Thematic organisation in academic article abstracts

Mohsen GHADESSY
University of Brunei Darussalam

ABSTRACT

Academic article abstracts (AAA) are examples of a recognised genre/register that has been analysed in terms of specific grammatical, lexical, and discoursal features. However, there are no published studies dealing with the thematic organisation of AAA. By using the Theme/Rheme system of analysis proposed by Halliday (1985,1994), the present paper explores the thematic organisation of 150 AAA coming from 30 different academic disciplines. The results show that despite grammatical and lexico-semantic differences in the Themes of the clauses of the abstracts, a common 'method of development' or 'point of departure' can be established by Theme selections. Correlations between features of the Themes in the same discipline and across disciplines are then used to show that it is possible to further sub-classify the abstracts of the 30 disciplines. It is concluded that thematic analysis is a powerful and down-to-earth method of text analysis that can contribute towards building a more complete profile of the AAA as well as other registers that have not yet been analysed in this way.

1. INTRODUCTION

The significance attached to the role of language in academic situations has increased in the last few decades as a result of an increasing demand on the part of the learners of this language to master its grammar, vocabulary, and discourse organisation in order to tackle the many academic tasks required for their education in colleges and universities all over the world. There are now many books and series available for the teaching of various EAP (English for

Academic Purposes) courses to native as well as non-native users of English; two of the more popular series being Bates and Dudley-Evans' the *Nucleus* series (1976) and Allen and Widdowson's the *English in Focus* series (1974).

On the other hand, the assumptions of the Communicative Language Teaching (CAT) in recent years has had a marked effect on instructional materials for EAP courses. These course are required to make students understand the structural and functional meanings of the language, to express functional meanings, and to understand and express the social meanings as well (Littlewood, 1981). Recent research has thus moved away from an atomistic view of EAP —local features— which resulted in providing statistics on single vocabulary items, individual morphological patterns and simple/complex grammatical structures (Barber,1962) to more global characteristics, i.e. patterns of organisation beyond the sentence, *Genre Analysis* (Swales, 1990).

Another approach to the description of EAP is through *Register Analysis* (Ghadessy, 1993a) which dissects this language in terms of the three metafuctions present in any coherent text, i.e. the ideational or what is it about, the interpersonal or who are the participants, and the textual or how the language is structured. The three metafunctions are related to the three aspects of the context of situation, i.e. field, mode, and tenor as defined by Halliday (1985:12).

> *The field of discourse* refers to what is happening, to the nature of the social action that is taking place: what is it that the participants are engaged in, in which the language figure as some essential component?
>
> *The tenor of discourse* refers to who is taking part, to the nature of the participants, their statuses and roles: what kinds of role relationship obtain among the participants, including permanent and temporary relationships of one kind or another, both the types of speech role that they are taking on in the dialogue and the whole cluster of socially significant relationships in which they are involved?
>
> *The mode of discourse* refers to what part the language is playing, what is it that the participants are expecting the language to do for them in that situation: the symbolic organisation of the text, the status that it has, and its function in the context, including the channel (is it spoken or written or some combination of the two?) and also the rhetorical mode, what is being achieved by the text in terms of such categories as persuasive, expository, didactic and the like.

It is beyond the scope of this paper to discuss the similarities and/or differences between Genre Analysis and Register Analysis. Interested readers are referred to (Swales, 1990:40-42, Leckie-Tarry, 1993:26-42). The present analysis of thematic organisation of academic abstracts, however, is based on a system of selection provided in the textual metafunction of language in the Hallidayan approach. Abstracts have already been analysed extensively by others (Bazerman 1984, Graetz 1985). The present analysis can complement

these studies and suggest further possible research in this area. After defining an abstract in section 2, Halliday's system of analysis is briefly discussed in section 3. Section 4 elaborates on the data and section 5 lists the features chosen for analysis. Then sections 6 and 7 give the results of the analysis for grammatical and lexico-semantic properties of abstracts respectively. Section 8 provides information on one way of sub-classifying abstracts and section 9 deals with the conclusions.

2.   WHAT IS AN ABSTRACT?

The Longman Dictionary defines an abstract as 'a shortened form of a statement, speech, etc.' This definition does not tell us (1) where we can find abstracts, (2) who writes them, (3) who reads them, (4) what kinds of information they contain and (5) what are their grammatical and discourse features. The only qualification is that they are short, and that they relate to language, i.e. 'statement, speech, etc.' As members of an 'academic community', we know the answers to some of the above questions; the most difficult to answer are probably questions 4 and 5 which deal with the content, grammar and organisation.

One way of describing academic articles abstracts (AAA) is to determine their field, tenor, and mode as defined by Halliday above. We can ask the following questions,

1. What is the nature of the social action?
2. Who are the participants and what is the role relationship between them?
3. How is the language organised?

and then provide a possible definition for an abstract.

(1) An AAA is a condensed and compressed piece of language (Sinclair, 1988) that gives information on the content and organisation of a larger piece of written communication. In other words it is discourse on discourse.
(2) An AAA is usually written by the author of the article for a community of professional people who have a vested interest in the subject matter of the abstract. The writer may or may not be known to the readers.
(3) An AAA is an independent minimal text that can stand on its own. There is almost no redundancy in the language. Information is packed especially in the nominal groups of clauses. The organisation of the abstract reflects the organisation of the academic article.

Having placed the abstract in its context of situation, the next step is to define thematic organisation and then analyse the data accordingly.

