Topical themes and thematic progression: the “picture” of research articles

Leong, Ping Alvin

2015

Alvin, L. P. (2015). Topical themes and thematic progression: the “picture” of research articles. *Text & Talk*, 35(3), 289-315.

<https://hdl.handle.net/10356/81170>

<https://doi.org/10.1515/text-2015-0001>

© 2015 De Gruyter. This paper was published in *Text & Talk* and is made available as an electronic reprint (preprint) with permission of De Gruyter. The paper can be found at the following official DOI: [<http://dx.doi.org/10.1515/text-2015-0001>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 21:21:20 SGT

Leong Ping Alvin

Topical themes and thematic progression: the “picture” of research articles

Abstract: Although much has been written about the features of academic writing, there is a lack of research attention on macro issues related to the development of ideas, particularly in the writing of research articles. A concept that is useful in investigating such issues is the Hallidayan notion of theme. However, the thematic structure of research articles has received only modest attention over the years. It is also rare for thematic diagrams to be used even though they can be helpful in clarifying the thematic structure of the text. In this exploratory study, the patterning of topical themes in research articles was investigated using a diagrammatic approach. Twenty biology-related research articles were divided into t-units and analyzed for topical themes. Thematic diagrams were generated for all the articles. The diagrams revealed a progressive thematic pattern in the introduction sections of all the articles. At the whole-text level, an anchored-development pattern was observed in the majority of the articles. These findings suggest that research articles at the macro level share similarities in their thematic structure. They also shed light on how authors achieve focus in the writing through the systematic use of clause-initial elements.

Keywords: theme, thematic progression, research articles, academic writing, biology

DOI 10.1515/text-2015-0001

1 Introduction

In their study on student writing, Yates et al. (2005: 36) aptly note: “Science is fundamentally about communication. Un-communicated science in essence does not exist.” The need for clear communication in scientific writing is of particular importance to postgraduates, many of whom aspire to academic or research positions after graduation. Many universities are acutely aware of the

Leong Ping Alvin, Language & Communication Centre, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332, E-mail: alvin.leong@ntu.edu.sg

writing needs of their postgraduates and have in place various programs to accustom the students to the language features and conventions expected in scholarly writing. At the University of Sydney's Learning Centre, for instance, the postgraduate research program offers 16 workshops on various aspects of academic writing, and a further eight language-specific workshops for students from non-English-speaking backgrounds (Postgraduate Research Program 2013). A survey of the websites of major universities in America, the United Kingdom, Australia, and Canada reveals the presence of similar postgraduate-level courses.

Reference guides and textbooks on academic writing are also widely available, benefiting not just students but instructors, who rely on them for designing classroom activities. These guides and textbooks focus on a myriad of issues, from the structural features of research articles and dissertations (Craswell and Poore 2012) to conventions of style, referencing, and format (Bailey 2011).

Increasingly, pedagogy has also been guided by research findings on the rhetorical elements of the major sections of research articles (Kanoksilapatham 2005; Samraj 2005; Swales 1981; 1990; Williams 1999) and variations across disciplines (Hyland 2000, Hyland 2008; Hyland and Bondi 2006; North 2005). In particular, the work of Swales (1981, 1990) on the introduction section has been valuable. His analysis of articles from a range of disciplines resulted in the *Create-a-Research-Space* (CARS) model, which characterizes the information in the introduction section as typically developing from general to specific issues. Over the years, the CARS model has become very influential in writing pedagogy at both the undergraduate and graduate levels (Bawarshi and Reiff 2010: 182).

While such research efforts have been insightful, they have tended to be limited to structural and language issues. My experience in teaching postgraduates, however, reveals a growing need to address wider, macro concerns involving focus and the development of ideas in the writing. This is crucial to postgraduates since such issues are more challenging than matters concerning language features or formatting (Phakiti and Li 2011). Indeed, given the rise in the number of foreign/second-language students in English-speaking universities (First-time enrollment 2012), an understanding of idea development in scholarly work will go a long way in helping them write academic papers with greater focus.

In recent years, however, research interest in this area with respect to academic writing has been modest. Among the comparatively few studies available, Rogers's (2004) work on the coherence in the essays of business students comes closest in addressing such concerns. Rogers (2004: 137–138) relied on the notion of *topic*, defined as “the psychological subject” of a clause. In the analysis, this was taken to mean either the grammatical subject, or a clause-

initial nominal element that was commented on by the rest of the clause. Any topic that was deemed to be positioned too far from the beginning of the clause was regarded as a delayed topic. This framework allowed Rogers to quantify topic shifts and topic delays in the corpus she analyzed. As a quantitative study, though, Rogers's scope of research did not include the question of how topics were developed and distributed in the text. The notion of topic shift itself was restricted to only the conceptual paragraph (Harris 1990). The question of how ideas are initiated and developed in research articles remains unclear.

A modified approach is therefore crucially needed. In contrast to Rogers's work, the present study, involving twenty research articles in biological database development and DNA research, is characterized by two major features:

1. Instead of topic, the Hallidayan notion of *theme* was used (Halliday and Matthiessen 2014). This was needed to give the analysis a firmer theoretical basis, and so ensure consistency in the identification of themes. By contrast, Rogers (2004: 138) had to resolve certain problems on her own as her study was based on a "somewhat ambiguous" definition of topic offered by Williams (2000).
2. Instead of a quantitative approach, this study employed a modified version of thematic progression (TP) to track the development of themes in each article analyzed. This approach was admittedly tedious as it required each theme to be mapped onto a diagram for the entire length of the article. While quantitative studies do offer one way of viewing the overall trends in a corpus, they tend to obscure and even hide patterns that are perhaps better represented in the form of diagrams. The TP diagrams in this study offer a "picture" of research articles that statistics alone cannot.

Specifically, this study sought to provide answers to the following questions:

- How are themes patterned in the research articles analyzed in this study?
- Is/are there any observed difference(s) in the way themes are patterned in research articles in biological database development and DNA research?

