Alvin Ping Leong*

Marked Themes in academic writing: a comparative look at the sciences and humanities

https://doi.org/10.1515/text-2022-0188
Received December 8, 2022; accepted January 11, 2024; published online January 23, 2024

Abstract: Differences between science writing and humanities writing often appear as glosses in guidebooks, but empirical studies comparing these two genres of writing are uncommon. This study investigated the use of a highlighting mechanism – the Hallidayan notion of the marked Theme (MT) – to understand how the sciences and humanities foreground contextual information, and what this implies about the nature of writing in these two broad disciplines. The corpus comprised 80 research articles, 40 each from the sciences and humanities. MTs were analyzed for their grammatical forms and functions using the Hallidayan framework. The findings revealed that while both genres of writing had roughly the same proportions of MTs used, they differed in their use of thematized clauses. More non-finite clauses were found in science writing, and more finite clauses in humanities writing. Science writing favored the use of Cause MTs, whereas humanities writing used more Contingency and Angle MTs. These findings suggest that science writing values brevity and authorial presence. Humanities writing, by contrast, prefers a more elaborate writing style, with a focus on establishing the conditions needed for the authors' interpretations, and integrating the viewpoints from other scholars. Suggestions for further research involving other disciplines and multi-disciplinary fields of study are recommended.

Keywords: marked Themes; systemic-functional grammar; academic writing; genre; sciences; humanities

1 Introduction

Over the years, empirical studies on academic writing have yielded much insight into its purpose and structure. These have included invaluable work tracing the

***Corresponding author: Alvin Ping Leong**, Language & Communication Centre, Nanyang Technological University, 48 Nanyang Avenue, Singapore 639818, Singapore,
E-mail: alvin.leong@ntu.edu.sg

development of scholarly writing through the ages (Bazerman 1988; Kronick 1962) as well as the ground-breaking studies by Swales (1981, 1990) on the rhetorical structure of research articles. A cursory survey of the literature, however, reveals that such efforts have tended to focus more on science writing. For instance, Halliday and Martin (1993) have shown how science writing relies heavily on nomenclature (e.g., "chemotaxis", "relaxion dark matter") and nominalization (e.g., "a lack of identification of transformation products"), resulting in writing that is dense and frequently inaccessible to the lay reader. Indeed, its focus on concepts, entities, and outcomes characterize it as being object- and thing-oriented (Kinneavy 1980; Wilkinson 1992).

Similar work on humanities writing has been more modest, and descriptions about this latter variety of writing are often reduced to glosses in commentaries and guidebooks. The following, for instance, is found in a short paper by Grech (2019) on the broad differences in styles between the two varieties of writing:

> The sciences attempt to objectively and incrementally explore the material universe through controlled and reproducible experiments. A novel and single interpretation of observed phenomena is sought on which new experiments are then generated. […] The humanities, on the other hand, explore the self, emotions, and humanity by analysing and re-analysing complete units (such as narratives). Subjective self-expression is privileged, with style being of overriding concern, believing that there may be many truths. (Grech 2019: 96)

Such glosses characterize science writing and humanities writing as different genres. A genre is understood here as a schema of language use that reflects the shared norms and practices of the discourse community in which it is employed (Hyland 2015; Swales 1990). But such sentiments about science writing and humanities writing naturally raise the question about the specifics of these differences. To what extent are these two genres different from each other, and in which specific areas? Over the years, some attempt has been made to shed more light on various aspects of these two genres, such as the construction of identity (Hyland 2015), the use of the passive voice (Leong 2021a), and clause complexing (Leong 2021b). But such comparative studies have remained few.

Given the paucity of such studies, more comparative work of this nature is needed. In fact, common characterizations about the two genres of writing – such as the one by Grech (2019) above regarding the compactness of science writing and the diffuseness of humanities writing – raise rather interesting questions. How do these two genres express their explanatory and interpretive goals? In what way is science writing conceptually driven (cf. "thing-oriented"), and humanities writing text-driven (MacDonald 2010: 21–22)?

This study sought answers to these questions by investigating the message structure of the clause using the Hallidayan framework (Halliday and Matthiessen 2014). It focused on one particular aspect of the message structure, the marked

Theme (hereafter MT), in view of its role in establishing, and so foregrounding, shifts in topic areas. The specific research questions are as follows:

(1) Do scholarly papers in the sciences and humanities differ in their use of MTs?
(2) Do these differences, if present, account for how we generally conceive of writing in the sciences and humanities?

The rest of the paper is organized as follows. Section 2 presents the literature review, including the theoretical concepts of (marked) Theme and Rheme, and offers a review of related studies on MTs. Details about the corpus and method of analysis are provided in Section 3. This is followed by the analytical findings and discussion in Section 4. The concluding section summarizes the study and recommends areas for further research.

2 Literature review

2.1 Theme and Rheme

The research questions raised above concern the thematic structure of the text. This is distinct from the rhetorical structure, which is obviously dissimilar in the sciences and humanities because of the marked differences in their research approaches (Thelwall 2019). Even the classic introduction-methodology-results/discussion rhetorical structure in scientific papers is not uniform. For instance, *Science* and *Nature*, two of the most prestigious journals in science, do not strictly follow this classic structure in an effort to make their articles more accessible to non-specialist readers (Cargill and O'Connor 2009).

The thematic structure of the text relates to the way the clausal message is contextualized and developed. In language, this structure is realized by the Hallidayan notions of Theme and Rheme (Halliday and Matthiessen 2014), which together constitute the two-part message structure of the clause. (These notions are capitalized, following the convention in the Hallidayan framework.) The Theme establishes the starting point and context of the message, which is then developed by the Rheme in the rest of the clause. Unlike the rhetorical structure, where differences are obvious, variations in the thematic structure are more subtle, and thus warrant a more careful look.