3. THEMATIC ORGANISATION

One of the systems in the textual metafunction of language as defined by Halliday (ibid), is the Theme/Rheme options made for each individual clause in spoken and written language. Each clause conveys a message that has two parts, i.e. what comes first or the Theme, and what comes last or the Rheme. The Theme usually contains given information and the Rheme new information. It should, however, be pointed out that the system of Theme/Rheme is different from the system of Given/New in the Hallidayan approach. Halliday (1985:39) distinguishes between Theme/Rheme and Topic/Comment (Given/New) by saying that ,

> ...the label 'topic' usually refers to only one particular kind of Theme ... and it tends to be used as a cover term for two concepts that are functionally distinct,one being that of Theme and the other being that of given.

A consideration of the Themes of all of the clauses in spoken and written English can establish what has been called the 'method of development' of a text. (Fries, 1983).

> The choice of clause Themes plays a fundamental part in the way discourse is organised; it is this in fact, which constitutes what is often known as the 'method of development' of the text.   (Halliday, 1985:62).

In his research on thematic organisation, Peter Fries (1983:119) argues that the method of development deals with the lexico-semantic content of Themes,

> Thematic content correlates with the method of development of a text (and the nature of the text).

More recently, Fries (1992a:1) has shown that 'Thematic content correlates with different genre' and that 'Thematic content correlates with different generic structures within a text'. Following some recent research to identify the thematic development of clauses in several other registers of English (Fries, 1992b, Ghadessy, 1993b), the purpose of the present paper is (a) to establish the grammatical and lexico-semantic features of the Themes in a number of EAP abstracts, (b) to find out if there are any correlations between some of the selections made for the abstracts and 5 other registers and (c) to sub-classify the abstracts on the basis of the features across the 30 different academic disciplines. Although abstracts have been analysed in terms of grammatical, lexical and discoursal features, to this writer's knowledge, no published work is available on their thematic organisation.

The analytical tool for establishing the thematic development is provided by Halliday (1985, 1994). In order to show how this is done, we give below a summary of different Theme types and discuss the rationale for selecting each. The unit of analysis is the clause or simple sentence and, unless otherwise stated, all the examples come from the analysed texts.

It frequently happens that the Theme in a clause (in bold) coincides with the subject as in example 1 (Mathematics). If this happens, we have an 'Unmarked Theme'.

1. **Much of the recent work dealing with quasi-random methods**	has been aimed at ...
Theme	Rheme

But it is also possible to have another constituent placed initially in a clause. If this is the case, we have a 'Marked Theme' (in bold and underlined) as in example 2 (Chemistry).

2. **For the analysis of hepatic microsomal or cytosolic incubates,**	the coefficient of ...
Theme	Rheme

The rationale for dividing Themes into marked and unmarked is that there is a default option in presenting information in a clause. The Theme, usually given information, takes first position and the Rheme, usually new information, comes last. The Theme is the point of departure of the message and oriented towards the writer/speaker, i.e. this is how I want to start my message. The Rheme contains the point of focus of the message and is oriented towards the reader /listener, i.e. this is what I want you to pay attention to. Any change in this order will create a different thematic organisation which results in different methods of development.

Themes can also be classified according to their internal structure. Some are 'Simple' as in examples 1, 2, and 3 (Sociology) below.

3. **Four issues**	emerge as requiring greater attention: ...
Theme	Rheme

Some are 'Multiple' as in example 4 (Physical Education). The rationale for the structure of multiple Themes is that every clause in a text serves three concurrent functions. It must show the ideational connection of the present message with one or more than one of the messages that have preceded it. The point of departure then must create some kind of idea chains. The Theme of the clause also shows, overtly or covertly, the logical relationship between the clauses through conjunctions such as *and, but, then,* etc. and finally some kind of evaluation by the writer/speaker may be included in the Theme, e.g. *maybe,*

generally, on occasion, etc. As will be discussed below, these options are exercised with different frequencies in different registers.

4. ... **and**	**t-scores**	were calculated for differences between subgroup means.
textual	topical	
Theme		Rheme

The components of the multiple Theme above include the ideational Theme *t-score* and the textual element *and*. Halliday (1994) labels the ideational Theme as the *topical* or the *experiential* Theme. We can also have a combination of interpersonal +ideational as in example 5 (Political Science).

5. **Generally,**	**Conservative MP's actions in the Commons**	had no effect on their ...
interper	ideational	
Theme		Rheme

It is also possible to have a combination such as textual+interpersonal+ ideational as in example 6 (Political Science).

6. , ... **and**	**in any case**	**most MP's**	vote along party lines
textual	interper	ideational	
Theme			Rheme

It should be pointed out, however, that examples like 5 and 6 were very rare in the data. Every Theme then must have an ideational element (obligatory) but interpersonal and textual Themes may be missing (optional). To summarise the possibilities, they are, textual^interpersonal^ideational, textual^ideational, and interpersonal^ideational. (^ means followed by). This implies that all simple Themes are made of one ideational Theme.

Other Theme types used in the analysis are 'Predicated Theme', example 7 (The few examples in the data were of this type which are different from Halliday's examples.); 'Ellipted Theme', example 8; 'Existential Theme', example 9; 'Clause as Theme', example 10,11; and 'Thematic Equative, example 11.

7. **It was hypothesised**	that subjects would show diminished cognitive performance and ...
Theme	Rheme

8. **This author**	has analysed numerous oriental martial art styles
and -	has synthesised their various philosophies with western ...
Theme	Rheme

9. **Moreover,**	**there**	was a significant difference (p<0.01) between the ...
textual	ideational	
Theme		Rheme

10. <u>**As there are frequent scatterings of moving charge carriers with phonons,**</u> helicon ...	
Theme	Rheme

11. **What is most striking since the mid 1980s**	is that terms of political dialogue ...
Theme	Rheme

Examples of clause as Theme, numbers 10 and 11 above, were analysed at two levels, i.e. at the higher level of one clause complex, and at the lower level of two separate clauses with each having a Theme and Rheme.

