A GENRE-BASED INVESTIGATION OF THEME:
PRODUCT AND PROCESS IN SCIENTIFIC RESEARCH
ARTICLES WRITTEN BY NNS NOVICE RESEARCHERS

Thesis submitted in accordance with the requirements
of the University of Liverpool for the degree of
Doctor in Philosophy by
Hugh Robert Martin Gosden

JANUARY 1994

<u>To my wife Eleine</u>
For her constant inspiration from our early days
together at UCNW, Bangor; for her patient support
and understanding when it was hard to integrate
other things; for her insights in fields in which
we have come to share a common interest.

<u>To my parents Bob and Mary Gosden</u>
For their care and interest over many years and
long distances.

A GENRE-BASED INVESTIGATION OF THEME: PRODUCT AND PROCESS IN SCIENTIFIC RESEARCH ARTICLES WRITTEN BY NNS NOVICE RESEARCHERS

Hugh Gosden

ABSTRACT
This multi-method study presents an exercise in applied discourse analysis conducted within the broad framework of systemic-functional linguistics. The theoretical part of the work explores the relationship between one functional component of language, Halliday's notion of Theme, and the characterisation of a particular genre, the scientific research article (RA). Relevant literature on a variety of views of genre and Theme in the traditions of English for Academic Purposes (EAP) and systemic-functional linguistics is reviewed. The integration of these two levels of functional description is used as a basis for exploring the way in which Theme and generic structure relate to 'successful' processes and products of scientific research communication. The particular educational setting for the applied part of this work is the writing of first scientific RAs in English by NNS (non-native speaker) novice researchers.

The primary method of study adopted here is corpus-based and initial discourse-functional analysis and description of marked and unmarked thematic choices are based on a corpus of 36 published RAs in the physical and life sciences written by 'experienced' NSs. This corpus represents a base 'norm' of thematic usage against which other corpora are compared, namely, published RAs written by 'experienced' NNS scientists and unpublished first and final RA drafts written by NNS novices. Major findings indicated that 'appropriate' thematic selections in the RA genre are constrained by the changing rhetorical purposes, signalled by means of *moves*, which operate throughout the different stages of scientific RA discourse; thus, the textual metafunction of Theme plays a significant role in the characterisation and dynamic within-text structuring of the scientific RA genre. Furthermore, background surveys by means of questionnaires and interviews of the participants in the process of international research communication, in particular, of 'expert' NS journal editors, confirmed that 'appropriate' thematic control was clearly associated with the judgement of the merits of NNSs' RAs, and thereby, their 'successful' publication. With the pedagogical application of such theoretical insights in mind, the use of the teaching/research tool of Propositional Clusters (PCs) was explored in the EAP classroom as a heuristic for raising NNS novices' awareness about the manipulation of Theme in drafting and redrafting RA sections. Data collected from PCs exercises indicated their potential to raise awareness about the role of 'appropriate' thematic control in helping to create 'successful' texts.

This study contributes to our understanding of aspects of the functional relationship between elements of discourse structure and lexico-grammatical components such as Theme/Subject. In addition, reflecting the social-semiotic perspective of a systemic-functional framework, this work strongly emphasises the social-constructionist nature of the processes involved in international research communication through the medium of the scientific research article.

ACKNOWLEDGEMENTS

LIST OF CONTENTS

Title Page
Abstract
Acknowledgements
List of Contents
List of Tables
List of Figures

CHAPTER 1   INTRODUCTION

PART I
DESCRIPTION AND ANALYSIS - FOCUS ON GENRE/PRODUCT

CHAPTER 2   GENRE IN LINGUISTICS AND LANGUAGE EDUCATION

CHAPTER 3   THEME IN SCIENTIFIC RESEARCH ARTICLES

<u>PART III</u>
<u>APPLICATION</u> – <u>FOCUS ON GENRE/PROCESS</u>

CHAPTER 6   INVESTIGATION OF THE SCIENTIFIC RESEARCH
ARTICLE WRITING PROCESS: PROPOSITIONAL CLUSTERS
AS A HEURISTIC FOR RAISING NNS NOVICES'
AWARENESS ABOUT THE MANIPULATION OF THEMATIC
STRUCTURE IN REDRAFTING RA SECTIONS

LIST OF TABLES

LIST OF FIGURES

Fig.29 Example #2 – DISCUSSION: FINAL draft. Thematic flow of
information   266

CHAPTER 1 INTRODUCTION

1.0 BACKGROUND TO THIS CHAPTER

This thesis presents an exercise in applied discourse analysis conducted within the broad theoretical framework of systemic-functional linguistics. The aim of the theoretical part of the study is to explore the relationship between one functional component of language, Theme, and the characterisation of a particular genre, the scientific research article (RA). The applied part of this work employs the resulting integration of these two levels of functional description as a basis for exploring the way in which Theme and generic structure relate to 'successful' processes and products of research writing in a particular setting. In order to illustrate the multi-layered nature of this account, sections 1.1-1.5 of the introduction to the theoretical underpinnings and pedagogical background of this work highlight the five components which make up the thesis title: "A genre-based investigation of Theme: product and process in scientific research articles written by NNS (non-native English speaking) novice researchers". These first sections present an overview of this multi-layered study and contextualise the various key concepts which will be investigated in greater detail in subsequent chapters.

Since multiple layers of research imply multiple research methods, there is also the need for an overview of the general research design of the study and of the rationale for underlying methodologies employed throughout the various stages of research. To this end, section 1.6 briefly summarises the general research hypotheses which arise from the explication of the various strands of the thesis title and section 1.7 highlights the general research design of the study. This introductory Chapter One concludes with a detailed outline

of the contents of the six chapters to follow (section 1.8) which together form the three major theoretical and applied parts of this work.

As indicated above, the various layers which make up the present research are reflected in the thesis title and these five elements will now be taken up in turn:

- "A genre-based investigation.."
- "..of Theme:"
- "product and process.."
- "..in scientific research articles.."
- "..written by NNS novice researchers."

1.1 "A GENRE-BASED INVESTIGATION.."

In recent years in the fields of linguistics and language education, genre has become a controversial and highly politicised issue. Much of the current debate is associated with various schools of educational thought in North America, Europe and particularly in Australia, under the influence of the systemic-functional school of linguistics pioneered by Michael Halliday (1967, 1970a/b, 1973, 1977, 1978, 1979, 1985a/b, 1989a/b) and others. It is amongst those educationalists guided by the systemic-functional perspective that 'genre-consciousness' has perhaps gained the strongest and most coherent position as regards the agenda for language education (Green 1987). This is exemplified by research in both English first language (L1) and second language (L2) contexts, with a focus on the development of language skills at all ages from children to those in higher education, and in a variety of educational, professional and business settings [see, for example, Martin & Rothery (1986); Christie (1986); Brown & Herndl (1986); Hasan

(1989); Berry (1989a); Swales (1990b); Berkenkotter *et al.* (1991)].

In the fields of English for Specific Purposes (ESP), and related specialisms such as English for Academic Purposes (EAP) in the area of higher education, increasing degrees of 'genre-consciousness' and a resulting interest in genre-based approaches to curriculum design have become evident. This is particularly pertinent for the design of writing syllabuses intended for one of the principal populations which EAP caters for, namely, NNS science and engineering students at the postgraduate level who are typically required to write academic papers and theses in English, the dominant language of international research communication (Baldauf & Jernudd 1983a/b; Swales 1985a/b). Davies states that,

> 'the goal of much recent genre-based research is that of developing a heuristic for identifying and describing differences among text-types and genres, as a basis for planning courses in writing' (1991:1).

Over the last decade, the practical potential of a genre-based curriculum has been clearly demonstrated by the work of John Swales (1981, 1983, 1986a/b, 1987b, 1990a/b) and others in the field of EAP. Swales (1990a) notes that a genre-based approach owes its origins to a variety of applied and non-applied fields, both within linguistics and language education and outside. Particular influences have been:

- *variety studies*: early quantitative studies of the linguistic properties of functional varieties or registers of a language, such as scientific English, for example, Halliday *et al.* (1964);
- *skill and strategy studies*: the traditional categorisation of language learning activities into the four skills, and thereby, specialisation in particular

skill areas such as L2 writing. Swales comments that 'one specific area of investigation which holds particular promise is the tapping into the processing strategies of experts in a specific genre' (1990a:14) and studies such as Bazerman (1985) are valuable in revealing, in this case, target-like reading behaviours of a group of scientists;

- *discourse analysis*: for example, the early work of Sinclair & Coulthard (1975) on discourse acts and the establishment of macro-patterns in text, such as, Hoey's work (1983) on Problem-Solution macrostructure;

- *situational approaches*: research into situation-specific skills and strategies, for example, Ventola (1983) on Service Encounters;

- *notional/functional approaches*: the considerable influence of the in-built commitment of such approaches to communicative purpose and learner need, for example, the Casualty Department setting of Candlin *et al.* (1976);

- *writing context studies*: studies in the social contexts of academic writing as an outgrowth of a tradition of rhetorical scholarship, for example, Bazerman (1988) and Miller (1984);

- *sociolinguistics*; the frameworks and categorisations offered by major figures of ethnographic or systemic persuasions, such as Hymes (1974) and Halliday (1978);

- *cultural anthropology*: the work of Geertz on the necessity of 'local knowledge' in understanding 'what is really going on out there where men and women are thinking about things and writing down what it is they think' (1983:7).

It can be suggested that the field of EAP has been particularly successful in exploiting the pedagogical application of theoretical insights gained from research into genres from a synthesis of such a broad range of eclectic viewpoints. In the case of EAP-driven research, as with the current work, a great deal of research has

been based on the standard scientific genre commonly referred to as the research article (RA). One important practical reason for the success of a genre-based approach in EAP may be the accessibility, to teachers and students alike, of Swales's original (1981) discourse-based research on patterns of RA rhetorical structure, the latter term here taken as referring to the structure of 'the unfolding argument' (Halliday 1988:168) of scientific discourse.

Swales's (1981) study focuses on the patterning of what he calls rhetorical *moves* which may be broadly related in functional terms to other differing descriptions of discourse patterns, or staging (Martin 1992a), such as Hasan's (1989) *elements of structure* and Ventola's (1988) *schematic structures*. Swales's work clearly reveals the dynamic staging of rhetorical moves and the manner in which the moves inherent to each of the clearly delineated and labelled RA sections - typically Introduction, Methods, Results, Discussion (IMRD) - vary according to changing discourse demands and purposes. This scaffolding of macrostructural components has been eagerly adopted as a framework in recent functionally-oriented Academic Writing courses aimed particularly at novice NNS researchers [see, for example, Davies (1987), Weissberg & Buker (1990)]. Indeed, the motivation for the present research initially arose from teaching such courses in Academic Writing to NNS postgraduate science and engineering students at the English Language Unit, University of Liverpool and from supervising their research writing activities. Subsequently, this motivation developed in the natural process of attempting to refine and improve the effectiveness of L2 and research support for this and other audiences.

It may be suggested at the outset that a genre-based approach to the teaching and learning of scientific research communication skills is indispensable. After all,

it is evident that countless novice NNS researchers around the world who lack access to formal EAP support must somehow learn how to write academic papers, that is, learn the new genre/process skills required, by reading academic papers, that is, by imitating the genre/product model. [This is not to suggest that this is a desirable state of affairs, nor that novices will thus be able to 'successfully' write academic papers in the L2]. For those NNS novices who do have access to formal EAP support, the kind of approach proposed by Swales (1987b) to the teaching of the research paper is based on four bodies of literature which he suggests need to be utilised in concert: (a) the sociology of science (b) citation analysis (c) technical writing and (d) English for Academic Purposes. Thus, it is quite natural that, in designing Academic Writing courses and in teaching novices the skills of this new genre, academic papers themselves are a primary resource for the EAP teacher-researcher. They function both as a research tool and as EAP classroom teaching materials which are employed to encourage novices' awareness of L2 research writing processes in an attempt to bring to consciousness explicit knowledge about genre.

Stainton (1992) mentions two particular advantages she sees for the relevance of genre theory and thus for a genre-based approach to language education and these are of great relevance for the EAP classroom. Firstly, genre provides an economical way of talking about text and facilitates the development of metalanguage. For students to be able to appreciate the differences between the variety of language features, explicit knowledge about language has to be developed as part of their language learning. Using genre as a basis, however, means that the field of learning about language is narrowed to a particular set of features. Secondly, genre provides a means of comparing the demands made by different tasks. It provides a framework for the design of tasks and a

controlled means of evaluating the appropriacy of response
to the task specifications.

However, if genre theory is to be used as 'a way in'
(Martin 1993:156) to, for example, raising the awareness
of NNS novices about the L2 research writing process, then
following Stainton (1992), there are still basic questions
which require more detailed investigation: "how do we
identify, characterise and distinguish one genre from
another?" and "what distinguishes a 'successful' from an
'unsuccessful' instance of a given genre?".

The descriptive work of Swales and many other scholars who
have followed his lead in EAP-driven classroom-based
research demonstrate that, in one small corner of language
education, the field of EAP is a particularly strong and
coherent example of genre-based curriculum design and the
current work aims to build on this considerable tradition.
However, the fundamental strength of this tradition may
also be viewed as its fundamental weakness, that is that
it draws from such an eclectic array of theoretical
perspectives and bases that there is a need for more
clearly systematic descriptions of genres (such as the
scientific RA) and their linguistic realisations.
Resulting analyses would enable us to address central
questions such as those above and to systematically apply
valuable insights in the EAP classroom. Since the social
and cultural context of the present research is the
scientific discourse community, a synthesis of EAP-driven
interests with the strongly sociological perspective of
systemic-functional theory and its analytic tools provides
an ideal framework within which to investigate these
research questions further. Chapter Two will discuss more
fully current systemic approaches to the modelling and
description of genre in linguistics and language education
and will explore the potential to characterise the
interrelated social contexts of genre and register in

relation to the linguistic aspects of their functional
organisation and lexicogrammatical realisation.

We now turn our attention to the component in the thesis
title which, it has been suggested, plays a central role
in the linguistic realisation of generic character.
Section 1.2 introduces the notion of Theme and outlines
its overall place within Halliday's functional
organisation of language.

1.2 "..OF THEME:"

In broaching an area of inquiry, such as, "which features
of language help us distinguish one genre from another?",
Sinclair (1985b, 1989) has strongly emphasised the
importance of the continuing need for detailed description
of language. Consequently, studies of the linguistic
properties of varieties of written texts are a natural
prerequisite for a genre-informed approach to curriculum
design. Swales (1990a) comments that the focus of many
early studies was primarily on what he characterises as
the 'discrete-item surface feature assembly of data'; in
contrast, more recent analyses have generally become
'narrower and deeper' in focus. Halliday reminds us that,

> 'there are many different purposes for analyzing a text, and the
> scope and direction of the analysis will vary accordingly. Often
> we may want to scrutinize only one or two features, but to
> follow them through to a considerable depth' (1985b:54).

In ESP-type analyses, Swales suggests that this narrowing
of textual scope has been compensated for by an interest
in providing a 'multi-layered textual account' and, as a
result, there is a growing interest in 'assessing
rhetorical purposes, in unpacking information structures
and in accounting for syntactic and lexical choices'

8

(Swales 1990a:3). Eiler suggests that we know, intuitively or abstractly, that in order to convey and redeem meaning in texts, writers select from syntactic and lexical choices, given the genre constraints of their work, and readers interpret them in the light of conventional uses,

'but how can we capture as much analytically and empirically? How can we make heuristic generalizations regarding text design that in fact reflect actual writers' choices?' (1986:49).

In indicating a functional orientation for applied discourse analysis, Eiler (1986) suggests that one approach to these questions is to make case studies of texts, observing writers' products as specific incidents of linguistic choice and correlating such choice with use and function. In her own case study, Eiler's analysis of a scientific textbook chapter suggested that,

'genre is realized within the linguistic system that governs text formation itself - language's textual component. In fact, analysis of THEMATIC CHOICE and DISTRIBUTION - one aspect of the textual component as realized in the features of a specific text - can reveal heuristic structures defining a genre' (1986:49).

This work by Eiler, as well as an increasing number of studies which have adopted a genre-based functional approach to text analysis, raises interesting questions about the relationship between genre and Halliday's notion of Theme as 'the point of departure of the message' (1985a:38).

In recent systemic-functional research, Martin (1992a, 1993) presents a stratified model of genre and its realisation through the register configurations of field, tenor and mode which, in turn, are systematically associated with the linguistic system through functional components such as the thematic structure of Theme-Rheme. This important area of research is discussed in detail in

Chapter Two; however, it must be noted at the outset that
Theme is just one system of Halliday's functional
organisation of language and so the rationale for a
detailed focus on Theme in the present investigation, and
not on other linguistic systems, must clearly be
outlined. An introductory discussion is, therefore,
necessary here in order to contextualise the underlying
tenets of a systemic-functional framework in relation to
traditionally non-functional areas of interest and methods
of analysis in ESP/EAP studies.

Berry (1989b) comments that what has always distinguished
systemic-functional linguistics from other types of
linguistics is an interest in the interrelations between
language and the contexts of situation and culture and a
belief that such interrelations can be described
systematically. Thus, Halliday's theory of language use
(1967, 1970a/b, 1973, 1977, 1978, 1979, 1985a/b, 1989a/b)
is intended to highlight the internal structure of
language and thereby explain why language is patterned as
it is and not in other ways; of central interest therefore
is 'how the linguistic features of a text relate
systematically to the features of its environment'
(Halliday 1985a:xv), that is, the Malinowskian (1923)
contexts of situation and culture. An underlying
fundamental concept is that these patterns of language use
can be interpreted as 'networks of interlocking options'
(Halliday 1985a:xiv) with structures generated as output
by these system networks which together constitute the
grammar.

In an overview of Halliday's view of language as a
semiotic system, Hasan (1987) summarises as follows:

- language consists of three strata: semantics,
  lexico-grammar and phonology;
- these strata are related by 'realisation': meanings are
  coded as wordings, wordings are coded as sound patterns;

- each stratum is describable as a network of options: the description is therefore paradigmatic, with environments for options also being defined paradigmatically;
- the semantic stratum is organised into four meta-functional components: experiential and logical (as sub-components of the ideational meta-function), interpersonal and textual;
- each meta-function specifies a particular (set of) option network(s) as its output at the lexico-grammatical stratum;
- each act of choice, the selection of each option, contributes to the formation of a structure;
- a unified structure in its totality is the output of selections from four distinct (sets of) lexico-grammatical networks, specified by the meta-functions.

Thus each of the three main metafunctions of the semantic stratum (modes of meaning), which Halliday sees as language universals, generates a different kind of structural mechanism as its output or realisation (modes of expression); 'these different types of structure are non-arbitrarily related to the kinds of meaning they express' (Halliday 1979:61); however, 'the structural tendencies.. may differ very considerably from one language to another' (Halliday 1979:70). The three metafunctional components of the semantic system may be described as follows:

The first of these is 'language as representation: the semantic system as expression of experience' (Halliday 1979:59): IDEATIONAL meaning is 'meaning in the sense of "content". The ideational [experiential] function of the clause is that of representing what in the broadest sense we can call "processes": actions, events, processes of consciousness, and relations' (1985a:53) - the relevant options are from within the system of transitivity and give rise to constituent structures such as Actor,

Process, Goal. By ideational [logical] meaning, 'perhaps
the most difficult to interpret' (Halliday 1979:73), is
meant abstract relations between propositions, the
recursive complex structures generated by means of
parataxis and hypotaxis.

The second is 'language as interaction' (Halliday
1979:60): INTERPERSONAL meaning is 'meaning as a form of
action: the speaker or writer doing something to the
listener or reader by means of language. The interpersonal
function of the clause is that of exchanging roles in
rhetorical interaction: statements, questions, offers and
commands, together with accompanying modalities'
(1985a:53) - relevant options are from within the systems
of mood and modality and 'the interpersonal mode of
meaning is seen to be reflected in structures that
permeate the clause' (Gregory 1987:98).

There is a third component, TEXTUAL meaning, 'whereby the
meanings of the other two kinds take on relevance to some
real context' (Halliday 1979:60), both the preceding and
following text and the context of situation. 'The textual
function of the clause is that of constructing a message'
(1985a:53), it 'enables the speaker or writer to construct
"texts", or connected passages of discourse that are
situationally relevant; and enables the listener or reader
to distinguish a text from a random set of sentences'
(Halliday 1970a:143); 'in modern jargon we might refer [to
the textual component] as the ecology of the text'
(Halliday 1979:60). Textual meaning is realised within the
Theme-Rheme structure; in addition, thematic structure is
closely linked to, though distinct from, information
structure which imposes the structures of Given and New.
Since in spoken discourse the assigning of information
structure derives from tonic prominence, in written
discourse, in particular, 'the demarcation of the New is
not entirely straightforward' (Davidse 1987:70). However,
Halliday notes that what the two text-forming systems

realised by thematic and information structure do 'is to
organise discourse into a succession of message units,
quanta of information such that each has its own internal
texture, provided by the two systems of prominence'
(1979:68).

In summarising his metafunctional organisation of
language, Halliday (1979) stresses the need to talk in
terms of tendencies and not rules. Thus,

- experiential structures tend to be more elemental in
  character and tend to generate CONSTITUENT-like
  structures;
- interpersonal structures tend to generate PROSODIC or
  CUMULATIVE structures, extending over long stretches;
- textual structures tend to generate PERIODIC or
  CULMINATIVE structures; they are elements which occur at
  the boundaries of significant units and they give a kind
  of periodicity to the text, which is part of what we
  recognise as "texture". Their semantic movement is
  wave-like, or pulse-like with peaks of prominence
  [Theme/New] and troughs of non-prominence [Rheme/Given]
  (Halliday 1979, 1985a).

The technical analysis of Theme, along with a discussion
of other current issues in research on Theme, will be
fully exemplified in Chapter Three. However, from this
brief introduction to Halliday's metafunctional
organisation of language, it is clear that Theme
represents but one system of language choices. Halliday
himself considered that no single one of the three
metafunctions was more basic than any other. So, why the
focus on Theme here and not, for example, on the systems
of transitivity or modality? I can suggest three reasons
for this:

(1) Halliday emphasises that it is the essential nature of
a functional approach that the three metafunctional

strands of meaning are all interwoven in the fabric of the discourse: 'we cannot pick out one word or one phrase and say this has only experiential meaning, or this has only interpersonal meaning' (1989b:23). Thus, every sentence in a text is multifunctional. Nevertheless, a clear distinction can be made between, on the one hand, ideational and interpersonal and, on the other, textual 'modes of meaning'. Ellis first points out this distinction in general terms,

> 'the textual function is alone among the functions of systemic grammar in having devoted to it, outside the systemic model, a whole new discipline, a branch of linguistics generally which goes under the names of "text grammar", "discourse analysis" etc; on the other hand, systemic grammar is alone.. in its treatment of the other functions, and in its perspective of integrating the textual with them' (1987:114/5).

Gregory (1987) and Matthiessen (1992) further clarify this distinction: the ideational and interpersonal metafunctions are yoked together as *extrinsic* (Halliday 1978), that is, dealing in different ways with what goes on in behaviour and ultimately referring to natural and social reality. These two stand in contrast to the 'enabling' textual function, which refers *intrinsically* to semiotic, symbolic reality, 'the reality brought into existence by language itself (Matthiessen 1992:42). It is the textual metafunction 'which breathes relevance into the other two' (Halliday 1985a:xiii) and therefore, since the Theme-Rheme structure makes up the functional configuration of the clause as message, it is only because we can select the desired form of the message that we can also use language effectively, both to represent an experience (experiential) and to interact (interpersonal) with those around us:

> 'This enabling, second-order character of the textual metafunction is reflected in the fact that it employs the modes

of organization engendered by the other metafunctions as carriers of textual waves' (Matthiessen 1992:42).

In terms of syntagmatic relations, each metafunction, in turn, offers choices from within all three metafunctions. Thus, the dynamic textual waves of Themes through discourse may themselves be composed of several functionally organised strands of ideational, interpersonal and textual meaning as complex, multiple Themes. Consequently, any study of the relationship between thematic selections and generic structures in a corpus of real texts is likely to include a focus on the discourse-functional roles of distinct ideational, interpersonal and textual components of thematic selections and their individual contribution to the overall structuring of a given text. Ellis emphasises this point in commenting that,

> 'at various levels of analysis and ranks the textual function is entwined also with the interpersonal and the experiential, and a full account of the textualness of a text will tend to be an account of all its functions with the emphasis on the textual' (1987:123).

Despite some criticisms that the criteria for distinguishing between Halliday's metafunctional components appear vague and impressionistic (Berry 1982), Gregory suggests that we should continue to explore their use 'and see what they can do for us. They have already shown themselves to be helpful organising and investigating concepts' (1987:104). This review suggests the value of a detailed exploration of textualness as viewed through the textual metafunction in relation to the dynamic organisation of the scientific research article.

(2) The second reason why the present study focuses on the textual metafunction is that for a number of years now, there has been increasing interest in Theme and

consequently a growing number of international conference
workshops and body of empirical research on thematic and
information structure in a variety of spoken and written
texts. The work of Eiler (1986) above and other
researchers working within a broad systemic-functional
framework on the relationship between genres, elements of
discourse structure and their linguistic realisations
lends considerable support to the potential of thematic
structure as a significant generic discriminator [see
Berry (1987a, 1989a/b, to appear), Davies (1988a/b, 1989,
1991), Plum (1988); Francis (1989a/b, 1990), Fries (1983,
1992, to appear), Fries & Francis (1992)]. To continue
this established tradition of research into Theme and
genre, which is an underlying objective of the present
study, Fries & Francis (1992) stress the need to further
examine real texts in detail and to relate thematic
choices to other types of information so that we can
explain why choices are made and what effects they have on
our interpretations.

(3) Thirdly, informal insights and intuitions gained from
working closely in the role of participant-observer with
novice NNS researchers writing their first RAs in English
have over time become more formalised hypotheses about the
importance of thematic control and thematic
appropriateness in the writing and subsequent revision of
'successful' texts, success being judged in this context
and, of course, on only one level, by acceptance for
publication in international journals. In the vital
processes of revising RA manuscripts following review and
critique, effective manipulation of thematic selections
appears essential since this process enables the writer to
more clearly indicate relevance to co-text and the
contexts of situation and culture and thereby construct
more coherent and cohesive texts.

On a theoretical level, then, it is the aim of the present
work to address some of the issues raised in previous

research on thematic analysis, much of which reflects a deep commitment to the social-semiotic perspective which drives Halliday's functional theory of language. Such research and text description aims to contribute to 'the most important work to be done in systemics, in order to delineate the relationships between the meta-functions and the semantic and lexico-grammatical strata' (Gregory 1987:103). However, the research dynamic necessarily draws from difficulties in working with and applying some of these same ideas in specific educational contexts; it thus reflects the spirit of inquiry engendered in the work of scholars such as Christie (1986, 1987), Berry (1987b), Martin *et al.* (1987) in aiming to correlate facts and explanations about aspects of language use, in this case, thematic structuring in the scientific RA genre and its relation to the broader situational and cultural contexts of the scientific discourse community and its shared set of communicative purposes.

I have indicated that a cornerstone of systemic-functional linguistics is its central concern with language as a social phenomenon. Yet, how does this concern relate to the methodologies of teaching writing skills to a group of L2 learners, the population which is the focus of pedagogic interest here. Recent years have seen combative debates on various process- *versus* product-oriented approaches to the teaching of writing, with L2 education having been strongly influenced by changes in L1 fields. However, particularly at the level of science research education, shared L1/L2 interests have also led to an increasing awareness of the interactive nature of 'writing as social action', thereby aligning this field more closely with the research potential of a systemic-functional framework and social semiotic perspective. These changes in viewpoint are contextualised for the present study below.