This paper is organized as follows. Section 2 presents an outline of the Hallidayan framework on theme; a review of relevant studies involving the application of theme in academic writing is also offered. Section 3 provides the broad details about the corpus, method of analysis, and the diagrammatic representation of TP. The results of the analysis are discussed in Section 4, with emphasis on the pedagogical implications of the findings. The concluding section reiterates the main findings of the study and recommendations for further research.

2 Hallidayan notion of theme, and related studies

2.1 Theoretical concepts

The Hallidayan framework, also known as systemic-functional grammar, views language as functional in the sense that “it serves basic human functions to represent experience and knowledge, to construct relationship, and to create meaningful messages” (Christie 2007: 5). These constitute the three major functions (or *metafunctions*) of language – the ideational, interpersonal, and textual. While the ideational and interpersonal metafunctions construe experience or establish interpersonal relations, it is through the textual metafunction that such experiences and interpersonal relations are packaged and meaningfully conveyed through language. The textual metafunction, that is to say, performs a facilitating or enabling role. As Halliday and Matthiessen (2014: 31) note, this metafunction “build[s] up sequences of discourse, organizing the discursive flow and creating cohesion and continuity as it moves along.” One of the concepts that are central in the textual analysis of any discourse is *theme*.

Theme is a clause-internal element and, in English, it is always positioned first in the clause. Theme is glossed 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 recognizes three types of themes – *textual*, *interpersonal*, and *topical* themes – mirroring the textual, interpersonal, and ideational metafunctions of language. Of the three theme types, the most important is the topical theme. In any clause, the topical theme is obligatory, and is realized by the very first linguistic element in the clause that functions as the participant or circumstantial adjunct of the verb, or in the case of imperatives, the main verb itself. The topical theme, in other words, serves as the *anchor* of the clause, providing the experiential base upon which the rest of the clause is developed. By contrast, textual and interpersonal themes are optional; they are clause-initial elements that, respectively, connect textual units or express opinions and degrees of subjectivity. Each type of theme is realized by specific linguistic items from the start of the clause. These are represented in Table 1.

The thematic portion of a clause is delimited as follows:

The guiding principle of thematic structure is this: the Theme contains one, and only one, of these experiential elements. This means that the Theme of a clause ends with the first constituent that is either participant, circumstance or process. We refer to this constituent, in its textual function, as the topical Theme. (Halliday and Matthiessen 2014: 105)

Table 1: Textual, interpersonal, and topical themes (adapted from Halliday 1994; Halliday and Matthiessen 2014: 105–114).

Theme	Linguistic items
Textual theme	Continuatives Conjunctions or conjunctive adjuncts <i>Wh-</i> relatives
Interpersonal theme	Vocatives Modal adjuncts Finite operators <i>Wh-</i> question words/phrases (content interrogatives)
Topical theme	First participant, first circumstantial adjunct, or main verb

The remainder of the clause (i.e., the non-thematic portion) is known as *rheme*.

In the clause, the topical theme may appear alone, or be accompanied by textual and/or interpersonal themes. The following examples illustrate the theme–rheme structure of various clauses; all examples are taken from the corpus used in this study (the article reference in each case is enclosed within brackets; see the Appendix).

- (1)

(DB02)

Each phenotype variation

is linked to zero or more documents from each of the text sources.

Participant

Topical theme

Rheme
- (2)

(DB09)

Since its inception in 1999,

gene ontology (GO) has become the standard for functional annotation ...

Circumstantial adjunct

Topical theme

Rheme
- (3)

(DNA04)

Most probably,

angiosperm PRRs

have lost their phosphorelay function through the course of evolution ...

Modal adjunct

Participant

Interpersonal theme

Topical theme

Rheme

2.2 Thematic progression

Since the topical theme is the only theme type that carries experiential content, it is crucial in helping researchers identify the ideas that serve as the points of departure in the text. A further understanding of the development and patterning of topical themes in the text is facilitated by thematic progression (TP).

In TP, the themes and rhemes of clauses are compared to find out how they are connected with each other. TP was first proposed by Daneš (1970, 1974), who identified various TP patterns in written discourse. These include the simple linear TP and the constant TP, both of which are illustrated below with examples from the corpus (the topical themes are underlined).

Simple linear TP

(4) (DB01; see Figure 1)

The primary source of data in hERGAPDbase is published scientific literature. A paper on hERG assay and another on APD assay were obtained from PubMed. In PubMed, we searched for the keywords ‘hERG’, ‘IC50’ and the phrase ‘patch clamp’...

$T_1 \rightarrow R_1$

↓

$T_2 \rightarrow R_2$

↓

$T_3 \rightarrow R_3$

Figure 1: Simple linear TP.

Constant TP

(5) (DNA06; see Figure 2)

Jatropha curcas L. is a plant belonging to the family Euphorbiaceae that is endemic to tropical America. It is now grown commercially in tropical and subtropical Africa and Asia. Jatropha has considerable potential for various uses including biofuels.

$$T_1 \rightarrow R_1$$

↓

$$T_2 \rightarrow R_2$$

↓

$$T_3 \rightarrow R_3$$

Figure 2: Constant TP.

It should be noted that in actual discourse, TPs are rarely manifested in such clearly defined patterns (Dubois 1987; Leong 2005; McCabe 1999). Instead, actual TPs are frequently far more complex, and this may cause some to hesitate using the TP approach in thematic analysis. The complexity, in large part, is brought about by including *both* the topical theme and rheme when forming connections.

However, while the TP approach is effortful – as any detailed analysis is – the diagrammatic representation can be extremely useful in providing a macro view of the text. In capturing, and thus showing, the thematic “picture” of the text, such diagrams present another aspect of the analysis that may not be easily achieved using merely statistics. Nonetheless, the problem of complexity is a real issue, which probably explains the dearth of recent TP-related studies in the literature. In Section 3.3, I present an amended, simpler version of TP, and show how it can be used to illustrate the patterning of ideas in the text.