4.   DATA AND METHOD OF ANALYSIS

A total number of 150 abstracts, written by both native and non-native speakers, from 30 different academic disciplines at Hong Kong Baptist University were collected for the present study. The abstracts all came from reputed refereed journals. In order to create a small data-base of written materials for EAP, heads of departments at Baptist University were asked to send some examples of the kind of academic texts that the students are normally required to read. It was suggested to them to confine the list to one academic article, one chapter from a book, samples of the handouts given to students by lecturers, examinations, and five abstracts. Several departments sent the required materials. In order to have a representative sample, additional abstracts for other disciplines , five for each, were also collected from the College library.

Despite several changes proposed for thematic analysis by other systemisists (Martin, 1992; Berry, 1995,1996; Downing and Locke, 1992) Halliday's 1985 system was used for the analysis in the present work. The Theme was considered as the first ideational element in the clause, e.g. ***Many expensive failures in developing and marketing software that is designed to support groups*** *are not due to ....* It could be a Participant (usually a nominal group) or a Circumstance (usually a prepositional phrase) in a declarative clause. Unlike in other spoken and written texts , the abstracts did not contain many examples of imperative or interrogative clauses. In imperatives, the first ideational element, usually the verb, was considered as the Theme, e.g. ***Begin*** *by asking* ... There were no examples of *Let's, You, Don't,* etc. In interrogatives with WH-word, the WH-word, with or without the accompanying noun, was the Theme, e.g. ***What*** *is included?* In Yes-No questions the auxiliary and the

subject of the clause were considered the Theme, the auxiliary being the interpersonal element and the subject the ideational element of the Theme, e.g. *Is it being used to support groups?* Only the Theme of the main clause was selected for analysis.

5.　RESULTS: GRAMMATICAL AND LEXICO-SEMANTIC FEATURES

Probably the most significant feature of an abstract is that it deals with a summary of some events that took place for some reason, under certain specified conditions, and with some results. This indicates that an abstract should at least provide answers to: (1) what was done? (2) why was it done? (3) how was it done? and (4) what was the result? The abstract is then 'exophoric' in its reference to the outside world. The other equally significant characteristic is that the abstract is a summary of the contents of another piece of language and therefore it is 'endophoric' in its reference to the other text. It represents an overview of the discourse structure of the academic article and is discourse on discourse.

It is evident from the small number of examples above that almost all types of participants and circumstantial elements can occupy the Theme position. However, due to the function of abstracts in the larger context of the academic article, some participant and circumstantial types are more frequent than others. In this section we first discuss the grammatical properties of the Themes and then elaborate on the lexico-semantic features.

Based on the discussion in section 3, we can give the following statistics for the frequency of the grammatical properties of Themes. The statistics were obtained by multiplying the total number of one specific feature in 5 abstracts of the same discipline by 100 and then dividing the result by the total number of clauses in the same five abstracts. The abbreviations in the first row are as follows: Simple Theme (ST), Multiple Theme (MT), Textual Theme (TT), Interpersonal Theme (IT), Unmarked Ideational Theme (UIT), Marked Ideational Theme (MIT), Clause as Theme (CT), Ellipted Theme (ET), Predicated Theme (PT), Thematic Equative (TE), Grammatical Metaphor (GM), Number of Clauses (Ncl).

Grammatical Metaphor was added to the Theme features especially because the written language of science and social sciences makes frequent use of this process for packaging information and placing it in the thematic position which in the majority of cases constitutes given information. Then the new information would be given against the background set up by the ideational Theme of the clause.

TABLE 1

Grammatical properties of Themes in academic article abstracts

Props Abstract	ST 1	MT 2	TT 3	IT 4	UIT 5	MIT 6	CT 7	ET 8	PT 9	TE 10	GM 11	Ncl 12
1	64	36	32	4	88	12	0	2	0	0	42	50
2	82.6	17.4	17.4	0	73.9	26.1	0	8.7	0	0	13	23
3	71.1	28.9	28.9	0	81.6	18.4	2.6	0	0	0	52.6	38
4	68.3	31.7	29.3	2.4	75.6	24.4	4.9	2.4	0	0	24.4	41
5	85.3	14.7	14.7	0	82.4	14.7	2.9	2.9	0	0	38.2	34
6	75	25	25	0	80.6	19.4	0	2.8	0	0	22.2	36
7	64.9	35.1	33.3	1.8	89.5	10.5	3.5	1.8	1.8	0	17.5	57
8	77.3	22.7	22.7	0	88.6	11.4	0	6.8	0	0	38.6	44
9	51.3	48.7	48.7	0	92.3	7.7	5.1	5.1	0	0	46.1	39
10	78.2	21.7	21.7	0	86.9	13	0	0	0	0	21.7	23
11	70.6	29.4	25.5	3.9	70.6	29.4	0	5.9	1.9	1.9	11.8	51
12	47.4	52.6	50	2.6	92.1	7.9	0	2.6	0	0	36.8	38
13	86.4	13.6	13.6	0	81.8	18.1	4.5	0	0	0	27.3	22
14	80.8	19.6	17.7	1.9	98.1	1.9	0	0	0	0	38.5	52
15	72.2	27.8	25	2.8	72.2	27.8	0	2.8	0	0	19.4	36
16	71.9	28.1	25	3.1	87.5	12.5	3.1	9.4	0	0	25	32
17	71.7	28.3	28.3	0	79.2	20.8	3.8	9.4	0	0	26.4	53
18	52.1	47.9	43.7	4.2	89.6	2.1	0	2.1	6.2	0	18.7	48
19	63.3	36.7	33.4	3.3	93.3	6.7	3.3	3.3	0	0	36.7	30
20	71.4	28.6	25	3.6	82.1	14.3	3.6	3.6	0	0	32.1	28
21	70.9	29	29	0	87.1	12.9	0	3.2	0	0	29	31
22	75.7	24.3	24.3	0	83.8	16.2	0	8.1	2.7	0	8.1	37
23	62.8	37.2	34.9	2.3	88.4	11.6	0	2.3	2.3	0	18.6	43
24	83.1	16.9	14.1	2.8	91.5	8.4	1.4	1.4	2.8	0	23.9	71
25	68.6	31.4	31.4	0	85.7	14.3	0	2.8	0	0	22.8	35
26	90	10	10	0	82	18	0	6	0	0	38	50
27	58.5	41.5	39	1.5	70.7	29.3	2.4	2.4	0	2.4	24.4	41
28	62.5	37.5	37.5	0	81.2	21.9	6.2	3.1	0	0	31.2	32
29	72.7	27.3	27.3	0	100	0	0	0	0	0	27.3	33
30	65.3	34.7	34.7	0	89.8	10.2	2	8.2	4.1	0	18.4	49