1.3 "PROCESS AND PRODUCT.."

It is well known that there has over the last twenty years been a major paradigm shift in composition theory and research whereby the emphasis has moved from the *product* to the *process* of writing. Witte & Cherry (1986) note that the rediscovery of process has become so important that one of the classic texts of the history and philosophy of science, *The Structure of Scientific Revolutions* by Thomas Kuhn (1970 2nd. edition), is regularly invoked to characterise the significance of this paradigm shift. Kuhn's thesis is that the assimilation of a new theory requires the reconstruction of prior theory and the re-evaluation of prior fact, an intrinsically revolutionary process which produces,

> 'a consequent shift in the problems available for scientific scrutiny and in the standards by which the profession determined what should count as an admissible problem or as a legitimate problem-solution' (Kuhn 1970:6).

According to Silva (1990), the theory behind the traditional product-centred approach stressed learning to write as a form-focused exercise in habit formation; the writing context was the classroom with negligible concern for audience or purpose. The writing process itself was seen as linear, determined by writers before they started to write (Connor 1987). In contrast, the process-centred paradigm emphasises recursiveness in the writing process; it considers audience, purpose and context of writing and teaches strategies for invention, drafting, revising and editing. The writer is the centre of attention - someone engaged in the discovery and expression of meaning. 'The text is a product - a secondary, derivative concern, whose form is a function of its content and purpose' (Silva 1990:16). Within the process camp, the cognitivist or 'writing as problem-solving' viewpoint (Flower 1979; Flower & Hayes 1977, 1980, 1981a/b, 1983) has been

particularly influential in L1 research and Johns (1990) comments on its enormous influence upon L2 classrooms. However, Johns also warns that 'we may be doing our students a disservice by strictly adhering to all tenets of this [process] approach' (1986:251) and indeed much of the criticism levelled against an overconcentration on process comes from scholars and educationalists working within the genre-based framework of English for Academic Purposes (EAP) and from a systemic-functional linguistic perspective.

Reid (1984a/b) suggests that the cognitive process approach neglects variations in writing processes due to differences in writing tasks and situations and, in particular, it neglects the development of schemata, or rhetorical scaffolding, for academic discourse. Horowitz (1986b) is most succinct in his criticism: (i) the process approach does not realistically prepare students for academic work since it creates a classroom environment which bears little resemblance to the situations in which students' writing will be exercised (ii) it ignores certain types of academic writing tasks (iii) it gives a false impression of how academic writing will be evaluated and reflects what Swales (1990a) calls the 'soft' process which protects students from the exigencies of external criteria (iv) a basic tenet of the process approach, that 'content determines form', is not necessarily true of academic discourse (v) it overemphasises the individual and neglects the sociocultural context, that is, 'the realities of academia' (Silva 1990:17).

Bizzell (1982) has also most strongly emphasised this latter point, namely, that academic writing cannot be viewed solely as an inner-directed activity, but that it must also be seen as an acquired response to the preferred modes of communication and discourse conventions of a particular academic community, Swales's (1988, 1990a) 'hard', 'norm-developing' process - 'hard' since it

anticipates and countenances the reactions of the intended readership and 'norm-developing' since it is connected to the persuasive reporting task of the outside professional world [as against 'norm-developed' where it is a question of students showing familiarity with a body of knowledge with little need for subtle negotiation with an outside audience]. Martin (1985) considers that the process writing approach, which he believes mystifies what has to be learnt to produce effective written products, strongly disadvantages 'outsider' groups. In his Australian context, he cites migrant and Aboriginal children as potential examples; the same disadvantage is evident for those who wish to, or who are required to participate in their international research community through the dominant L2 written medium of English (Baldauf & Jernudd 1983a/b; Swales 1985a/b). Such groups have the status of 'outsiders' on two counts; firstly, as L2 writers dealing with the challenge of a new genre and secondly, in their apprenticeship as novices in their fields of academic research.

Because of these many criticisms, Johns (1990) emphasises the important role of another paradigm for composition theory and research which attempts to fill this sociocultural vacuum, a *social constructionist* perspective. This view of the writing process has much in common with the fundamental origins of systemic-functional linguistics with its emphasis on the Malinowskian (1923) contexts of situation and culture. Here, the product is considered a social act that can take place only within a specific context and audience; the knowledge, the language and the nature of discourse are determined by the discourse community for which it is written. Indeed, in recent sociological studies of the writing activities of the academic community, for example, of the initial drafting and revising of scientific papers by researchers in response to feedback and criticism, scholars such as Bazerman (1988) and Myers (1985, 1988) have strongly

emphasised the very nature of these writing and negotiation processes as social action.

However, in commenting on the harsh realities of becoming new members of a particular discourse community through academic publications, Johns (1990) comments that its implicitly shared goals and discourse conventions may well be difficult for 'outsider' novices to fathom. The many unwritten 'rules of the game' of academic discourse manifest themselves textually in a multitude of linguistic and rhetorical subtleties - how do these come to be appreciated and 'appropriately' imitated by novices? Berry notes that there has been little attempt 'to study communicative competence systematically, to find out exactly what it is that we know about the relations between language and situation, to find out how such knowledge is structured' (1989b:48). For practitioners of English for Academic Purposes whose primary pedagogic concern is the initiation and 'successful' participation of novice NNS 'outsiders' in their respective international discourse communities, it is clearly important to understand more fully factors which contribute to the development of novices' *academic* communicative competence (Berkenkotter *et al.* 1991) and thereby their ability to produce 'successful' RAs.

A basic research problem then is to combine the two conceptions of the text as product and as process and to relate them both from a social-semiotic perspective to the metafunctional organisation of language structures and systems (for example, thematic selections in declarative statements) and aspects of the contexts of situation and culture (scientific RA genre/register) that lie behind them. The present study of the textual organisation in scientific RA discourse from a functional, social-constructionist perspective aims to explore one important area of this fundamental research.

Thus to re-invoke Kuhn, with a shift in the research paradigm towards one firmly embedded in 'writing as social action', a different range of admissible problems and legitimate problem-solutions is thus opened up to teacher-researchers and discourse analysts. However, as we move away from the traditional interests and methodologies of the L2 domain into interdisciplinary areas which have been of concern to social scientists for many years, how are such problems and solutions to be explored? Echoing Sinclair's (1985b, 1989) earlier reminder that an important starting point is a detailed description of discourse and the linguistic properties of language varieties, Berry comments that,

> 'teachers need to make much more explicit precisely what it is that has to be learnt in order to produce effective written products; .. Making explicit what people know about effective texts might be thought to be particularly the job of the discourse analyst or textlinguist. .. Yet discourse analysts have been slow to focus on the question of what makes a text successful or not, and even slower to investigate this question in the context of the specialised language varieties of the workplace' (1989a:64).

This brings us then to the last two elements in the thesis title which require initial description prior to detailed investigation: the specialised language variety of the scientific research article (section 1.4) and the workplace of the science laboratory where NNS novice researchers are typically required to write up and negotiate the publication of their research results in English as a second language (section 1.5).

1.4 "..IN SCIENTIFIC RESEARCH ARTICLES.."

Potter & Wetherell (1987) comment that there are good reasons for studying scientists' language:

> 'It might seem odd that discourse analysts have taken scientific language as one of their central topic areas. You might think that science is an abstract, technical and precise realm, and as such can tell us little of value about the social and psychological processes operating in language. Yet it is exactly these features which make scientists' discourse such an attractive research topic. If interesting discourse processes can be found even in this rarefied environment it is extremely likely that they will also be found, probably in more exaggerated forms, in everyday conversation, newspaper stories and in all other kinds of talk we will encounter. Science is a useful hard case where discourse analysis can hone its claims' (1987:64).

The standard product of the 'manufacture of knowledge' (Knorr-Cetina 1981) which functions as a primary mechanism of communication between members of the international scientific discourse community is the research article (RA). The RA genre is thus highly conventionalised, yet, at the same time, it has been shown that it is also constantly evolving (Bazerman 1984, 1988; Swales 1990a). As has been suggested above, the RA and a whole range of other research process genres are of interest to a variety of interdisciplinary L2 and L1 fields - on the one side English for Specific/Academic Purposes and Applied Discourse Analysis; on the other Composition, Technical Communication, Sociology of Science and Research Communication Studies. It is, therefore, not surprising that the literature that explores the textual properties of RAs is quite extensive (Swales 1990a). Moreover, Swales notes that, within this eclectic body of research, there is considerable variation in terms of:

- the *coverage* [all of the RA: Adams Smith (1984) or one
  of its sections; Swales & Najjar (1987) - results
  statements];
- the *scale of research* [corpus size: Swales (1981) - 48
  RAs; Hill *et al.* (1982) - 1 RA];
- the *level of analysis* [features selected for study:
  Heslot (1982) - tense; Tarone *et al.* (1981) - voice;
  Crookes (1986) - structure];
- the *fields* [Hill *et al.* (1982) - psychology; Adams Smith
  (1984) - medicine; Swales (1981) - undifferentiated];
- the *methodological and linguistic approaches* [the
  underlying philosophical and conceptual principles
  devised or drawn upon: Weissberg (1984) - Functional
  Sentence Perspective; Stanley (1984) - Clause Relations;
  Tarone *et al.* (1981) - Rhetorical-Grammatical Approach].

Perhaps partly due to Swales's (1981) own seminal applied
discourse analysis of RAs, there appears to have been an
entrenched research focus on RA Introductions in much of
the work that has followed: for example, in Swales's
(1990a) 37 page review of major RA sections (Introduction;
Methods; Results, Discussion and Conclusion), 29 pages
deal with research on features associated with
Introductions. There are, of course, good reasons for this
concentration as rhetorically multifunctional
Introductions 'are known to be troublesome.. and present
the writer with an unnerving wealth of options' (Swales
1990a:137). However, there is clearly a need for breadth
of coverage, as well as depth of analysis. Harris observes
that recent work which has attempted to define generic
properties of texts from within a systemic-functional
perspective, 'has so far been confined to "simple",
cooperative and relatively brief [spoken encounters or]
examples of written discourse' and that 'it would
certainly be extremely limiting to confine work in genre
to small, self-contained [verbal] interactions where
nobody has anything very much at stake and the most

interesting social and ideological issues simply do not arise' (1987:35).

Following these criticisms, and in an attempt to combine breadth of coverage with depth of analysis, it can be suggested that there is a clear need to adopt a holistic, ecological approach to the study of genres. As staged, communicative events, genres have beginnings, middles and ends and a systematic study of aspects of these entire progressions is desirable in order to fully reflect their communicative potential. Moreover, a study of exemplars of genres, in which writers *do* have much at stake, is likely to provide more valuable insights for pedagogic applications. These are primary objectives in the present work.

To the best of my knowledge, Nwogu's (1990) research is the only study to date which has attempted to characterise complete written scientific genres from a dynamic functional perspective. His research examines the differences in the organisation of discourse in three parallel genres of written medical texts, the Abstract, the research article and journalistic reports of RAs in terms of cohesion, thematic progression and rhetorical moves. Analysis of the Theme-Rheme concept is based on the theoretical approach of the Prague School of linguistics known as Functional Sentence Perspective, in particular the work of Daneš (1970, 1974) on thematic progression (see Figure 27 below for an illustration of Daneš's three major patterns). Nwogu's valuable study highlights common mechanical patterns of thematic progression across the three related genres and indicates significant variations between professional and journalistic accounts. However, in his study, thematic analysis is not approached from a discourse-functional perspective and hence there is little indication of how thematic selections are constrained by, and thereby characterise, top-down rhetorical structures in their staging and realisation of genre-specific goals.

Consequently, the current work aims to add to Nwogu's study on discourse variation in medical texts and to the growing body of work on scientific discourse. However, an important objective is to go further in the integration of levels of linguistic description and to explore the progressive and dynamic textual development of complete exemplars of scientific discourse, thus following the systemic-functional tradition of correlating variation in language form with variation in the contexts of situation and culture.

The exemplars of scientific discourse to be investigated here are written by both 'expert' and 'novice' RA writers, and by both NSs and NNSs of English. It is one particular permutation of these groupings, namely, NNS novice researchers and RA writers, which is the focus of interest in the applied part of the present work and we therefore turn now to the final component of the thesis title.

1.5 "..WRITTEN BY NNS NOVICE RESEARCHERS."

Baldauf & Jernudd (1983a/b) have shown that the predominance of English within the international discourse community has been steadily increasing over the past thirty years. Judging from English language research papers and authors' first language, Swales (1983, 1985a/b, 1987b) confirms the scale of the increasing pre-eminence of English and suggests that at least 50% of the several million RAs published annually are in English; in some fields, for example, Engineering, this may be as high as 80%. Moreover, Swales comments that 'the role of the NNS in this Anglocentric research environment remains rather obscure' (1987b:42), with well-trained NNS researchers remaining quasi-invisible due to a lack of exposure through international research reporting. The corollary of the lack of attention paid to this obscurity gives rise to

sensitive socio-political issues which cross the boundaries of economic and pedagogical considerations.

On the one hand, Dhaif is quite right to point out the negative effects of cultural and linguistic imperialism and observe that NNS researchers, especially from the so-called developing countries, should not need to 'justify [their] research worthiness by publishing in English' (1985:225). However, Swales's point above is this: if sponsoring governments and agencies around the world invest in scholarships for young NNS researchers, from a simple economic point of view, as well as from an overwhelming intellectual one, there is clearly a responsibility for the continued support and assistance of NNSs in being able to fully participate in their international research communities *on an equal research basis*.

Perhaps ironically, there are many NNS researchers who are aided with language support at the outset of their academic careers whilst graduate students, for example, through scholarships to English-speaking research environments in the U.K., North America, Australia etc. where participation in specialised EAP programmes is often a prerequisite for higher studies. As returnees, these young researchers then become immersed in their research communities once this professional threshold has been reached. However, it is possibly just at this time when assistance may be most required in maintaining confidence in their English language proficiency as the processes of submitting first RAs for publication and of gaining recognition and status may commence.

As mentioned above, the motivation for the present study comes from working over a number of years with two groups of such NNS novice researchers described above: firstly, with overseas postgraduate students at the English Language Unit, University of Liverpool who were

participating in specialist EAP courses, and subsequently, with doctoral students at a science and technology university in Japan. Clearly, there are some differences between these two audiences; with the former, we are conducting intensive short courses in the target L2 medium and environment but, in general, prior to students' main studies and, thereby, in anticipation of later research writing activities and participation in their respective discourse communities through publications. With the latter group, the educational context implies the greater immediacy of providing drip-feed L2 support to NNS novice researchers as they write their first RAs in English *in situ*, in the workplace of their laboratories [for example, it is commonly expected in Japan that doctoral candidates publish scientific RAs *prior* to graduation - these RAs typically represent separate chapters of a doctoral thesis]. With the desire here to explore aspects of RA drafting and redrafting processes in response to continual feedback and criticism, the latter group of NNS novices provides the data on which the applied part of the study is based.

The above explication of the five components of the thesis title has now brought us to the position where we can draw together the various strands and summarise general hypotheses on which this research is based (section 1.6). In turn, we can outline the general research design of the study set up to address and investigate those research areas (section 1.7). As already previously indicated, research questions and methodologies outlined in brief in this introduction are taken up in greater detail at later stages. Section 1.8 outlines a guide to the contents of the chapters to follow.

1.6 GENERAL RESEARCH HYPOTHESES

The testability and testedness of theories is important in all academic traditions and Berry emphasises that the setting up and testing of hypotheses in systemic-functional linguistics research is also indispensible: moreover, 'hypotheses must be stated in a form [so] that it is possible to review the evidence for and against them' (1989b:23). The general hypotheses on which this study is based fall into six major areas:

GENERAL RESEARCH HYPOTHESIS #1: Halliday's notion of Theme plays a significant role in the characterisation and dynamic within-text structuring of one particular genre - the scientific research article (RA).

GENERAL RESEARCH HYPOTHESIS #2: Such Themes can be categorised according to their discourse-functional roles at given stages of RA discourse and throughout the RA as a whole.

GENERAL RESEARCH HYPOTHESIS #3: Thematic selections in this genre are constrained by the changing rhetorical purposes, signalled by means of moves, which operate throughout the different stages of RA discourse, here, taken as formally separated sections [Introduction, Experimental, Results, Discussion].

GENERAL RESEARCH HYPOTHESIS #4: The 'appropriate' usage (as highlighted through analysis of a significant corpus of NS RAs) of thematic selections across RA sections appears to contribute significantly to the perceived coherence and cohesion of a research article. It is thus a significant measure of academic communicative competence and 'success', as defined and validated by acceptance of an RA for publication by an international English-language scientific journal.

GENERAL RESEARCH HYPOTHESIS #5: An examination of the RA
writing procedures and strategies, and, in particular, of
the progressive drafts of written products of novice NNS
researchers in response to feedback and criticism
illustrates the development of this crucial aspect of
academic communicative competence.

GENERAL RESEARCH HYPOTHESIS #6: A pedagogic focus on the
manipulation of genre-specific thematic choices in RA
writing activities in the EAP classroom may raise novices'
awareness about the potential of thematic manipulation to
help create 'successful' texts.

These and further related research questions and
hypotheses will be raised at various stages throughout the
study and discussed in greater detail as appropriate; the
next section 1.7 outlines the general research design of
the study set up to explore these basic hypotheses.

1.7 GENERAL RESEARCH DESIGN OF THE STUDY

Potter & Wetherell comment that any observation of the
physical or social world is imbued with theoretical
interpretation, 'even the simplest scientific description
is dependent on a whole variety of theoretical
assumptions' (1987:158). It is evident from the above
introduction that the current research involves a
combining of various theoretical and applied traditions in
language research. Thus, as regards research orientation,
it is quite natural that multiple methods of research will
drive the investigation of the multiple layers and stages
that make up this research project. The general research
design and philosophical underpinnings of this
multi-method account, therefore, reflect an eclectic
synthesis of theoretical assumptions and research

paradigms. By the term paradigm, Kuhn means to suggest that,

> 'some accepted examples of actual scientific practice - examples
> which include law, theory, application and instrumentation
> together - provide models from which spring particular coherent
> traditions of scientific research' (1970:10).

Paradigms tell the practitioner what is important, legitimate and reasonable; they also tell the practitioner what to do without the necessity of long existential or epistemological consideration - but all paradigms have their strengths and weaknesses (Patton 1978).

Silverman (1985) indicates the polar oppositions inherent in the two approaches of positivism and interpretivism; Grotjahn (1987) outlines in greater detail the basic distinctions by means of which these paradigms can be characterised and notes the two 'pure' forms of:
(i) the positivist *analytical-nomological* paradigm, using an experimental or quasi-experimental design, a quantitative form of data and a statistical method of analysis, where the testing of hypotheses is the central concern of empirical science, whilst the formation of hypotheses plays only a minor role;
(ii) the *exploratory-interpretative* paradigm, using a non-experimental design, qualitative data and interpretative analysis, where the division between hypothesis testing and hypothesis formation is often suspended.

In discussing the need for a coherent research methodology in systemic-functional linguistics, Berry (1989b) suggests the need to maintain what is thought of as the 'classic' Popperian (1959, 1976) approach of (i) above to scientific investigation, a hypothetico-deductive approach which aims to systematically test and *refute* generated hypotheses with counter-examples, rather than simply *verify* them.

[Incidentally, the strict necessity for this tradition is countered by Kuhn who comments that a falsification process 'might equally be called verification since it consists in the triumph of a new paradigm over the old one' (1970:147)]. However, Halliday & Fawcett note, perhaps somewhat tongue-in-cheek, that a quasi-Popperian approach,

> 'has about it an intellectually attractive rigour. But it is not, in reality, how progress is made in linguistics, or any other science. Indeed, an excess of rigour may lead to rigor - that is, to rigor mortis' (1987:1).

Lincoln & Guba (1985) comment that three central tenets of the Popperian positivist paradigm are concerned with (a) objectivity (b) hypothetico-deductive theory and (c) the separation of facts from meaning. In contrast, postpositivist critics [for example, Hesse (1980), Harre (1981)], characterise the post-empirical account of natural science in relation to these same three points above as follows:

> (a) 'in natural science data is not detachable from theory, for what count as data are determined in the light of some theoretical interpretation, and the facts themselves have to be reconstructed in the light of interpretation';
> (b) 'in natural science theories are not models externally compared in nature to a hypothetico-deductive schema, they are the way the facts themselves are seen';
> (c) 'meanings in natural science are determined by theory; they are understood by theoretical coherence rather than by correspondence with facts' (Hesse 1980:172 cited in Lincoln & Guba 1985:29).

Evidently, these accounts of an exploratory-interpretative postpositivist paradigm more clearly reflect the concerns of a social semiotic perspective of language use. Furthermore, a central tenet of the systemic-functional

model has parallel relevance for the general research design of this study. Berry (1975/1989) points out that language is so extremely complex that it is difficult to be cut-and-dried when analysing language. Consequently, in the systemic-functional tradition, linguistic description is frequently formulated in terms of dynamic points along progressive continua or clines - more towards pole *x* than pole *y*. Likewise, research paradigms themselves may be oriented more towards one pole than the other - more towards the qualitative end of the continuum than the quantitative, more inductive/heuristic than deductive, more applied than theoretical, more exploratory than confirmatory, more functional than formal. The corollary of adopting these positions is that underlying research methodologies may be naturally integrated on both perspectives; indeed a blending of paradigms is often seen as desirable in the social sciences where the simultaneous testing and formation of hypotheses [paradigm (ii) above] is more commonly conducted by means of naturalistic data (Kirk & Miller 1986).

In systemic-functional studies which are clearly more exploratory in nature, such as here, it is evident that the positivist approach of analytic deduction of the Popperian tradition is not entirely relevant or appropriate. Indeed, since the textual focus of this work is on scientific writing, what is highly relevant is that recent postpositivist sociological studies and analyses of scientists' discourse [see Latour & Woolgar (1979); Knorr-Cetina (1981); Gilbert & Mulkay (1984)] have cast doubt on the continued applicability of the 'classic' scientific method to the way scientists actually carry out and report on their work, primarily due to 'our hopelessly idealized view of the scientific enterprise' (Potter & Wetherell 1987:159). Other sociologists and philosophers of science have claimed more radically that the more plurality in theory and method the better [see, for example, *Against Method* Feyerabend (1975)] and that now

practically 'anything goes'. Whilst this position may have liberating benefits, Berry (1989b) is careful to point out that this approach should not suggest an *absence* of method.

In addition to the two 'pure' paradigm forms noted above, Grotjahn (1987) lists 'mixed' forms of interrelated paradigms, notably the exploratory-*qualitative*-statistical and the exploratory-*quantitative*-statistical paradigm. Working within such a mixed paradigm in systemic-functional linguistics, Nesbitt & Plum note that,

> 'the integration of both quantitative and qualitative description in modelling language provides potential insight into the relation between system and process, and the process of language change. It provides the conceptual apparatus to investigate the dialectic between the linguistic system and linguistic process. The problem can be addressed as follows: "To what extent do the quantitative patterns to be observed in linguistic process, the statistical properties of text, explain the qualitative patterns of linguistic difference?"' (1988:9).

In answer to this question, the findings of the study by Nesbitt & Plum suggest that it is only through a 'mixed' form integration of quantitative and qualitative methods that studies of language as a dynamic, open system may reveal types of linguistic patterning which would otherwise remain hidden.

In view of the broad base of the present research indicated by the thesis title, and given that initial research hypotheses (section 1.6) are closely related to the kind of question posed by Nesbitt & Plum, it was considered that an integration of diverse research concepts and methodologies into the kind of 'mixed' form noted above was most appropriate for the present exploratory study. However, approaches which strongly emphasise the qualitative, interpretative poles of the

research paradigm employing process-oriented, naturalistic data should equally aim to be as structured as their more normative paradigm counterparts.

As noted earlier in section 1.2, initial *a priori* hypotheses chiefly arise from addressing (rather than aiming to directly verify or refute) the various theoretical standpoints taken in previous work on Theme and genre. This appears to be a natural starting point - on the one hand, Fries & Francis suggest the need to eliminate preconceived assumptions in approaches to research on Theme and genre; however, even a process of discovery in analysis 'would require that the investigator begin with a very good idea as to the aspects of Theme which might be relevant to genre' (1992:50). In addition, pedagogical intuitions and insights gained as an EAP teacher-practitioner have contributed significantly to the research impetus. The overall approach to the initial theoretical analysis of Theme here may therefore be typified as quasi-deductive; however, in tune with the concept of dynamic continua or clines, the paradigm progressively changes throughout this work to become one more oriented towards an inductive/heuristic focus as the pedagogical application of initial thematic analysis to EAP classroom activities progresses.

In terms of adopted methods of research, one major problem in an applied study such as this is the nature of the core investigative tool of discourse analysis itself. It may be thought that there is in fact no *method* to discourse analysis, but rather a 'broad theoretical framework concerning the nature of discourse and its role in social life, along with a set of suggestions about how discourse can best be studied and how others can be convinced findings are genuine' (Potter & Wetherell 1987:175). Halliday (1985b) comments on the essentially interpretative nature of text analysis which presents fundamental difficulties for the text analyst, since there

are relatively few absolute and clearcut categories in language, and there are many tendencies, continuities and overlaps. Berry also warns that descriptive frameworks in the fields of textlinguistics and discourse analysis are notoriously problematic and thus 'before we can engage in this type of work it will be necessary to sharpen the tools that we propose to apply' (1989a:77). In emphasis of the application of pedagogic insights gained from text analysis, Davies comments that a fundamental assumption,

'is that curricula design needs to be underpinned by the development of methods of analysis which do justice to the full range of texts available for study and which have practical application for the classroom' (1991:1).

As illustrated by Davies's own work across a wide variety of different genres, the approach to linguistic description should be flexible enough to cover all distinctive varieties of a language, whereby, 'the varieties can be compared with each other on reliable criteria' (Sinclair 1985a:18). Hence, there is the prior need to establish clear analytical frameworks and to avoid the vague gloss of indistinct and ambiguous labels (Berry 1989b). Equally clear is the requirement to adequately explain methodological approaches taken in such research when applied to the classroom, since insights arising from this work must be of benefit to fellow researchers, EAP classroom practitioners and NNS students alike. There is little interest here in the discussion of theoretical issues for their own sake; rather, in Berry's (1987b, 1989b) terms of facts in search of explanations and *vice versa*, problems have been raised by interesting linguistic data and systematic solutions are then being sought for practical purposes. The central interest lies in the optimal transferability of insights gained from theoretical discussion to the applied situation in the development of effective writing curricula for the EAP classroom.

Since the present work is a multi-layered study, it is required in turn to draw upon multiple methods of study. It is more appropriate to discuss in detail the rationale for the adopted methodologies and the respective procedures undertaken at the relevant stages throughout this thesis; however, the remainder of this section now presents a brief overall summary of the methodologies selected in this study and these are discussed in relation to initial general hypotheses noted above (section 1.6) under the following headings:

(a) sources of data collection;

(b) the criteria for the selection of a main corpus of 36 published scientific RAs written by NSs of English and of a supplementary corpus of 36 published NNS RAs;

(c) the establishment of a theoretical framework and analytic procedures for the investigation of Theme and thematic structure in the (b) corpora;

(d) a synthesis from previous research of genre-specific rhetorical *move* patterns [that is, using Swales's (1981, 1990a) terminology – systemicists' *elements of structure*] in scientific RA sections;

(e) the mapping of dominant Themes onto the move structure of discrete RA sections following (c) and (d);

(f) background surveys of NS and NNS participants in international science research communication with a view to investigating (i) the influence of thematic structure in judging the merits of RAs submitted by NNSs (ii) wider L2 research writing procedures, processes and strategies;

(g) product- and process-oriented analysis of a corpus of NNS novice researchers' FIRST and FINAL RA drafts to be submitted to international academic journals;

(h) the establishment of Propositional Clusters as a research and teaching tool to explore thematic manipulation in NNS RA drafts.