2.3 TP-related studies

As alluded to in the earlier section, there are not many recent studies focusing on TP in relation to academic texts. References to TP studies typically draw on work done more than a decade ago, some dating as far back as 26 years ago. A frequently cited work is Dubois’s (1987) study on biomedical texts. Her analysis of these texts led to an expansion of Daneš’s basic TP patterns to include *gapped* patterns, where the progression of a theme is interrupted by a short passage, and multiple development patterns, where a theme is developed in numerous ways in the text. Nwogu and Bloor (1991) also looked at medical texts; their

corpus, however, included not just research articles and abstracts, but journalistic reports as well. They discovered that while the simple linear and constant TPs appeared frequently in the genres they analyzed, the simple linear TP occurred more often in journalistic reports, and the constant TP, in research articles. Abstracts made roughly even use of both TPs. In the case of Spanish and English history textbooks, McCabe (1999: 203) found similarities in the use of the simple linear TP in both corpora, attributing this to “the analytical, or explicative, nature of the texts, which was expected due to the pedagogical purpose of the texts.”

The few recent studies on academic texts have continued on this comparative path. Jalilifar (2010), for instance, made a comparison between international and local norms in published articles. He analyzed applied linguistics articles from two journals – an international journal and an Iranian journal – and found that both sets of articles were largely similar in their TP patterns. In another study, Williams (2009) compared the thematic patterns in Spanish medicine articles and Spanish translations of comparable English articles; he found that linear TPs were largely used in the former, and constant TPs, in the latter.

The survey of the TP studies above reveals two interesting observations. First, TP variations tend to surface when comparisons are made between different genres (e.g., research articles vis-à-vis journalistic reports) or between source and translated texts. This is perhaps to be expected because different genres serve different rhetorical purposes, and translated texts may retain the thematic norms of translated texts that may be quite different from those in source texts. Second, when the same genre is analyzed for thematic patterning (e.g., history textbooks), similarities in TP tend to be observed instead. This is not to say that no differences were detected – differences are bound to occur when the same genre is produced under different contexts – but the similarities in TP are suggestive of a generic feature, one that is characteristic of that particular genre.

This naturally raises the question about the TP of research articles as a whole. Jalilifar’s work is tantalizing, but as he provided no diagrammatic representations in his work, it is as yet unclear how themes are typically sequenced in a research article. This paper is an attempt to provide a tentative answer to this question using a small sample of research articles from the field of biology, but from two distinct sub-fields – biological database development and DNA research. A brief description of the corpus is provided in the following section, which also includes an amendment to Daneš’s TP framework to allow the patterning of topical themes in each research article to be more clearly represented.

3 Corpus and methodology

3.1 Corpus

The corpus comprised twenty research articles. An equal number of articles were taken from the journals *Database* and *DNA Research* (DNAR), both published by Oxford Journals (Oxford University Press). All the articles were published in 2011, and consisted of 93,321 words (*Database* = 45,988 words; *DNAR* = 47,333 words). They were the ten most recent articles in each journal to be downloaded at the time of writing.

Both journals address different aspects of biological research – *Database*, on the archiving and analysis of biological (including genetic/genomic) datasets, and *DNAR*, on the structures and functions of genes and genomes. The journals are also internationally peer reviewed and well received. The 2013 impact factor for *Database* and *DNAR* was 4.457 and 4.975, respectively (the 2014 figures were not known at the time of writing). According to the 2013 version of Journal citation reports, *Database* was ranked fifth (out of 52 journals) in the “Mathematical and computational biology” category, and *DNAR* was ranked 27th (out of 165 journals) in the “Genetics and heredity” category.

The articles in these two journals reflect the nature of scientific writing in a field of immense significance in the life sciences. The field is also highly visible. In the latest announcement by the journal *Science*, biology and related fields accounted for four of the ten scientific breakthroughs for the year 2014 (Breakthrough of the year 2014). The two sub-fields selected for the study – biological database development and DNA research – also permitted comparisons to be made between two distinct areas of focus. The bibliographic details of the articles are listed in the appendix.

3.2 Method of analysis

The unit of analysis was the t-unit, short for “minimal terminable unit,” an expression first used by Hunt (1965). A t-unit comprises only one independent clause, which may be accompanied by any dependent clause(s) associated with it. Such dependent clauses include embedded clauses, such as relative clauses and noun clauses. The t-unit is widely used in text-based analyses, including the studies surveyed in Section 2.3 (e.g., Jalilifar 2010; McCabe 1999; Williams 2009). As Fries and Francis (1992: 47) argue, looking at t-units allows the analyst to “discern the method of development and thematic progression of a text” more

easily, since “the structure of beta [dependent] clauses, including their thematic structure, tends to be constrained by the alpha [independent] clauses.” The division of the articles into t-units resulted in 2,184 t-units in the *Database* articles, and 1,910 t-units in the *DNAR* articles.

In the analysis, only the topical themes of the independent clauses, following the framework outlined in Section 2.1, were identified. In the case of constructions with empty grammatical subjects (e.g., existential clauses), provided they were not preceded by circumstantial adjuncts, the complement of the verb was regarded as the topical theme. This approach differs from the mainstream Hallidayan approach, which regards the empty subject as the topical theme. The position of this paper is that since topical themes are by definition the first elements in the clause that carry experiential content, it would be incongruous for them to be realized by empty subjects. As Leong observes:

A thematic portion that is solely made up of empty words is unacceptable, both theoretically and intuitively, since this amounts to saying that the rheme develops nothing. Empty subjects such as *there* and *it*, therefore, cannot make up the theme in its entirety. (Leong 2004: 202–203)

Examples illustrating the thematic analysis of empty-subject constructions are provided below:

- (6) (DB01)
... there is increased risk of ventricular tachyarrhythmia.
- (7) (DNA06)
There are 40 genes for terpenoid synthase (AtTPS) in *A. thaliana* that are most closely related to JcCS phylogenetically.