Subjects : 1. Accounting, 2. Applied Linguistics, 3. Biology, 4. Broadcasting, 5. Business Management, 6. Chemistry, 7. Communication, 8. Computer Studies, 9. Economics, 10. Education, 11. Film and Cinema, 12. Finance, 13. Geography, 14. Geology, 15. History, 16. Information Systems, 17. Journalism, 18. Language, 19. Law, 20. Marketing, 21. Mathematics, 22. Music, 23. Philosophy, 24. Physical Education, 25. Physics, 26. Plant Pathology, 27. Political Science, 28. Social Science, 29. Sociology, 30. Theology

6. DISCUSSION: GRAMMATICAL PROPERTIES

A noticeable feature of all the abstracts is the frequency of the simple Themes (ST) with geography having the highest (86.4) and finance the lowest (47.4). This means that there are no overt textual and/or interpersonal elements in the majority of the Themes of the 150 abstracts analysed here. Logical relations between consecutive clauses are 'internal' and evaluation elements, except when they are packed in the nominal group functioning as the Themes, are to be found in the verbal groups and the nominal groups included in the Rheme of each clause.

Multiple Themes (MT) in the abstracts include, in addition to the ideational Theme, textual (TT) and /or interpersonal Themes (IT). As such there is a high negative correlation between column 1 and column 2. In (MT) category, the highest % belongs to finance (52.6) and the lowest to plant pathology (10). One significant difference between the use of overt 'external' textual elements and Interpersonal elements is that the frequency of the former in all the cases is higher than the frequency of the latter. These statistics are shown in columns 3 and 4. Indeed in the majority of cases, there are no interpersonal elements in the selected Themes. This means that there is a very high positive correlation between the categories of multiple Themes, column 2 and the category of textual Themes, column 3. The correlation between columns 2 and 4 on the other hand is negative. Among the abstracts, finance has the highest use of overt textual elements (52.6) and plant pathology the lowest (10). The use of interpersonal elements in the Themes is negligible with language having the highest (4.2) %.

An even more characteristic feature of all the abstracts is the very frequent use of unmarked ideational Themes (UIT). This means that in the majority of cases the Theme is the subject of the clause, i.e. a participant in the transitivity system. In this category, sociology has the highest % (100) and film and cinema the lowest (70.6) %. There is a negative correlation between column 5 and column 6, marked ideational Themes (MIT), as the latter includes statistics on the use of ideational elements that are not the subject of the clause. They are either circumstances or processes in the transitivity system placed in the thematic position. In this category film and cinema has the highest % (29.4) and sociology the lowest % (0).

Of the next four columns 7,8,9, and 10, column 8 is noteworthy. In the majority of the abstracts there are ellipted Themes (ET) to be found. The statistics are very low with the highest belonging to information systems and journalism (9.4). In almost half of the abstracts, there are also a few examples of clause as Theme (CT) with social science having the highest % (6.2). The statistics show that predicated Themes (PT) and thematic equatives (TE) are very rare in the abstracts.

A noticeable feature of the abstracts is the use of what Halliday (1985, 1994) has called 'grammatical metaphor' (GM). This concept is based on the assumption that in the use of language one can use a more 'congruent' as opposed to a 'non-congruent' or metaphorical version. In all the cases, the latter has derived from the former and especially in written language there is a constant attempt to present information metaphorically. One of the grammatical resources used for this purpose is nominalisation, i.e. changing verbs, adjectives, and adverbs into nouns (Eggins et al., 1993). Nominalisation packs a lot of information in the nominal group which can be used as the Theme of a clause. Such information is presented in clause initial position as background 'old information' upon which 'new information' can be presented. Although there are various ways of creating grammatical metaphor, in the present research only the nominalisation of verbs, adjectives , and adverbs was taken into consideration. As such, biology had the highest statistics on GM (52.6) and music the lowest (8.1).

The information in the last column relates to the total number of clauses used in the five abstracts of each subject. Individually, the shortest abstracts, made of only 2 clauses, came from Education, Social Science, Physics, and Information Systems. The longest came from Physical Education, made of 24 clauses. Collectively, Geography had the shortest abstracts with an average length of 4.4 clauses, while Physical Education had the longest abstracts with an average of 14.2 clauses. The average clause number for all the 150 abstracts was 7.3.