Halliday's theory of language, first articulated in a series of seminal papers in the 1960s (Halliday 1967a, 1967b, 1968), views language as performing three broad functions, what he terms "metafunctions". According to the framework, language allows us to construe our experiences of the world (ideational metafunction), express our attitudes (interpersonal metafunction), and organize the text as a unified whole

(textual metafunction). The notions of Theme and Rheme belong to the textual metafunction of language.

The Theme and Rheme organize the clause as a message unit, and are sequenced in that order; in English, Theme is always positioned first, followed by Rheme. Theme serves as the "point of departure of the message; it is that which locates and orients the clause within its context" (Halliday and Matthiessen 2014: 89). The Hallidayan framework identifies three types of Themes – textual, interpersonal, and topical. Textual Themes are generally cohesive expressions, whereas interpersonal Themes are interactional in nature or express the subjective opinions of the speaker/writer. The most important Theme type is the topical Theme. Of the three, the topical Theme is the only obligatory element of the thematic segment of the clause; the other two Theme types are optional. The topical Theme is realized by the first constituent in the clause that functions as the participant or adverbial, or in the case of imperatives, the main verb itself. The topical Theme, as it were, grounds the clause, without which "the clause lacks an anchorage in the realm of experience" (Halliday and Matthiessen 2014: 112). The clause-initial constituents realizing each Theme type are listed in Table 1.

In English, the thematic portion extends from the beginning of the clause up to and including the topical Theme. The topical Theme may be preceded by textual and/or interpersonal Themes, but it is the topical Theme that ends the thematic segment of the clause. The rest of the clause following the topical Theme is regarded as Rheme.

The following example illustrates the thematic structure of a clause. All examples are taken from the corpus; the topical Theme is boldfaced, and the source of each example is bracketed in italics.

(1) Alternatively, [textual Theme] perhaps [interpersonal Theme] **the crux of your identity** involved drug-unrelated social roles and relationships. *(Philo 6)*

Table 1: Textual, interpersonal, and topical Themes.

Theme	Clause-initial constituents
Textual Theme	Continuatives
	Conjunctions or conjunctive adjuncts
	Wh- relatives
Interpersonal Theme	Vocatives
	Modal adjuncts
	Finite operators
	Wh- question words/phrases (content interrogatives)
Topical Theme	First participant, first circumstantial adjunct, or main verb

Adapted from (Halliday and Matthiessen 2014: 105–114).

2.2 Unmarked and marked Themes

What gets positioned at the front of the clause is fundamentally a matter of choice. If the choice is a typical one, the Theme is said to be unmarked (Webster 2019: 41). For instance, it is typical for the grammatical subject of a declarative to be in the clause-initial position. Hence, in the case of declaratives, the grammatical subject functioning as the topical Theme is regarded as an unmarked Theme; all other realizations of the topical Theme (e.g., adjuncts) are considered marked.

It is also important to note that MTs are *obligatory* Themes, i.e., topical Themes (Fontaine 2013; Halliday and Matthiessen 2014; Thompson 2014). The obligatoriness of topical Themes allows the writer/speaker to make a conscious choice about fronting non-typical constituents, making them "unusual enough to draw attention to themselves" (Thompson 2014: 149). Textual and interpersonal Themes, being optional, are neither unmarked nor marked.

As alluded to above, whether a Theme is marked depends on the grammatical mood of the clause (i.e., declarative, interrogative, or imperative). Different grammatical moods in English have different structures, and so what is considered typical in one may be highly unusual in another. For instance, it is usual for imperatives to begin with a verb (e.g., '**Talk** quietly'), but this would be highly marked in declaratives (e.g., '**Talk**, you should not!'). The unmarked Themes in different grammatical moods are listed in Table 2, with accompanying examples from the corpus:

Table 2: Unmarked Themes in different grammatical moods.

Grammatical mood	Examples
Declarative	**The efficacy population** included 133 participants in the pembrolizumab group and 131 in the placebo group. *(Nature 20)*
Yes-no interrogative	Is [interpersonal Theme] **this** a traumatic repetition of vigilance or a viable premonition? *(Lit 7)*
Wh- interrogative	**What** is truth? *(VPA 10)*
Imperative	**Imagine** a friend of yours has just come into a large sum of money. *(Philo 10)*

At the clause level, all other realizations of the topical Theme not listed in Table 2 (e.g., adverbials) are considered marked. At the sentence level, the framework also considers fronted subordinate clauses (referred to as β-clauses) as MTs. As Halliday and Matthiessen (2014) explain:

> In general, thematic β-clauses serve to set up a local context in the discourse for the α-clause: they re-orient the development (as in the staging of a narrative), often distilling some aspect of what has gone before to provide the point of departure for the dominant clause, […] (Halliday and Matthiessen 2014: 551)

An example of a thematized subordinate clause is given in (2). Such thematized clauses, like topical Themes, are also boldfaced.

(2) **Although Higman did not necessarily mean for this pattern to be generalized throughout Caribbean history**, elements of these differing perspectives on land and land use can be traced in the present example. *(Hist 3)*

2.3 Review of literature on marked Themes

Research on Theme and Rheme has included investigations into the thematic structure of student writing (Chen 2019), research articles (Leong et al. 2018), and websites (Stoian and Dejica 2016), among others. While these studies have been insightful, detailed discussions regarding MTs are not common. This is curious, given the role of MTs in signaling changes or shifts in the progression of the message; an understanding of the use of MTs in science writing and humanities writing can shed light on how the sciences and humanities handle message salience and whether this reflects the larger modus operandi of each discipline.

Corpus-based studies with a specific focus on MTs in academic texts suggest that such Themes constitute between 27 % (Rørvik and Monsen 2018) and 32 % (Gosden 1992) of all Themes. These figures, however, need to be further verified primarily because of a dearth of similar research in this area.