These stages will now be discussed briefly to give a general overview of the progression of the research project as it is reported here, with reference to later sections which report in specific detail. It is natural that the thesis has been organised so as to present the most coherent ordering of research components; however, this may quite naturally not reflect the concurrent nature of some of the research activities.

(a) <u>sources of data collection</u>: Grotjahn (1987) comments that the selection of a specific data collection method does not necessarily determine the choice of analytical procedure and *vice versa*. The present research is driven by both (i) product-oriented, quantitative-statistical treatment involving the calculation of means, frequency distributions, standard deviations and correlation coefficients (Pearson product-moment r and Spearman rank r_s); (ii) process-oriented, qualitative-interpretative treatment. A description of the sources of data collection may be organised according to various paradigms e.g. whether they are naturally occurring or elicited. However, since the product-process orientation is an important one in this research, this distinction has been selected here to describe data collected from the following sources:

(i) PRODUCT-oriented:
- published NS and NNS research articles (RAs) from international English-language scientific journals;
- unpublished NNS drafts of RAs to be submitted to international English-language scientific journals;
- drafts of NNS RA writing elicted in the EAP classroom;
- Theme concordances from unpublished NNS RA drafts;
- surveys of NS and NNS participants in the research writing process by means of written questionnaires;

(ii) PROCESS-oriented:
- fieldwork as a participant-observer of novice NNSs' RA writing and redrafting processes;

-       transcribed protocols from an oral interview with NNS
        novice RA writers;

-       analysis of successive drafts of NNS novices' RA
        writing tasks elicited in the EAP classroom.

(b) <u>corpus selection</u>: The method of study adopted here is
primarily corpus-based and thereby principally
data-driven. The initial analysis and description of
thematic choices are based on a corpus of 36 refereed,
published research articles in the physical and life
sciences which appeared in U.K. and North American
professional-academic journals, ostensibly written by
English native speakers (NSs). The aim of this initial
analysis was to establish a 'norm' base of thematic usage
as manifested by 'experienced' RA writers (i.e. in the
present context, published = experienced); moreover, this
'norm' has been validated by the 'expert' editorial
gatekeepers of the discourse community. (Full details of
the corpus and selection criteria are given in section
3.4.1 and APPENDIX A).

(c) <u>theoretical framework</u>: following Berry (1987a) and
Davies (1988a), the method of thematic analysis adopted
here identifies the grammatical Subject as an obligatory
component of Theme, that is, as unmarked Theme;
consequently, any optional elements appearing before
Subject were taken as marked Themes and labelled Context
Frames (section 3.2 fully outlines and exemplifies the
rationale for this method of thematic analysis and section
3.4.2 highlights particular criteria for analysis). The
analysis of sentence-initial elements (N=4358), led to the
primary classification of Themes according to nine
suggested discourse-functional categories of marked Theme
(section 3.5) and four major domains of unmarked Theme
(section 3.6). In the initial stages of analysis, some of
these discourse-functional categories were combined,
separated, refined or abandoned, according to empirical
evidence as data collection and analysis progressed.

(d) <u>rhetorical move patterns</u>: as well as attention to the lexicogrammatical system of Theme in RAs, the study focuses on common patterns of macrostructural, rhetorical *moves* in each of the four main RA sections in order to correlate these two levels of linguistic description. These macrostructural elements of discourse structure represent a synthesis of findings from previous research into scientific RA move structure (section 3.10.2).

(e) <u>Theme-Move maps</u>: by combining elements of (c) and (d) above, we aim to describe at some degree of delicacy the range of selections available within the system of Theme across the IMRD macrostructure. For example, thematic choices in the RA Introduction section can be identified and mapped onto the commonly-established rhetorical *move* structure of that particular section. In turn, a complete pattern of thematic distribution across the dynamic range of the entire RA can be established as a genre-specific characterisation of these language features (section 3.10.1 and Figures 11-14). This pattern of within-text structuring would again function as a 'norm' against which to compare other corpora of published NNS RAs (section 5.1) and NNS RA drafts (section 5.2.3).

(f) <u>background surveys</u>: in addition to the quasi-deductive approach of published corpus analysis, the study employs progressively more inductive/heuristic approaches by means of naturalistic, elicited data. As background to the investigation of a group of sixteen NNS novices' research writing procedures and strategies, but still with a particular interest in Theme and textual development in mind, a series of both oral and written surveys was conducted with participants in the scientific research publication process: journal editors (NSs section 4.2; NNSs section 4.3) and NNS research supervisors and students (section 4.4). In addition to the structured or semi-structured written responses of mail-shot and classroom-conducted surveys, one particular questionnaire

utilises verbal reporting [Ericsson & Simon (1985, 1987); Faerch & Kasper (1987)], where informants are asked to 'think-aloud' about their perceptions of their own individual RA drafting procedures and L2 writing strategies (section 5.2.2.2). Collated findings from transcriptions aim to contribute to a closer understanding of developmental phenomena in the L2 research writing process. By incorporating such data, we follow the more flexibly data-driven approach of analytic induction which allows for the emergence and development of hypotheses as the classroom-based observational study progresses. This implies that hypotheses are suggested and addressed as a result of continuing data collection and as emerging patterns in the data become evident. Such an approach reflects the natural interplay between observing phenomena and theorising on their underlying concepts.

(g) NNS novices' RA drafts: the applied part of thematic analysis in the study focuses on the same group of as yet unpublished, NNS 'novice' RA writers (that is, in this context, unpublished = inexperienced = 'novice') who had all taken a course in Academic Writing taught by myself at Tokyo Institute of Technology. Further to published RA corpora, here additional product-oriented data collected include FIRST and FINAL drafts of novices' first RAs to be published in English (section 5.2.3). However, this corpus represents a primary source of process-oriented data which are generated from a participant-observer viewpoint by tracing textual modifications in these same RA drafts in response to external feedback and criticism (section 5.3). This group of novice writers is required to make such modifications to their RA drafts in order to have them accepted for publication by the gatekeeping, editorial 'experts' of the academic community. By means of both textual and statistical comparison with the base norm NS corpus of (b) above, aspects of thematic manipulation and any correlation with 'success' in RA writing (as measured by acceptance for publication) can be explored.

(h) <u>Propositional Clusters</u>: the final methodological
approaches employed in this multi-method account are the
exploratory research and teaching tools of Propositional
Clusters (PCs) and reformulation (Chapter 6). Following
previous stages of research outlined above, PCs were
developed as a heuristic in the EAP classroom in an
attempt to raise NNS novice RA writers' awareness about
the potential of thematic manipulation to create more
'successful', cohesive and coherent RA drafts. Clusters
serve as a basis for text production and they owe their
origins to research on propositional analysis, primarily
by Kintsch (1974, 1977, 1985). Preparation for PCs
exercises and classroom procedures adopted are fully
outlined; in parallel with previous analyses in this work,
both process- and product-oriented data are gathered in
EAP classroom PCs writing activities and compared.

1.8 OUTLINE OF THE THESIS

Given the multi-layered nature of this account, each main
chapter is prefaced by a brief section entitled
'background to this chapter' which contextualises the
overall report to follow. Similarly, 'concluding remarks'
offer a summary of chapter contents and point the way
forward to the following layer of research. The
theoretical framework of a genre-based approach to the
analysis of Theme and thematic distribution (Part I), the
background surveys of the various participants in the
process of international research communication and
relevant textual analysis (Part II), and the applied
pedagogical focus on L2 RA research writing activities and
teaching/research tools to investigate RA writing
processes and products (Part III), naturally suggest a
three-part presentation for the main components of this
research as follows:

Chapter Two gives an evaluation of current approaches to genre-based studies in linguistics and language education and, thereby, the rationale for the underlying approach in this study. It pays particular attention to previous research into genre and register in the two major traditions drawn upon in the present work, namely, studies in the applied fields of English for Specific/Academic Purposes and theoretical insights gained from the school of systemic-functional linguistics. Of particular interest are descriptions of how written genres are thought to be realised on both micro- and macro-levels and how such realisations are configured to create the recognised distinctiveness of a given written genre. Chapter Three introduces the theoretical focus-on-genre/product of the study itself, namely the discourse-functional investigation of Theme in scientific RAs and thematic distribution across components of this genre. The rationale for this focus on Halliday's textual metafunction of language is first put forward and followed by a discussion of current issues in research on Theme which reviews and synthesises previous work conducted within a broad systemic-functional framework. Following a detailed description of the corpus and methods on which analysis is based (section 3.4), sections 3.5 and 3.6 present the detailed analysis and categorisation of marked and unmarked Themes; in turn, sections 3.7, 3.8 and 3.9 review the major findings from analysis of the 36 RAs by means of interpretive commentary and statistical analysis. To complete this theoretical focus and to highlight the relevance of the textual metafunction for the realisation of rhetorical structure, elements of research reported in Chapter Two (generic move structure) and Chapter Three (thematic distribution) are mapped onto each other to suggest Theme-Move maps (section 3.10).

PART II: Application - Background surveys and analysis

Chapter Four introduces Part II of the study with a focus
on a series of questionnaires conducted with the various
groups of participants involved in the process of
international research communication. The first
questionnaire surveyed a group of NS scientific journal
editors (N=116) and was designed in part to investigate
the influence of thematic structure on the judgements of
the merits of RAs written and submitted by NNS
researchers. The results of the same questionnaire with a
group of NNS editors (N=45) are also compared. A modified
version of this questionnaire was used with a group of NNS
research supervisors (N=38) and their doctoral students, a
group of NNS novice researchers (N=40), to investigate
more broadly NNS RA writing processes, procedures and
strategies as background to EAP classroom-based research
activities. Following the presentation of data in
preceding chapters on the usage of Theme in a corpus of NS
RAs and background surveys on L2 RA writing processes,
Chapter Five follows with a comparison of data generated
from textual analysis of a parallel corpus of 36 published
NNS RA products (section 5.1) and a mini-corpus (N=10) of
unpublished FIRST and FINAL drafts written by NNS novice
researchers (section 5.2). As background data, one section
here (5.2.2) investigates by means of verbal reports
aspects of the L2 RA writing processes of a core group of
sixteen NNS novice researchers. With an increasing
interest in the RA redrafting procedures and strategies
which may contribute to 'success' in RA writing, section
5.3 employs a more qualitative approach to the
investigation of changes in thematic selections and
patterns of thematic progression between these FIRST and
FINAL unpublished drafts. In this way, we trace the
dynamic textual development of a scientific paper from its
state as a relatively immature, unpublishable and hence
'unsuccessful' RA draft to one which is a recognisably
more mature and 'successful' piece of research reporting,

now acceptable to the expert members of the international discourse community.

PART III: Application - focus on genre/process
In Chapter Six, the third part of this research project attempts to apply much of the work previously reported in Parts I and II, with the research setting having moved more clearly into the research laboratory and the EAP classroom, and with the adopted research roles of EAP teacher-practitioner and participant-observer. As a result of the accumulation of both theoretical and applied insights above, this chapter explores the use of the teaching/research tool of Propositional Clusters as a heuristic for raising NNS novices' awareness about the manipulation of Theme in drafting and redrafting RA sections. Finally, Chapter Seven presents a summary of the present research, in the light of initial general research questions and hypotheses, and concludes with suggestions for future work in related areas.

PART I

DESCRIPTION AND ANALYSIS - FOCUS ON GENRE/PRODUCT

CHAPTER 2 APPROACHES TO GENRE IN LINGUISTICS AND LANGUAGE EDUCATION

2.0 BACKGROUND TO THIS CHAPTER

As indicated by the title and the earlier introductory discussion (section 1.1), this work presents a *genre-based* investigation of Theme in scientific research articles. We start Chapter Two therefore with an evaluation of current approaches to genre analysis in the relevant fields of linguistics and language education. It has been noted that, due to underlying methodological and pedagogical considerations, this study seeks to combine the traditional research interests of the field of English for Academic Purposes (EAP) with the analytic potential of a systemic-functional theoretical framework. This chapter aims therefore to more clearly contextualise the present genre-based study within this research framework.

Section 2.1 first presents an introduction from a broader research perspective to the notion of genre and positions taken on definitions of genre. Section 2.2 focuses more particularly on the positions taken by those in the systemic-functional school on the interrelated concepts of genre and register and their place in relation to Halliday's metafunctional organisation of language, and thereby, to the 'enabling' textual function of Theme (section 1.2). Section 2.3 focuses on a genre- and register-specific description of the scientific research article (RA) which will be employed in the present multi-layered analysis; the final section (2.4) rounds off with brief concluding remarks on the discussion presented in Chapter Two and indicates its relevance for the corpus analysis to follow.

2.1 INTRODUCTION

'Genre is one of the most contentious topics in curriculum
theory today, and important practical issues are at stake. This
is especially so for language learning, and for writing
development in particular, from the earliest years of schooling
right through to post-secondary education' (Reid 1987:1).

Many of the current debates about genre in the fields of
linguistics and language education are associated with
distinctions between overlapping strands of educational
thought and are influenced by their respective student
populations. Most notably this relates to the Writing
Across the Curriculum (WAC) and English for Academic
Purposes (EAP) movements with their predominant concern
for secondary and post-secondary populations of NSs and
NNSs, respectively, predominantly in the U.S., Britain and
elsewhere. In language education in Australia, the
systemic-functional school of linguistics pioneered by
Michael Halliday and others has been particularly
influential since, as Swales (1992) notes, genre provides
much of the curricular scaffold from elementary through to
high school. Green (1987) thus suggests that in the
impassioned conflict with the process approach (section
1.3), 'genre consciousness' has in recent years gained a
strong position as regards the agenda for literacy
education in schools and teacher education programmes.

As indicated by Swales's (1990a) and Ventola's (1989)
overviews, any focus on genre will quite naturally
incorporate substantial insights from those interested in
genre in a wide range of research fields such as
cognitive psychology, rhetoric, anthropology, ethnography
and, of course, various branches of linguistics. It is a
natural consequence of approaching the study of genres
from this broad range of research perspectives that each
tradition would use its own analytic approaches for
modelling genre and this of course leads to a variety of

models and positions as regards definitions of genre and generic structure. For example, in the field of linguistics and language education, Davies suggests that,

> 'implicit or explicit in all discussions of genre is the assumption that the notion of genre is one which is universal and is utilised in everyday life by the lay person, and that this recognition of generic differences across both spoken and written texts is manifest in the labels which are given to different types of communication' (1991:2).

Whether or not the notion of genre is universal across interested fields of study, Davies suggests that definitions have come to share the assumption that a genre is strongly identified by the specific communicative 'job' it is doing in a specific social context. Swales (1986b) comments that an exemplar of a particular genre is often identifiable through reference to its source: he mentions newscasts, recipes, research papers and identifies written sources such as newspapers, health and safety regulations and text-books; Halliday (1989a) lists 36 varieties of written English including a novel, a play, a road sign, a graffito and a telephone directory; Martin *et al.* (1987) and Kress (1989) include jokes, letters to the editor, anecdotes, essays and interviews; Brown & Yule (1983) list fairy stories, salesmen's routines and chats; Miller (1984) lists ransom notes, user manuals and the eulogy; Couture (1986), Martin & Rothery (1986) and Christie (1986) include sermons and service encounters, as well as what they describe as a wider sense of genre with explanations, narratives and exposition, thus blurring the distinction between reference to text-source and text-type.

With applied purposes in mind, Swales's working definition of genre is widely adopted by teacher-practitioners in the fields of English for Specific/Academic Purposes: genre is 'a recognised communicative event with a shared public

purpose and aims mutually understood by the participants within that event' (1986b:13). More recently, Swales defines genre in more ecological terms as comprising,

'a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognised by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains the choice of content and style' (1990a:58).

In focusing on implicitly shared communicative purposes, Miller (1984) observes that what we learn when we learn a genre is not just a pattern of forms or a method of achieving our own ends, we learn, more importantly, what ends we may have. Researchers thus emphasise that appreciation of this dynamic process of defining goals or ends is important to a rhetorically sound definition of genre which must be centred, not on substance or form of discourse, but on the shared communicative purposes and goals of the discourse community and the action it is used to accomplish. These are important insights and intuitions about genre and this underlying social perspective is quite naturally shared by many working in a systemic-functional tradition. For example, Christie comments that learning a genre implies the tendency to,

'structure situations towards the satisfactory achievement of socially defined goals or ends. It means we structure and organise our behavioural patterns, including most crucially our linguistic patterns' (1987:28).

In focusing more closely on the linguistic patterns of genres from a systemic-functional perspective, the question arises as to how the notion of genre and generic elements of structure and linguistic patterns systematically relate to the social semiotic model of

language use described by Halliday and others (section 1.2) whose theoretical framework underpins the present study. The notion of genre has long been a part of academic tradition (Ventola 1989); what is new is,

> 'the attempt to characterise genre in terms which are clearly definable as a set of formal properties generally applicable across a large number, if not all, genres' (Harris 1987:36).

This question will now be taken up in the next section in a review of relevant systemic-functional views of genre theory and generic structure within a social semiotic model of language.

2.2 A SYSTEMIC-FUNCTIONAL VIEW OF GENRE

Martin *et al.* (1987) state that current thinking on genre underlying systemic-functional approaches was developed by Hasan (1978) Kress (1982), Martin (1985) and others as an extension of earlier work on register by systemic linguists including Halliday and Gregory (1967). However, genre theory differs from register theory in that it particularly stresses social purpose as a determining variable in language use (Martin *et al.* 1987). Genre is seen as a *staged, goal oriented social process* - a *social process* because members of a culture interact with each other to achieve genres, *goal oriented* because they have evolved to get things done, and *staged* because it usually takes more than one step for participants to achieve their goals. As with Swales's (1990a, 1992) recognition that definitions of discourse communities, and thereby genres, need to clearly reflect evolutionary processes, Martin *et al.* emphasise the way that genres as evolved systems introduce both stability and flexibility into a culture at one and the same time.

So far then, as in previous discussion, the stress on
social purpose is fundamental to a functional definition
of genre. However, compared to other broader perspectives
on genre, the systemic-functional school offers a
theoretical basis for the integration of genre into an
overall model of language use. Martin (1992a, 1993)
reviews the place of genre within Halliday's (1978) model
of language as a social semiotic system: language plays an
instrumental role in construing the broader social
contexts of culture and situation [terms taken from
Malinowski (1923)] in which we live - at the same time,
language is construed by these social contexts and thus
'the relationship is symbiotic - one of mutual
engendering' (Martin 1993:142). The social context can be
characterised as a system of genres by its construal in
linguistic terms as a semiotic system contextualising
language, that is, as one semiotic system expressed in
terms of another semiotic system. Figure 1 below
represents Martin's (1992a, 1993) view of language as the
realisation of social context. Of fundamental importance
then is how the 'higher' order systems and structures of
the social context of genre are systematically related to
and realised by 'lower' order linguistic systems and
structures. These interrelated planes will now be
addressed proceeding top-down with genre and register.

Ventola (1989) suggests that the nature of genre system
networks is still obscure, as is the nature of other
proposed system networks, notably, the metafunctional
semantic networks. However, subcategorisations of related
genres, for example, the science research process genres
of RAs, Letters, Notes, Communications, Abstracts etc.,
may be seen as initial system choices at the level of
genre. On the other hand, the nature of generic elements
of structure has been studied from a broad perspective and
relevant research positions on these will be reviewed
shortly. Firstly, however, in order to be able to
contextualise generic elements of structure, let us deal

Figure 2. Modeling tenor, field and mode as the realisation of genre (Martin 1993)

Figure I. Language as the realisation of social context (Martin 1993)

briefly with the plane of register which in Halliday's (1978) view is said to realise genre.

Register is defined as a configuration of meanings that are typically associated with a particular situational configuration of the three variables FIELD, TENOR and MODE. Halliday (1978, 1985a) describes this relationship between the text and the wider social contexts of situation and culture as follows:

FIELD is the social action, 'what is actually taking place' in the situation; it is realised by the *ideational* metafunction - the management of ideas;

TENOR is the role structure, 'who is taking part' in the situation; it is realised by the *interpersonal* metafunction - the management of personal relations;

MODE is the symbolic organisation, the 'role assigned to the language' in the situation; it is realised by the *textual* metafunction - the management of discourse.

Register variables are systematically associated with the 'lower' order linguistic system through the metafunctional components (ideational, interpersonal, textual) of the semantic stratum. Thereby, the plane of register correlates linguistic and recurrent cultural and situational features realised in the 'higher' order system of genre. Of recent research interest [see Martin (1992a) for review] are questions as to how the plane of genre itself is systematically related to register variables. Martin proposes that 'it is useful not to associate genre too closely with any one register variable (e.g mode in Halliday's work or field in Hasan's)' (1992a:506); thus, genre is formulated as a pattern of register variables which is itself not metafunctionally organised. Expanding on Figure 1, Figure 2 above represents Martin's (1992a, 1993) holistic view of these relationships in a

stratified model of metafunctions and register variables as the realisation of genre. Let us now return to descriptions of generic elements of structure and relate these to aforementioned register variables; this will lead us on to consideration of the specific schematic structure of the scientific research article (RA) which is of central interest here.

Hasan's (1978, 1989) definition of genre is particularly concerned with a more rigorous functional description of the crucial relationship between the social contexts of situation and culture and text structure. Hasan, like Brown & Yule (1983), emphasises the predictability of relevant features of context within particular types of communicative event. Hasan aims to show how features of the social context can be used to predict some elements of the structure of appropriate texts; 'all else being equal, the presence of those elements of the text's structure would "construct" those same features of the context' (1989:55). In order to be able to identify any given text as an instantiation of a given genre, Hasan refers to the configurations of very specific values of register variables Field, Tenor and Mode which indicate the resulting *contextual configuration* (CC) [for example, see Hasan (1989:59) for the CC of a service encounter].

Hasan stresses that the notion of CC is crucial since it is the specific features of a CC that permit statements about the text's structure. Specific contextual configurations are the primary criterion for genre-membership and they permit predictions about both obligatory and optional elements of a text's structure and sequence which will make up the CCs and the genre. The obligatory elements define the genre to which a text belongs 'and the appearance of all these elements in a specific order corresponds to our perception of whether the text is complete or incomplete' (1989:62). The total range of these elements and their order represent the

structure potential of this genre, what Hasan refers to as
the generic structure potential (GSP):

> 'genre bears a logical relation to CC, being its verbal
> expression. ..for some given texts to belong to one specific
> genre, their structure should be some possible realisation of a
> given GSP'(1989:108).

Hasan's correlation of CC values and generic structure is
clearly a useful basis for the purposes of analysis of
discourse [for example, Harris (1987)]. However, there
are, quite naturally, competing opinions from those above
amongst researchers interested in discourse and linguistic
accounts of genre and some of these will be briefly
mentioned below. Such rival theories naturally provide a
strong dynamic research base for ongoing investigation
and, as Berry (1989b) so strongly emphasises, the need for
the development of a systematic framework of analysis
appropriate for studies in language education.

Perhaps the most radical departure from above positions
concerns the very rationale for generic elements of
structure. In discussing story structures, Morgan &
Sellner stress the importance of a writer's goals but
argue that insights about events and their relations
should follow naturally from an understanding of what
stories are for, '*not* from any kind of internalized
schema' (1980:189), thus challenging the suggestion that
social purposes are actually encoded in schematic elements
of structure. Of particular relevance for the present
research context is that studies in the sociology of
sciences [see Latour & Woolgar (1979); Knorr-Cetina
(1981); Gilbert & Mulkay (1984)] suggest that Morgan &
Sellner's view of text-as-observed-reality is certainly
likely to underplay the social processes at work in the
genre of the scientific RA and thereby the socio-
rhetorical nature of research writing.

Butler (1985, 1989) and Ventola (1989) point out that there are also more specific differences amongst systemicists regarding the interpretation of register variables Field, Tenor and Mode [see Gregory (1967); Gregory & Carroll (1978); Berry (1989b)], principally due to lack of clear evidence that each register variable may in fact determine only one type of meaning (that is, that Field determines only ideational meaning, that Tenor determines only interpersonal meaning, that Mode determines only textual meaning), as suggested in Figure 2. For example, Berry (1989b) shows that, rather than one register variable determining only one metafunctional component, one type of meaning [namely, the Identification of participants when making a medical appointment] is jointly controlled by all three variables. Importantly, Ventola comments that register variables 'do not remain constant throughout elements, but rather that all or some values are renegotiated' (1989:141). Stressing the assumption that genre as a *staged, goal oriented social process* (Martin *et al.* 1987) controls choices on the register plane, Ventola points out that the register variables of Field, Tenor and Mode 'vary as the social process - the genre choices - unfolds as a structure' (1989:71). Thus, Ventola feels that Hasan's GSP and CC impose a static linearity which does not accurately reflect the dynamic stage by stage unfolding of certain discourses. Consequently, a more flexible stratified representation of Martin's (1992a, 1993) model of register as the realisation of genre (Figure 2), one which is an elaboration of Ventola (1988), is suggested in Figure 3.