3.3 Representation of thematic patterns

As this study is focused on the development of topical themes at the text level, the use of the traditional TP diagram – which incorporates both themes and rhemes – may complicate the representation unnecessarily. While there is merit in including rhemes in the diagram, particularly in in-depth analyses of shorter text segments, this may be counter-productive when the entire text is considered. The insertion of rhemes and lines indicating connections with preceding themes and/or rhemes may cause the broader picture of the text to be obscured.

The simplified method used in this study plotted only the topical themes in the diagram. The Microsoft Excel program was used to create the diagrams, as follows:

1. Each row in the spreadsheet corresponded to a t-unit. All t-units were numbered.
2. The width of each column was narrowed to 1.89 character units, with the default font set at Calibri font size 8. This was done to create cells resembling small squares.
3. The semantic content of each t-unit's topical theme was indicated at the header row of the spreadsheet, and the cell corresponding to that t-unit was colored black. This same procedure was repeated for all the other topical themes in the research article. New semantic labels were added as necessary. Topical themes were placed under an existing semantic label if they met the following criteria proposed by Martin:

Identity chains are based on co-referentiality, which is realised through pronominal cohesion, instantial equivalence, the definite article and demonstratives (or lexical repetition if the reference is generic); similarity chains [...] are realised through substitution and ellipsis, lexical repetition, and relations of synonymy, antonymy, hyponymy and meronymy. (Martin 1992: 419)

The above criteria, which have their basis in Halliday and Hasan (1976) and Hasan (1984), have been used in a number of studies, including those involving scientific texts (Parsons 1991) and research articles (Zmrzlá 2013), among others. In the present study, the assigning of semantic labels to the topical themes was checked by two doctoral students specializing in the biological sciences and serving as content experts. In cases where it was difficult to decide if a new semantic label was needed, I deferred to the views of the experts. Such cases, however, were few and far between.

The thematic diagram is exemplified in Figure 3. The diagram is based on the opening paragraph of DNA01, given in (8). The numbers in square brackets refer to the numbers assigned to each t-unit.

- (8) [1] Glutathione (GSH) is the tripeptide γ -glutamylcysteinyl-glycine
 [2] and [glutathione] plays a central role in the processes of detoxification and redox buffering.
 [3] Glutathione S-transferases (GSTs, EC. 2.5.1.18) catalyze the conjugation of GSH to an electrophilic substrate.
 [4] Plant GSTs have been actively investigated during last decades.
 [5] Currently, large numbers of GST genes have been identified or annotated from at least 17 plant species.

- [6] In Arabidopsis, the complete identification in a genome-wide level revealed at least 53 GST genes.
- [7] In rice, 59 GST genes have been identified.
- [8] However, their works were carried out by BLAST searches against the rice expression sequence tag (EST) database;
- [9] thus, some members may escape from their collection.
- [10] Recently, 81 GST genes have been genome-widely identified in *Populus trichocarpa*
- [11] and they exhibit extensive functional diversification.
- [12] However, no other data have been reported on the genome-wide identification of the GST family, although at least 20 plant genomes have been completely sequenced.

Figure 3: Thematic diagram based on the opening paragraph of DNA01.

The theme-only diagrams used in this study are simpler and easier to create as compared to Daneš's TP diagrams. More crucially, by isolating the topical themes, they allow the patterning of the points of departure in the text to be more clearly presented. In Figure 3, for instance, a progressive pattern is clearly visible, with perhaps an early indication of some focus on GST genes, as indicated by the constant pattern observed in t-units 9–11.

4 Results and discussion

The results are presented in two parts. The first focuses on the introduction sections of the *Database* and *DNAR* articles. The second examines the different thematic patterns observed in the articles in their entirety.

4.1 Thematic pattern in the introduction section

Only the introduction sections of the articles were compared. This was because the structures of the *Database* and *DNAR* articles were not the same. While the *DNAR* articles followed the conventional Introduction–Methodology–Results–Discussion structure that is typical of scientific writing, the *Database* articles did not. Instead, there was a mix of sections detailing the background, features, capabilities, and updates of databases and database tools. This flexibility is perhaps warranted in view of the nature of database development, which can be very different from the experimental nature of DNA research. The only section that was common to both sets of articles was the introduction section.

A comparison of the introduction sections revealed uniformity across all the articles in the corpus. The diagrams in Figure 4(a)–(f) illustrate the thematic patterns in six representative articles (DB01, DB04, DB06, DNA02, DNA03, DNA09).

As can be seen, a general progressive thematic pattern is displayed in each of the diagrams. On the surface, this appears to suggest a continual introduction of new ideas in the section. What gave rise to such a pattern, however, was not simply the introduction of new ideas *per se*, but a narrowing of the description in the writing from general to specific issues.

Figure 4: Thematic diagrams of the introduction sections of six representative articles.

Consider, for instance, the following introduction section in DB04:

- (9) [1] Theileria parva is a tick-transmitted haemoprotozoan parasite that causes an acute and often fatal disease of cattle, East Coast fever (ECF).
 [2] ECF severely constrains the livelihoods of poor livestock-keepers in sub-Saharan Africa.
 [3] Current control methods include use of acaricides to limit tick populations, drug-treatment of cattle exhibiting clinical symptoms and deployment of a live vaccine that involves infection with a potentially lethal dose of cryo-preserved sporozoites and simultaneous treatment with long-acting oxy-tetracycline.
 [4] A subunit vaccine will provide a long-term solution to this socio-economically important constraint to livestock development in Eastern and Southern Africa.

- [5] The completion of the genome sequence of T. parva Muguga represents an important milestone in research on the parasite biology,
- [6] and [the completion of the genome sequence of T. parva Muguga] has contributed to the identification of candidate schizont antigens for vaccine development targeting this stage of the parasite.
- [7] It is also an important resource for apicomplexan comparative genomics, in particular with *Plasmodium falciparum*, which causes malaria in humans and *T. annulata*, the cause of tropical bovine Theileriosis, a related disease of cattle that has a wide range in North Africa and South and East Asia.
- [8] To date the utilization of the *T. parva* genome and associated information by interested scientists has been limited by the lack of a user-friendly interface that provides access not only to genome data but the large set of expression data from MPSS and EST data for the schizont stage and additional unpublished and published microarray expression data.
- [9] In addition, the current system of access does not easily allow for updating of annotation and curation as new data becomes available.