A further complication is the varying definitions of MTs adopted by some studies. For instance, in a study comparing academic texts written by native and non-native speakers of English, Green et al. (2000) regarded logical connectors (such as “and” and “furthermore”) as examples of MTs; the Hallidayan framework, by contrast, classifies these elements as optional textual Themes instead. Similarly, Gosden (1992) and Chan and Ebrahimi (2012) also differed from the Hallidayan framework in the way they identified MTs. Both studies drew inspiration from the work of Davies (1989), who grouped MTs according to three broad ‘context frames’ (CFs), as illustrated in Table 3.

Relying on these CFs, Gosden’s (1992) study explored the use of MTs in scientific research articles in the disciplines of physics, chemistry, and biology. He found that the most pervasively used MTs were those performing addition-emphatic and contrastive/concessive functions, both of which are CF1 Themes.

Table 3: Context frames of MTs.

Context frame	Examples
CF Type 1 (CF1): conjunctions, conjunctive/ modal adjuncts	Therefore, briefly, certainly, unfortunately, …
CF Type 2 (CF2): preposition or adverbial phrases	In 2019, among these techniques, in the case of complex processes, …
CF Type 3 (CF3): Subordinate clauses, including non-finite clauses	Unless otherwise stated, if the polarity is reversed, …

Adapted from (Chan and Ebrahimi 2012: 1149).

By contrast, in their study comparing the abstracts of applied linguistics and economics articles, Chan and Ebrahimi (2012) found more similarities than differences between the two disciplines. Only differences in the use of MTs serving a validating function were found to be statistically significant; these appeared more in applied linguistics abstracts than economics abstracts (20 % vs. 5 %, $p = 0.0027$). While Chan and Ebrahimi (2012) also relied on Davies's (1989) notion of CFs, their analysis however did not separate the three CFs, but grouped them as a whole. For instance, the 'validation' label included all examples in the corpus that were deemed to validate the respective studies, e.g., "evidently" (CF1) and "as noted earlier" (CF3).

While insightful, these studies are not aligned with the theoretical underpinnings of this article. In the Hallidayan framework, CF1 Themes are textual and interpersonal Themes, which, being optional, have no marked or unmarked status; as Halliday and Matthiessen (2014: 107) note, "textual or interpersonal [Themes play] no part in the experiential meaning of the clause". Seen in this light, the occurrences of MTs in the corpora of Gosden (1992) and Chan and Ebrahimi (2012) are overstated.

Studies that are more closely based on the Hallidayan framework include the work of Kies (1988), Kong (2004), and Rørvik and Monsen (2018). These studies involved a diverse variety of genres. Kies's (1988) analysis of two television interview programs revealed presentational, connective, and contrastive pragmatic functions of MTs. In his study, Kong (2004) highlighted the interactive function of MTs in abstracts, advertisements, and government documents, showing how thematic variations in each genre are guided by its communicative goals. Rørvik and Monsen's (2018) contrastive study focused on the functions of MTs in English and Norwegian academic texts in the field of didactics. They discovered that MTs functioning as place markers were more frequently used in Norwegian texts; by contrast, English texts contained more MTs related to the instruments of the study, concession statements, and purpose.

Apart from the work of Rørvik and Monsen (2018) – and, in part, Kong (2004) – there is a lack of studies involving scholarly texts, particularly those related to the

sciences and humanities. Rørvik and Monsen's (2018) work is essentially a contrastive study of English and Norwegian, and there is no frame of reference regarding the distribution and function of MTs in English scholarly articles across disciplines. While Kong did include abstracts (from discourse analysis and sociology), these are not representative of extended scholarly writing. Specific statistical data in relation to the frequency of MTs as a proportion of the total number of Themes are also lacking. This present study is an attempt to fill this research gap.

3 Data and methodology

3.1 Corpus

The corpus comprised 80 research articles, 40 each from the sciences (236,111 words) and the humanities (304,936 words). All the selected articles were published in 2021, and were the most recent publications at the time of analysis. The research articles in the sciences were taken from *Nature*, a prestigious journal covering all fields of science and technology, ranging from genetics to astrophysics.

Unlike *Nature*, however, there is no single scholarly journal covering the full spectrum of topics in the humanities. As a workaround, four journals were selected to represent some of the major fields in the humanities (Drees 2021) – (a) visual and performing arts, (b) history, (c) literature, and (d) philosophy. These journals were highly ranked according to SCImago (2020), ranging from rank 2 in philosophy to rank 10 in history and literature. Ten articles from each journal were selected for inclusion in the corpus. The four humanities journals were as follows:

(a) Visual and performing arts (VPA) – *Empirical Studies of the Arts*
(b) History – *American Antiquity*
(c) Literature – *New Literary History*
(d) Philosophy – *Noûs*

The specific nature of the corpus – i.e., research articles published in scholarly journals – distinguishes the texts from other genres where topics in the sciences and humanities may also be covered (e.g., book reviews).

3.2 Method of analysis

The unit of analysis was the "minimal terminable unit" (hereafter "t-unit"), a term first used by Hunt (1965). A t-unit comprises only one main clause, which may be accompanied by any subordinate clause(s) associated with it. The t-unit is widely

favored in analyses involving extended texts (e.g., Leong 2022; Rørvik and Monsen 2018; Williams 2009); it allows the analyst to easily discern the overall development of the message in the text since "the structure of beta [subordinate] clauses, including their thematic structure, tends to be constrained by the alpha [main] clauses" (Fries and Francis 1992: 47). In total, there were 10,686 and 13,518 t-units in the science and humanities articles, respectively.

Each t-unit was analyzed for MTs, identified according to the procedure outlined in Section 2.2. We may recall that MTs are clause-initial elements that do not match the norm for a particular grammatical mood (e.g., a declarative beginning with an adverbial rather than the grammatical subject). Each MT was coded for its grammatical form and function. The grammatical forms of MTs are given in Table 4.