The statistics on the grammatical properties of Themes in AAA would not be meaningful unless they are compared and contrasted with statistics on other registers of written English. Unfortunately, there are not many such studies in the literature and if there are, the data used is very limited (Francis, 1990; Backlund, 1991). We can, however, compare the results with the results of a similar study carried out by the author on thematic organisation of 37 written sports commentaries on football (soccer) from The Times of London (Ghadessy,1993b) and also with some findings reported by Fries (1992a) for Obituary, Programme, Narrative, and Exposition.

As statistics given by others do not cover all the grammatical and lexico-semantic properties of the Themes, only a number of features are compared. In the following chart, averages for some of the categories are used. It shows the similarities and differences among the above 6 registers in relation to the grammatical properties of the selected Themes. Comparison of the lexico-semantic properties are shown in Chart 1 in the following section.

CHART 1

Lexico-semantic properties of 6 registers compared
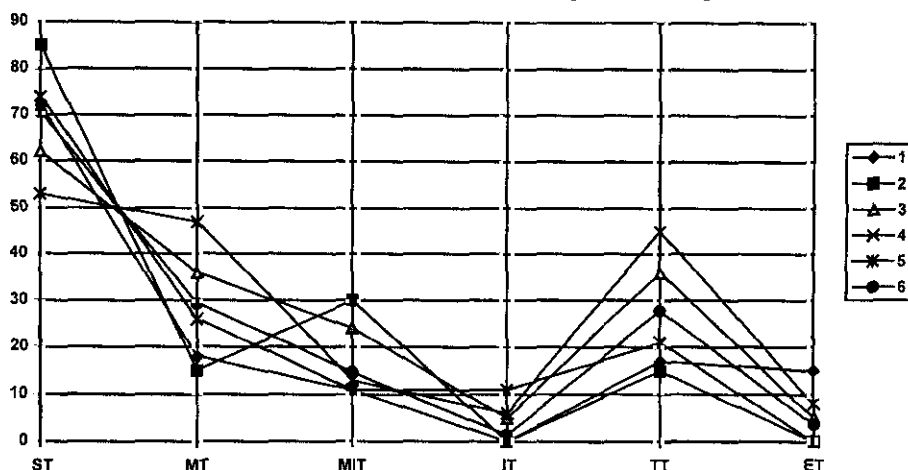
Legend: ST (simple theme), MT (multiple theme), MIT (marked ideational theme), IT (interpersonal theme), TT (textual theme), ET (ellipted theme)
1. Obituary, 2. Programme, 3. Narrative, 4. Exposition, 5. Sports Commentary, 6. Abstracts

Correlation coefficients were then calculated to show the exact correspondences. The results are presented in the following table.

TABLE 2

Correlations based on grammatical properties of the registers

	Obituary	Programme	Narrative	Exposition	Sports Com	Abstracts
Obituary	1.00	0.93	0.85	0.94	0.68	0.95
Programme		1.00	0.87	0.93	0.61	0.93
Narrative			1.00	0.92	0.91	0.97
Exposition				1.00	0.78	0.96
Sports Com					1.00	0.85
Abstracts						1.00

The above table shows that there is a strong correlation between some of the grammatical properties of Themes in the abstracts and the other 5 registers; we have only compared the choice of Simple Themes, Multiple Themes, Marked Ideational Themes, Interpersonal Themes, Textual Themes, and Ellipted

Themes . This may be a surprising result. However, as we have emphasised, a comparison of all the grammatical features of abstracts with the same features in the other 5 registers may alter the picture. For example, statistics for grammatical metaphors are not available for the other registers. This, we believe, is a crucial feature of abstracts which sets them apart from the other 5 text-types under discussion. On the other hand, a consideration of the lexico-grammatical features of the 6 registers shows that there are significant differences in the choices made. These are indicated in table 4 below.

The following table gives the statistics on the lexico-semantic properties of the abstracts. An explanation of providing two statistics in some of the boxes is provided in the following discussion section. The abbreviations in the first row are as follows : Speaker/Hearer (Writer/ Reader) (SH), Main Text Participants (animate) (MTP), Object/Portion of Scenes (inanimate) (OPS), Abstract Concepts (ABS), Process (PRO), Time (TIM), Location (LOC), Manner (MAN), Cause (CAU), Condition and Concession (CON), Number of Clauses (Ncl).

TABLE 3

Lexico-semantic properties of Themes in academic article abstracts

pros subjs	SH 1	MTP 2	OPS 3	ABS 4	PRO 5	TIM 6	LOC 7	Man 8	CAU 9	CON 10	Ncl 11
1	0	24	0	48 16	0	2 0	0 2	4 0	0 2	2	50
2	8.7	0	0	43.5 21.7	0	8.7 0	0 8.7	8.7 0	0 0	0	23
3	0	5.3	23.7	55.3 0	0	5.3 2.6	2.6 0	7.9 0	0 0	0	38
4	0	21.9	7.3	26.8 14.6	2.4	7.3 0	0 0	0 4.9	0 4.9	9.7	41
5	0	14.7	0	61.8 8.8	0	0 0	0 0	0 0	8.8 5.9	0	34
6	0	0	30.6	38.9 8.3	0	2.8 0	0 0	5.6 0	8.3 0	5.6	36
7	1.8	17.5	1.8	54.4 12.3	7	1.8 0	0 0	0 0	0 0	3.5	57
8	6.8	2.3	2.3	61.4 13.6	6.8	0 0	0 2.3	0 0	2.3 2.3	0	44
9	12.8	7.7	5.1	53.8 12.8	0	0 0	0 0	0 2.6	0 2.6	2.6	39
10	0	0	13	56.5 4.3	0	0 0	0 17.4	4.3 0	0 4.3	0	23