The description begins with the specific mention of a parasite (“*Theileria parva*”), the disease it causes, and the current control methods. It then narrows to the genome sequence of a particular strain of the parasite (“*T. parva Muguga*”), before highlighting the current situation (“to date”) concerning the lack of a user-friendly interface.

The general pattern observed in Figure 4(a)–(f) matches the observations of Swales (1981, 1990) in his CARS model. As briefly surveyed in Section 1, the CARS model characterizes how the introduction sections of research articles are typically written. In particular, the articles in the corpus adhere to the norms of what Swales and Feak (2004: 44) term “general-specific texts,” which “move from broad statements to narrower ones.”

4.2 Thematic pattern at the whole-text level

At the whole-text level, a distinct pattern, termed *anchored development*, was observed in both the *Database* and *DNAR* articles. This takes the form of one or a small number of vertical lines running through the text. In graphical terms, these lines resemble (vertical) anchor rods, holding the text in place as it were, and it is for this reason that the term *anchored development* was used. Anchored development was clearly observed in eight *Database* and seven *DNAR*

Figure 5: Anchored developments in DB01 and DNA02.

articles. This is illustrated by DB01 and DNA02 in Figure 5(a) and (b); the horizontal line across each diagram marks the introduction section of the article.

Two points should be noted about the vertical developments, or *anchors*. First, the anchors correspond to the research objectives of the respective articles. In the case of DB01, the authors' focus was to introduce a unique database, hERGAPDbase, to allow scientists easy access to experimental data. In DNA02, the authors focused on soybean genes in different experimental conditions. Seen in this light, anchored development represents a thematic organization that highlights the focus of the writing. The reader is continually reminded of the important ideas in the text through the use of topical themes.

The reader might note the somewhat lengthy break in the vertical line in the materials and methods section in Figure 5(b). This indicates that the central idea – “soybean/soybean genes” – is not mentioned in the topical themes. This is to be expected since the focus of this section is on the various procedures and instruments used in the experiments. In the results and discussion section in Figure 5(b), however, the anchor is clearly visible and extends to the end of the text.

The second point is that the anchors are not realized by identical words or phrases, but a range of expressions referring to the ideas or points in focus. Indeed, repeating the same words or phrases as the topical themes in the entire text is likely to result in some awkwardness. In DNA02, for instance, the soybean (genome) anchor is realized by a number of alternative expressions, such as “global soybean (*Glycine max*) production”, “soybean cv. Williams 82 seeds”, “12-day-old plants”, “seedlings”, and “the majority of soybean TCS genes”, among others.

Apart from the clear vertical lines running through the texts, the diagrams in Figure 5 also show the origin of the anchors. Each anchor starts from the tail end of the introduction section, where the writing for that section is at its most focused. In DB01, the anchor – the database hERGAPDbase – is introduced toward the end of the introduction section. Similarly, in the case of DNA02, references to soybean and the soybean genome appear only at the end of the introduction section.

As previously noted, anchored development was clearly observed in fifteen of the twenty articles in the corpus. It was less obvious, however, in the remaining five articles. A case in point is the thematic development in DNA04, as shown in Figure 6.

Figure 6: Thematic development in DNA04.

The research focus of DNA04 was on the characterization of certain pseudo-response regulator (PRR) genes of moss for the purpose of defining the change in plant PRRs. No clear anchor, however, could be detected for “genes”. Instead, Figure 6 displays a scattering of topical themes in the latter half of the text; apart from “genes”, these topical themes also relate to ideas such as “evolution”, “proteins”, and “prior studies”, among others. This appears to run against the observations made so far about anchored developments in research articles.

A closer look at DNA04, however, reveals that the central idea of the article is carried largely by the rhemes of the t-units, particularly in the results and discussion sections. Two extracts from DNA04 illustrate this (the “gene” expressions in the rhemes are marked by angled brackets).

- (10) [161] In 12:12LD, <all genes> showed high-amplitude mRNA rhythms with a period of ~1 day, which peaked in the latter half of the light phase.
 [162] In DD, <all genes> showed endogenous rhythms with damping.
 [163] These rhythms in 12:12LD or DD showed phase relationships roughly similar to <A. thaliana PRR3, PRR5 or PRR7 genes>.
 [164] In LL, in contrast, <all the genes> exhibited no hint of circadian regulation
 [165] and [all the genes] were arrhythmic as demonstrated for <PpPRR1>.

In (11), the abbreviated form PpPRR stands for “Physcomitrella patens pseudo-response regulator”. The genes that were investigated in the research were PpPRR1, PpPRR2, PpPRR3, and PpPRR4.

- (11) [175] The rates of induction by white light are 8.7 for <PpPRR1>, 2.6 for <PpPRR2>, 6.2 for <PpPRR3> and 2.5 for <PpPRR4>.
 [176] The rates of induction for PpPRR1 and PpPRR3 are the highest and second highest, respectively, for any colour of light;
 [177] this observation does not contradict the idea that there is an intraspecific divergence between <PpPRR1/PpPRR3> and <PpPRR2/PpPRR4>.
 [178] The rates of induction by blue light and red light are lower than those of white light for <any PpPRR gene>.
 [179] The results of the current study help to understand the origins of <the angiosperm PRR genes>.

In fact, by incorporating all the gene-related rhemes into the thematic diagram, a clear vertical line corresponding to “genes” emerges. This is shown in Figure 7,

Figure 7: Theme–rheme diagram of DNA04, with gene-related rhemes.

which captures the theme–rheme development of the results and discussion sections (the rhemes are indicated by the crossed boxes).

