive genre structures in methods sections of research articles: A corpus study. *Journal of English for Academic Purposes* 7: 38-54.

Bruce, I. 2009. Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes* 28: 105-124.

Casal, J.E., and X. Lu, X. Qiu, Y. Wang, G. Zhang 2021. Syntactic complexity across academic research article part genres: A cross-disciplinary perspective. *Journal of English for Academic Purposes* 52: 1-12.

Charles, M. 2000. The role of an introductory *it* pattern in constructing an appropriate academic persona. In P. Thompson (ed.), *Patterns and Perspectives: Insights into EAP Writing Practice*, 45-59. CALS: The University of Reading.

Charles, M. 2004. *The construction of stance: a corpus-based investigation of two contrasting disciplines*. Unpublished PhD thesis, University of Birmingham.

Collins, P. 1994. Extraposition in English. *Functions of Language* 1(1): 7-24.

Cotos, E., and S. Huffman, S. Link 2017. A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes* 46: 90-106.

Del Saz Rubio, M.M. 2011. A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes* 30: 258-271.

Donner, P., and P-S. Chi, V. Aman 2014. *Bibliometric Study for German National Academy of Sciences Leopoldina in the Disciplines Public Health and Epidemiology*. Berlin: iFQ.

Fetzer, A. 2011. "I think this is I mean perhaps this is too erm too tough a view of the world but I often think...". Redundancy as a contextualization device. *Language Sciences* 33(2): 255-267.

Fišerová, H. 2016. The vagaries of subject *it*: can *it* serve as a style marker? *Linguistica Pragensia* 1: 35-47.

Francis, G., and S. Hunston, E. Manning 1996. *Grammar Patterns I: Verbs*. London: HarperCollins.

Francis, G., and S. Hunston, E. Manning 1998. *Grammar Patterns II: Nouns and Adjectives*. London: HarperCollins.

Groom, N. 2005. Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes* 4: 257-277.

Herriman, J. 2000a. Extraposition in English: A study of the interaction between the matrix predicate and the type of extraposed clause. *English Studies* 81(6): 582-599.

Herriman, J. 2000b. The functions of extraposition in English texts. *Functions of Language* 7(2): 203-230.

Hewings, M., and A. Hewings 2002. *It is interesting to note that…*: A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes* 21: 367-383.

Huddleston, R., and G.K. Pullum 2016. *The Cambridge Grammar of the English Language*. 9th edition. Cambridge: Cambridge University Press.

Hunston, S., and G. Francis 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: John Benjamins.

Hunston, S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London and New York: Routledge.

Hyland, K. 2019. Metadiscourse: Exploring Writing in Interaction (2nd ed.). Bloomsbury Publishing.

Kanoksilapatham, B. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24(3): 269-292.

Larsson, T. 2016. The introductory *it* pattern: Variability explored in learner and expert writing. *Journal of English for Academic Purposes* 22: 64-79.

Larsson, T. 2017. A functional classification of the introductory *it* pattern: Investigating academic writing by non-native-speaker and native-speaker students. *English for Specific Purposes* 48: 57-70.

Larsson, T. 2019. A syntactic analysis of the introductory *it* pattern in non-native-speaker and native-speaker student writing. In V. Wiegand and M. Mahlberg (eds.), *Corpus Linguistics, Context and Culture*, 307-338. Berlin/Boston: De Gruyter.

Martínez, I.A. 2005. Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing* 14(3): 174-190.

Moghaddam, R., and S.F. Ebrahimi 2019. Subject *it-extraposition* in applied linguistics research articles: Semantic and syntactic analysis. *Discourse and Interaction* 12(1): 29-46.

Mur-Dueñas, P. 2018. Exploring ELF manuscripts An analysis of the anticipatory *it* pattern with an interpersonal function. In P. Mur-Dueñas and J. Šinkūnienė (eds.), *Intercultural Perspectives on Research Writing*, 277-297. Amsterdam/Philadelphia: John Benjamins.

Parkinson, J. 2011. The Discussion section as argument: The language used to prove knowledge claims. *English for Specific Purposes* 30: 164-175.

Peacock, M. 2011. A comparative study of introductory *it* in research articles across eight disciplines. *International Journal of Corpus Linguistics* 16(1): 72-100.

Quirk, R., and S. Greenbaum, G. Leech, J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rodman, L. 1991. Anticipatory *it* in scientific discourse. *Journal of Technical Writing and Communication 21(1):* 17-27.

Rowley-Jolivet, E., and S. Carter-Thomas 2005. Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting. *English for Specific Purposes* 24: 41-64.

Samraj, B. 2005. An exploration of a genre set: Research article introductions in two disciplines. *English for Specific Purposes* 24(2): 141-156.

Scott, M. 2012. *WordSmith Tools (version 6.0)*. Stroud: Lexical Analysis Software.

Swales, J., and C.B. Feak 2012. *Academic Writing for Graduate Students: Essential Tasks and Skills*. 3rd ed. Ann Arbor, MI: University of Michigan Press.

Thompson, D.K. 1993. Arguing for experimental "facts" in science: A study of research article results sections in Biochemistry. *Written Communication* 10(1): 106-128.

Yang, A., and S. Zheng, G. Ge 2015. Epistemic modality in English-medium medical research articles: a systemic functional perspective. *English for Specific Purposes* 38: 1-10.

Zhang, G. 2015. *It is suggested that…* or *it is better to…?* Forms and meanings of subject *it*-extraposition in academic and popular writing. *Journal of English for Academic Purposes* 20: 1-13.

Zhang, G. 2022. The citational practice of social science research articles: An analysis by part-genres. *Journal of English for Academic Purposes* 55: 1-14.

Appendix A. List of abbreviations with explanations

Abbreviation	Explanation
ADJ	An adjective
adv	An adverb group
be V-ed	The lemma BE followed by a past participle
det	A determiner
it	Introductory *it*
-ing	A clause beginning with the *-ing* form of a verb
like	A finite clause beginning with *like*
n	A noun group
N	A noun
poss	A possessive determiner
prep	A prepositional phrase
that	A that clause
to-inf	A clause beginning with a *to*-infinitive form of a verb
V	A verb (group)
wh	A finite clause beginning with a *wh*-word/*to what extent*/*degree*

Appendix B. Patterns of introductory *it* in the IMRD part-genres with raw frequencies

Pattern	Introduction	Methods	Results	Discussion
it V ADJ to-inf	34	22	32	152
it V ADJ that	9	8	10	133
it V ADJ wh	6	1	4	14
it V ADJ for n to-inf	3	4	2	6
it be V-ed that	35	17	19	67
it V that	1			22
it V det N that	1			
it V prep that	3			1
it V prep to-inf	5	1		2
it V ADJ –ing		1	1	10
it be V-ed ADJ for n to-inf		1		
it V n to-inf		1		
it V adv for n to-inf		1		
it be V-ed to-inf		2	1	2
it be V-ed prep that		2	1	5
it V like			1	
it V ADJ to n that				1
it V ADJ to n wh				2
it V ADJ by n that				1
it V to-inf				5
it V n that				6
it V det N to-inf				1
it V n for n to-inf				1
it V n among n to-inf				1
it V poss N that				1
it be V-ed wh				1