TABLE 3 *(continued)*

Lexico-semantic properties of Themes in academic article abstracts

pros subjs	SH 1	MTP 2	OPS 3	ABS 4	PRO 5	TIM 6	LOC 7	Man 8	CAU 9	CON 10	Ncl 11
11	17.6	11.8	13.7	11.8	5.9	9.8	0	0	0	11.8	51
				13.7		0	0	1.9	1.9		
12	7.9	7.9	2.6	52.6	0	0	5.3	0	0	0	38
				21.1		0	0	0	2.6		
13	0	0	0	50	0	4.5	0	0	0	4.5	22
				31.8		9.1	0	0	0		
14		0	42.3	57.7	0	0	0	0	0	0	52
				0		0	0	0	0		
15	2.8	22.2	8.3	27.8	2.8	0	0	13.9	0	5.6	36
				11.1		2.8	0	2.8	0		
16	18.7	0	0	50	0	0	0	6.2	3.1	3.1	32
				15.6		0	3.1	0	0		
17	0	15.1	13.2	43.4	0	3.8	1.9	11.3	0	0	53
				9.4		0	0	0	1.9		
18	10.4	20.8	0	37.5	0	0	0	0	0	0	48
				31.2		0	2.1	0	0		
19	0	3.3	0	63.3	0	3.3	0	0	0	3.3	30
				26.7		0	0	0	0		
20	10.7	3.6	0	35.7	0	0	0	0	0	07.1	28
				35.7		3.6	3.6	0	0		
21	9.7	3.2	0	51.6	0	0	0	3.2	0	3.2	31
				19.4		0	6.4	0	0		
22	2.7	2.7	18.9	31.4	2.7	2.7	0	2.7	0	2.7	37
				8.1		0	5.4	0	0		
23	11.6	4.6	6.9	55.8	0	0	0	4.6	0	2.3	43
				2.3		0	2.3	2.3	4.6		
24	5.6	14.1	0	64.8	0	2.8	1.4	1.4	1.4	2.8	71
				4.2		0	0	0	1.4		
25	8.6	2.8	2.8	48.6	0	0	0	2.8	0	0	35
				25.7		0	0	0	8.6		
26	0	0	42	34	6	6	6	0	0	4	50
				2		0	0	0	0		
27	14.6	7.3	0	39	0	14.6	2.4	2.4	0	7.3	41
				9.8		0	0	0	2.4		
28	0	25	3.1	43.7	0	3.1	3.1	0	0	3.1	32
				15.6		0	0	0	3.1		

TABLE 3 *(continued)*

Lexico-semantic properties of Themes in academic article abstracts

pros subjs	SH 1	MTP 2	OPS 3	ABS 4	PRO 5	TIM 6	LOC 7	Man 8	CAU 9	CON 10	Ncl 11
29	3	3	9.1	69.7	0	0	0	0	0	0	33
				15.1		0	0	0	0		
30	0	16.3	10.2	38.8	0	0	2	0	0	4.1	49
				18.4		2	6.1	0	2		

7. DISCUSSION: LEXICO-SEMANTIC FEATURES

Academic article abstracts reflect a large part of the activities in academia; they are condensed *reports* on such activities. The information given in column 1 relates to references to the writer or reader in the Theme position. Almost all such references are to the writer(s) in the form of *I, we,* and *the author(s).* On the whole percentages for such references are very low with the highest belonging to information systems (18.7).

The second statistics given in each box of column 4, 6, and 7 relate to the use of language in the Theme position in order to make the text reader-based. In column 4 such abstract language includes examples like *this study, this research, this work, the results, the findings, the conclusions,* etc. Abstracts in marketing have used the highest number of these 'representational' examples (35.7). The second statistics in each box of column 6 relate to time adverbials in thematic position that refer to time sequence in the research process itself, e.g. *After outlining different forms of aid and their importance to recipient countries,* the major doners and recipients are identified, and ... and again in column 7 the second statistics in each box refer to the location of information in the research article, e.g. *The first part, The second part, The third section,* etc. The important point here is the use of these 'discourse markers' to facilitate the comprehension of the text by the reader.

It should be emphasised here that whether a text is writer-based or reader-based cannot be determined by looking at the themes of the clauses only. Relevant information can also be given in the Rhemes of such clauses. However, it is clear from the given statistics that the majority of the Themes in all the abstracts refer to the subject matter and the relevant ingredients discussed below.

Of the three major categories of participants, circumstances, and processes, the Theme selection in the abstracts is mainly confined to participants. Column 2, major text participants, (MTP) shows the statistics on animate participants in thematic position. Although the majority of abstracts use this category, the percentages are not very high. The highest being (25) % belonging to social

science. There are fewer participants in the category object/portion of scene (OPS). These are inanimate Themes with the highest statistics belonging to geology (42.3)%.

By far, the majority of Themes belong to the category abstract concept (ABS) that all abstracts have used, the first statistics given in boxes in column 4. The abstract nature of academic discourse is nowhere clearer than in the abstracts analysed in this study. Sociology has the highest statistics (69.7) % and film and cinema the lowest (11.8) % for this category. The main reason for the frequency here is the process of grammatical metaphor that was referred to in the discussion on the grammatical properties of the Themes.