Texts such as DNA04 represents an *exception*, requiring rhemes to be included in the interpretation rather than just topical themes. As shown in the majority of the articles in the corpus, the propensity is for authors to use topical themes, rather than rhemes, to express the focus in the writing. This can be

explained by the close association between thematic structure and information structure, the latter of which comprises given information and new information. In the unmarked case, Halliday and Matthiessen (2014: 120) note that “the Theme falls within the Given, while the New falls within the Rheme,” and that “Theme + Rheme is speaker-oriented, whereas Given + New is listener-oriented.” Given that the main purpose of research articles is to convey new information about the research focus, it therefore makes sense for the focus to be largely realized by topical themes.

An awareness of the role of topical themes and the typical thematic patterns in scholarly articles is useful in helping students better understand the norms of scholarly writing. In her book on scientific writing skills, Laurens (2007: 7) lists “a lack of focus on the topic” as one of the main problems facing students in academic writing. Familiarizing students with the concepts of topical theme, rheme, and anchored development can go some way to minimize this problem by helping them to keep track of the major ideas being introduced in the text, and so ensure that the writing is focused.

For researchers and scholars in the field, the finding that both *Database* and *DNAR* articles display similar thematic patterns would also be of interest. In the literature, research articles are commonly referred to as a genre, even though disciplinary variations have long been observed by various scholars (e.g., Hyland 2008). For instance, with respect to mathematics research articles, Harris remarks:

[...] there are articles that introduce new structures, articles that carry out intermediate calculations, articles that establish relations between different structures, [and] articles that solve longstanding problems. Each description defines a corresponding genre, and the list is far from exhaustive. (Harris 2012: 149)

Such differences are hard to ignore, and Harris’s reservation is perfectly understandable. The present study, however, has offered suggestive evidence that there is perhaps another aspect – concerning (the patterning of) topical themes – that is common in research articles. Anchored development fits well with the goal of research articles to report results and address a limited number of issues related to the research focus. We have seen this in the majority of the articles in the corpus, notwithstanding the different methodologies used in *Database* and *DNAR*. Hence, even though the rhetorical structures of the articles may vary in view of the different methodological orientations, the inner core of these articles remains similar. Insofar as theme is an integral component of the message structure of the clause, the patterning of topical themes reflects this deeper message structure of the larger text.

5 Conclusion

This exploratory study focused on the topical themes in twenty research articles in the sub-fields of biological database development and DNA research. It explored how topical themes were patterned in these articles, and whether any observed differences in the thematic patterns could be detected. Thematic diagrams, based on an amended version of Daneš's TP, were generated. The diagrams and analysis revealed the following:

1. In the introduction section, a progressive thematic pattern was observed, reflecting the gradual narrowing of the written description from general to specific issues. This pattern was found in all the introduction sections of the articles.
2. At the whole-text level, an anchored-development pattern was observed in fifteen of the twenty articles, with little difference between the *Database* and *DNAR* articles. The vertical developments, or anchors, were related to the central ideas in the writing.
3. Exceptions were noted in the remaining five articles. In these articles, the central ideas were realized largely by rhemes, rather than topical themes.

As this study has tried to show, the topical theme plays an important role in shaping the ideas in scientific writing. As previously noted, the Hallidayan framework regards the topical theme as the first clausal element with experiential content. In the words of Halliday and Matthiessen (2014: 112), the topical theme provides “an anchorage in the realm of experience.” At the text level, the thematic diagrams in this study suggest that topical themes can also provide an anchorage in terms of the central ideas in the writing.

For students, a good understanding of topical themes goes beyond the awareness that they are merely clause-initial elements or that they carry experiential content. It is also important for them to realize that topical themes, as points of departure, *contextualize* the development of the message in the rest of the clause (Halliday and Matthiessen 2014: 89). Hence, choosing what to put into these points of departure should be carefully considered as it has a bearing on whether the author is able to develop them appropriately in the rest of the writing. In turn, it also determines whether the text as a whole is appropriately focused on the topic.

The findings of this study, of course, are tentative in view of the small number of articles analyzed. Much more remains to be done. As this research involved only articles belonging to two sub-fields of biology, further comparative work with a much larger corpus is crucially needed to determine if the findings are also generally true of articles in other fields, both science

and non-science related. When considering research articles from diverse disciplines, it is perhaps natural to focus on the differences and identify discipline-specific norms. This work has provided some indications that research articles may share similarities at the macro level. It is in understanding not just the differences but also the similarities that the true picture of research articles will emerge.

Appendix: Articles analyzed in the study

Articles from Database

- DB01 Hishigaki, Haretsugu & Satoru Kuhara. 2011. hERGAPDbase: A database documenting hERG channel inhibitory potentials and APD-prolongation activities of chemical compounds. *Database* 2011. doi: 10.1093/database/bar017.
- DB02 Green, Jason M., Jaturon Harnsomburana, Mary L. Schaeffer, Carolyn J. Lawrence & Chi-Ren Shyu. 2011. Multi-source and ontology-based retrieval engine for maize mutant phenotypes. *Database* 2011. doi: 10.1093/database/bar012.
- DB03 Harper, Lisa C., Mary L. Schaeffer, Jordan Thistle, Jack M. Gardiner, Carson M. Andorf, Darwin A. Campbell, Ethalinda K. S. Cannon, Bremen L. Braun, Scott M. Birkett, Carolyn J. Lawrence & Taner Z. Sen. 2011. The MaizeGDB genome browser tutorial: One example of database outreach to biologists via video. *Database* 2011. doi: 10.1093/database/bar016.
- DB04 Visendi, Paul, Wanjiku Ng'ang'a, Wallace Bulimo, Richard Bishop, James Ochanda & Etienne P. de Villiers. 2011. TparvaDB: A database to support Theileria parva vaccine development. *Database* 2011. doi: 10.1093/database/bar015.
- BD05 Williams, G. W., P. A. Davis, A. S. Rogers, T. Bieri, P. Ozersky & J. Spieth. 2011. Methods and strategies for gene structure curation in WormBase. *Database* 2011. doi: 10.1093/database/baq039.
- BD06 Blankenberg, Daniel, Nathan Coraor, Gregory Von Kuster, James Taylor & Anton Nekrutenko. 2011. Integrating diverse databases into an unified analysis framework: A Galaxy approach. *Database* 2011. doi: 10.1093/database/bar011.