As regards the functions of MTs, this study used the categories offered by the Hallidayan framework for circumstantial adjuncts. Diagnostic probes to identify these categories are covered in detail in Halliday and Matthiessen (2014: 310–332), and also appear in workbooks based on earlier editions of Halliday's work (e.g., Fontaine 2013; Martin et al. 1997). The functional categories, with brief descriptions and the accompanying diagnostic probes, are presented in Table 5.

Two points of clarification are needed with regard to the functional labels in Table 5. First, with respect to Location and Time, the Hallidayan framework regards the latter as being a sub-type of the former – "Location construes the location of the unfolding of the process in space-time: the place where it unfolds or the time when it unfolds" (Halliday and Matthiessen 2014: 316). While I concede that the expression

Table 4: Grammatical forms of MTs.

Grammatical form	Examples
Preposition phrase (PP)	**In this article**, we used Ngram analysis to trace the frequency of "big questions" regarding art and beauty over the last two centuries. *(VPA 10)*
Adverb/adverb phrase (AdvP)	**Here**, Cov and Var denote the covariance and the variance, respectively. *(Nature 3)*
Adjective/adjective phrase (AdjP)	**Unreadable in itself**, Dermisache's writing testifies at once to the massacre and to the impossibility of its inscription. *(Lit 5)*
Noun/noun phrase (NP)	However, [textual Theme] **several decades after the Spanish invasion**, stone tools were still manufactured, even with the adoption of metal tools. *(Hist 5)*
Subordinate clause (SubCl)	**Using the Storey–Tibshirani method of computing the false-discovery rate**, we estimate that the results identified as significant at the 5 % level have less than a 5.07 % chance of being a true null. *(Nature 38)*

Table 5: Functions of MTs.

Function	Examples
Extent. The degree and scope of the unfolding event, including frequency – *how far/long/extensive/frequent?*	**On average**, they had 23.50 (SD 13.00) years of experience in the domain of sculpting […] *(VPA 3)*
Location. The place, both physical and abstract, of the event – *where?*	**In this section**, we give more details on the dark-count physical origin. *(Nature 30)*
Time. The time of occurrence of the event – *when?*	**At time t_n** it addresses agent A, enjoining them not to φ. *(Philo 4)*
Manner. The means by which the event is actualized – *how?*	**Using the EDXRF data, lithic technological organization, and historical narratives**, we assess the credibility and likelihood of the different hypothesized models of prehispanic SW/NW-Mesoamerican interaction, and obsidian use by the Mexican Indian allies. *(Hist 5)*
Cause. The reason and purpose for the event – *why?, for what purpose?, who/what for?*	**To further evaluate the possible sources of deep mantle Kr**, we investigate the nature of the accretionary material. *(Nature 35)*
Contingency. The condition(s) upon which the actualization of the event depends – *in the case/event of what?*	**If the act of reading presupposes a relation (a relation that is learned) between words on the page and the things (or concepts, as in Saussure) beyond them**, then the potential in question requires the impossible of us, that we unlearn to read. *(Lit 5)*
Accompaniment. The inclusion or addition of an element in the unfolding of the event – *with whom/what?*	**In addition to the association with social reward and social cachet**, addiction can create structure and a sense of purpose. *(Philo 6)*
Role. The attribute or value assigned to an event or entity – *as/into what?*	**As a practice of language**, critical thinking is far from independent; […] *(Lit 1)*
Matter. The issue singled out for mention – *about what?*	**With respect to the transformation products**, using the above sensitivity assumptions results in MDLs for the 10 products in Fig. 3 in the range of 0.6–7.8 pg m^{-3} (Supplementary Table 4). *(Nature 34)*
Angle. The source or viewpoint of the information expressed in the Rheme – *according to whom/what?*	**According to Sauer (1950b)**, the seeds and leaves of this plant were not eaten, […] *(Hist 7)*

"location in time" is not uncommon, its use is metaphorical; grouping Location and Time as a single category may appear counter-intuitive to some since the former is (often) tangible, but the latter is not. For clarity, I have therefore decided to use two labels instead.

Second, the labels in Table 5 pertain, technically, to circumstantial adjuncts and thematized subordinate clauses, not complements. In the corpus, there was only a single instance of a thematized complement, as shown in (3).

(3) **The first approach** we can immediately set aside. *(Philo 2)*

Although the MT in (3) is not a circumstantial element, it can nevertheless be analyzed as an instance of Matter in the sense that it draws the reader's attention to the very issue that is to be ignored. This is the only exception encountered in the corpus.

3.3 Statistical analysis

Real Statistics Resource Pack (Zaiontz 2023), a Microsoft Excel add-in, was used for all statistical analyses. The Student's t-test was conducted to investigate if the observed differences in the means between the scientific and humanities articles were statistically significant. The significance level for all statistical tests was $\alpha = 0.05$. Significant differences in this article are indicated as follows – a single asterisk for a significant result (i.e., $p < 0.05$), and double asterisks for a highly significant result (i.e., $p \leq 0.01$).

4 Findings and discussion

4.1 Proportion of marked Themes

MTs constituted about a fifth of all t-units – 20.66 % in science writing, and 22.82 % in humanities writing. The observed difference between the two broad disciplines was not statistically significant. As compared to the proportions reported by Gosden (1992) and Rørvik and Monsen (2018) – i.e., 32 % and 27 %, respectively – these numbers are much lower. As noted in Section 2.3, Gosden (1992) analyzed MTs according to CFs, a major departure from the mainstream Hallidayan framework. It therefore comes as little surprise that his figure is comparatively higher since he included textual Themes, which are optional in the framework, as MTs. By contrast, the figure reported in Rørvik and Monsen's (2018) study, which is more aligned with the Hallidayan framework, is closer to the figures reported here. The difference could be disciplinary in nature, for unlike the present study, Rørvik and Monsen's (2018) research involved articles in the field of didactics.