Circumstances are used much more frequently in thematic position than processes. Indeed the statistics on processes, given in column 5, process (PRO), are negligible in comparison with the statistics for circumstances. The use of time adverbials, the first statistics in column 6, time (TIM), is more frequent than the use of place adverbials, the first statistics in column 7, location (LOC). In column 8, the first statistics relate to manner (means) (Man), and the second to manner (comparison) (Man). In column 9, the first statistics relate to cause (purpose) (CAU) and the second to cause (reason) (CAU). The statistics in column 10 relate to conditions and concessions (CON). The statistics for all these categories are low indicating a preference for participants as Themes rather than circumstances.

CHART 2
Lexico-semantic properties of 6 registers compared
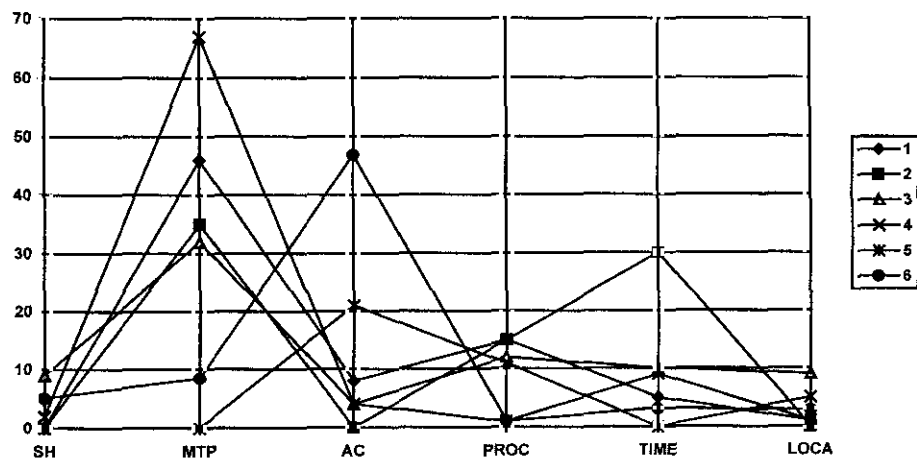
Legend : SH (speaker/hearer), MTP (major text participant - animate), AC (abstract concept), PROC (process), Time, LOCA (location)
1. Obituary, 2. Programme, 3. Narrative, 4. Exposition, 5. Sports Commentary, 6. Abstracts

The above chart shows similarities and differences for the lexico-sematic properties of the 6 registers considered. Correlations between registers are given in table 4 below.

TABLE 4

Correlations between the lexico-semantic features of the 6 registers

	Obituary	Programme	Narrative	Exposition	Sports Com	Abstracts
Obituary	1.00	0.71	0.93	-0.17	0.94	-0.03
Programme		1.00	0.75	-0.48	0.72	-0.35
Narrative			1.00	-0.50	0.95	-.32
Exposition				1.00	-0.38	0.80
Sports Com					1.00	-0.06
Abstracts						1.00

Unlike the statistics given in table 2 for correlations between grammatical properties of Themes in the 6 registers, the above statistics show that there *are* differences in the lexico-semantic features of the Themes. Except for a high correlation with Exposition (.80), selection of Themes in the abstracts is negatively correlated with the other text-types. Again we would like to emphasise that only 6 lexico-semantic features were considered for the comparisons as statistics on the other features were not available . The features considered were references to Speaker/Hearer (Writer/Reader), Major Text Participants (animate), Abstract Concepts, Process, Time, and Location. In comparison with the other registers, abstracts are more like Exposition (.80 correlation), and less like Programma (-.35 correlation).

8. CORRELATIONS BETWEEN CATEGORIES AND ACROSS ABSTRACTS

In order to find out if the abstracts for the 30 academic disciplines could be further classified, the K-means clustering method (Jiu-en Shi, 1991) was used in this study. It was possible to cluster the abstracts into 2, 3, 4, or 5 groups based on the presence/absence of grammatical or lexico-semantic properties in each abstract type. The following table shows this classification. From the left, column 1 gives the 30 disciplines. The subsequent columns provide the information for each abstract type and its inclusion in groups that form the clusters. The first row at the top gives the clusters of groups for the grammatical and lexico-semantic features. To give an example of how the table should be

interpreted, we look at the abstracts for chemistry, nº. 6. With a cluster of 2 groups for grammatical features, chemistry belongs to group 1; with a cluster of 3 groups, it belongs to group 1; with a cluster of 4 groups , it belongs to group 1; with a cluster of 5 groups, it belongs to group 3. The same interpretation can be given for the groups of the clusters based on the lexico-semantic features.

TABLE 5

Sub-classification of abstracts

Discipline/Groups	Grammatical				Lexico-semantic			
	2	3	4	5	2	3	4	5
1. Accounting	2	2	1	1	1	1	1	1
2. Applied Lingu	1	1	2	2	1	1	2	2
3. Biology	1	3	1	1	2	2	3	3
4. Broadcasting	1	1	1	3	2	3	4	4
5. Business Manag	1	3	2	2	1	1	2	2
6. Chemistry	1	1	1	3	2	2	3	3
7. Communication	2	1	1	3	1	1	1	1
8. Computer Stud	1	3	4	1	1	1	2	2
9. Economics	2	2	3	4	1	1	2	2
10. Education	1	1	2	2	1	1	2	2
11. Film and Cinem	1	1	1	3	2	3	4	5
12. Finance	2	2	3	4	1	1	2	2
13. Geography	1	3	2	2	1	1	2	2
14. Geology	1	3	4	5	2	2	3	3
15. History	1	1	1	3	2	3	4	4
16. Information Sys	1	1	1	1	1	1	2	2
17. Journalism	1	1	1	3	2	3	1	1
18. Language	2	2	3	4	1	1	1	1
19. Law	2	2	1	1	1	1	2	2
20. Marketing	1	1	1	1	1	1	2	2
21. Mathematics	1	1	1	1	1	1	2	2
22. Music	1	1	1	3	2	2	3	3
23. Philosophy	2	1	1	3	1	1	1	1
24. Physical Educa	1	3	4	5	1	1	1	1
25. Physics	1	1	1	1	1	1	2	2
26. Plant Pathology	1	3	4	5	2	2	3	3
27. Political Sci.	2	1	1	3	1	3	4	5
28. Social Science	2	1	1	1	1	1	1	4
29. Sociology	1	1	1	1	1	1	2	2
30. Theology	2	1	1	3	1	3	1	1