From this systemic stratified model, and previous discussion (section 1.2) on Halliday's textual metafunction, it can be seen that the central interest of the present study lies in a detailed correlation of the 'higher' order generic elements of structure (far left in Figure 3) of the scientific research article with the 'lower' order lexicogrammatical structure of Theme (far
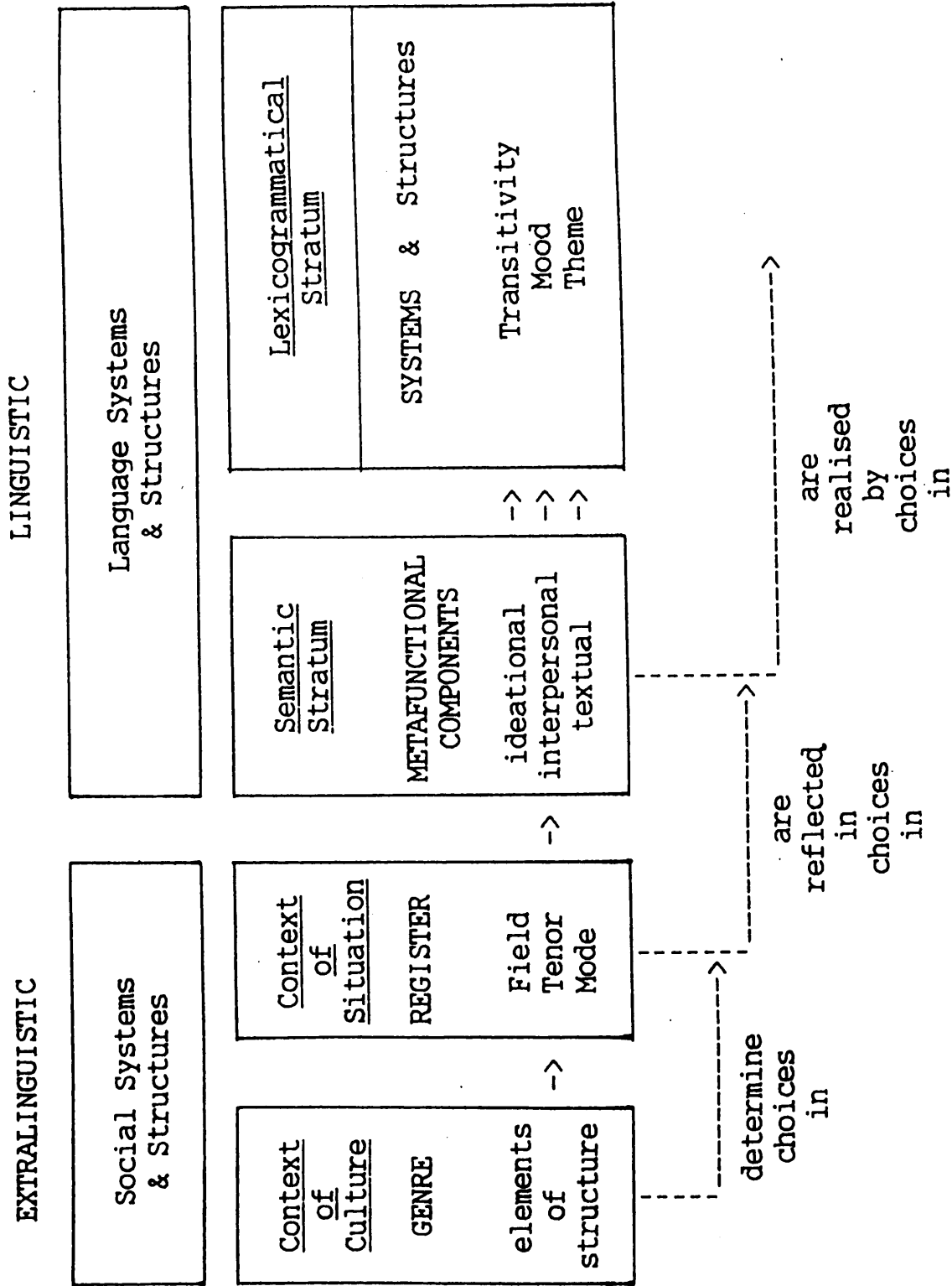
Figure 3. Systemic stratified view of the extralinguistic contexts of genre, register, and the linguistic strata of the semantics and the lexicogrammar [after Ventola (1988) and Martin (1993)]

right) through the more abstract filters of register and the metafunctional organisation of language. Although progressively staged elements of generic structure are expected to determine varying values of register variables, Figure 3 above indicates that Field, Tenor and Mode choices will tend to be reflected in ideational, interpersonal and textual components and the major clusters of the systems and structures of transitivity, mood and theme, respectively [the phonological stratum is excluded here as our interest is only written discourse]. Considering above comments, however, the option is open that for scientific RA writers to stage the 'appropriate' balance of *topic-based* and *interactional* discourse (section 3.1), the registral variable of Tenor may be seen to reflect this changing balance in the textual metafunction (as realised in Theme choices), as well as in other interpersonal realisations.

The next section now considers the above general discussion on the extralinguistic contexts of culture and situation in relation to the specific register of scientific English and the specific cultural context of the scientific research article (RA) genre which is the textual focus of interest here.

2.3 THE CONTEXTS OF CULTURE (GENRE) AND SITUATION (REGISTER) OF THE SCIENTIFIC RESEARCH ARTICLE

Of somewhat controversial interest regarding Hasan's (1989) approach to the analysis of text structure is her statement regarding the metalanguage for generic descriptions that 'irrespective of the differences between genres, the considerations relevant to a description of text structure will be the same at a certain degree of abstraction' (1984:74). Harris (1987) suggests that the

case for the applicability of a general metalanguage to all genres is inconclusive, possibly since Hasan deals with very 'simple' interactive situations (purchasing groceries, telephoning for a medical appointment) which allow for a succinct statement of the Contextual Configuration in terms of Field, Tenor and Mode. Detailed statements of CCs for quite different complex situations may, Harris comments, be another matter.

The contexts of culture (genre) and situation (register) from which derives the scientific research article (RA) appear very different and complex when compared to those spoken exemplars exemplified by Hasan. On the one hand, scientific RAs commonly give rise to formally delineated and labelled elements of generic structure, Introduction, Methods, Results, Discussion (IMRD) which internally structure RA textual dynamics. On the other hand, these elements of structure, like Hasan's CC, specify conditions for beginning, continuing and ending a text. These Aristotelian beginnings, middles and ends emphasised by Couture (1986) as generic properties highlight the attention given to the discreet stages of the purposeful, social processes of genres.

Clearly, systemicists' labels of *elements of structure* or *schematic structures* have much in common with the *rhetorical moves* arising from the applied discourse analysis of text structure in the tradition of English for Academic Purposes (EAP). The latter term, used most notably by Swales (1981), has been widely adopted by others in the fields of EAP due to the relative accessibility of analysis, description and pedagogical application of move structure in certain genres. As with Hasan's GSP and CC, sequences of obligatory and optional communicative moves may identify particular genres by means of the rhetorical goals which drive the various stages of discourse and their signalled move transitions as the discourse proceeds. This is most clearly

exemplified in Swales's (1981) seminal analysis of research article Introductions which has inspired a great deal of EAP-related research into genre and genre-based curricula.

In combining the traditions of EAP and systemic analysis, we can now indicate the contexts of culture and situation for the scientific RA. First, the register we are dealing with here. Halliday (1988:162) comments as follows on the Field, Tenor and Mode of scientific English [with values of less relevance for the written RA in square brackets]:

FIELD: 'extending, transmitting or exploring knowledge in the physical, biological [or social] sciences';

TENOR: 'addressed to specialists, to learners [or to laymen], from within the same group (e.g. specialist to specialist) [or across groups (e.g. lecturer to students)]';

MODE: '[phonic or] graphic channel, most congruent (e.g. formal "written language" with graphic channel) [or less so (e.g. formal with phonic channel)] and with variation in rhetorical function - expository, [hortatory], polemic, [imaginative] and so on'.

This description evidently captures general registral values of scientific research articles. However, as suggested above by Ventola (1989), registral values may not be linearly static but rather dynamically determined by the changing rhetorical demands of the 'higher' order elements of generic structure. [In Figure 3, this more flexible relationship between elements of structure and register values is marked by the absence of specifically corresponding horizontal arrows (cf. the relationship between components of the semantic and lexicogrammatical strata)].

Let us now turn to the plane of genre and indicate the typical schematic structure of the scientific RA as traditionally described by researchers working in the field of English for Academic Purposes [for example, Swales (1981, 1990a), Weissberg & Buker (1990) - see section 3.10]. The far left-hand side of Figure 4 below suggests the scientific written genre system network with research process genre subcategorisations of RAs, Abstracts, Notes etc. All sub-genres are realised through specific elements of structure (from now on, referred to as *moves*); for example, the RA has the surface elements of IMRD which, like Hasan's (1989) CC, contain a number of distinct obligatory and [optional] moves.

The basis of the above move patterns in the IMRD sections of the RA will be indicated and then drawn upon at a later stage (section 3.10) in order to map onto them, at some degree of delicacy, the potential range of marked and unmarked thematic choices available within this system of the lexicogrammar.

2.4 CONCLUDING REMARKS

Chapter Two has outlined the rationale for a systemic-functional view of genre theory within the general framework of a social-semiotic model of language use and has indicated the specific elements of the context of culture which realises the RA genre and which it is suggested ultimately determine available choices within the 'lower' order system of Theme across IMRD sections. The present work shares the views of genres and their instantiation in texts as *staged, goal oriented social processes* (Martin *et al.* 1987) and aims to build upon the analysis, description and pedagogical application of genre-based research currently in progress across many disciplines:

—Establishing a Territory
—Establishing a Niche
—Occupying the Niche

—Materials & Procedures

—[Statement of Location of Results]
—Statement of Important Findings
—Comment on Results

—Statement of Most Important Findings
—Comparison with Previous Research
—Explanations & Speculations for Results
—[Limitations of Present Research]
—[Recommendations for Future Research]

I   M   R   D

moves

elements of structure

schematic structure

ABSTRACT   RA   NOTE   LETTER

Figure 4. SYSTEM and structure in the context of culture of the scientific written genre - rhetorical moves in the four sections (IMRD) of the research article [after Swales (1990a) and Weissberg & Buker (1990)]

'Needless to say, genre theory depends on a careful analysis of the textual features of genres. Any generic categories.. need to be substantiated by looking in detail at.. a large range of agnate texts' (Martin *et al.* 1987:62).

Consequently, by means of a corpus-based analysis of published scientific research articles, the next chapter investigates the dynamic range and distribution of thematic selections available across different stages of RA discourse and their relation to the generic moves outlined here. Once established, these genre-specific Theme-Move patterns will allow for comparative analysis and description of pre-generic writing in particular pedagogic settings, for example, of novice NNS researchers writing their first academic papers in English, and how their written RA drafts approach the target rhetorical and linguistic norms of the culture/genre as defined by the experts of the discourse community.

CHAPTER 3   THEME IN SCIENTIFIC RESEARCH ARTICLES

3.0 BACKGROUND TO THIS CHAPTER

As previously indicated in the introductory section 1.2, Chapter Three presents the description and results of analysis of Theme and thematic distribution in a corpus of published scientific research articles. The initial question "why look at Theme?" (section 3.1) will be re-addressed in relation to Halliday's functional theory of language introduced earlier (section 1.2) and, in particular, the tripartite organisation of language metafunctions, and set against a background of research conducted in the traditions of systemic-functional linguistics and English for Academic Purposes. Current issues in research on Theme (section 3.2) focus on problems with the identification of thematic structure and resulting theoretical positions which may be adopted; the rationale for both the formal and the discourse-functional approaches to Theme analysis taken here is subsequently explained. Initial research hypotheses relevant to this Chapter (section 3.3) are reiterated along with a description of the methods of analysis of Theme (section 3.4). The detailed discourse-functional categorisation and exemplification of marked (section 3.5) and unmarked Themes (section 3.6) from the corpus is reported prior to a summary of major findings (sections 3.7, 3.8, 3.9). The penultimate section 3.10 in this Chapter represents an important synthesis of work reported so far with the mapping of major patterns of thematic distribution (Chapter Three) onto RA rhetorical *moves* - systemicists' *elements of structure* (Chapter Two), thereby, characterising the dynamic within-text structuring of this genre. Concluding remarks (section 3.11) address earlier hypotheses in the light of Chapter Three and point forward to the pedagogical application of this work on Theme in subsequent chapters.

3.1 WHY LOOK AT THEME?

One focus of interest in recent studies of the writing of professional communities within a variety of academic, scientific and business settings is the perception that writing is not only socially embedded but that it is socially constructive:

> 'Writing structures our relations with others and organises our perceptions of the world. By studying texts within their contexts, we study as well the dynamics of context building. In particular, by understanding texts within the professions, we understand how the professions constitute themselves and carry out their work through texts' (Bazerman & Paradis 1991:3).

From many sociological studies of scientific discourse in the academic-professional community [see Myers (1988), Swales (1987a) for review], it appears that one key element for perceived success in written communication is a writer's ability to judge the 'appropriate' balance between more 'human face' *interactional* discourse and more impersonal *topic-based* discourse, the 'dynamics of context building' being determined by the requirements of different written tasks and genres.

Martin (1986) indicates that the thematisation, or foregrounding, of certain types of information is an important resource in achieving this 'appropriate' balance. He notes that topic-based thematisation produces a text which is well oriented to a reader looking for information, requiring the reader to possibly only pay attention to thematic content to get the gist. However, such a text may become impersonal and alienating; a greater degree of interactional thematisation is likely to improve the 'human face' of the text, but the resulting text may have the disadvantage of being less easy to skim read for information. From this distinction, Berry (1989a) surmises that, in certain genres, an 'appropriate' balance

between interactional and topic-based information is vital for 'successful' communication. Moreover, this balance must be one which progressively varies as the discourse proceeds with changing rhetorical goals. It is this progressively varying pattern of textual dynamics which typifies scientific research reporting and, in particular, the genre of the scientific research article (RA) with its formally delineated and labelled sections - commonly Introduction, Methods, Results, Discussion - and their attendant rhetorical goals.

As is evident from relevant studies [see Naerssen & Kaplan (1987), Swales (1990a) and section 1.4 for review], this clearly delineated discourse structure of the RA facilitates analysis of the many lexicogrammatical features which help to realise the character of the distinct RA sections. Thereby, top-down, extralinguistic and bottom-up, linguistic levels of information (see Figure 3 above) may be integrated and mapped onto each other. This approach is exemplified by the influential analysis of RA Introductions by John Swales (1981, 1990a). For instance, the transition between his rhetorical *moves* indicating (i) a review of previous research and (ii) a gap in that previous research, is clearly signalled by *However* in the example (Swales 1981:54) below:

> *Brown's (7) compilation of handwriting research is more recent and thorough. However, it lacks the detail necessary to draw the sort of conclusions aimed at here.*

Likewise, in building on Swales's approach, it may be possible to assign many sentences or propositions in scientific discourse to their respective RA sections, not simply due to global content, but from specific linguistic features which realise that content and signal the rhetorical goals of the distinct RA sections. I have selected four examples from the present corpus of published RAs to illustrate this point:

-I- *Recently, we have provided evidence that calpactin or p36 plays an important role in exocytosis[15].*

-M- *Both ECGS and collagenase were dissolved in medium and filtered through a Millipore GVS 0.22 filter to ensure sterility.*

-R- *As shown in Fig.3a, cycloheximide completely abolished the anoxia-induced increase of PMN adhesion.*

-D- *The most likely possibility is some sort of receptor-activated Ca^{2+} channel [e.g. 24,25].*

From these examples it appears that, as part of a writer's available linguistic resources, the choice of first position in the sentence is significant. Thus the thematising of certain types of information according to the genre is one means by which writers can achieve both local discourse goals, for example, the signalling of rhetorical moves within one RA section and, more globally, the vital interplay of interactional and topic-based Themes throughout RA discourse.

In the first of the four examples above, in the Introduction (-I-), the use of discourse participant *we*, contextualised by the temporal frame *recently*, focuses strongly on the writer's current involvement in and potential contribution to the research community, an intrinsic Public Relations angle of Introductions. In contrast, the Methods (-M-) sentence-initial element contains purely topic-based information, one item of a checklist of procedures. In Results (-R-), the foregrounding of a discourse-internal entity, *Fig.3a*, and the writer's appeal to the evidence therein, creates stronger reader-writer interaction. The modality and explicit hedging of the Theme *the most likely possibility* in Discussion (-D-) further strengthens the 'human face' dimension as the writer moves towards concluding remarks on research outcomes.

Thus it appears that, as well as contributing to a more general structuring and flow of information, these

sentence-initial thematic choices help a writer stage the 'appropriate' flow of social interaction throughout this specific genre. Indeed, a number of recent studies (Fries 1983; Martin 1985, 1986; Eiler 1986; Davies 1988a; Berry to appear; Francis 1989a, 1990; Stainton to appear) suggest that generic distinctiveness is itself encoded in thematic choices. This is not to dismiss the potential contribution of other non-thematic components to generic characterisation; however, with reference to the scientific RA genre and its distinct sections, the primacy of Theme appears to be strongly borne out by these four examples above.

As well as this increasing interest in research on Theme as a generic discriminator, the earlier brief introduction to the "Theme" component in the thesis title (section 1.2) suggested other reasons for a focus on textual development in written texts:

- informal insights and intuitions gained from working with novice NNS researchers in the writing and revision of RA drafts have suggested the importance of 'appropriate' thematic manipulation in the creation of 'successful' RAs;

- in the introductory section 1.2 above, we presented a general description of Halliday's functional organisation of language, with particular attention being paid to the three main metafunctions which typically generate different structures at the lexico-grammatical stratum as their output (Halliday 1979). We recall also that the textual metafunction is seen as instrumental to the other two, since the effective representation of experiential (ideational) meaning and effective interaction (interpersonal) with those around us correlate with the selection of the desired form of the message (textual). Thus, the 'enabling' textual metafunction is entwined with 'extrinsic' interpersonal and the experiential modes of meaning - a full account of the textualness of a text will

tend to be an account of all its functions as viewed through the textual function.

Thus, various strands of Theme have the potential to reveal not only the overall *textual* distinctiveness of a given genre in terms of within-text structuring, but they also have the potential to reveal generic distinctiveness in relation to the dynamic staging of *ideational* ('topic-based') and *interpersonal* ('interactional') thematic elements. The relevance of this latter point will now be clarified with an illustration from Halliday's (1985a) *Introduction to Functional Grammar* (hereafter IFG) of the notion of complex, multiple Themes and their metafunctional organisation (based on IFG:55 Fig.3-14c):

On the other hand	maybe	on a weekday	it would be less crowded
conjunctive Adjunct	modal Adjunct	topical Circumstantial	
textual Theme	interpersonal Theme	ideational Theme	
THEME			RHEME

From the above example, three distinct elements of Theme can be seen to derive from each of the three metafunctions. Thus, Halliday suggests potentially tripartite functional roles for multiple Themes, realised by the three strands of ideational (experiential), interpersonal and textual meaning. The typical unmarked sequence within complex Themes is as in the example above, that is, textual^interpersonal^ideational, [^ = followed by] with the ideational element being the only obligatory component in Theme; according to Halliday, whatever

follows the first ideational element of the clause is automatically part of Rheme.

In terms of *textual* Theme, the above example indicates that the textual element within a complex Theme which 'enables' the formulation of the message may be realised by a conjunctive adjunct - *on the other hand*. We see that the accompanying modality of an 'interactional', *interpersonal* Theme may be realised by a modal adjunct - in this case *maybe*. The experiential sub-component of the ideational is labelled the topical Theme - here Halliday identifies the Circumstantial element *on a weekday*. The topical Theme 'typically corresponds to what the nearby textual environment is concerned with, its subject matter' (McGregor 1990), 'the writer's topic area' (Brown & Yule 1983:141) and 'topic-based' thematisation (Martin 1986), although, as we shall see in the next sections, the identification of topical thematic elements is not in fact straightforward.

Having now indicated the strands of meaning that may individually contribute to thematic structure, we can now more clearly exemplify the potential relevance that ideational ('topic-based') and interpersonal ('interactional') functional components of Theme may have for the dynamic characterisation and staging of scientific discourse.

Recent research across a variety of written genres emphasises the extent to which the 'appropriate' balance of ideational, interpersonal and textual thematic choices appears to strongly correlate with perceived 'success' in texts. For example, Berry (1989a) looked at the correlation between the respective distribution of topic-based and interactional Themes in a small sample of children's writing. Her personal judgements of 'success' indicated the clear relationship between these types of thematisation and the exhibited degree of genre-awareness

and, thereby, the 'appropriate' and 'successful' generic character of the written text. Berry (1990) employed Halliday's traditional tripartite metafunctional categorisation to account for informants' judgements of the relative 'success' of two letters of apology. Gibson's (1992) research also investigated the metafunctional organisation of Themes in a corpus of scientific Abstracts in relation to their perceived 'success', as validated in this case by genuine consumers of Abstracts. With the clear caveat that thematic structure cannot be be expected to account entirely for 'success' in writing, it is simply one major system of choices, Gibson argues in his study that Theme is nevertheless a relatively good predictor of informants' success judgements.

As indicated earlier (section 1.2), the underlying approach taken in these above studies is of course guided by the social-semiotic perspective of Halliday's model of language use which is concerned with the relationships between language and social structure:

'for us, then, the perspective primarily adopted - ..because this is where we look first to seek our explanations for linguistic phenomena - is the social one. ..for the questions we are interested in, especially educational questions, the social dimension seems particularly significant - and it is the one that has been the most neglected in discussions of language in education' (IFG:4).

Thus, the way into understanding about language lies in the study of real texts written for real audiences since analysis from a social-semiotic perspective, 'can uncover the links between the structure and the processes of a text and the structure and processes of the larger social system in which that text participates' (Brandt 1986:93).

The current approach to text analysis is also informed by the work of scholars such as Sinclair and an interest in

'a descriptive system of language that is designed to bring out the underlying similarities of structure in all text and discourse' (1985a:13). Sinclair's focus is on a dynamic model of discourse, an integrated description of language in use which sees all discourse, written or spoken, as 'a continuous movement from one state of affairs or *posture* to another. .. Hence the unfolding or existential quality of discourse description' (1985a:15). Thus, a dynamic, social-semiotic orientation in text analysis aims to highlight the purposeful directionality of discourse.

The present EAP-driven study aims to add to this established functional tradition of work on the textual dynamics of discourse. However, whereas the field of EAP research has prescribed a major focus of interest on areas, such as, discourse communities, skill and strategy studies and the use of expert informants (section 1.1), underlying methodologies in EAP research are often less explicit as regards the seemingly indispensible social-semiotic orientation to linguistic analysis of academic discourse. The opening comments to this section on the perception that writing is not only socially embedded but that it is socially constructive support the need for a methodology derived from a systemic-functional perspective which provides an appropriate framework for an integrated description of language in use.

The functional framework adopted here in relation to the analysis and description of Theme is now outlined. A brief historical overview of research on thematic structure prefaces a discussion of current issues regarding various positions on the identification of thematic position and the rationale for a discourse-functional approach to thematic description and analysis in the scientific research article.

3.2 CURRENT ISSUES IN RESEARCH ON THEME

3.2.1 Introduction

Historically, the current interest in a social interpretation of language draws heavily on the work of major influences in the field, the London School of Malinowski [see, for example (1923)] and Firth (1957), the Copenhagen School of Hjelmslev (1961) and Prague School linguists' work on Functional Sentence Perspective [see Mathesius (1961); Vachek (1972); Fried (1972); Firbas (1974); Daneš (1970, 1974)]. What these approaches have in common is that 'functional principles inform the very outline and details of their model of language' (Davidse 1987:39). However, Michael Halliday, as the principal architect of the systemic-functional school of linguistics, has drawn together a number of bridging concepts such as system network, metafunction and register; thereby, 'Halliday offers the outlines of a theory that relates language, situation and culture systematically' (Davidse 1987:74).

In relation to the present research topic, Halliday drew in particular on Prague School terminology, notably borrowing the functional terms Theme and Rheme (Halliday IFG:38) and Fries (1983) extended Daneš's earlier work on thematic progression. One of Halliday's major contributions in this area is considered to be his separation of thematic structure from information structure, that is, the functions of Theme-Rheme and Given-New, which had previously been combined in Prague School terminology.

However, it is natural that, in continuing earlier scholars' work over some decades, clear differences have emerged as regards systemicists' interpretations and descriptions of language, which itself has resulted in a 'potentially confusing range of meaning' (Brown & Yule 1983:154) associated with the terminology when applied to

analysis in a variety of spoken and written contexts. Researchers may be referring to a variety of different intended meanings when using terms such as 'theme' and 'topic' and hence a basic difficulty in discussing these issues arises. As is clear from a valuable review of problems in research on Theme by Fries & Francis (1992), many of these issues and approaches to investigation of the textual structure of discourse are exploratory in nature and hence the subject of considerable debate. The next sections highlight some of these current issues with relevance to the present study.

3.2.2 Theme as 'point of departure'

With a focus on the clause in English, Halliday described Theme as the element which serves as 'the point of departure of the message (1967:212); it is 'the peg on which the message is hung' (1970a:161) and it is 'that with which the clause is concerned' (IFG:38). Halliday points out that Theme cannot be defined as such and the descriptions above are intended to say how Theme is to be recognised: 'the meaning of Theme is to be found in its manifold realisations.. At best it can be semantically glossed, it can be explained' (Fries & Francis 1992:2).

However, researchers have pointed out certain difficulties with such glosses. Downing indicates that there are often initial elements which 'are not even remotely concerned with what the clause is about' (1991:124). McGregor comments that 'what comes first is necessarily the point of departure of the message, whatever the language. It is hard to imagine a Theme final language in which the starting point of a clause is at the end!' (1990:7). McGregor suggests a description of Theme as an 'anchor point' for the clause, a reference point for tying an utterance down to its context, although it is not clear how this description is any less metaphorical than descriptions such as 'point of departure'. Fries states

that Theme 'correlates with the method of development of a text' (1983:119), and 'relates strongly to the perceived flow of information of that text' (to appear:3), indicating that 'point of departure' must mean something more than simply coming first as the clause-initial element. Further to this, Martin comments that,

'analysis of Theme in clause structure has important repercussions for other levels of discourse organisation. The Theme^Rheme structure proposed by Halliday for the English clause is echoed in larger units in such a way that patterns of Theme selection in the clause tend to be predicted by hyper-Themes [paragraph] which may in turn be predicted by macro-Themes [text]. This resonance across layers of text structure is the basis for a powerful discourse interpretation of clause Theme' (1992b:161).

Thus Martin considers that elements which come first as 'points of departure' are choices motivated by discourse considerations and therefore correlate with other aspects of discourse organisation. The investigation of these particular issues is central to the present work.

3.2.3 The scope of thematic position
With regard to the potential of a 'powerful discourse interpretation' of Theme, the role of thematic position and, in particular, the actual identification of its scope have caused considerable debate [Quirk *et al.* (1985); Eiler (1986); Hudson (1986); Lowe (1987); Huddleston (1988, 1991, 1992); Davies (1988a, 1991); Berry (1989a); McGregor (1990); Matthiessen & Martin (1991); Downing (1991); Gibson (1992); Matthiessen (1992); Martin (1992b); Martin & Matthiessen (1992); Stainton (to appear)]. In short, up to which component does 'point of departure' actually hold and how elastic may its interpretation be? Researchers, many predominantly working within the Hallidayan tradition, may disagree about exactly which

elements realise Theme. Hudson (1986) specifically comments on the difficulties in picking out parts of the clause which define 'what it is going to be about'. In their analyses, Brown & Yule (1983) assume that Theme is a formal category and Huddleston suggests that 'it is not clear that "point of departure" or "starting-point" can sustain an interpretation that is independent of syntactic structure' (1988:158).

The present work does not aim to add further to this controversial, at times somewhat combative debate with delicate analysis of thematic choices; rather, in aiming here for at least internal consistency of identification and discussion, reasons for the approach to Theme adopted are explained in the light of analytical, coding and descriptive difficulties arising from previous work on Theme. These are illustrated below and predominantly focus on the attempt to apply an operationally ostensive description of Theme, and thereby Theme-Rheme boundary.

3.2.4 Markedness

In Hallidayan terms (IFG *passim*), the typical unmarked thematic form in a declarative clause in English is one which conflates three separate and distinct functions: SUBJECT, the grammatical subject, 'that of which something is predicated'; the THEME, the psychological subject, 'that which is the concern of the message'; the ACTOR, the logical subject, 'the doer of the action'. Since Theme is realised as the element which serves as the 'point of departure' of the message, the conflated grammatical/psychological/logical subject functions as unmarked Theme unless a writer has a good reason for choosing something else. The majority of Themes in the declarative statements of scientific written discourse conform to the predominant unmarked pattern of example [1] below and this therefore suggests a powerful discourse role for the grammatical Subject:

[1] *The Curie temperature of the alloy* is about 80K..

Marked Theme choices in declarative statements are where the grammatical Subject (GS) is preceded by a wide range of adjuncts which function as contextualising elements. As Brandt comments, 'when themes do not correspond with the grammatical subject, they may provide the reader with an interpretive framework for the rest of the sentence' (1986:98):

[2] *In this paper*, we describe the Al spectra in detail..

[3] *However*, a complete sign change occurred..