4.2 Distribution of marked Themes by form and function

Summary statistics concerning the forms and functions of MTs are presented in Table 6. Only the *p* values for significant differences are reported.

As regards the grammatical forms of MTs, there was little to separate science writing from humanities writing in the use of phrasal elements. Apart from AdjPs, the observed differences in relation to the other phrasal elements (i.e., PPs, AdvPs, and NPs) were not statistically significant. While the observed difference for AdjPs was statistically significant, such MTs were uncommon and involved fairly typical expressions such as "similar to" or "consistent with". The two genres differed markedly only in the use of thematized clauses.

Where the functions of MTs are concerned, statistically significant differences were found in four of the ten categories. However, as the proportion in Accompaniment was small, only three categories – Cause, Contingency, and Angle – could be taken to meaningfully separate the two genres.

Table 6: Use of MTs in the corpus by form and function.

	Sciences (%)	Humanities (%)	
Form			
PP	10.31	10.29	
AdvP	1.18	1.11	
AdjP	0.25	0.47	$p = 0.03^{*}$
NP	0.23	0.29	
SubCl (finite)	3.19	6.93	$p = 5.30e^{-08**}$
SubCl (non-finite)	5.50	3.73	$p = 8.50e^{-05**}$
Function			
Extent	0.97	1.09	
Location	3.04	3.32	
Time	1.70	1.30	
Manner	1.40	1.97	
Cause	5.87	2.19	$p = 3.19e^{-11**}$
Contingency	4.38	8.01	$p = 3.07e^{-07**}$
Accompaniment	0.54	0.32	$p = 0.03^{*}$
Role	0.71	1.05	
Matter	1.02	1.29	
Angle	1.03	2.28	$p = 1.08e^{-04**}$

The figures are reported as a proportion of total t-units for each genre. *Significant result (i.e., $p < 0.05$), and **highly significant result (i.e., $p \leq 0.01$).

The key grammatical forms and functions in Table 6 are discussed more fully in the following sub-sections.

4.3 Distinctive features – grammatical forms and associated functions

As shown in Table 6, the oft-cited difference between science writing and humanities writing – that the former is dense and the latter, diffuse (MacDonald 2010) – is reflected not so much in the use of phrasal elements, but clausal elements. The effect created by such thematized clauses is not small – the non-finite thematized clauses in science writing (26.98 %) and the finite thematized clauses (29.32 %) in humanities writing each constituted more than a quarter of all MTs.

In science writing, denseness is typically seen in the examples below, where the non-finite clauses capture the purpose (4) and manner (5) of the action succinctly:

(4) **To control for housing-related confounders**, littermate controls were subjected to the same confined smoking chamber environment, without smoke exposure, in all experiments (termed Non-SMK or NS). *(Nature 18)*

(5) **Using detailed maps from Leakey and Harris**, we identified the proboscidean trail adjacent to the bipedal footprints. *(Nature 36)*

In the condensed nature of such clauses, only the barest of information is typically contained. All other details – to do with time, person, and number – require the reader to rely on the co-text or main clause for interpretation (Carter and McCarthy 2006: 546).

Two further points may be made as regards the use of thematized non-finite clauses in science writing. First, it is of course possible to reword (4–5) without resorting to non-finite clauses. For instance, a finite clause and preposition phrase can serve as workable alternatives, as seen in (6–7).

(6) **As we needed to control for housing-related confounders**, littermate controls were subjected to the same confined smoking chamber environment, without smoke exposure, in all experiments (termed Non-SMK or NS).

(7) **Through our use of detailed maps from Leakey and Harris**, we identified the proboscidean trail adjacent to the bipedal footprints.

These revisions, however, result in MTs that are longer and not common in a genre of writing where "[b]revity is appreciated and actively encouraged" (Grech 2019: 96). The conscious use of non-finite clauses, it would appear, is guided by the norms and

expectations of the scientific discourse community, since alternatives to non-finite clauses *are* available, but they are less favored.

Second, the use of such non-finite clauses is almost formulaic. In the science articles, authors' actions are frequently preceded by thematized clauses. This is shown in (8), where only the thematized non-finite clauses are boldfaced for clarity.

(8) a. **To detect TetO-array-labelled ecDNA molecules**, we used the TetR-eGFP construct as described previously.
b. **To reduce the dimerization potential associated with wild type eGFP**, we generated the A206K point mutation according to a previous report.
c. Tet-eGFP-labelled hubs have a slightly smaller size compared to monomeric TetR-eGFP(A206K)-labelled hubs, potentially owing to eGFP dimerization effects (Extended Data Fig. 2c),
d. but the number of ecDNA hubs per cell is not significantly different with Tet-eGFP versus TetR-eGFP(A206K) (Extended Data Fig. 2d).
e. We transiently expressed TetR-eGFP or TetR-eGFP(A206K)
f. and performed imaging experiments two days after transfection.
g. **To image BRD4**, we stained the cells with 200 nM HaloTag ligand JF646 for 30 min followed by washing 3 times in culture medium, each for 10 min.
h. **To monitor ecDNA dynamics within the nucleus**, the COLO320-DM TetO-eGFP cell line was transfected with the PiggyBac vector expressing H2B–SNAPf and the super PiggyBac transposase (2:1 ratio) as described previously51. *(Nature 21)*

As can be seen, in six t-units mentioning the authors' actions (8a–b, e–h), non-finite clauses appear in four of them (8a–b, g–h), and all these clauses are thematized. As MTs, the clauses draw attention to themselves, and so highlight the reasons for the authors' actions. They add clarity to the methodology, and ensure that the writing is "understood well enough to allow replication, either in principle or in fact" (Endler 2015: 465).

By contrast, humanities writing lays emphasis on elaboration, "with style being of overriding concern" (Grech 2019: 96). Thematized clauses are thus lengthier, containing more contextual details. This is evident in (9–10), where both the thematized clauses function as instances of Contingency.