The above table shows that irrespective of the number of groups for grammatical and lexico-semantic properties, a few abstracts stick together in the same group, i.e. (1 1 1 1). It can be concluded that these abstracts are closer together in terms of the features considered. There are five such abstracts in the grammatical section, i.e. Information Systems, Marketing, Mathematics, Physics, and Sociology and five in the lexico-semantic section, i.e. Accounting, *Communication, Language, Philosophy, and Physical Education. Other group* patterns can be established, i.e. (1 1 2 2 or 2 2 1 1 etc.).

9. CONCLUSIONS

The building blocks of all meaningful texts are clauses which have a structure in terms of what comes first, the Theme, and what comes last, the Rheme. If the Theme is to be taken as the point of departure for each clause, then it is possible to establish a pattern of thematic organisation which contributes to the method of development of each text. Studies of local and/or global features of abstracts have so far concentrated on linguistic characteristics either below or above the clause level. Thematic analysis shows that it is possible to relate choices in one part of the clause to patterns of organisation above the clause.

Halliday (1982:231) emphasises the similarities between the organisation of a clause and the organisation of a text.

> It is no accident that it is possible to illustrate so many of the relations in a text by reference to relations in a clause. ... In showing that the text has structure, coherence, function, development and character, we cannot help at the same time showing that a clause has all these things too, though in an interesting variety of different ways. Presumably this is how clauses evolved — as the most efficient means of encoding text.

Our understanding of how texts are created and interpreted would be much poorer without the concept of Theme and thematic organisation. The findings of the present study confirms Halliday's insights especially when the results are compared and contrasted with the findings for other registers. Thematic analysis is not a panacea for all our questions about a register. It is, however, a powerful and down-to-earth way of establishing some of the crucial characteristics of the texts that we encounter everyday in our professions.

Department of English Language and Literature
University of Brunei Darussalam

REFERENCES

Allen, J.P.B. and Widdowson, H.G. (1974). *English in Focus*. Oxford: Oxford University Press.

Backlund, I. (1991). Theme in English telephone conversation. Paper delivered at the 17th International Systemic Congress, Stirling, Scotland.

Barber, C.L. (1962). Some measurable characteristics of modern scientific prose. *Contributions to English syntax and philology*. Gothenberg Studies in English No. 14, pp.1-23, Stockholm: Almquist and Wiksell.

Bates, M. and Dudley-Evans, A. (1976). *Nucleus: English for science and technology*. Harlow, England: Longman.

Bazerman, C. (1984). The writing of scientific non-fiction: contexts, choices, constraints. PRE/TEXT, 5:39-74.

Berry, M. (1995). Thematic options and success in writing. In *Thematic Development in English Texts*. M. Ghadessy (ed.), London: Pinter Publishers.

Berry, M. (1996). What is Theme? *Meaning and Form: Systemic Functional Interpretations*. M. Berry, C. Butler, R. Fawcett, G. Huang (eds.) Norwood, New Jersey: Ablex.

Downing, A. and Locke. Ph. (1992). *A University Course in English Grammar*. London: Prentice Hall.

Eggins, S.; Wignell, P. and Martin, J.R. (1993). The discourse of history: distancing the recoverable past. In *Register Analysis: Theory and Practice*. M. Ghadessy (ed.), London: Pinter Publishers.

Francis, G. (1990). Theme in the daily press. *Occasional Papers in Systemic Linguistics*. Vol. 3: 199-244.

Fries, P.H. (1983). On the status of Theme in English: arguments from discourse. Papers in Textlinguistics, Vol. 45, János, S. Petöfi and Emel Sözer (eds.), *Micro and Macro Connexity of Texts*. Haburg: H. Buske.

Fries, P.H. (1992a). Themes, method of development, and texts. Unpublished manuscript.

Fries, P.H. (1992b). Information flow in written advertising. Proceedings of Geogetown University Roundtable on Language and Linguistics. Washington: Georgetown University.

Ghadessy, M. (1993a). *Register Analysis: Theory and Practice*. London: Pinter Publishers.

Ghadessy, M. (1993b). Thematic development and its relationship to registers and genres. *Occasional Papers in Systemic Linguistics*. Vol 7: 1-25.

Graetz, N. (1985). Teaching EFL students to extract structural information from abstracts. In Uligin, J.M. and Pugh, A.K. (eds.) *Reading for Professional Purposes*. Leuven, Belgium: ACCO.

Halliday, M.A.K. (1982). How is a text like a clause. In Allen, S. (ed.), *Text Processing*. Stockholm: Almqvist and Wiksell.

Halliday, M.A.K. (1985,1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

Jiu-En, Shi. (1991). *Dynamic and Static Data Processing*. Beijing: Chinese Meteorological Press.

Leckie-Tarry, H. (1993). The specification of a text: register, genre, and language teaching. In Ghadessy (ed.) 1993a.

Littlewood, W. (1981). *Communicative Language Teaching: An Introduction.* Cambridge: Cambridge University Press.

Martin, J.R. (1992). *English Text: System and Structure.* Amsterdam: John Benjamins.

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