- BD07 Vellozo, Augusto F., Amélie S. Véron, Patrice Baa-Puyoulet, Jaime Huerta-Cepas, Ludovic Cottret, Gérard Febvay, Federica Calevro, Yvan Rahbé, Angela E. Douglas, Toni Gabaldón, Marie-France Sagot, Hubert Charles & Stefano Colella. 2011. CycADS: An annotation database system to ease the development and update of BioCyc databases. *Database* 2011. doi: 10.1093/database/bar008.
- DB08 Magrane, Michele & UniProt Consortium. 2011. UniProt Knowledgebase: A hub of integrated protein data. *Database* 2011. doi: 10.1093/database/bar009.
- DB09 Costanzo, Maria C., Julie Park, Rama Balakrishnan, J. Michael Cherry & Eurie L. Hong. 2011. Using computational predictions to improve literature-based Gene Ontology annotations: A feasibility study. *Database* 2011. doi: 10.1093/database/bar004.
- DB10 Bluhm, Wolfgang F., Bojan Beran, Chunxiao Bi, Dimitris Dimitropoulos, Andreas Prlić, Gregory B. Quinn, Peter W. Rose, Chaitali Shah, Jasmine Young, Benjamin Yukich, Helen M. Berman & Philip E. Bourne. 2011. Quality assurance for the query and distribution systems of the RCSB protein data bank. *Database* 2011. doi: 10.1093/database/bar003.

Articles from DNA Research

- DNA01 Chi, Yunhua, Yansong Cheng, Jeevanandam Vanitha, Nadimuthu Kumar, Rengasamy Ramamoorthy, Srinivasan Ramachandran & Shu-Ye Jiang. 2011. Expansion mechanisms and functional divergence of the glutathione s-transferase family in sorghum and other higher plants. *DNA Research* 18(1). 1–16.
- DNA02 Le, Dung Tien, Rie Nishiyama, Yasuko Watanabe, Keiichi Mochida, Kazuko Yamaguchi-Shinozaki, Kazuo Shinozaki & Lam-Son Phan Tran. 2011. Genome-wide expression profiling of soybean two-component system genes in soybean root and shoot tissues under dehydration stress. *DNA Research* 18(1). 17–29.
- DNA03 Kenny, Elaine M., Paul Cormican, William P. Gilks, Amy S. Gates, Colm T. O'Dushlaine, Carlos Pinto, Aiden P. Corvin, Michael Gill & Derek W. Morris. 2011. Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection. *DNA Research* 18(1). 31–38.

- DNA04 Satbhai, Santosh B., Takafumi Yamashino, Ryo Okada, Yuji Nomoto, Takeshi Mizuno, Yuki Tezuka, Tomonori Itoh, Mitsuru Tomita, Susumu Otsuki & Setsuyuki Aoki. 2011. Pseudo-response regulator (PRR) homologues of the moss *Physcomitrella patens*: Insights into the evolution of the PRR family in land plants. *DNA Research* 18(1). 39–52.
- DNA05 Garg, Rohini, Ravi K. Patel, Akhilesh K. Tyagi & Mukesh Jain. 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research* 18(1). 53–63.
- DNA06 Sato, Shusei, Hideki Hirakawa, Sachiko Isobe, Eigo Fukai, Akiko Watanabe, Midori Kato, Kumiko Kawashima, Chiharu Minami, Akiko Muraki, Naomi Nakazaki, Chika Takahashi, Shinobu Nakayama, Yoshie Kishida, Mitsuyo Kohara, Manabu Yamada, Hisano Tsuruoka, Shigemitsu Sasamoto, Satoshi Tabata, Tomoyuki Aizu, Atsushi Toyoda, Tadasu Shin-i, Yohei Minakuchi, Yuji Kohara, Asao Fujiyama, Suguru Tsuchimoto, Shin'ichiro Kajiyama, Eri Makigano, Nobuko Ohmido, Nakako Shibagaki, Joyce A. Cartagena, Naoki Wada, Tsutomu Kohinata, Alipour Atefeh, Shota Yuasa, Sachihiko Matsunaga & Kiichi Fukui. 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha Curcas* L. *DNA Research* 18(1). 65–76.
- DNA07 Gourcilleau, Delphine, Catherine Lenne, Claudia Armenise, Bruno Mouliat, Jean-Louis Julien, Gisèle Bronner & Nathalie Leblanc-Fournier. 2011. Phylogenetic study of plant q-type C2H2 zinc finger proteins and expression analysis of poplar genes in response to osmotic, cold, and mechanical stresses. *DNA Research* 18(2). 77–92.
- DNA08 Doorduyn, Leonie, Barbara Gravendeel, Youri Lammers, Yavuz Ariyurek, Thomas Chin-A-Woeng & Klaas Vrieling. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research* 18(2). 93–105.
- DNA09 Sayama, Takashi, Tae-Young Hwang, Kunihiko Komatsu, Yoshitake Takada, Masakazu Takahashi, Shin Kato, Hiroko Sasama, Ayako Higashi, Yumi Nakamoto, Hideyuki Funatsuki & Masao Ishimoto. 2011. Development and application of a whole-genome simple sequence repeat panel for high-throughput genotyping in soybean. *DNA Research* 18(2). 107–115.
- DNA10 Lamprea-Burgunder, Estelle, Philipp Ludin & Pascal Mäser. 2011. Species-specific typing of DNA based on palindrome frequency patterns. *DNA Research* 18(2). 117–124.