(9) **Although domesticated amaranth seeds have been previously reported from Hohokam sites in Arizona that likely predate the seeds from Aztec North**, we currently know of no previous reports of this cultivar at other Chacoan outliers or, indeed, at any Ancestral Pueblo site in the San Juan Basin. *(Hist 7)*

(10) **When evidence for performing an action accumulates sufficiently so that it reaches a certain critical threshold, for example, when estimates of overall benefits exceed costs, or exceed them by a certain margin**, a person-level action ensues. *(Philo 9)*

Contingency finite clauses in the humanities articles were commonplace, accounting for almost three quarters of the thematized finite clauses (72.63 %). This finding lends support to the observation of Henderson (2023) about writing in the humanities. She identifies "the art of critique" as a constitutive feature, manifesting the "curiosity, creativity and openness to difference" (Henderson 2023: 46) among humanities scholars. This approach to research and writing naturally requires authors to establish situations and contingencies clearly in the writing. As we have seen in the cited examples above, these include not merely actual conditions (9), but imagined ones as well (10).

On the other hand, the key functions of thematized non-finite clauses in science writing are restricted to Manner (17.68 % of the thematized non-finite clauses) and, far more pervasively, Cause (59.79 %). Collectively, Manner and Cause non-finite clauses constituted more than three-quarters of the thematized non-finite clauses (77.47 %), with Cause clearly being the more dominant of the two. The use of thematized clauses in science writing, as we have seen above, appears to lay emphasis on the actions of authors by specifying *how* and *why* the described actions were undertaken. Authorial presence in science writing is well documented in the literature (e.g., Hyland 2002; Kuo 1999). Past studies have in fact noted an increasing trend in science writing toward "authorial intervention, argumentation, and personalization" (Martínez 2005: 182). For instance, in their work on the thematic structure of scientific research articles, Leong et al. (2018) found that the first-person pronoun "we" featured prominently as the topical Theme in 29 of the 30 articles analyzed. Admittedly, these topical Themes, being the grammatical subjects in declaratives, are unmarked. But as the result in this present study suggests, even the MTs in science writing orient the reader toward the actions of authors.

4.4 Distinctive features – functions

As Table 6 shows, more instances of Accompaniment and Cause were found in the science articles. However, Accompaniment MTs were not common, accounting for only about half a percentage point of the total t-units in the science articles. The difference observed in the use of Accompaniment MTs was also weakly significant ($p = 0.03$); in a study with a larger corpus, this result may be overturned. In view of this, Cause MTs, whether realized by phrasal or clausal elements, appear to be a

distinctive feature of science writing. Almost three in ten MTs in science writing (29.90 %) performed this function. Like the Cause MTs in non-finite clauses (see Section 4.3), phrasal Cause MTs also contextualize the actions of the researchers (11–12):

(11) **For operation in the TEM**, the photonic structure was packaged via ultrahigh numerical aperture (UHNA) fibres (Fig. 1c and Methods). *(Nature 7)*

(12) **For LD computation**, we retrieved LD matrices provided by gnomAD v.2.1.1.141 for each population analysed in this study (except for admixed American, Middle Eastern and South Asian genetic ancestry populations, for whom data are not available). *(Nature 37)*

In the case of humanities articles, Angle and Contingency MTs were more widely used. We have already alluded to the pervasive use of Contingency MTs in Section 4.3. While Angle MTs did not occur as frequently as Contingency MTs, the proportion per t-units of the former in humanities writing was more than twice of that in science writing (2.28 % vs. 1.03 %). Like Contingency MTs, the analysis suggests that the use of Angle MTs is characteristic of humanities writing. Angle MTs are exemplified in (13–14) below.

(13) **As William Carlos Williams memorably insisted**, long before Roland Barthes made the point, "If it ain't pleasure, it ain't a poem." *(Lit 2)*

(14) a. **According to O_B-coherence**, the set of live possibilities is [w_1, w_2], for both [w_1] and [w_2] are such that no stronger proposition entails them.
b. But $^{\text{textual Theme}}$ **according to $\cap B$-coherence**, the set of live possibilities is the empty set, for there's no world common to every proposition in the set. *(Philo 1)*

Hogsette (2019), in fact, regards the integration of viewpoints from other scholars as an important attribute of humanities writing. In the humanities corpus, this is done in at least two ways. The first is seen in (13), where the MT assigns legitimacy to the author's larger point. Here, the author thematizes the clause to draw the reader's attention to the poet Williams, using his words to support an earlier point. The second, as seen in (14), uses the MT to raise a significant issue related to it, and so enables the authors to establish their own position on the issue. The MTs in (14) highlight two perspectives (O_B-coherence and $\cap B$-coherence) to illustrate a contradictory result. As Hogsette (2019: 375) notes, "humanities scholars engage the views and perspectives of other scholars and critics, using information to help sustain their claims while also distinguishing their unique perspective from that of other scholars." This view is shared by Hyland (2015: 35), who attributes this to the less

linear and dispersed nature of humanities research, and "so writers can't presuppose a shared context to the same extent, but have to build one far more through citation."

This is not to say, of course, that science writing integrates the viewpoints and work of other scholars less extensively. This is unlikely to be true since the literature review is a critical part of any academic article, whether in the sciences or humanities. Instead, the integration is done differently in science writing. Unlike humanities writing, which favors the use of MTs, science writing instead relies more on unmarked Themes and paraphrases, as illustrated in (15–16).