[4] *Since n is nearly constant*, the temperature dependence for the transformation reaction can be modelled..

Thus the semantic focus of the GS in these examples follows from an initial emphasis on: location in discourse space [2]; contrast [3]; cause-reason [4]. As frameworks for the sentence, such marked Themes play an important pivotal role for the organisation and dynamic development of discourse.

Other rarer examples of marked Themes in scientific discourse are complements, or interrogative and imperative clauses:

[5] *Particularly relevant to present investigation* are the XPS observations of Monnier et al.[20].

[6] *Note* that this sudden crossover between Q and P appears only as a subtle effect in all measurements..

[7] *How extensive* is proliferation within the nervous system..?

The conflated Subject/unmarked Theme pattern, the presence of marked contextualising adjunct(s) + Subject at different stages of discourse, and the virtual absence of other non-declarative Themes, typically characterise scientific writing and the RA genre.

3.2.5 Degrees of Theme markedness

As seen in Figure 5 below, within the potential range of marked Themes, varying degrees of markedness are evident in declarative clauses according to the amount of actual choice involved in the ordering of thematic components within the clause and extended to the clause complex.

MORE   Conjunctions

WEAKLY   *But, So, Yet, And*

MARKED

THEME   Conjunctive & Modal Adjuncts

     ┊   *Here* (deictic forms = more naturally thematic)

     ┊   *Finally, Clearly, Interestingly,*

     ┊

     ┊   Circumstantial elements

     ┊   *In this paper...* (deictic forms)

     ┊   *After 3 hours...*

     ┊

     ┊   Subordinate clauses

MORE   *When x=0.92*, we have Al atoms...

STRONGLY

MARKED   Complements

THEME   *Of particular interest* is mechanism...

Figure 5. Degrees of Theme markedness in declarative clauses in scientific research articles

<u>Conjunctions</u>: some cohesive devices such as the small group of coordinating and subordinating conjunctions, for example, *but, so, yet, then, when, if,* are obliged to be clause-initial. They are thus 'inherently thematic' (Halliday IFG:51), and are consequently described by Fries (1983) as weakly thematic.

<u>Conjunctive and modal Adjuncts</u>: (i) conjunctive Adjuncts and (ii) modal Adjuncts comprise a vast range of discourse markers. Conjunctive adjuncts 'are those which relate the clause to the preceding text' (Halliday IFG:49), for example, *however, finally, nevertheless.* Modal Adjuncts 'are those which express the speaker's judgement regarding the relevance of the message' (IFG:49), for example, *perhaps, interestingly, amazingly.* Both types of Adjunct are frequently clause-initial, but are not required to be, and can therefore be seen as more strongly thematic since their position in the clause involves greater speaker/writer choice. Halliday comments that they are, however, natural points of departure if present.

<u>Circumstantial elements</u>: a wide range of adverbials and prepositional phrases functions as Circumstantial elements (IFG:137-144). (1) Spatial and Temporal Extent/Location: *prior to growth, on a glass substrate*; (2) Manner: *by means of comparison*; (3) Cause: *in view of difficulties*; (4) Accompaniment: *besides this*; (5) Matter: *concerning this problem*; (6) Role: *as a catalyst.* These components have more varied room for movement within the clause and consequently the fronting of such elements indicates more strongly marked choices.

<u>Subordinate clauses</u>: within the clause complex, the fronting of dependent subordinate/non-finite clauses represents strongly marked choices, for example, *When x=0.92*, we have Al atoms.. . Such elements can be 'treated holistically as thematic units for the larger sentence' (Eiler 1986:54) and this holistic approach is followed by

other analysts. Fries (1992), for example, employs the T-Unit (cf. Hunt 1965) - an independent clause together with all hypotactically related clauses which are dependent on it - in his analyses of thematic structure.

In terms of traditional Hallidayan structural relationships, all of the above examples can be classified under the heading of Adjunct. Other categorisations will be found in other grammars, most notably Quirk *et al.* (1985), where interpretation of the range of Adjunct functions is more restricted. However, the wider usage of the term is employed here as applied to the above structures in their role as fronted thematic choices.

<u>Complements</u>: finally, the most strongly marked Themes in the predominantly declarative statements of RA discourse are Complements, 'the least likely to be thematic' (Halliday IFG:45), for example, *Of particular interest* is mechanism x.

As a result of recognition of varying degrees of markedness outlined above, both Fries and Halliday suggest that in cases where Theme is weak, for example, with conjunctions, because their thematic potential is partly in-built, 'when one of them is present it does not take up the whole of the thematic potential of the clause. ..whatever item is selected to follow will still have thematic force' (Halliday IFG:51). In multiple Themes, then, the following component (Subject, Complement or Adjunct) will also be thematic.

Davies (1991) points out that possibly one of the most problematic areas in Theme analysis is being able to distinguish between overlapping semantic categories and realisations which appear to serve more than one metafunction, yet which clearly have the same discourse function. The problem most clearly affects Halliday's treatment of Circumstantial elements (IFG:137-144) and

conjunctive/modal Adjuncts listed above. In turn, difficulties with their analysis affect the metafunctional status that is assigned to these components and thus present certain difficulties in defining the scope of Theme. Downing (1991) also indicates that particular problems of analysis occur with Adjuncts and Circumstantials when taken as topical Themes.

The use of an example from a scientific research article may illustrate some of these potential difficulties and inconsistencies in working Halliday's thematic analysis of metafunctional roles:

[8] *However, recently, Fraas[10] reported a decreasing carbon concentration.*

From Halliday's (IFG:53) comments on multiple, complex, Themes, the sentence-initial thematic constituents of [8] can be analysed as follows:

[8a]	*However,*	*recently*	*Fraas[10] reported*
metafunction:	textual	ideational	
discourse function:	adversative	temporal location	
grammatical function:	conjunctive Adjunct	Circumstantial element	
thematic structure:	THEME		RHEME

According to Halliday (IFG:138), *recently* may function as a Circumstantial element (temporal location), serving the ideational metafunction as part of the system of transitivity which 'in principle.. is anything

representing a process (person, thing, institution, etc.)
or a circumstance attendant on that process (time, place,
manner, etc.)' (IFG:54). Halliday's position regarding
multiple Themes is that 'the ideational element is always
the final one - whatever follows the first ideational
element of the clause is automatically part of the Rheme'
(IFG:54). In our example [8], therefore, if the writer had
chosen (moreover, due to generic constraints is led) to
front *recently*, the following thematic structure is
implied:

[8b]	*Recently,*	*however,*	*Fraas*[10] *reported...*
THEME			RHEME

The Adjunct *Recently* then is said to realise the
obligatory 'content' thematic element. Furthermore, from
Halliday's comments (IFG:82), it appears that the
conjunctive Adjunct *however*, now with decreasing thematic
force or flavour, falls 'between' Theme and Rheme; this
does not appear to be a particularly satisfactory
functional explanation of such cases. Further comparison
of some of Halliday's examples reveals other apparent
difficulties; for example, in one case, the whole phrase
on Sunday perhaps we.. (1979:72 Fig.13) is taken as
thematic (labelled Themes[1-3]). In another example, quoted
above (IFG:55 Fig.3-14c), however, *on the other hand maybe
on a weekday..* realises Theme. Thus, in one example, *on a
weekday* is the first topical element and therefore the
Theme/Rheme boundary; in the other, the similar
(functionally speaking) temporal element *on Sunday* appears
not to qualify as the first obligatory ideational
component.

As mentioned above, Davies (1991) has highlighted some of
the problems for text analysts which arise in identifying

potential cut-off points between metafunctions, in particular, when the same realisation appears to serve both ideational-experiential and textual/interpersonal metafunctions. This question is particularly relevant when referring to some of Halliday's more fully lexicalised examples, for example under conjunctive/modal Adjuncts (IFG:50): *under the circumstances*, *with this in mind*, *as far as that's concerned*, *to my surprise*; and under Circumstantials (IFG:137-144): *as for John*, *with a view to promotion*, *a long time ago*. Davies suggests that to identify a cut-off point between the ideational-experiential component in Circumstantials and that in conjunctive or modal Adjunct is extremely problematic, leading to much greater differences in analysis in sentence Themes such as [8a] and [8b] than their functional role would seem to suggest. Alternative views may be explored therefore.

In [8] above, *recently* may also be coded as serving its common function as temporal conjunctive Adjunct (IFG:50), along with many other common 'minimals' of Location in Time and Space, such as, *now*, *then*, *there*, *here*. As indicated above (section 3.1), the notion behind *recently* clearly has relevance to the discourse-functional context in which RA writers are developing their research context, as well as meaning in the sense of topical 'content'. Thus elements, which may be seen to be inherently Circumstantial [i.e. ideational], serving the micro-functional purposes of expressing notions of Time, Cause, Concession, Probability etc. (Davies 1991) may also, from a discourse-functional perspective, be seen to have textual or interpersonal thematic flavour in terms of their local text-specific functions. Below are listed a variety of examples (taken from the present main corpus of 36 research articles) which Halliday lists under five categories of Circumstantial elements (IFG:137-144).

- <u>Extent and Location</u>

    [a] *Here*, F_{CENT} allows for weak collision effects..

    [b] *There*, the power had to be reduced..

    [c] *Recently*, we reported on the growth of..

- <u>Manner</u>

    [a] *With this construction*, we were able to perform..

    [b] *Like villin*, fimbrin was localised..

- <u>Cause</u>

    [a] *Because of their important roles in these fields*,
       laboratory experiments..have concentrated on..

    [b] *For the present purposes*, we have adopted..

- <u>Accompaniment</u>

    [a] *Besides Bi₄m*, the other distinction..is..

- <u>Matter</u>

    [a] *As for the other proteins*, a low level..was..

As marked sentence-initial thematic components, these Circumstantial elements have powerful discourse contextualising and text-organising functions, that is, they can be seen to serve the overall, message-enabling *textual* metafunction of language - of course, it must be noted that internal elements of marked Themes may, in turn, include strands of interpersonal and ideational Theme. The suggested reason for a dominant discourse-functional role for both marked and unmarked Themes in scientific research articles is outlined below as part of an overview of the discourse-functional approach to Theme adopted in this study.

3.2.6 <u>A discourse-functional approach to Theme</u>

An increasing number of recent studies on Theme emanating from the systemic-functional school represent both minor and major departures from Halliday's analysis of thematic components. Fries (to appear) comments that there appear to be two key issues in the development of the concept. He poses the question whether there is a single consistent function to Theme and suggests the need for a better

semantic description, a more 'operationalizable description' of Theme. Fries proposes that,

> 'it seems to me more useful (and more approachable) to explore the reasons why one chooses certain items of information as peg and one does not choose others in specific contexts. This, then, will eventually become a sort of definition by example' (to appear:7).

Perhaps partly due to some of the challenging difficulties in analysis mentioned above, and with optimally transferable pedagogical applications in mind, other researchers [notably Eiler (1986); Berry (1989a); Davies (1991); Downing (1991); Gibson (1992); Fries & Francis (1992)] have indicated an interest in the potential value of a broader-based discourse-functional categorisation of Theme. Furthermore, a more operationally-defined treatment of Theme may give a more transparent view of textual development in written scientific discourse. Although Martin himself closely follows Halliday's own analysis of Theme, he suggests that 'for purposes of textual analysis and interpretation... it generally proves more practical to draw a categorical line between Theme and Rheme' (1992b:151); the spirit of this statement guides the present approach.

As an example of this position and following Davies (1991), in [8] above, *However* expressing contrast can be viewed as the primary semantic notion behind the writer-selected and genre-constrained point of departure - it contextualises both preceding stretches of discourse and sets up expectations for following statements; the element expressing temporal location *recently* is a secondary contextualising notion. At this point in the discourse, the two Adjuncts together set the scene for the following thematic/rhematic development. Reversal of these two naturally implies a subtle shift of emphasis, but in a broad sense, there appears to be little significant

functional difference between the contexts of *However, recently* and *Recently, however.* In Halliday's analysis, if *recently* is seen to exhaust thematic potential as the obligatory ideational component, following Hudson's (1986) criticism, it may be a point of departure as the clause-initial element, but it is very hard to see it as a contribution to the definition of what the clause is going to be 'about'. On the other hand, with both *however* and *recently* as marked thematic choices, then, the Theme in [8c] below would include the unmarked Subject *Fraas* as the first ideational component. [This is, in effect, similar to Halliday's (IFG:64) notion of a 'displaced' Theme; but this description is noted by Gibson (1992:266) as a 'particularly unsatisfactory' construct and by Downing (1991:126) as 'unnecessary'].

[8c]	*However,*	*recently*	*Fraas[10]*	*reported*
thematic function:	primary context	secondary context	Topic	
discourse function:	contrast	temporal location	participant	
grammatical function:	Adjuncts		Subject	
thematic structure:	THEME			RHEME

Thus, the sentence-inital Theme is 'about' setting up a contrast between previously mentioned research and what the latest from *Fraas[10]* has to report; in this way, the 'contextualising frame' (cf. Chafe's (1976) use of 'frameworks' and Lowe's (1987) 'situational frameworks') serves to move the discourse rhetoric dynamically forward at this point (in this case) of the Introduction section.

In her analysis, Berry implies that the scope of Theme may be further extended beyond the Subject:

'practice varies widely over how much of a clause to count as the beginning. ..I have erred on the side of generosity, as it were, and included in the theme everything that anyone working in the Hallidayan tradition has ever to my knowledge advocated including. This means that I have treated as theme everything that precedes the verb of the main clause. Where a subordinate clause precedes the main clause, this too has been included' (1989a:71).

If we consider example [8d] below, we have:

[8d]	*Recently,*	*Fraas[10],*	. *however,*	*reported*
thematic: function:	primary context	Topic	secondary context	
discourse function:	temporal location	participant	contrast	
grammatical function:	conjunctive Adjunct	Subject	conjunctive Adjunct	
thematic structure:	THEME			RHEME

By means of Berry's extension of Theme up to the main clause verb (here *reported*), [8d] suggests that the contextualising Adjunct placed after the Subject as obligatory ideational element carries thematic flavour. However, as a more embedded 'point of departure', the contribution of this secondary context *however* to the method of development of the text of which it forms a part is likely to be more limited. [In connection with the examples we are presently using, Tim Johns (pc) notes from concordanced samples of scientific writing that there may be a strong preference for *however* to appear in non-initial position. This was not confirmed by

the present corpus - of 124 instances of *however*, 86% were sentence-initial (s-i), as in example [8a], 3% were pre-Subject non s-i [8b], 2% were post-Subject pre-verb [8d], with 9% post-verb as in *The observed transition behaviour was, however, still reproducible...*]. In attempting to gain a more workably transparent view of textual development in the scientific RA, this approach indicates the potential for functional discourse analysis of taking Subject as the obligatory thematic component.

Previous studies on Theme have raised questions regarding the relation between grammatical function and discourse function of Theme. Halliday rejects the view that Subject has a purely 'grammatical' function, since all elements are considered semantic in origin - being the Subject 'means' something, in particular it is held responsible for the functioning of the clause as an interactive event; it is seen as a structure imposed by the interpersonal metafunction through the system of Mood. In scientific discourse, it can be suggested that the predominant conflated Theme/Subject pattern may have particular significance for within-text structuring and the balancing of 'interactional' and 'topic-based' Themes.

On the other hand, although Fries (1983) agrees with Halliday that thematic content correlates with the method of development of a text, he feels that the function of the Subject is not relevant for this development in his analyses. Berry comments that on the one hand the arguments for the discourse functions of Theme are convincing, 'nevertheless it seems a little premature to dismiss the grammatical subject as irrelevant' (1987a:71). Consequently, Berry argues that both Subject and Theme are important for both the notion of 'topic' (that is, what a given stretch of discourse is 'about') and the method of development of a text. In general research terms, Fries & Francis (1992) comment that it is important to get a sense of the functional environment in which thematic

information works and therefore we need to collect more data and compare and contrast Theme with concepts such as the grammatical Subject in order to gain more insight into the meaning of Theme.

In order to capture generalisations about both marked and unmarked Theme discourse functions, Davies (1988a) has also adopted the position of including the Subject as an obligatory element in her analyses of Theme and the merits of this approach have also been suggested by other researchers (Enkvist 1973; Downing 1991). Davies's rationale for this proposed approach is as follows:

> 'Subject is equated with the intuitive notion of "what the clause is about". In discourse... the repeated occurrence and re-occurrence of the same topical element or a related topical element as Subject is seen not only to specify topic... but also to be the primary means by which the continuity of coherent discourse is achieved' (1988a:3).

Davies states that since Subjects and their thematic function as 'topical elements' are recurrent components in discourse, their interest derives not only from their prominent structural position, but also from their functional-semantic roles in creating coherent discourse. In order to explore these roles, Davies (1988a) puts forward a two-part analysis of Theme, namely obligatory TOPIC realised by the grammatical Subject (GS) and optional CONTEXT FRAME (CF) realised by any elements preceding the GS. Thus Davies stresses both the formal and functional properties of the term TOPIC; this approach is similar to Givón's entity as the 'most likely to be coded as the primary topic - or grammatical subject - of the vast majority of sequentially-ordered clauses/sentences comprising the thematic paragraph' (1982:8). This gloss can also be compared with other descriptions of the term 'topic' which focus more on 'aboutness': according to Lautamatti (1987), it is the main idea discussed

(discourse topic) or subordinate ideas (sub-topic); for
Kaplan (1983), it is the dominant notion that governs a
sequence of discourse.

Following Davies, the example in [9] of Topic (unmarked
Theme) may be described as follows:

[9]	*The Curie temperature of the alloy*	*is*
grammatical function:	Subject	
discourse function:	Real World Entity	
Theme component:	TOPIC	
thematic structure:	THEME	RHEME

Furthermore, example [10] below indicates a possible
analysis for a binary Theme, namely, marked Context Frame
+ unmarked Topic/Subject.

[10]	*In this paper,*	*we*	*describe*
grammatical function:	Adjunct	Subject	
discourse function:	Location in Discourse Space	Discourse Participant	
Theme component:	CONTEXT FRAME	TOPIC	
thematic structure:	marked THEME	unmarked Theme	RHEME

Davies (1989) suggests that, seen from a functional
perspective, as well as often representing common notions
such as concession, cause, condition etc., fronted marked
elements in scientific discourse frequently indicate
notions such as validation of internal evidence, *according
to our results*, location in discourse time and space, *in
brief/in Figure 2b* and writer viewpoint, *interestingly*.
Investigation of marked choices from the perspective of
the primary semantic notions they serve to introduce

implies that the Theme of a fronted *if* subordinate clause is its real or hypothetical conditionality; that the foregrounding of *moreover* introduces the primary notion of additive emphasis; that *in agreement with Keuch*[5] performs the functional role in RA discourse of validation of external evidence. The idea of a primary semantic notion is particularly important in multiple Themes illustrated by example [8] above: *However, recently, Fraas [and co-workers*[10]*] reported...* where contrast is seen as the primary notion behind the selected point of departure, chosen ahead of the secondary marked thematic notion of time. [However, this would not be the case with *But, recently* - since the order cannot here be reversed - and thus differing degrees of Theme markedness (see Figure 5 above) are involved].

To what extent are thematic choices such as these illustrated determined by the writer at any particular point in discourse and to what extent are they guided by generic constraints? Fries & Francis comment on research into correlations between the elements of structure of a genre and Theme choice and indicate that 'there remains an important question here as to whether thematic choices are truly local choices or whether they are part of more global operations' (1992:55). In addressing this point, let us consider a writer's available choices in the framework of Swales's influential Four Moves pattern (1981) [or later CARS (Create A Research Space) model (1990a) - see Figure 4 above] in an RA Introduction. At a given point, for example, in indicating gaps in previous research [Swales's Move Three (1981)], in Establishing a Niche [CARS Move Two (1990a)], a writer typically needs to realise a shift in discourse focus through a fronted element indicating contrast:

[11]   *SiHn has been widely applied in semi-conductor technology and used extensively in research as a doping material. However, SiF₄ has so far received little attention.*

The same contrastive/concessive function may be realised
by the structurally more complex:

[12]   *Although SiHn has been widely applied in semi-conductor
       research and has been used extensively as a doping
       material, SiF₄ has so far received little attention.*

Clearly the juxtaposed components of 'SiHn usage +
contrast + SiF₄ neglect' are the local rhetorical focus of
the writer's intended message at this point. The marked
notion of contrast serves as a point of departure of that
move in both cases. More globally within the textual
macrostructure of a carefully crafted introductory
section, the writer is setting up 'a link in an
expectation chain' (Thompson 1985:61), thus stressing the
predictive nature of language and the role of
'probabilities in the form of typical or recurrent
patterns' (Fine 1992:10). Nesbitt & Plum (1988) stress
that the relation of language and context is one of mutual
predictability and, in this case, we can predict that RA
writers will attempt to fill those perceived gaps in
existing research by means of explicit comment on their
own research.

To address the earlier question by Fries & Francis (1992),
examples [11] and [12] above suggest that local thematic
choices are strongly determined and constrained by the
global within-text structuring of the RA genre and hence
are part of a predictable dynamic progression. Moreover,
within the present context of scientific discourse,
thematic choices would be expected to reflect the highly
conventionalised hypothetico-deductive method inherent in
conducting scientific research and this would consequently
be reflected in its reporting. In this way, the
'appropriate' interplay between generic (for example,
rhetorical moves) and text-type (for example, description
of processes) top-down constraints and bottom-up thematic

selections are expected to contribute greatly to the perception of texts as 'successful'; this interplay helps to fulfill the expectations of readers of a familiar genre since 'this close match provides a powerful boost to coherence for the reader and a powerful boost to organisation for the writer' (Swales 1990a:190).

The purpose of the following section is to clarify initial hypotheses that have arisen as a result of this review of current issues in research on Theme and which are thus central to the present study.

3.3 INITIAL HYPOTHESES

From the above discussion, it is hypothesised that both unmarked and marked components of Theme perform a variety of discourse functions. Consequently,

[HYPOTHESIS #1]
a detailed functional analysis of both Topics and Context Frames in a corpus of refereed, published scientific research articles will highlight the changing options for, and constraints on, thematic choices across RA sections as the discourse proceeds with developing rhetorical goals. More specifically,

[HYPOTHESIS #1a]
a functional analysis of marked Context Frames will eludicate their potential to describe the rhetorical organisation and conceptual macrostructure of scientific RAs. In addition,

[HYPOTHESIS #1b]
a functional analysis of unmarked Topics/Subjects will give insight into how experienced writers use such Theme choices to balance interactional and topic-based thematic

93

components throughout RA discourse. This would help clarify the social construction of the RA product and how it is developed as an instrument of social interaction.

Delicacy refers to the degree and depth of detail in which given structures are described and the binary Context Frame + Topic approach to the analysis and description of Theme which has been adopted in this study is clearly less delicate in some respects - however, the required degree of delicacy is itself dependent on the overriding purpose of description. As the current research is an example of EAP-driven discourse analysis,

[HYPOTHESIS #2]
by means of integration and mapping of levels of linguistic and rhetorical description, the 'tendency of certain types of theme to cluster in given components of a text' (Eiler 1986:54) will suggest Theme-Move configurations which can be seen as integral to the characterisation of the RA genre.

This investigation of the textual characteristics of the scientific RA genre is explored through analysis of a corpus of RAs which is outlined in the next section (3.4). Hypotheses outlined above are addressed in the following sections with detailed description of marked Context Frames (3.5) and unmarked Topics/Subjects (3.6); major findings from their functional analysis are presented in 3.7 and 3.8, respectively. The next section (3.9) highlights examples of binary Theme choices (Context Frames + Topics) with comments on their occurrence and relevance. The final two sections represent a synthesis of work so far: 3.10 suggests how marked and unmarked thematic choices can be mapped onto the generic moves they help to signal and characterise, thus creating Theme-Move maps; 3.11 offers concluding remarks resulting from the corpus analysis which will lead to the pedagogical

application of insights gained from this study of Theme choices in scientific research articles.

3.4 RESEARCH METHODOLOGY

3.4.1 <u>Selection of corpus</u>

It is clear that studies which aim to describe generic characteristics should not be based on a small number of hand-picked, highly-valued or possibly prototypical (Swales 1987a, 1990a) examples of the genre under investigation. Texts must be seen to represent the authentic discourse of the particular community under investigation. Consequently, three major issues which should guide the identification of sources for texts are representativity, reputation and accessibility (Nwogu 1990) and in order to satisfy these conditions the corpus was selected as follows.

The research articles selected for the present corpus [see APPENDIX A for references] are based within three major disciplines of the physical and life sciences - Physics, Chemistry and Biological Sciences, although with the nature of overlapping research fields these may include areas such as physical chemistry, chemical physics, biochemistry etc. The rationale for this selection was as follows: the background to the present study involves teaching support courses in English for Academic Purposes to groups of novice NNS (non-native English speaking) researchers in these above fields - in this case, Japanese doctoral students at a science and technology university. Consultation with their research supervisors, a wide range of experienced Japanese academics in the above departments who thereby acted as subject specialist informants, produced a list of international English-language journals (i) which were considered to be widely-read mainstream research publications in their respective fields; (ii)

which were also common destinations for their own submissions of RAs and (iii) which were publicly available accounts, in this case from the university library.

For the present exploratory purposes, 36 RAs taken from a broad selection of twelve journals (the top four journals listed in each of the three major scientific fields, but for the main NS corpus, journals which were not published in Japan) with three articles taken from each appeared suitable [4 journals x 3 fields x 3 RAs = 36]. The RAs appeared in professional, refereed academic journals (see TABLE 1) published in the U.K. (6), the U.S. (5) and Canada (1) and were ostensibly written by English NSs affiliated to institutions in those countries, as indicated on RAs. The processes of peer review and editorial scrutiny ensure that published RAs have satisfied both the scientific and linguistic standards of the discourse community. Six of the journals are published monthly, one weekly, four fortnightly and one bimonthly.

TABLE 1
NS Corpus Journals

Journals of Physics: Condensed Matter
Journal of Materials Science
Canadian Journal of Physics
Journal of Applied Physics
Journal of the Chemical Society: Faraday Transactions
The Biochemical Journal
Journal of Catalysis
Journal of Chemical Physics
Cellular Signalling
Journal of Cell Science
Journal of Cell Biology
Cell Motility and the Cytoskeleton

Limitations on the selection of RAs were as follows: all RAs appeared in 1989/1990 volumes as regular papers - not Letters or Notes - and only had the common structure: Introduction, Experimental (Materials & Methods in Biological Sciences), Results, Discussion/(Conclusion), in

this order and with no other sections. Decisions whether there are separate Results and Discussion sections are often a question of journal style rather than author preference, although this can depend on the nature of results and how they can be discussed, which itself may be decided by the author. Length of article was not a selection limitation. Because of these imposed limitations, the procedure for selection of the 36 RAs was thus not as random as outlined by Crookes (1986). However, as refereed, published articles, these products of a highly conventionalised genre may be taken as representative of the socially embedded models which are presented to scientific RA readers and writers.

3.4.2 Method of analysis

The study of Theme presented here involved a quantitative analysis of corpus data with discrete stages of classification of Themes, counting and tabulation of data and interpretation of thematic patterns throughout RA sections and of their significance for discourse structure.

The classification of (i) unmarked Themes (Topic/grammatical Subject = GS) according to their perceived discourse-functional roles, and (ii) marked Themes (Context Frame = CF) according to both discourse functions and grammatical form, entailed the defining of variables and the setting of parameters for membership of a particular thematic choice in a specific category. Rather than putting forward *ad hoc* categories for thematic analysis, the present approach aimed to build more systematically on previous research, particularly the work of Davies (1988a, 1989, 1991) and the functional categories set up in the present work represent a modification and extension of this earlier work.