References

- Bailey, Stephen. 2011. *Academic writing: A handbook for international students*, 3rd edn. Oxford: Routledge.
- Bawarshi, Anis S. & Mary Jo Reiff. 2010. *Genre: An introduction to history, theory, research, and pedagogy*. West Lafayette, IN: Parlor Press, WAC Clearinghouse.
- Breakthrough of the year. 2014. <http://news.sciencemag.org/scientific-community/2014/12/breakthrough-year-top-10-scientific-achievements-2014> (accessed 22 December 2014).
- Christie, Frances. 2007. Ongoing dialogue: Functional linguistic and Bernstein sociological perspectives on education. In Frances Christie & James R. Martin (eds.), *Language, knowledge, and pedagogy: Functional linguistic and sociological perspectives*, 3–13. London: Continuum.
- Craswell, Gail & Megan Poore. 2012. *Writing for academic success*, 2nd edn. London: Sage.
- Daneš, František. 1970. One instance of Prague school methodology: Functional analysis of utterance and text. In P. L. Garvin (ed.), *Method and theory in linguistics*, 132–140. The Hague: Mouton.
- Daneš, František. 1974. Functional sentence perspective and the organization of the text. In František Daneš (ed.), *Papers on functional sentence perspective*, 106–128. The Hague: Mouton.
- Dubois, Betty Lou. 1987. A reformulation of thematic progression typology. *Text* 7(2). 89–116.
- First-time enrollment of international graduate students up 8 percent. 8 November 2012. <http://www.cgsnet.org/first-time-enrollment-international-graduate-students-8-percent> (accessed 12 May 2013).
- Fries, Peter H. & Gillian Francis. 1992. Exploring theme: Problems for research. *Occasional Papers in Systemic Linguistics* 6. 45–59.
- Halliday, Michael A. K. 1994. *An introduction to functional grammar*, 2nd edn. London: Arnold.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. & Christian M. I. M Matthiessen. 2014. *Halliday's introduction to functional grammar*, 4th edn. Oxfordshire: Routledge.
- Harris, David P. 1990. The use of “organizing sentences” in the structure of paragraphs in science textbooks. In Ulla Connor & Ann M. Johns (eds.), *Coherence in writing: Research and pedagogical perspectives*, 67–86. Alexandria, VA: TESOL.
- Harris, Michael. 2012. Do androids prove theorems in their sleep? In Apostolos Doxiadis & Barry Mazur (eds.), *Circles disturbed: The interplay of mathematics and narrative*, 130–182. Princeton, NJ: Princeton University Press.
- Hasan, Ruqaiya. 1984. Coherence and cohesive harmony. In James Flood (ed.), *Understanding reading comprehension*, 181–219. Newark, DE: International Reading Association.
- Hunt, Kellogg. 1965. *Grammatical structures written at three grade levels*. Champaign, IL: National Council of Teachers of English.
- Hyland, Ken. 2000. *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1). 4–21.
- Hyland, Ken & Marina Bondi (eds.). 2006. *Academic discourse across disciplines*. Frankfurt: Peter Lang.

- Jalilifar, A. R. 2010. The status of theme in applied linguistics articles. *Asian ESP Journal* 6(2). 7–39.
- Journal citation reports. 2013. http://wokinfo.com/products_tools/analytical/jcr/ (accessed 22 December 2014).
- Kanoksilapatham, Budsaba. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24(3). 269–292.
- Laurens, Amanda. 2007. *Scientific writing skills: Guidelines for writing theses and dissertations*. Stellenbosch: Sun Press.
- Leong, Ping Alvin. 2004. *Theme and rheme: An alternative account*. Bern: Peter Lang.
- Leong, Ping Alvin. 2005. Talking themes: The thematic structure of talk. *Discourse Studies* 7(6). 701–732.
- Martin, James R. 1992. *English text: System and structure*. Amsterdam & Philadelphia: John Benjamins.
- McCabe, Anne M. 1999. *Theme and thematic patterns in Spanish and English history texts*. Birmingham: Aston University PhD thesis.
- North, Sarah. 2005. Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education* 30(5). 517–533.
- Nwogu, Kevin & Thomas Bloor. 1991. Thematic progression in professional and popular medical texts. In Eija Ventola (ed.), *Functional and systemic linguistics: Approaches and uses*, 369–384. Berlin & New York: Walter de Gruyter.
- Parsons, Gerald. 1991. Cohesion coherence: Scientific texts. In Eija Ventola (ed.), *Functional and systemic linguistics: Approaches and uses*, 415–429. Berlin & New York: Walter de Gruyter.
- Phakiti, Aek & Lulu Li. 2011. General academic difficulties and reading and writing difficulties among Asian ESL postgraduate students in TESOL at an Australian university. *REL C Journal* 42(3). 227–264.
- Postgraduate research program. 2013. http://sydney.edu.au/stuserv/learning_centre/pgradjun+jul.shtml (accessed 9 May 2013).
- Rogers, Sandra H. 2004. Evaluating textual coherence: A case study of university business writing by EFL and native English-speaking students in New Zealand. *REL C Journal* 35(2). 135–147.
- Samraj, B. 2005. An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes* 24. 141–156.
- 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.
- Swales, John M. & Christine B Feak. 2004. *Academic writing for graduate students: Essential tasks and skills*, 2nd edn. Ann Arbor, MI: University of Michigan Press.
- Williams, Ian A. 1999. Results sections of medical research articles: Analysis of rhetorical categories for pedagogical purposes. *English for Specific Purposes* 18(4). 347–366.
- Williams, Joseph M. 2000. *Style: Ten lessons in clarity and grace*, 6th edn. New York: Longman.
- Williams, Ian A. 2009. Discourse style and theme–rheme progression in biomedical research article discussions. *Languages in Contrast* 9(2). 225–266.
- Yates, Simeon J., Noel Williams & Ann-Florence Dujardin. 2005. Writing geology: Key communication competencies for geoscience. *Planet* 15. 36–41.
- Zmrzlá, Petra. 2013. *Co-referential cohesive chains in research articles*. Brno: Masaryk University PhD dissertation.

Bionote

Leong Ping Alvin

Leong Ping Alvin 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-length publications include *Transforming Literacies and Language* (co-editor; Continuum, 2011) and *Theme and Rheme* (Peter Lang, 2004). His research interests are in systemic-functional grammar, discourse analysis, and literacy studies.