(15) **The CDC** has noted the discrepancies between their own reported vaccine uptake and that of the Census Household Pulse. *(Nature 15)*

(16) **Several ideas** have been put forward as to how different candidates of dark matter can directly couple to gravitational-wave detectors, ranging from scalar field dark matter[4,16] to dark photon dark matter[17] and to clumpy dark matter coupling gravitationally or through an additional Yukawa force[18]. *(Nature 28)*

5 Conclusions

This article addressed this gap by analyzing a corpus of 80 research articles published in scholarly journals in the sciences and humanities using the Hallidayan framework. Specifically, this paper sought to investigate the extent to which the articles differed in terms of the grammatical forms and functions of MTs, and whether these differences, if any, validate or complicate current characterizations of science writing and humanities writing. The broad conclusions are as follows:

(a) There was no statistically significant difference in the overall proportions of MTs in science writing and humanities. About one in five t-units contained an MT.

(b) Science writing and humanities writing differed in their use of thematized clauses. More non-finite clauses than finite clauses were used in the science articles; the reverse was true in the humanities articles. Given the condensed structure of non-finite clauses, as opposed to the full structure of finite clauses, this difference aligns with the general observation of science writing being dense, and humanities writing being more elaborate.

(c) Cause MTs were more commonly found in science writing. These MTs underscore the role of the author in the research process, clarifying the reason for or purpose of the action.

(d) Contingency MTs were used more in humanities writing. These MTs characterize the interpretive nature of humanities writing; they specify the conditions, real or imagined, in support of the authors' arguments.
(e) Angle MTs were also used more in humanities writing, although not as extensively as Contingency MTs. The integration of viewpoints from other scholars – in an effort to either affirm or challenge them to establish the author's own position – further characterizes humanities writing as being interpretive.

The findings reported in this study quantify the (extent of the) differences between science writing and humanities writing in their use of MTs. These differences offer us a glimpse of how we may make sense of the anecdotal accounts of the characteristics of these two genres. As atypical Themes, MTs draw attention to themselves, and so serve as a highlighting mechanism. Here, they showcase how each broad discipline foregrounds important contextual information as the point of departure of the message in the Rheme, and what this implies about the nature of the writing. As the findings suggest, MTs in science writing hint at the value placed on brevity and authorial presence. MTs in humanities writing, on the other hand, imply a more reflective, elaborate style of writing, with a focus on establishing the conditions needed for the authors' interpretations.

Crucially, knowing these key features helps writers, particularly novice writers, to equip themselves with the knowledge that is expected of members within the discourse community. This knowledge is not always made explicit. As noted in Section 1, while various aspects of science writing have been investigated, similar comparative research interest in humanities writing is more modest. Some linguistic features, such as MTs in this case, are also not as immediately obvious as, say, the headings used in the rhetorical organization of texts (see Thelwall 2019). A better understanding of these features can be helpful to novice writers to guide them on the norms of the discourse community. As Borg (2003) aptly observes:

> If discourse communities are seen as stable, with experts who perform gatekeeping roles, then their genres are normative, and novices must conform to the expectations of the community in order to enter it. Other writers […] suggest that this view takes power away from learners, and instead [they propose] that conventions and rules should be deconstructed, with novices encouraged to appropriate the discourse of the community, both for their own purposes and for the renewal of the discourse community itself. (Borg 2003: 400)

The findings of the present study have relevance in both scenarios. They show not just how the compactness or diffuseness in writing is linguistically realized via MTs, but the extent to which these linguistic forms and functions occur in each genre.

These serve as guideposts for novice writers to fit in with the community. But as the discourse community is a group of people, the discourse norms are bound to

change over time. For instance, studies have shown how the use of the passive voice in science writing has declined from 1980 to the present (Leong 2020). But as the discourse community is also a "community of assent" (Morris 1996: 238–245), changes tend to be gradual. The statistics reported in this study thus also serve as the baseline figures from which such deviations occur. They continue to serve as guideposts, albeit shifting ones.

This work can be extended in at least two ways. The first is to include a larger group of disciplines, particularly those from the social sciences. Many view the sciences, social sciences, and humanities as existing along a continuum (e.g., MacDonald 2010), and it would be interesting to see how writing in the social sciences straddles the sciences and humanities. Second, given the current shift toward inter- and multi-disciplinary studies, articles that bring together diverse perspectives and approaches – in fields such as aural architecture, multimedia studies, and digital humanities, among others – can shed further light on emerging genres of writing.

References

Bazerman, Charles. 1988. *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.

Borg, Erik. 2003. Discourse community. *ELT Journal* 57(4). 398–400.

Cargill, Margaret & Patrick O'Connor. 2009. *Writing scientific research articles: Strategies and steps*. West Sussex: Wiley-Blackwell.

Carter, Ronald & Michael McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Chan, Swee Heng & Seyed Foad Ebrahimi. 2012. Marked themes as context frames In research article abstracts. *GEMA Online Journal of Language Studies* 12(4). 1147–1164.

Chen, Qi. 2019. Theme-rheme structure in Chinese doctoral students' research writing – from the first draft to the published paper. *Journal of English for Academic Purposes* 37. 154–167.

Davies, Florence. 1989. Developing competence in academic discourse: The role of language awareness. *University of Liverpool English Language Unit Working Paper No. 1*. Liverpool: University of Liverpool.

Drees, Willem B. 2021. *What are the humanities for?* Cambridge: Cambridge University Press.

Endler, John A. 2015. Writing scientific papers, with special reference to Evolutionary Ecology. *Evolutionary Ecology* 29. 65–478.

Fontaine, Lise. 2013. *Analysing English grammar: A systemic functional introduction*. Cambridge: Cambridge University Press.

Fries, Peter H. & Gillian Francis. 1992. Exploring theme: Problems for research. *Occasional Papers in Systemic Linguistics* 6. 45–59.

Gosden, Hugh. 1992. Discourse functions of marked theme in scientific research articles. *English for Specific Purposes* 11(3). 207–224.

Grech, Victor. 2019. Writing a scientific paper (WASP): An overview of differences in styles between the sciences and the humanities. *Early Human Development* 129. 96–99.