In order to interpret counted examples of Theme in any statistically reliable way, categories should be mutually exclusive. However, problems initially arose in the coding of some Themes due to the apparent semantic blending of several of the categories which had been set up. Consequently, the refinement of discourse-functional categories depending on the required degree of delicacy of analysis and the subsequent reclassification of Themes can be seen as a natural corollary of an exploratory data-driven approach. In addition, there were a very few cases of Themes with dual coding potential [for example, the Context Frame *then* could be coded in the temporal, causal or conditional sense]. In these few rare cases, three NS subject specialist informants (in addition to the NNS research supervisors mentioned above in section 3.4.1) were consulted with the individual texts and examples prior to final coding for advice.

In the end, four major classifications of discourse-functional domains were established for (i) unmarked Themes and these were further subdivided into related categories according to data. Nine major categories of discourse functions were established for (ii) marked Themes, some with subdivisions; in addition, three major types of Context Frame according to grammatical form were highlighted. Although the main focus of attention here is on a discourse-functional analysis of Theme within a semantic framework, it was thought that a secondary analysis of Context Frames according to one measurement of grammatical complexity [for example, the 'minimal' thematic choice of coordinate *However..* versus the complex subordinate *Although..*] may shed some light on developmental phenomena in novice NNS researchers' writing of RA drafts.

The number of Themes was tallied and collated data were examined by means of elementary statistical analyses, percentage tables and standard deviations, and use was

made of graphs in the interpretation of evident patterns
of thematic distribution according to discourse functions
throughout the four major sections of the RA. An attempt
was initially made to utilise the potential of computer
concordancing with the main corpus of 36 published NS RAs.
However, due to the extensive length of many
sentence-initial Themes to be coded (several lines of
text), the juxtaposition of marked Context Frames with
unmarked grammatical Subject headwords on screen proved
technically difficult and so binary Themes (CF+GS) and
their collocations were manually tallied and examined
(section 3.9). [However, computer concordanced Themes are
used as data elsewhere in this study - see Chapters Five
and Six].

Further criteria as regards the coding of examples of
Themes in the main corpus of 36 NS RAs were as follows:

- The analysis of sentence-initial elements (SIEs)
[N=4358] produced a range of unmarked Themes
(Topics/grammatical Subjects) [N=4337], of which a number
were preceded by a wide variety of marked Themes (Context
Frames) [N=1408]; these were coded separately. As is
evident from the above total N counts, 21
non-Subject/Context Frame Themes (complement, imperative,
interrogative) were recorded but excluded from the
analysis. Apart from the necessity to refine initial
categories of Theme in the light of emerging data and some
minor difficulties noted above, the overwhelming presence
of declarative statements in scientific RA discourse, and
thus the relatively limited thematic range of this genre
(compared to, for example, literary texts or spoken
discourse), accounted for the generally straightforward
task of coding Themes here.

- The study examines only Themes in sentence-initial
clauses since the intention is to give a clearer picture
of thematic patterns throughout RA discourse without

secondary organisation. Halliday confirms that in the method of development of a text, it is the thematic structure of independent clauses that makes the main contribution; Sinclair also comments that 'boundary utterances have a more far-reaching role than medial ones' (1985a:15). The analysis of only main clause Themes was the approach taken in studies by Brandt (1986) and Berry who comments that 'it would seem sensible to concentrate on those themes generally agreed to be significant for text organisation and genre-awareness' (1989a:71).

- Included in this analysis are examples of anticipatory *it* and existential *there* Subjects - both labelled hereafter 'empty Theme' choices after Quirk *et al.* (1985). The thematic interpretation of such examples has created difficulties for discourse analysts, for example, Eiler (1986) could offer no text-specific explanation for their occurrences in her study. However, following Davies's (1988a) examples of 'invisible viewpoint', *it/there* may be coded according to the semantic function of the predicated Theme they serve to introduce.

- Where there were multiple Context Frames, for example, *However, in the present paper,..* (in ca. 9% of total N), analysis and coding was limited to the primary semantic notion involved (in this case, one of contrast). Following Halliday's notion of gradience, the principle behind this is that the fronting of the notion inherent in *however* is still as a result of choice.

- If a separate Conclusions/Summary section was present in the RA, for the purpose of analysis, this was counted with the Discussion section which in any case frequently carries concluding remarks. The Abstract was excluded from analysis as it is not considered to be an integral part of RA discourse but a separate distilled version of it.

The following five sections now present in detail a discourse-functional description and analysis of Context Frames (marked Themes) and Subjects (unmarked Themes) in scientific RAs.

3.5 DESCRIPTION OF CONTEXT FRAMES (MARKED THEMES)

The first section (3.5.1) describes Context Frames (marked Themes) according to a grammatical paradigm with the next section (3.5.2) focusing on their potential discourse-functional categorisation.

3.5.1 Context Frames - grammatical form

As indicated above, it is suggested that a formal analysis of Context Frames according to one measurement of grammatical complexity may be useful in the investigation of developmental phenomena in novice NNS researchers' writing of RA drafts. The analysis of marked Themes in the corpus suggested three major Context Frame (CF) types according to grammatical form:

CF Type 1: Conjunctive/Modal Adjuncts and Conjunctions:
'minimals' - conjunctive adjuncts *briefly, however, moreover, therefore, likewise*; modal adjuncts *certainly, evidently, unfortunately, surprisingly*; conjunctions *but, and, yet, nor, so,*

CF Type 2: Prepositional and Adverbial Phrases:
simple prepositional phrases with optional modification, for example, *in this paper, among these techniques, from these results, (50 cm) upstream, (even) after six hours* - complex prepositional and adverbial phrases, for example, *because of the similarities, in the case of x=0.92, in contrast to p36p33, in light of these results,*

CF Type 3: <u>Subordinate Clauses and Non-finite Clauses</u>:
while the fraction was 59%, when fk=1 is applied, if the polarity is reversed, summarizing the behaviour, referring to data of Haber[21], unless otherwise stated, as shown in Figure 1,

3.5.2 <u>Context Frames - functional categories</u>

From analysis of their primary semantic functions, marked thematic choices fell into nine major categories some with subdivisions. These categories represent a modification of those originally suggested by Davies (1989) in the light of analysis of the present corpus of published scientific RAs. Examples are given for each CF Type where they have been identified:

1a. <u>Location in Time - Real World Entity (RWE)</u>

CF1 *Recently, Previously, Subsequently, Later, As yet,*

CF2 *After the removal of IFNs, At fertilization,*

CF3 *As the temperature increases, When observing LIF,*

1b. <u>Location in Time - Discourse Entity (DE)</u>

CF1 *In summary, In brief, Initially, First, Second,*

CF2 *Throughout the discussion,*

2a. <u>Location in Space - Real World Entity (RWE)</u>

CF1 *Here, There*

CF2 *On a glass substrate, Below the spectra, Near 540K,*

CF3 *Where uptake occured,*

2b. <u>Location in Space - Discourse Entity (DE)</u>

CF1 *Here*

CF2 *In this paper, In Figure 1b, On the figure,*

Categories 1a/2a focus on orientation within the external real world of existing scientific knowledge in terms of matter, processes and procedures and their chronological and spatial location. On the other hand, 1b/2b describe

the internal reader-oriented medium through which the act of manufacturing knowledge occurs, that is, the presented argument of the writer, specifically the written paper, report, figure etc. and CFs here are commonly expressed through deictic forms helping to create textual cohesion and interaction. In some cases, the same realisation may occur in both categories, for example, *First*, *Here*, referring to location in both real-world and discourse processes; this usage is generally unambiguous.

3a. <u>Addition - appositive</u>
    CF1 *For example, For instance, In other words, That is,*

3b. <u>Addition - emphatic</u>
    CF1 *Furthermore, Moreover, Indeed, Essentially,*
    CF2 *Besides BiM, In addition to localization,*
    CF3 *Besides being effective, Taken together,*

The additive functional layer serves (3a) to exemplify and elaborate through apposition or (3b) to expand on preceding statements through positive emphasis.

4. <u>Contrast/concession</u>
    CF1 *However, By contrast, On the other hand, Instead,*
    CF2 *Despite this information, Unlike the present data,*
    CF3 *Although there is controversy, While X was 59%,*

In contrast to the positive additive function of 3a/b, CFs of Contrast/concession are essentially for negative expansion. As such, they help create the polarising tension necessary for writers to set up and achieve certain rhetorical aims. For example, in introductory sections, CFs in 4 introduce the initial component of the problem-solution macrostructure inherent in the indication of gaps in previous research, and thereby serve to predict the rhetorical justification of research opinions and actions. CF categories 3 and 4 indicating positive and negative expansion and elaboration are certainly less

103

delicate than other analyses. For instance, according to other accounts [see Halliday & Hasan (1976); Halliday (1985a); Quirk *et al.* (1985); Morrow (1989)], examples in category 4 can alternately be taken as adversative, contrastive, concessive, replacive, antithetic - in other words, they involve varying degrees of semantic blending. For this reason, the aim in the present analysis is to describe a necessarily broader range of semantic functional categories which are nevertheless useful in the interpretation of RA discourse.

5a. Cause - reason/result
    CF1 *Thus, Therefore, Consequently, Hence, As a result,*
    CF2 *Due to crystal symmetry, Because of the similarities,*
    CF3 *Because t-PA binds fibrin, Since there is anisotropy,*

5b. Cause - purpose
    CF1 *To this end*
    CF2 *For evaluation of the gap, As a final check,*
    CF3 *In order to obtain HL cells,*

Closely coupled with contrast/concession, Cause CFs in this category contribute to the justification of research actions and hypotheses, often through introduction of the solution part of the recurrent rhetorical pattern indicated in 4 above.

6. Means
    CF1 *In this way,*
    CF2 *By further studies, With this construction,*
    CF3 *Using Fura-2, By using Henry's law constant,*

CFs of means introduce common processes and techniques of scientific investigation. They are frequently part of the solution component in CF cycles indicating Cause - reason/result and purpose above.

7a. <u>Condition - real</u>

    CF1 *In this case, Then, In practice,*

    CF2 *Under these conditions, In the case of x=0.92,*

    CF3 *If the polarity is reversed, Unless otherwise stated,*

7b. <u>Condition - hypothetical</u>

    CF1 *In principle, Perhaps, Conceivably,*

    CF2 *With such possibilities,*

    CF3 *Assuming that V is constant, If Cs⁺ works as CsOH,*

The criteria for distinguishing here are based on observed and hypothesised or assumed phenomena. Category 7a focuses on the reporting of real-world events and facts based on processes and procedures executed according to the Experimental section with resulting observed cause and effect. On the other hand, 7b highlights as yet not fully tested hypotheses resulting from observed phenomena and unexplained or partial data which suggest future clarification. Use of 7b is an example of how writers can imply personal opinions whilst remaining rhetorically objective and invisible.

8a. <u>Validation - external</u>

    CF1 *In general, Usually, Normally, Qualitatively*

    CF2 *In previous studies, From data of Noyce[23],*

    CF3 *As is well known, As noted earlier[19,23],*

8b. <u>Validation - internal</u>

    CF1 *Significantly, Clearly, Certainly, Evidently,*

    CF2 *From results in Figure 1, According to Figure 4,*

    CF3 *As shown in Figure 1, As expected,*

Researchers are naturally required to place their contributions within the framework of relevant existing knowledge. With 8a writers seek validation and confirming evidence for their own hypotheses and conclusions from within the wider scientific community. Validation of previous findings is anticipated as part of Introduction

sections and is naturally most strongly echoed in the
presentation and discussion of new results and findings.
However, a clear distinction can be made between 8a, the
arbitrating external community and current state of
knowledge and 8b, the internal discourse mechanisms
writers employ to validate and structure their own
evidence to help persuade the external community. In
scientific RAs, 8b will particularly draw readers'
attention to graphic evidence presented in figures,
equations and tables, although of course validation of new
data may not itself be self-evident without detailed
reasoning.

9.   Viewpoint
     CF1 *Interestingly, Unfortunately, Surprisingly, Apparently,*
     CF2 *In the broader view,*

In terms of visibility, the use of CFs expressing an overt
viewpoint temporarily gives writers a high discourse
profile, similar to a switch to the participant role of *we*
as grammatical subject.

This section 3.5 has presented a discourse-functional
description and categorisation of Context Frames as marked
thematic choices. Section 3.6 now turns to a parallel
analysis of the functional roles of unmarked grammatical
Subjects in the same main corpus of 36 published NS RAs.

3.6 DESCRIPTION OF SUBJECTS (UNMARKED THEMES)

This first section (3.6.1) describes the rationale behind
the model of Subject role domains (Figure 6) proposed
here; the following section (3.6.2) provides a detailed
description and exemplification of the
discourse-functional criteria for these Subject roles.

3.6.1 Subject role domains

An analysis of the functional roles of grammatical Subjects (GSs) from the 36 published research articles suggests that four main domains can be distinguished:

A. The Participant domain
B. The Discourse domain
C. The Hypothesised and Objectivised domain
D. The Real World domain

As can be seen in Figure 6 below, these four domains represent a continuum from the Participant to the Real World domain. Towards one end, it is typified by the increasingly overt presence of the writer as a visible participant in the research and research reporting process; towards the other, there is a greater focus on research-based, that is, real-world physical and mental entities and activities.

These two ends of the scale recall the earlier mentioned work (section 3.1) of Martin (1986) and the emphasis on a balance between more 'human face' *interactional* (Participant) thematisation and more impersonal *topic-based* (Real World) thematisation. The rationale for Figure 6 also follows Davies's (1988a/b) suggestion that Subject roles form a progressive cline of writer visibility, that is, means by which writers seek to present themselves and their viewpoints in the research community, with both very obvious and very subtle means of realisation. Sinclair (1985a) comments that much of the interactive quality of formal written language is covert and the model in Figure 6 reflects much of the work in social studies of science which dispels the myth that scientific writing is simply objective descriptive reporting. Swales (1990a) comments that the seemingly innocent issue of impersonality in scientific writing is indeed complex and notes,

PARTICIPANT DOMAIN	DISCOURSE DOMAIN	HYPOTHESIZED AND OBJECTIVIZED DOMAIN	REAL-WORLD DOMAIN
INTERACTIVE PARTICIPANT (Smith 1987)	INTERACTIVE DISCOURSE ENTITY (Previous studies)	HYPOTHESIZED ENTITY (The model)	
	EMPTY DISCOURSE THEME (It is concluded) (There are reports)	EMPTY H AND O THEME (It was evident) (There was evidence)	REAL-WORLD ENTITY (AlFeNi alloy)
	Micro DISCOURSE ENTITY (Figure 1b)		EMPTY REAL-WORLD THEME (It was found) (There was a delay) REAL-WORLD EVENT/PROCESS (Annealing)
		OBJECTIVIZED VIEWPOINT (One factor)	MENTAL PROCESS (Deduction)
	Macro DISCOURSE ENTITY (This paper)	HYPOTHESIZED VIEWPOINT (The possibility)	
PARTICIPANT VIEWPOINT (Our argument) DISCOURSE PARTICIPANT (We)	DISCOURSE EVENT/PROCESS (The conclusion)		

MORE EXTERNAL COMMUNITY-ORIENTED THEME ↑

MORE INTERNAL WRITER-ORIENTED THEME ↓

MORE INTERACTIONAL THEME ←——————————→ MORE TOPIC-BASED THEME

GREATEST WRITER VISIBILITY

LEAST WRITER VISIBILITY

Figure 6. Model of Subject role domains.

'we are far away from a world in which power,
allegiance and self-esteem play no part, however much
they may seem to be absent from the frigid surface of
RA discourse. And yet we find the research article,
this key product of the knowledge-manufacturing
industry, to be a remarkable phenomenon, so cunningly
engineered by rhetorical machining that it somehow
still gives an *impression* of being but a simple
description of relatively untransmuted raw material'
(1990a:125).

This model suggests how varying degrees of impersonality
and personality in scientific RA discourse may be clearly
revealed. The (horizontal) cline of writer invisibility
<-> visibility, that is, interactional <-> topic-based
thematisation is illustrated by examples taken from the
present corpus. As can be seen from Figure 6, within each
domain, there are several sub-categories, each related by
the primary focus of the domain. A secondary (vertical)
continuum, that of an 'internal' writer orientation versus
an 'external' research community orientation is also
suggested. Therefore, the Participant domain includes both
the 'internal' RA writers-as-researchers Subject *we*, and
also Subjects which originate in the 'external' world of
the academic community, such as cited author *Smith (1987)*.
Both GSs as Theme choices give research participants
maximum visibility and thereby implied authority. In
journals and disciplines where only number references are
commonly used, a switch to the occasional use of discourse
participant as Subject would represent a strongly 'marked'
choice.

Between the two poles of the continuum, there are two
further domains which represent more subtle realisations
of interactional thematisation. The Discourse domain
represents the research outcomes and products of
participants; they are manifested through the internal
discourse properties of the RA and are referred to by, for

example, *this paper* (a macro discourse entity); *figure 1b* (a micro discourse entity); or studies by other cited researchers *previous studies in this field*[1,2]... (an interactive discourse entity). The latter is another example of the 'external' community origin of the GS compared to the 'internal' RA writer orientation of the former. A shift from Participant to Discourse domain evidently makes the RA author less visible and the discourse at this point less interactional, with the focus on the research product rather than the producer. Nevertheless, behind *this paper reports on...* is the implication *we report on...* . However, the frequent use of Subject *we*, and thus the foregrounding of research participants rather than research entities and processes, may be seen as an overstatement of the perceived status of the writer-as-researcher. Such thematic choices may consequently be 'strongly modulated by perceptions of the anticipated reactions of peer-colleagues' (Swales & Najjar 1987:175)., as well as severely constrained by the rhetorical goals of the different generic components of RA discourse.

Between the relatively high writer-visibility presented through the Discourse domain and the seeming invisibility of the Real World domain, Subject roles in the Hypothesised and Objectivised (hereafter H&O) domain represent a wealth of perhaps the most subtle means by which writers' comments on hypotheses and viewpoints can be realised. This domain may therefore be seen to represent the most discreetly interactional Theme. The focus is primarily on research activities, their outcomes and hypothesised conclusions, for example, *the most probable cause for these two observations...*, *the most intriguing data seen in these experiments....* H&O Viewpoints are naturally seen as more 'internal' writer-oriented compared with the Hypothesised Entity, for example, *a model,* which has already been 'externally' validated by the research community.

Finally, Figure 6 shows that Empty Theme choices anticipatory *it* and existential *there* occur across three of the four domains. These are 'devices for providing some kind of dummy theme which will enable the originator to indicate the "new" status of a whole clause, including its subject' (Quirk *et al.* 1985:1402). In terms of the present research, they can be coded according to the functional role of the postponed Theme; for example *there are reports*[1,2] clearly suggests a Discourse domain role. They are extremely flexible thematic choices as they have the capacity to thematise different grammatical and semantic elements across different sections.

3.6.2 Criteria for Subject roles

This section describes more clearly the discourse-functional criteria on which the four domains and Subject roles are based and these are illustrated by examples. As with the previous description of Context Frames, components of these categories represent a modification of those originally suggested by Davies (1989) in the light of analysis of the present corpus of scientific RAs.

A. PARTICIPANT DOMAIN

(i) Discourse Participant (DP): clearly recognised through the use of *we*, even in cases where there is a single named author. Rarely, the participant may be unidentifiable e.g. pronominal *one* and hence more distanced.

   [a] *We* shall present the argument in this paper that...
   [b] *One* may write the intensity of a Cd line as $I_{Cd} = K_{Cd} N_{Cd}$.

(ii) Participant Viewpoint (PV): use of *our*, with a focus on research outcomes and activities i.e. our research rather than our reporting.

   [c] *Our data* must be extrapolated...

(iii) <u>Interactive Participant</u> (IP): researchers referred to by name in citations e.g. *Smith (1987), Keuch[5]*. This allows for the implicit validation and appropriate challenging of authorised published sources at a distance. As Myers comments, 'scientific discourse consists of interactions among scientists in which the maintenance of face is crucial' (1989:5). This Subject role is therefore a powerful means for writers to present their own stance through agreement/inclusion and disagreement/exclusion.

[d] *Leitnaker[12]* has reported that the role of molybdenum...

[e] *None of those authors* reported the optoelectronic changes...

[f] *All of these authors* observed good agreement...

B. DISCOURSE DOMAIN

(i) <u>Discourse Event/Process</u> (DEP): reference not to the investigative events and processes in research themselves as described in the Experimental section, but to the discourse acts and processes of their discussion and reporting e.g. *interpretation, conclusion, argument, explanation, observation, description*. These are frequently nominalised reporting verbs used in conjunction with Interactive Participants. Halliday (1988) commenting on the language of physical science notes that nominalisation is an essential resource for constructing scientific discourse. It achieves the important discoursal effects of packaging complex phenomena into single semiotic entities, thus rendering their rhetorical functions in terms of thematic and information structure fully explicit. Nominalised DEP Subjects/unmarked Themes acquire stative, factualised status as given information. In addition, nominalisation 'attracts lexis to itself' (Francis 1990:56) and adds to the flexibility of a writer's choices in structuring information because of the modification options it offers. This category also includes reference to writers' discourse goals in emphasis of the research/reporting process e.g. *the aim of this*

paper (that is, "the aim of doing this work and subsequently writing this paper").

[a] *The main conclusion on the work on Ni_3Al* is that...

[b] *The interpretation given here on the magnetic structure of re-entrant spinglasses in the FM state* gives some support for...

[c] *Similar arguments* could be applied.

[d] *The aim of this paper* is to study the origin of...

[e] *A detailed description of the deposition system and the MOMS growth technique* has been given elsewhere[2].

(ii) <u>Macro Discourse Entity</u> (MaDE): integral units of discourse e.g. *paper, report, communication, thesis.* This includes e.g. *work,* implying the reported entity as well as the research investigation.

[f] *This paper* describes similar experiments...

[g] *The present work, like that of Gorte and co-workers[2],* is a step in that direction.

[h] *This report* addresses the question of whether...

(iii) <u>Micro Discourse Entity</u> (MiDE): part, discourse-internal entities e.g. *figure, diagram,* and connected reference items.

[i] *Figure 2* shows the Fe spectra...

[j] *The broken curve in Figure 4* shows...

[k] *The graphs of ln[-ln(1-x)] against lnt* are approximately linear,..

(iv) <u>Interactive Discourse Entity</u> (IDE): referenced units which refer to previous, community-validated, macro discourse entities e.g. *previous studies* [6,7,8]. It also includes discourse entities and events/processes that relate to the current state of knowledge about a subject. Such examples can be taken as components of this discourse domain due to the implied written mode of communication within the community: e.g. *little is known* (i.e. "there have been few published studies").

[l] *A number of parallel studies* have recently... [7,8,9].

[m] *Little* is known about the parameters...

[n] *Much less attention* has been paid to...[2].

[o] *Much of the literature* has the v_1 and v_3 definitions reversed.

(v) <u>Empty Discourse Theme</u> (EDTh): in this and other domains, empty Theme choices introduced by *it/there* share any of the indicated domain criteria. They frequently introduce evaluative comment and therefore play a significant role in interactional thematisation.

[p] *It can be argued* that...

[q] *It can be seen [in figure 2b]* that...

[r] *There have been reports[5]* that...

[s] *There are several plausible explanations* for why...

C. HYPOTHESISED AND OBJECTIVISED DOMAIN

(i) <u>Hypothesised Viewpoint</u> (HV): refers to comment and judgement on research matters; it expresses writers' perceptions of degrees of uncertainty or explicit hedging, realised through obligatory modality e.g. *the possibility of variations, the most probable cause, the apparent contradiction*. The presence of hedging indicates a raised degree of writer visibility as it is partly introduced to counteract anticipated negative feedback and therefore allows for the presentation of writer viewpoint at a guarded distance.

[a] *One possibility* is that...

[b] *The most probable cause for these two observations* is...

[c] *The likelihood for its occurrence* increased with power...

[d] *This apparent anomalous variation* may be easily recognised...

[e] *Possible reasons for this behaviour* are discussed...

(ii) <u>Objectivised Viewpoint</u> (OV): a high degree of evaluative modification implies acknowledged or given status which is frequently reinforced by the use of timeless present tense. The broad range of adjectival and adverbial modification, sometimes in comparative/

superlative form, includes *a significant difference, the most striking influence, the surprising feature*, but compared with HV, OV is typified by its lack of modal qualification in the subject phrase. There are often strongly cohesive anaphoric references *one of the factors, a further contribution, further evidence* - this modification further enhances the perceived given status of the head noun. The rare examples (N=5) of equative theme (see [g/h] below) in the corpus appeared in this category. Many of these OV general headwords belong to Winter's category of Lexical Items of Connection (1977:20), such as *reason, feature, fact, difference, distinction, problem, matter* and near-synonyms which 'can function as exponents of a clause relation, and as such can have a predictive effect on the organisation of written discourse' (1977:1). Thus, Winter suggests that an important functional role of a phrase, such as, *a significant difference* lies in its reference as a connector of sentences and propositions. Although this vocabulary has features of open-system meaning, OV items appear to form a fairly stable closed-set vocabulary.

[f] *The most striking influence of the P-phase on the mid-infrared spectrum of NaNO₃* is...

[g] *What is particularly striking* are...

[h] *What is remarkable* is the fact that...

[i] *A significant difference* is...

[j] *The most intriguing data seen in these experiments* was...

[k] *Further evidence* indicates that...

(iii) <u>Hypothesised Entity</u> (HE): modes of testing and carrying out research and their means of expression. As such they are hypothetical/theoretical in nature e.g. *models, approaches, formulae* and open to continued re-evaluation.

[l] *A simple model to account for this effect* assumes...

[m] *The Landau theoretical approach* may...

[n] *The Davis and Mott formula* takes the form...

(iv) <u>Empty H&O Theme</u> (EH&OTh): refer to postponed Themes indicating a less visible viewpoint as with HV and OV above. They include many formulaic patterns.

[o] *It is interesting to note* that...

[p] *It is clear* that...

[q] *It should be noted* that...

[r] *It appears* that...

[s] *It is highly likely* that...

[t] *There are two possible contributions* to this line broadening.

[u] *There was no evidence* of localized failure-inititating flaws...

D. REAL WORLD DOMAIN

(i) <u>Mental Process</u> (MP): focus on intellectual processes and entities which are part of the investigative real-world research domain; hypothesised and objectivised viewpoints are often based on the outcomes of these mental processes which are again realised through nominalised forms e.g. *calculation, comparison, analysis, determination, extrapolation, evaluation, estimate, assumption, confirmation.* Additionally, *idea, insight, question.*

[a] *A prediction* was made...

[b] *The assumption of the predominance of such collisions* is basic to [2] above.

[c] *Comparison of the d-ferrite transformation curves for 304LW, 304W and BW15* indicate that...

[d] *The idea of a discrete Cu^{II}-amine-cellulose complex* is tentatively favoured.

(ii) <u>Real World Entity</u> (RWE): typically material entities and objects concerned with the physical world.

[e] *The AlFeNi alloy system* has been reviewed by Rivlin (1980).

[f] *All the specimens* were prepared...

[g] *The Curie temperature* is almost constant...

[h] *The value of TN for ordered Ni_3Fe* is plotted at 940K...

(iii) Real World Event/Process (RWEP): actions and procedures executed in or resulting from scientific research activities.