Green, Christopher F., Elsie R. Christopher & Jacquelin Kam Mei Lam. 2000. The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes* 19(12). 99–113.

Halliday, Michael A. K. 1967a. Notes on transitivity and theme in English (part 1). *Journal of Linguistics* 3(1). 37–81.

Halliday, Michael A. K. 1967b. Notes on transitivity and theme in English (part 2). *Journal of Linguistics* 3(2). 199–244.

Halliday, Michael A. K. 1968. Notes on transitivity and theme in English (part 3). *Journal of Linguistics* 4(2). 179–215.

Halliday, Michael A. K. & James R. Martin. 1993. *Writing science: Literacy and discursive power*. London: Routledge.

Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's introduction to functional grammar*, 4 edn. London: Routledge.

Henderson, Juliet. 2023. What might the 'art of critique' in humanities and social sciences academic writing look like? *Teaching in Higher Education* 28(1). 35–50.

Hogsette, David S. 2019. *Writing that makes sense*, 2 edn. Eugene: Resource Publications.

Hunt, Kellogg W. 1965. *Grammatical structures written at three grade levels*. Champaign: National Council of Teachers of English.

Hyland, Ken. 2002. Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics* 34(8). 1091–1112.

Hyland, Ken. 2015. Genre, discipline and identity. *Journal of English for Academic Purposes* 19. 32–43.

Kies, Daniel. 1988. Marked themes with and without pronominal reinforcement: Their meaning and distribution in discourse. In Erich H. Steiner & Roert Veltman (eds.), *Pragmatics, discourse and text: Some systemically inspired approaches*, 47–75. London: Pinter Publishers.

Kinneavy, James L. 1980. *A theory of discourse*. New York: W. W. Norton & Company.

Kong, Kenneth Chak Chung. 2004. Marked themes and thematic patterns in abstracts, advertisements and administrative documents. *Word* 55(3). 343–362.

Kronick, David A. 1962. *A history of scientific and technical periodicals: The origins and development of the scientific and technological press 1665–1790*. New York: Scarecrow Press.

Kuo, Chih-Hua. 1999. The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes* 18(2). 121–138.

Leong, Ping Alvin. 2020. The passive voice in science writing through the ages: A diachronic study. *Text & Talk* 40(4). 467–489.

Leong, Ping Alvin. 2021a. The passive voice in scholarly writing: A diachronic look at science and history. *Finnish Journal of Linguistics* 34(2021). 77–102.

Leong, Ping Alvin. 2021b. Writing in the sciences and humanities: A clause-complex perspective. *Word* 67(2). 137–158.

Leong, Ping Alvin. 2022. Thematic and rhematic progression in science writing: A pilot study. *Journal of Language and Literature* 22(1). 238–254.

Leong, Ping Alvin, Audrey Lin Lin Toh & Soo Fun Chin. 2018. Examining structure in scientific research articles: A study of thematic progression and thematic density. *Written Communication* 35(3). 286–314.

MacDonald, Susan Peck. 2010. *Professional academic writing in the humanities and social sciences*. Carbondale & Edwardsville: Southern Illinois University Press.

Martin, James Robert, Christian M. I. M. Matthiessen & Clare Painter. 1997. *Working with functional grammar*. London: Arnold.

Martínez, Iliana 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.
Morris, Paul. 1996. Community beyond tradition. In Paul Heelas, Scott Lash & Paul Morris (eds.), *Detraditionalization: Critical reflections on authority and identity*, 223–249. Cambridge: Blackwell Publishers.
Rørvik, Sylvi & Marte Monsen. 2018. Marked themes in English and Norwegian academic texts in the field of didactics. *Bergen Language and Linguistics Studies* 9(1). 43–68.
SCImago. 2020. SCImago Journal and Country Rank. http://www.scimagojr.com/ (accessed 23 October 2021).
Stoian, Claudia Elena & Daniel Dejica. 2016. Theme–rheme analysis of English and Romanian tourism websites. In Daniel Dejica, Gyde Hansen, Peter Sandrini & Iulia Para (eds.), *Language in the digital era. Challenges and perspectives*, 14–28. Warsaw: De Gruyter.
Swales, John M. 1981. *Aspects of article introductions*. Birmingham: Aston University.
Swales, John M. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
Thelwall, Mike. 2019. The rhetorical structure of science? A multidisciplinary analysis of article headings. *Journal of Informetrics* 13(2). 555–563.
Thompson, Geoff. 2014. *Introducing functional grammar*, 3 edn.. London: Routledge.
Webster, Jonathan J. 2019. Key terms in the SFL model. In Wendy L. Bowcher, Lise Fontaine & David Schoenthal (eds.), *The Cambridge handbook of systemic functional linguistics*, 35–54. Cambridge: Cambridge University Press.
Wilkinson, Antoinette Miele. 1992. Jargon and the passive voice: Prescriptions and proscriptions for science writing. *Journal of Technical Writing and Communication* 22(3). 319–325.
Williams, Ian A. 2009. Discourse style and theme–rheme progression in biomedical research article discussions. *Languages in Contrast* 9(2). 225–266.
Zaiontz, Charles. 2023. Real statistics using excel. http://www.real-statistics.com/ (accessed 11 March 2023).

Bionote

Alvin Ping Leong
Language & Communication Centre, Nanyang Technological University, 48 Nanyang Avenue, Singapore 639818, Singapore
alvin.leong@ntu.edu.sg

Alvin Ping Leong lectures at the Language and Communication Centre, Nanyang Technological University, Singapore. He obtained his PhD degree from the National University of Singapore in 2001 under a research scholarship. His book publications include "Transforming literacies and language" (co-editor with Caroline Mei Lin Ho and Kate T. Anderson; Continuum, 2011) and "Theme and rheme" (Peter Lang, 2004). His research interests are in grammar, systemic-functional linguistics, and discourse analysis.