> [i] *The preparation of the amorphous samples* has been described...

> [j] *This phase transition* has been attributed...

> [k] *Cessation of growth at some temperature between PM-FM and FM-SG transitions* should occur,..

(iv) Empty Real World Theme (ERWTh): postponed real world entities, research events/processes and reference to mental processes.

> [l] *There* is a small further *red shift* of the emission...

> [m] *It was found* that n was insensitive to this energy,..

These last two sections have presented in detail the theoretical description of marked and unmarked Themes which underlies the present exercise in Applied Discourse Analysis. We have therefore now reached the stage where we can use quantitative methods to reveal types of linguistic patterning which may otherwise remain hidden (Nesbitt & Plum 1988). The following sections present statistical analysis and interpretive commentary on the major findings of corpus analysis of Context Frames as marked Themes (3.7), unmarked Subjects (3.8) and binary combinations of Context Frames + Subjects (3.9).

3.7 ANALYSIS OF CONTEXT FRAMES (MARKED THEMES) - MAJOR FINDINGS

3.7.1 SIEs - marked vs. unmarked Themes

First of all, to give the general context for the study of CFs, TABLE 2 shows the total number of sentence-initial elements (SIEs) in the 36 texts where SIE = unmarked Theme, that is, grammatical Subject (GS), where

SIE = marked Theme, that is, Context Frame (CF) and where
SIE = neither GS nor CF (non-GS/CF) form, that is,
complement, imperative or interrogative [see APPENDIX B
for complete data]. Percentages throughout are rounded to
0.1%.

TABLE 2
Total N of Sentence-Initial Elements in the 36 NS Texts

SIE = GS	2929 (67.2%)	
SIE = CF	1408 (32.3%)	[8.2][1]
SIE = non-GS/CF	21 (0.5%)	
Total SIEs	4358 (100%)	

[1] *Standard Deviation of Context Frames*

Swales comments that 'the occasional recurrence of minor
dispreferred structures is itself of interest both
intrinsically and in terms of what it may reveal about the
rationale behind the major preferred ones' (1990a:146).
The predominant declarative statements of the RA genre
here account for the minimal usage of other non-GS/CF
dispreferred structures, that is, imperative (*Consider...
Note...*), direct questions (*How extensive are..?*),
complements and 'complementary' adjuncts with S-V
inversion (*Of obvious interest is the mechanism..., In
neither case was removal rapid.*) which make up only 0.5%
of SIEs.

3.7.2 GSs vs. CFs by RA section
In investigating the generic contraints on the
distribution of CFs throughout different stages of
discourse, the relative percentages of GSs and CFs in each
RA section are of particular interest and these are given
below.

The degree to which one section of RA discourse is more
rhetorically multifunctional in nature than another will

118

determine the higher frequency of topic shifting and hence
the desirability of signalling such shifts with CFs.

TABLE 3
Distribution of Subjects and Context Frames by RA Section

SIE =	GS	CF	nonGS/CF	[SD][1]
Int	66.1%	32.9%	1.0%	[15.3]
Exp	80.7%	19.2%	0.1%	[10.4]
Res	65.8%	33.8%	0.4%	[12.7]
Dis	60.5%	39.0%	0.5%	[13.2]

[1] *Standard Deviation of Context Frames*

This appears to be the case with 3 out of 4 sections
above where a greater usage of CFs may indicate a writer's
attempts to overtly create a more cohesive text. On the
other hand, the lowest percentage of CFs in the
Experimental section indicates the more matter-of-fact
statements of reporting scientific procedures, as Swales
(1990a) comments, understood to take more of a linear
checklist approach to contained information. As a result,
considering the rhetorical objectives of the Experimental
section, there are generally fewer topic shifts to
indicate and hence fewer examples of marked Theme.

3.7.3 CFs by function

Eiler suggests that although distribution of thematic
elements can indicate the relevance of a feature in
determining the genre, 'only when coupled with an analysis
of the discourse function and content can we draw
conclusions about heuristic value' (1986:56). With content
shaped by the rhetorical functions of each RA section, in
order to compare the use of CFs according to discourse
functions listed earlier, TABLE 4 indicates their
percentages distributed throughout the RA as a whole.

It can be seen from Figure 7 below and the data in TABLE 4
that only four functions (Contrast/concession,

Figure 7. Distribution of Context Frames by Function throughout the scientific RA
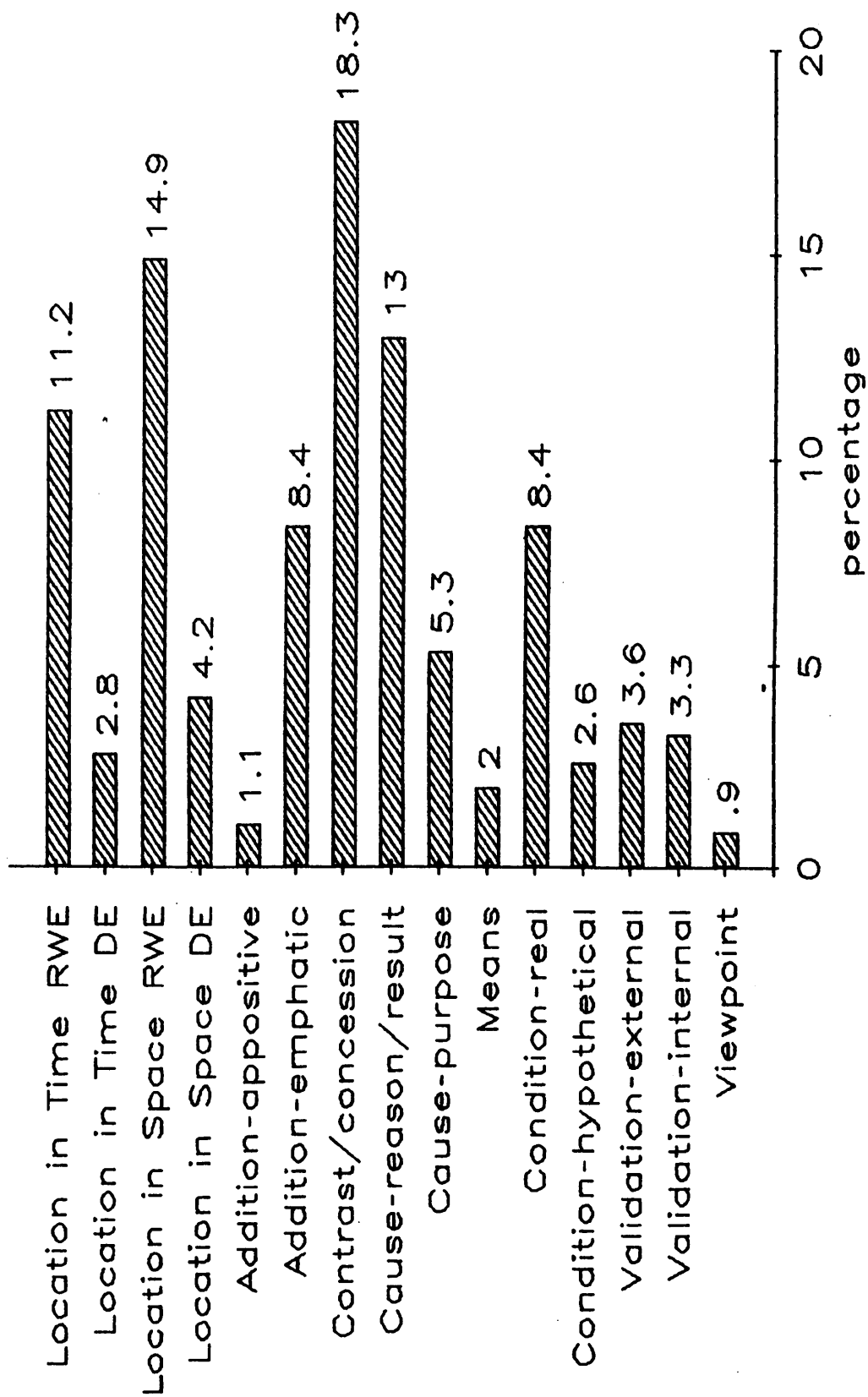
Cause-reason/result, Location in real-world Time and Space) account for 57.4% of all CFs.

TABLE 4
Distribution of Context Frames by Function throughout the RA

1a. Location in Time - RWE	11.2%
1b. - DE	2.8%
2a. Location in Space - RWE	14.9%
2b. - DE	4.2%
3a. Addition - appositive	1.1%
3b. - emphatic	8.4%
4. Contrast/concession	18.3%
5a. Cause - reason/result	13.0%
5b. . - purpose	5.3%
6. Means	2.0%
7a. Condition - real	8.4%
7b. - hypothetical	2.6%
8a. Validation - external	3.6%
8b. - internal	3.3%
9. Viewpoint	0.9%
	100%

However, the overall picture here can only show an isolated global dimension of CF usage, that is, a local context-independent one. Of greater significance are indications of local Theme choice-making as a reflection of rhetorical macrostructure and as part of the dynamic progression throughout an RA, from its introduction to its concluding discussion.

3.7.4 CFs by function and RA section

TABLE 5 shows the breakdown of percentages of Context Frames according to functional category and RA section. It should again be noted that the absence of dispreferred CF functions at different stages of RA discourse is as significant as the presence of others but comments on data in the following sections are limited to their noteworthy presence [see APPENDICES C1-3 for full data on CFs].

TABLE 5
Distribution of Context Frames by Function and RA Section

		Int N=218 (15.4%)	Exp N=177 (12.6%)	Res N=446 (31.7%)	Dis N=567 (40.3%)
1a.	Location in Time – RWE	11.5%	36.2%	8.7%	5.3%
1b.	– DE	1.4%	3.9%	1.8%	3.7%
2a.	Location in Space – RWE	12.8%	13.6%	19.1%	12.7%
2b.	–DE	8.7%	1.1%	6.5%	1.6%
3a.	Addition – appositive	2.8%	*[1]	*	1.1%
3b.	– emphatic	11.0%	3.3%	6.9%	10.4%
4.	Contrast/concession	22.0%	4.5%	18.4%	21.2%
5a.	Cause – reason/result	12.8%	11.3%	5.4%	16.0%
5b.	– purpose	4.1%	15.3%	10.1%	2.5%
6.	Means	1.4%	2.8%	1.6%	1.9%
7a.	Condition – real	4.6%	6.2%	10.5%	8.8%
7b.	– hypothetical	*	*	*	5.5%
8a.	Validation – external	6.0%	*	2.0%	4.9%
8b.	– internal	*	*	6.7%	3.0%
9.	Viewpoint	*	*	*	1.4%
		100%	100%	100%	100%

[1] *denotes value under 1.0%*

Introduction: This section focuses on locating previous research and real-world entities and processes within the temporal/spatial contexts of scientific procedures and processes; *recently/in Raman spectroscopy*. Furthermore, within the 4 sections, there is the greatest emphasis on discourse-related circumstances, most commonly through the use of *in this paper/figure*. Above all, there is the need for contrast with the current state of knowledge, hence the indication of gaps, problems and disagreements in existing research, which may be signalled in a recycled problem-solution sequence with additive emphasis on each cycle. This is first typically realised by *however* and *although* and subsequently linked with reason/result *thus/consequently*. In terms of Swales's CARS model, a writer creates a research niche and then aims to occupy it. Across sections, it would not be unexpected that frames contextualising external validation, for example,

in previous studies are most likely in the Introduction (but closely followed by the Discussion).

Experimental: The purpose of this section is to permit replicated data gathering and substantiation of experimental procedures and sequences [although Swales (1990a) indicates there is some doubt about the feasibility of this]. Hence there is an overwhelming focus on the description of real-world temporal sequences e.g. *after removal of IFNs* and, to a lesser extent, real-world spatial CFs. Cause CFs feature most strongly across the 4 sections, namely purpose e.g. *in order to obtain HL cells*, *for evaluation of the gap* and reason/result e.g. *thus*, *since M interferes with the protein.*

Results: CFs here renew focus on relocating experimental procedures and sequences, in this case emphasis is on real-world spatial rather than temporal Location. But in order to move this more discursive RA component forward dynamically from the Experimental stage, markers of Contrast/concession *however, by contrast* serve to contextualise a variety of situations; results contrasting with previous data; unexpected findings; technical problems and possible solutions which act as future gaps in research. This is combined with continued explanation of Cause - purpose behind procedures e.g. *in order to observe clear profiles in a micrograph.* From the repeated cycles of these CFs in Int, Res and Dis sections, it can be seen that such sequences are textual threads which run throughout RAs, not just as opening gambits for research justification and creating research space. Swales comments that 'research questions or unexplained phenomena are the life-blood of many research undertakings' (1990a:140) and the constant addressing and reviewing of such problems in their reporting are reflected in the marked thematic status of CFs highlighting these functions. This cycling is emphasised by the presence of clusters of CFs throughout the texts; 57% of all CFs occur

as consecutive sentence-initial pairs (or more); 32% in groups of threes; 17% in fours; 7% in fives; 3% in groups of six with 1% as seven consecutive CFs.

Two further points are worth noting. In Results, CFs are used to pick up on experimental procedures by means of resultant real (rather than hypothetical) Conditions of cause and effect, for example: *when the luminescent state decays through two channels,*. Also, Results CFs not surprisingly give the clearest indication of discourse internal Validation in an attempt to make a stronger case for evidence with reference to graphic data, for example: *as can be seen from Figure 1,*. Such marked thematic phrases seemingly become formulaic patterns.

Discussion: This section accounts for 40.3% of all CFs (Results and Discussion taken together account for 72.0%), which emphasises the greatest need here for skilled rhetorical manipulation of discourse. Research questions and objectives posed in the Introduction are commonly addressed in the Discussion, confirming the hour-glass symmetry of RA discourse suggested by Hill *et al.* (1982). Despite the limited evidence on which their suggestion is based (that is, one paper in psychology), the discourse symmetry of general context to particular and then back to general is reflected in marked thematic choices. Hence, as in the Introduction, we again find sequences of 3 CF functions indicating Contrast/concession + emphatic Addition + Cause-reason/result. Here, they account for 47.6% of this section. In emphasis of this, the highest totals of a single functional category of CF occur in this section, namely Contrast/concession with 120 examples (including 51 of *however*, the highest single count). This is followed by Cause-reason/result with 91 examples.

In parallel with the use of real Condition CFs in this (and the previous) section, with less clear data or where a researcher feels it is premature to judge data without

124

attracting damaging peer criticism, a more speculative
approach is taken. More indirect opinions and conditions
are presented as more hypothetical: *if we postulate that a
guest molecule is fully embraced in the CyD cavity,*. Such
choices contribute to the delicate balance and tension
intrinsic in scientific reporting as indicated by Myers:
'the writer must stay within a consensus to have anything
to say to members of his or her discipline, but must also
have a new claim to make to justify publication' (1989:5).

3.7.5 Thematic flow of CFs

How may the progressive textual flow through scientific RA
discourse of elements of marked Theme be summarised?
Figure 8 below attempts to visualise this progression, in
this case by comparing four CF functions, cycles of which
figured strongly in the above analysis: (3b)
Addition-emphatic; (4) Contrast/concession; (5a)
Cause-reason/result; (5b) Cause-purpose. From Figure 8, it
is clear that (5b) Cause-purpose has the opposite flow
throughout the RA compared to the other three functions,
with a low starting point and a steadily declining effort
from its key role in the Experimental section. Conversely,
as scientific RA discourse progresses with changing
rhetorical goals, CFs expressing positive (3b) and
negative (4) expansion and elaboration assist the
justification of opinions and hypotheses and actions
through (5a) Cause-reason/result. [Note: the line chart of
Figure 8 does not intend to suggest a graded flow from the
end of one RA section to the beginning of the next, but
simply indicates the direction of flow based on the four
points of measurement].

Figure 8. Dynamic Flow of Context Frames in RA Discourse

Addition-emphatic
Contrast/concession
Cause-reason/result
Cause-purpose
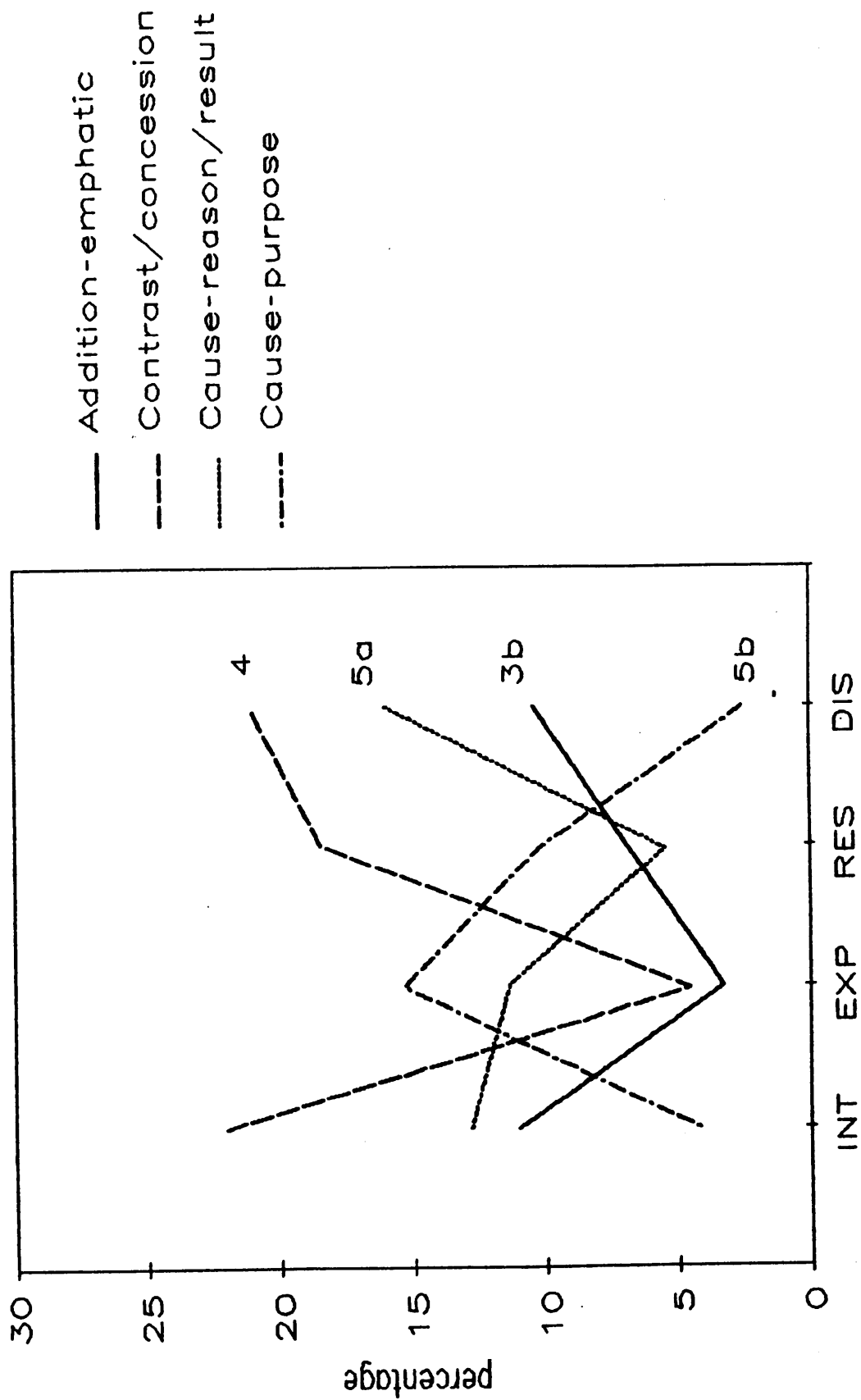
Of course, the perceived flow is dependent on the arrangement of RA sections as in the corpus. Some scientific academic journals in some disciplines may downgrade the Experimental section by publishing it after Results and Discussion. Even though RA sections may be written and read in a different order compared to their presentation, their final published sequence can be seen as an indication of the perceived textual flow of the genre.

3.7.6 CFs by function and type

It is not assumed that the 3 CF types presented above represent a continuum of grammatical/syntactic complexity. Nevertheless, some researchers (for example, Johns (1984)] have commented that CF Type 1 'minimals' may indeed be overused by NNSs and Rutherford suggests that apprentice writers have a tendency to rely more heavily on coordination structure, 'with a gradual shift, as learning progresses, towards the degree of subordination characteristic of the target language' (1987:51). However, classification of some CFs according to grammatical/syntactic structure may be deceptive. Working with apprentice NNS researchers in Japan (Gosden 1991, 1992a), it was noted that the ability to pick out 'smart expressions' from published NS 'model' texts was considered a valuable skill; St.John (1987) has also commented on a similar procedure as part of the RA writing process. [This issue of aspects of NNS novices' writing procedures and strategies will be developed in greater detail in section 5.2.2]. However, set phrases such as, *As shown in Figure 1*, *Given that x=y*, the kind of pattern that novice NNS writers find it valuable to 'borrow', are coded here as CF Type 3 – seemingly more 'complex', but not equal to the structural complexity involved in the choice of non-formulaic clauses beginning *however* vs. *although*.

Nevertheless, as we are interested in the overall potential that thematic analysis may have for the development of RA writing skills, a picture of CF usage according to both functional and grammatical paradigms may be useful and TABLE 6 presents a summary of this.

CF Type 1: Conjunctive/Modal Adjuncts and Conjunctions:

As anticipated from previous comments, three categories make up 71.2% of all Type 1 CFs: Contrast/concession (+ emphasis) + Cause - reason/result accounting for 'minimals' such as *however* (+ *in addition, furthermore, moreover*) + *thus*.

TABLE 6
CF Usage according to Functional and Grammatical Paradigms

		Type 1 N=528	Type 2 N=535	Type 3 N=345
1a.	Location in Time - RWE	7.0%	17.9%	7.2%
1b.	- DE	6.4%	*[1]	1.2%
2a.	Location in Space - RWE	1.1%	37.6%	*
2b.	- DE	*	10.8%	*
3a.	Addition - appositive	3.0%	*	*
3b.	- emphatic	18.9%	3.4%	*
4.	Contrast/concession	31.1%	3.9%	21.1%
5a.	Cause - result/reason	22.2%	2.2%	15.7%
5b.	- purpose	*	4.7%	13.9%
6.	Means	*	*	6.4%
7a.	Condition - real	1.7%	11.8%	13.3%
7b.	- hypothetical	1.3%	*	8.1%
8a.	Validation - external	2.3%	5.6%	2.6%
8b.	- internal	1.7%	1.1%	9.3%
9.	Viewpoint	2.3%	*	*
		100%	100%	100%

[1] * *denotes value under 1.0%*

CF Type 2: Prepositional and Adverbial Phrases:

Naturally, given the structure of these adverbial/prepositional clauses, the majority (55.5%) occur in temporal and particularly real-world spatial functions - *after removal of IFNs, on a glass substrate.*

128

Also noteworthy are CFs indicating real Conditions where set prepositional phrases, such as: *in the case of x=0.92, in the presence of 2m-GdmHCl, in the absence of these compounds*, predominate.

CF Type 3: Subordinate Clauses and Non-finite Clauses:
Here CFs mirror the functional usage of Type 1 CFs i.e. in the expression of Contrast/concession *although* + Cause-reason/result *since*, and specifically in indicating Cause-purpose *in order to,* which are by far the most common realisations. Fronted subordinate clauses indicating both real and hypothetical Conditions, *when fk=1 is applied, if C^{s+} works as CsOH*, also feature.

3.8 ANALYSIS OF SUBJECTS (UNMARKED THEMES) – MAJOR FINDINGS

3.8.1 Distribution of Subject roles
For details of the total N of Subjects in the corpus, refer to TABLE 2 above. The ranked distribution by domain of all Subjects (N=4337) is indicated in TABLE 7 below. Superscript letters refer to comments on details of interest which follow.

The empirical content of scientific research and its reporting account for the dominance of the Real World domain, specifically Real World Entities[a] with by far the highest percentage of 56.1% and Real World Events/Processes[b] with 17.9%, accounting for approximately three-quarters of all Subjects. Of the remaining quarter, Discourse Participants[c] (5.7%) and the realisation of H&O Viewpoint through empty Theme choices[d] (4.4%) are worth noting. The overwhelming dominance of the Real World domain in the scientific RA genre is taken as self-evident. However, the issues of how writers seek to explain and interpret facts and theories

and thereby present themselves and their viewpoints are of interest. The three non-Real World domains represent key elements of a writer's potential to create a balance between interactional and topic-based Themes and their usage is therefore highly significant. The following comments pay particular attention to these components.

TABLE 7
Distribution of Subject Roles by Domain

Subject Role Domain	Mean %
Participant domain	
Discourse Participant (DP)	5.7%[c]
Interactive Participant (IP)	2.9%
Participant Viewpoint (PV)	0.6%
Total this section	9.2%
Discourse domain	
Micro Discourse Entity (MiDE)	2.1%
Interactive Discourse Entity (IDE)	1.6%
Empty Discourse Theme (EDTh)	1.1%
Discourse Event/Process (DEP)	1.0%
Macro Discourse Entity (MaDE)	0.2%
Total this section	6.0%
H&O domain	
Empty H&O Theme (EH&OTh)	4.4%[d]
Objectivised Viewpoint (OV)	2.0%
Hypothesised Entity (HE)	0.6%
Hypothesised Viewpoint (HV)	0.6%
Total this section	7.6%
Real World domain	
Real World Entity (RWE)	56.1%[a]
Real World Event/Process (RWEP)	17.9%[b]
Empty Real World Theme (ERWTh)	2.0%
Mental Process (MP)	1.2%
Total this section	77.2%
	100%

3.8.2 Distribution of Subjects by RA section

A central hypothesis of the present work is that thematic choices are guided by the progressively changing rhetorical goals of RA discourse as it proceeds from the introduction to the concluding discussion. Therefore, TABLE 8 breaks down the percentage distribution of GS roles by RA section. This allows for comparison of the relative distributions of Subject roles both across and within each section.

TABLE 8
Distribution of Subjects by Domain and RA Section

	Int	Exp	Res	Dis
Participant Domain				
DP	7.8%[a]	1.7%	4.1%	8.6%[e]
PV	*[1]	*	*	1.4%
IP	8.8%[b]	*	*	3.9%
Total this section	16.9%	2.4%	4.5%	13.9%[c]
Discourse Domain				
DEP	1.7%	*	*	1.6%
MaDE	*	*	*	*
MiDE	*	*	5.3%[a]	1.7%
IDE	5.6%[c]	*	*	1.7%
empty Theme	1.4%	*	1.0%	1.6%
Total this section	9.9%	0.6%	7.1%	6.9%[f]
H&O Domain				
OV	2.6%	*	1.8%	2.8%
HV	*	*	*	1.6%
HE	*	*	*	1.6%
empty Theme	3.5%	*	3.6%	7.6%[d]
Total this section	6.8%	1.0%	6.1%[b]	13.6%[b]
Real World Domain				
MP	1.7%	*	*	2.4%
RWE	40.0%[d]	85.9%[a]	58.2%	43.1%
RWEP	21.7%	9.8%[b]	20.7%	18.2%
empty Theme	3.0%	*	2.8%	1.9%
Total this section	66.4%	96.0%	82.3%	65.6%[a]
	100%	100%	100%	100%

[1] * *denotes a value below 1.0%*

131

Introduction: across sections, Participant domain roles feature here most strongly, particularly Discourse Participant[a] and Interactive Participant[b] with cited researchers. The same is true for the Discourse domain; moreover, the presence of Interactive Discourse Entities[c] emphasises the task of the RA Introduction as one of interactive Public Relations between the individual writer(s) and the arbitrating academic community, emphasising what Tarantino calls the 'dialogic nature of scientific discourse' (1991:51). Thus, the multifunctional nature of Introductions implies a reduced focus on the Real World domain, particularly Real World Entities[d] - of course, at 40.0% by far the highest single count in the Introduction, but across RA sections, the lowest. With the integration of bottom-up linguistic and top-down rhetorical information, introductory moves suggested by Swales (1981, 1990a) can be mapped onto the three non-Real World domains in RA discourse. The function of the Participant and Discourse domains is to review previous research (IP/IDE) and announce present research (DP), or in Swales's (1990a) CARS model, Establish a Territory and Occupy a Research Niche; the H&O domain assists in the indication of gaps in previous research, in the establishment of a research niche.

Experimental: as expected from the rhetorical functions of this RA section, Real World domain roles predominate, in particular Real World Entities[a] with 85.9%, indicating the checklist nature of this section. In view of the fact that the function of this section is generally thought to be to permit the replication of experimental procedures, it is of note that Real World Event/Process[b] Themes are here at their lowest percentage (9.8%) across all sections.

Results: rhetorical effort begins to rise with an increase in the presence of the three non-Real World domain Subject roles, notably Micro Discourse Entities[a],

highlighting the internal validation of results through graphs, tables etc. A further increase in writer visibility and viewpoint comes through H&O discussion of research outcomes[b], particularly empty Theme choices, for example, formulaic patterns *it seems that..., it appears that... .*

Discussion: the Introduction<->Discussion symmetry of RA discourse is completed with declining Real World focus - in fact the lowest across sections at 65.6%[a]. There is a concomitant increase in other domains - H&O Viewpoint[b] is highest across RA sections at 13.6%, with Participant domain[c] at 13.9%. It is interesting that across RA sections, but particularly in Discussion, H&O domain is primarily expressed through empty Theme[d] roles, notably, *it is clear that..., it is interesting to note that... .* This, allied with the highest appearance of highly visible Discourse Participant[e] Subjects (8.6%), gives a powerful boost to interactional writer viewpoint in the Discussion. From the previous section, Discourse domain[f] focus remains stable.

Empty Theme choices across RA sections: Of 316 examples, 249 (78.8%) were *it* Empty Themes with 67 (21.2%) examples of *there* structures. By Subject domain, H&O made up 58.2% of all Empty Themes with 14.3% under the Discourse domain. From comments above, it is clear that Empty Themes are a valuable thematic resource and data confirm Davies's suggestion (1991) that their primary role is to project evaluative material, in this case, across the two domains, H&O and Discourse. In the present corpus, there were no Participant domain examples (such as *There are authors who...* or *It is our argument that...*). Real World Empty Themes accounted for 27.5% of all Empty Themes and represent 'bare' ontological statements which simply postulate the existence of some entity or entities (Quirk *et al.* 1985:1406), for example, *there is a shift..., it was not possible to derive an equation..* (NB: instances of

speculation by means of: *it is possible that their effects..* are classified here as Empty H&O Theme). Of the 316 examples, there were just two (Real World) instances of what Halliday refers to as predicated Themes (1985a:59), known in traditional grammar as cleft sentences (Quirk *et al.* 1985:1384): *it is the conservation of distinct isotope classes that suggests.., it is during this time that BBMI assumes...*

Amongst Empty H&O Themes, 37% of all examples (68/184) were variations on the following:
It is evident... It appears....... There is evidence......
It is clear..... It seems........ There is no evidence...
It is unclear... It would appear.. There was evidence.....
It is apparent.. It would seem.... There was no evidence..

These examples were in fact the largest single related function in all Empty Theme domains, with often just one or two examples of most other Empty Themes. The tense variation indicated in the examples above was also evident across all examples: 64.5% were timeless present: *it is important, there are no data*; 12.0% simple past: *it was of interest, there was an absence*; 9.2% present perfect passive: *it has been demonstrated, there have been no studies*; 13.0% modal: *it can be argued*; 1.3% will form: *it will be of interest*. However, there was some variation in these factors between the three domains:

	present	past	pres.perf.	modal	will	
Real World	58%	28%	14%	0%	0%	= 100%
H&O	71%	7%	2%	19%	1%	= 100%
Discourse	49%	2%	29%	16%	4%	= 100%

3.8.3 Thematic flow of Subjects

To conclude these comments, by using the data in TABLE 8, the resultant flow of Subjects as unmarked Theme choices from one RA section to the next can be visualised. In Figure 9 below, the data by section show a 'reverse hour-glass' pattern. (Note: for this figure, the data in TABLE 8 are read vertically, that is, Introduction section = 100%). Figure 9 clearly indicates the relative dominance of topic-based thematisation with Real World Themes and the concommitant reverse trend of the other three domains. This flow confirms Swales's (1990a) suggestion of a differential distribution across the four sections of linguistic and rhetorical features that divide the RA into 'simple' EXP and RES and 'complex' INT and DIS, although this appears to underplay the intermediate complexity of Results here.

In Figure 10 below, from the data by domain, we get a clearer picture of the decreasing and increasing scale of rhetorical effort involved through each section for each domain. (Note: for this figure, the data in TABLE 8 are read horizontally, that is, Participant domain = 100%). It shows particularly clearly that, by means of a greater number of Participant and H&O domain Subjects, which realise more highly interactional, writer-visible Themes, the RA gathers increasing rhetorical momentum towards its concluding discussion.

Figure 9. Dynamic Flow of Subject Role Domains by RA section

A: Participant
B: Discourse
C: H&O
D: Real World
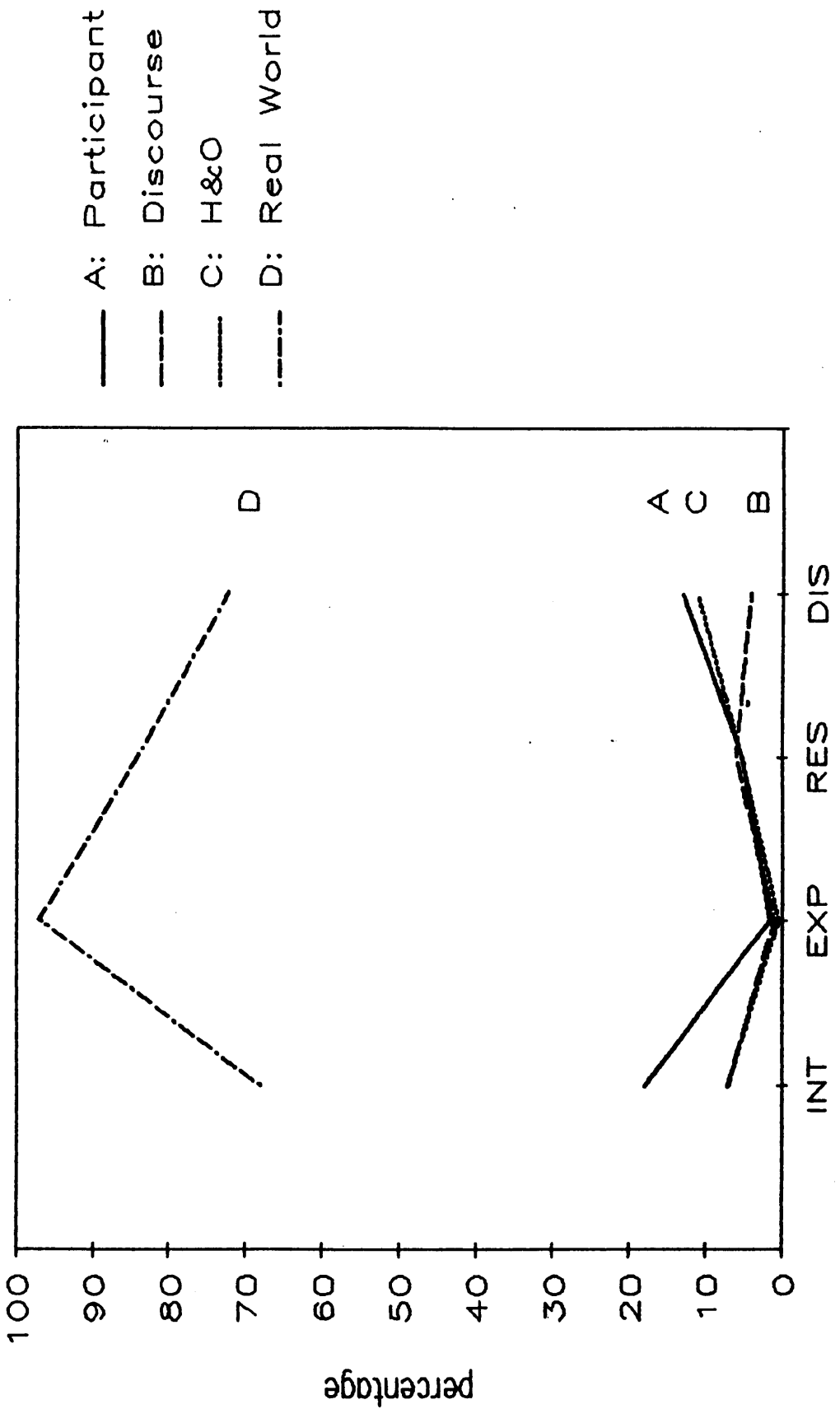
Figure 10. Dynamic Flow of Subject Roles by Domain

A: Participant
B: Discourse
C: H&O
D: Real World
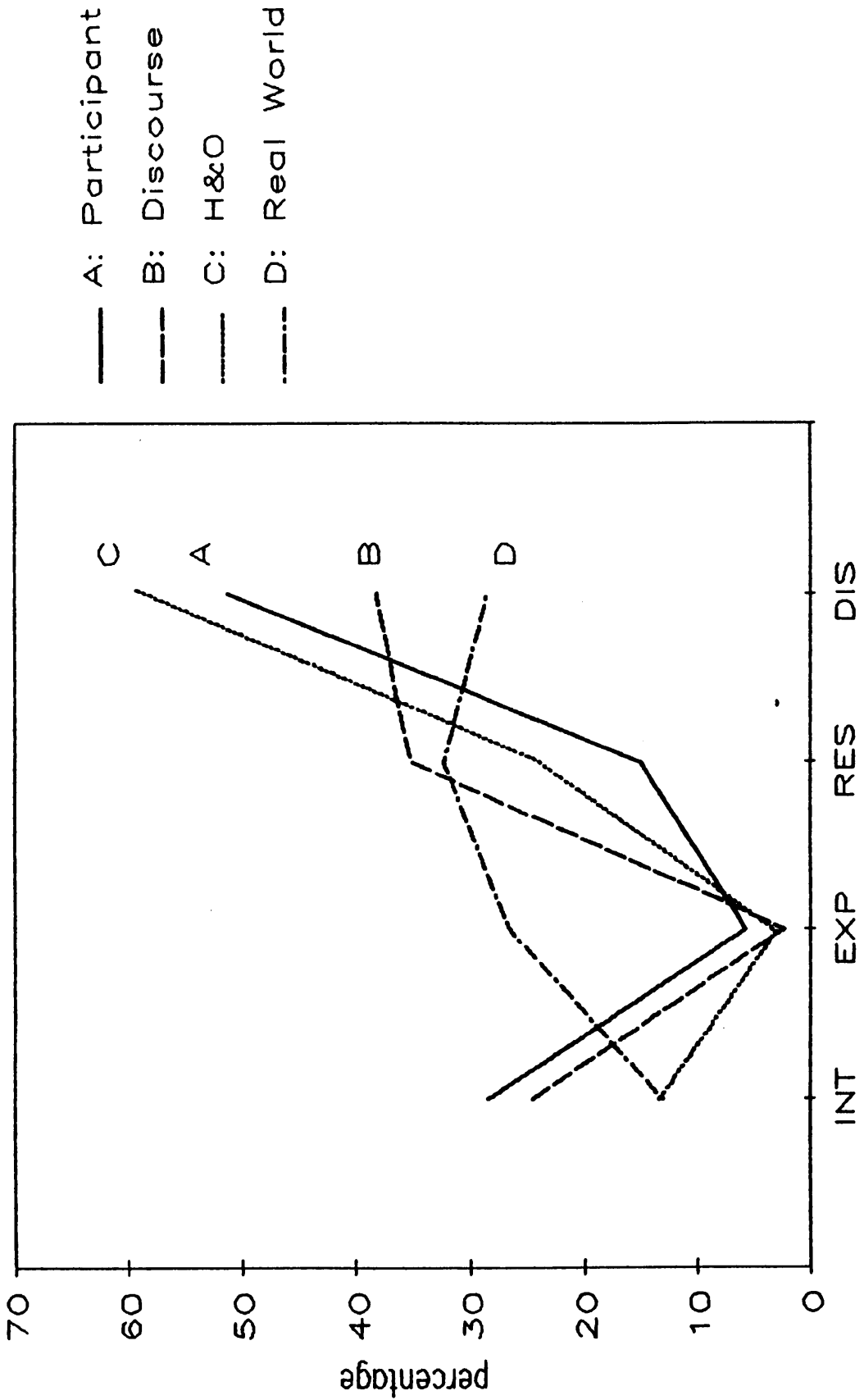
3.9 BINARY THEME CHOICES - CONTEXT FRAMES + SUBJECTS

3.9.1 Which CFs go with which Subjects?

In the corpus, 1408 out of 4337 (32.5%) Subjects were preceded by Context Frames. As has become clear from previous discussion, CFs evidently perform a variety of functions at different stages of RA discourse, signalling changes in real-world and discourse circumstances. Of interest is whether binary Themes are evenly distributed across Subject domains or whether certain Subject roles tend to combine with certain CF roles at specific stages of RA discourse. TABLE 9 shows the distribution of binary Themes for each Subject domain, followed by illustrative examples.

TABLE 9
Binary Theme Choices - Context Frames + Subjects (Total N=1408)

	SIE=CF+GS	SIEs		%
Participant Domain	188	393	=	48.7%[d]
Discourse Domain	60	265	=	22.6%[e]
H&O Domain	105	330	=	31.8%[b]
Real World Domain	1055	3349	=	31.5%[c]
TOTAL N	1408	4337	=	32.5%[a]

The overall average percentage for Subjects that are preceded by Context Frames is 32.5%[a]. Subjects belonging to two domains, H&O[b] and Real World[c], reflect this average. However, an above average number of Participant domain Subjects[d] are preceded by CFs. The reason for this is, I suggest, to 'soften' heightened writer visibility and to indicate that Participant domain Themes often represent changes in discourse moves which require contextualised signalling, for example, Swales's (1981) Move Four - Introducing Present Research. Conversely, the lower percentage of Discourse domain roles[e] implies that a focus on these Subject roles e.g. IDE in Introduction and MiDE in Results, requires or coincides with fewer

marked signals. Comments on and examples of the most common Context Frame + Subject collocations are given below.

PARTICIPANT DOMAIN: Subjects are quite widely distributed across CF categories. Of note, however, are combined examples with Cause and Location in Discourse and Real World Time and Space. It is perhaps interesting to note that the most powerfully obvious signalling of a writer's Viewpoint (CF category 9) is almost never combined with the overt presence of the writer-as-participant Subject (an example of this would be: *Interestingly, we..*).

Cause - result/reason + Discourse Participant

Accordingly, we have also examined a specimen of composition Ni_3FeMn.

Cause - purpose + Discourse Participant

To extend the investigation further, we have taken spectra...

Location in Time (Discourse Entity) + Discourse Participant

Finally, we turn to the intensities observed for varying substrate...

Location in Space (Discourse Entity) + Discourse Participant

In this paper, we describe the Al substituted spectra in detail...

Location in Time (Real World Entity) + Interactive Participant

While studying the absorption edge in silver bromide, Urbach[1] discovered that...

Location in Space (Real World Entity) + Interactive Participant

In their molecular dynamical simulations, Lynden-Bell et al. (1990) observed high-amplitude fluctuations at the F point..

DISCOURSE DOMAIN: of the few examples in this category, combination with Contrast/concession is of note.

Contrast/concession + Interactive Discourse Entity

Although there have been a number of indirect studies[6-11] of the CH_3O + NO_2 reaction, the only previous direct studies[12] have been in flow-discharge systems...

H&O DOMAIN: two functions, Contrast/concession and Cause - result/reason, predominate. Contrast CFs again function to create the dynamic tension necessary to promote the dialogic nature of scientific reporting. The H&O Subjects are therefore strongly contextualised through the justification of 'gap-filling' and Cause - result/reason.

Contrast/concession + empty H&O Theme
> *However, it is likely* that other components are required...

Cause - result/reason + empty H&O Theme
> *In pursuit of a concept of modulation of control mechanisms, it should be noted* that spermatozoa have no significant protein synthetic ability.

REAL WORLD DOMAIN: examples are widely distributed amongst this most populated domain. However, the highest totals are combined Location in Real World Time and Space; Contrast/concession; Cause - reason/result; Condition - real.

Location in Time (Real World Entity) + Real World Entity
> *At fertilization, the spermatozoon* releases its acrosomal contents in a Ca^{2+} dependent exocytotic process...

Location in Space (Real World Entity) + Real World Entity
> *In higher eukaryotes, tubulin* is encoded by multiple tubulin genes...

Contrast/concession + Real World Event/Process
> *However, the gradual assembly of the brush border that occurs over the course of three weeks in the embryonic chick gut* cannot fulfill...

Cause - result/reason + Real World Entity
> *Consequently, the 110K-CM complex* has been named BB myosin I...

Condition - real + Real World Entity
> *When the cephalic region chondrocytes were cultured in the presence of 2mM levamisole, the level of intracellular alkaline phosphatase activity* was reduced...

With this presentation of analysis of binary Themes, we
conclude the section on major findings of thematic
analysis.

3.10 MAPPING THEMES ONTO MOVES

3.10.1 Introduction

Previous sections of Chapter Three have presented a
detailed description of one element of 'textualness'
(Ellis 1987), as viewed through Halliday's system of
Theme. We recall also from Chapter Two (see Figure 3
above) that *moves* as generic elements of structure are
said to determine the configuration of meanings typically
associated with the register variables of Field, Tenor and
Mode, each being reflected in the three semantic
metafunctions: ideational, interpersonal and textual,
respectively. These in turn, realise the clusters of
systems and structures, transitivity, mood and theme in
the lexicogrammatical stratum.

In an attempt to integrate these bottom-up and top-down
levels of linguistic and extralinguistic description, the
next two sections focus on mapping the available range of
Themes onto generic moves (already indicated above in
Figure 4) in order to characterise one element of the
dynamic texture of the scientific RA genre.

3.10.2 Rhetorical move structure in scientific RAs

In the Introduction map (Figure 11 below), the three moves
are taken from Swales (1990a), which represent a
modification of the earlier (1981) Four Moves schema. The
first of three obligatory moves in Swales's (1990a) CARS
(Create A Research Space) model – Establishing a Territory
– now subsumes the earlier first two moves – Establishing
the Field and Summarising Previous Research – due to the

difficulty in separating these in analysis with the increasing practice of spreading references throughout the Introduction section. The second and third CARS moves aim to Establish and Occupy the Niche or gap in research.

In Experimental (Figure 12 below), the combined obligatory move of Materials and Procedures predominantly found throughout the present corpus is confirmed by the Methods section schema suggested by Weissberg & Buker (1990). However, they also indicate the potential presence of the following [optional] moves: [Overview of the Experiment], [Population/Sample], [Restrictions or Limiting Conditions], [Sampling Technique], [Statistical Treatment]; these optional moves may be more relevant to other research fields, but were not considered significant here.

Swales comments that the final two sections, Results and Discussion, generally present discourse analysts with greater difficulties since they may be coalesced or have additional or substituted sections such as Conclusions: 'more particularly, there is.. much variation in the extent to which Results sections simply describe results and the extent to which Discussion sections redescribe results' (1990a:170).

For the present purposes of mapping Themes onto Results moves (Figure 13 below), I have followed the schema suggested by Weissberg & Buker (1990): optional [Statement of Location of Results]; obligatory Statement of Results, with the integration of Results (R) + Comments (C), which may focus on the following three functions: generalisations about Results; explanations of Results; comparison with previous Results. The two possible patterns of ordering Results and Comments statements suggested by Weissberg & Buker - alternating (R1+R2+R3+C) and sequential (R1+C1, R2+C2, R3+C3) - are indicated in Figure 13.

As with Results, the suggested <u>Discussion</u> schematic structure (Figure 14 below) emphasises the potentially cyclical nature of inherent rhetorical moves. From a summary of the literature, Swales offers a distilled list of 8 more frequently found moves in this section: Background Information, Statement of Results, (Un)expected Outcome, Reference to Previous Research, Explanation, Exemplification, Deduction and Hypothesis, Recommendation. Thus,

'in strict contrast to Introductions, [Discussions] move during a cycle in an "inside-out" direction; they move from stating results themselves, to placing them within the established literature, to reviewing their general significance' (Swales 1990a:173).

The schema indicated in the Discussion map below - obligatory Statement of Most Important Findings [Move D1]; Comparison with Previous Research [Move D2a] + Explanations and Speculations for Results [Move D2b]; optional [Limitations of Present Research + Recommendations for Future Research] [Move D3] - thus represents a synthesis in the light of analysis of the present RA corpus.

The move patterning of combined Results & Discussion sections have not been considered here, although they will feature in NNS novices' RA drafts to be discussed at a later stage (section 5.2). It is, however, reasonable to assume that they will be characterised by components of the two individual sections, in particular by cycles of Results + Comments/Discussion indicated in Figures 13 and 14, which set the scene for eventual Conclusions.

143

CONTEXT FRAMES	INTRODUCTION MOVES	SUBJECTS
LOCATION IN TIME - RWE Recently, considerable effort................ LOCATION IN SPACE - RWE In such a device, the active electrode........ In chicken enzymes, the pre-sequences.......... ADDITION - EMPHATIC In particular, studies by Corbett[13-15]......... CAUSE - REASON/RESULT Thus, the repair of the elements................	ESTABLISHING A TERRITORY [MOVES 1 & 2]	REAL WORLD DOMAIN Stainless steels........ The lattice characteristics.......... It is well known that noble metals........... Crystalline sodium nitrate................ PARTICIPANT DOMAIN Urbach [1]............. Some researchers........ DISCOURSE DOMAIN Studies by Farrar[3].... Recent work in this laboratory[19-28].........
CONTRAST/CONCESSION Nevertheless, a major limitation.............. However, Smith and Tauc[14]............... Although relatively few studies of Cu[II] have been reported, several workers................. In spite of all this recent work, there is no consensus............... CAUSE - REASON/RESULT Consequently, there is little understanding.... Thus, the ratio........	ESTABLISHING A NICHE [MOVE 3]	REAL WORLD DOMAIN The role of type x collagen............... The mechanism.......... A full understanding of the cell biology........ These differences....... No recent observations.. PARTICIPANT DOMAIN DaSilva & Johnson[4].... None of those authors... DISCOURSE DOMAIN Much less attention..... Little.................
LOCATION IN SPACE - DE In this paper, we....... In this report, we...... In the work reported here, spectroscopy......	OCCUPYING THE NICHE [MOVE 4]	PARTICIPANT DOMAIN We.................... DISCOURSE DOMAIN The purpose of the present paper.......... Preliminary work from this laboratory[5]...... H&O DOMAIN It is clearly of value..

Figure 11.
INTRODUCTION
Theme-Move map

CONTEXT FRAMES	EXPERIMENTAL MOVES	SUBJECTS

CONTEXT FRAMES

LOCATION IN TIME - RWE
After rolling and machining, a narrow strip 70mm long........
Before experiments were begun, the medium.......
After 15 min, aliquots of the medium..........
At the end of the incubation period, the supernatants...........
Finally, an Mro I-Bgl II chimera................

LOCATION IN SPACE - RWE
On a glass substrate, a 200nm thick film........
In a second beam path, a discharge cell.........
In some cultures, calcium...............

CAUSE - PURPOSE
In order to see the effect, the substrate temperature............
For microstructural analysis, the films.....
To reduce the heat input to the sample, the measurement ....

CAUSE - REASON/RESULT
Since the new high-Tc superconductors are ceramics, electrical contact................
In view of these difficulties, we........

EXPERIMENTAL MOVES

MATERIALS

+
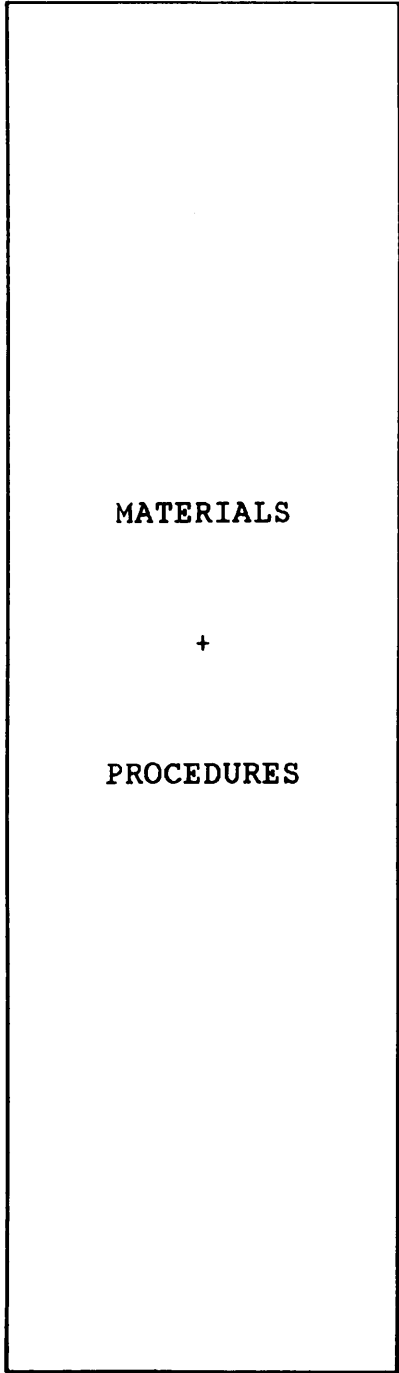
PROCEDURES

SUBJECTS

REAL WORLD DOMAIN
All the specimens.......

The foils and the Al powder.................

The sample of Au-14at %Fe...................

The preparation of the amorphous samples.......

Chromaffin cells........

p36.....................

The reaction...........

The source of cellulose.

Amines.................

The apparatus and procedures used in the present experiment......

The gases (B.O.C. methane, >99.99%; air, >99.9%; oxygen, >99.9%; nitrogen, >99.99%)......

Post-mortem human liver.

Both antibodies to villin.................

Similar concentrations of IgGs................

Aliquots of the solutions and samples of the c-a Cu^{II}...........

Figure 12.
EXPERIMENTAL
Theme-Move map

145

CONTEXT FRAMES	RESULTS MOVES	SUBJECTS

CONTEXT FRAMES   **RESULTS MOVES**   **SUBJECTS**

INTERNAL VALIDATION
As can be seen in Fig.7, the stepped face.......

LOCATION IN SPACE - DE
In Figure 1, the upper full curve.............

[STATEMENT of LOCATION of RESULTS]

DISCOURSE DOMAIN
Figure 7a..............
The broken curve in Figure 4...............
Table 1a................
Figures 3(a) and 3(b)...

LOCATION IN SPACE - RWE
In weld metal 304WL, these carbide films.....
In similar experiments, however, no difference..
At the interpass boundaries, reheating...

CONTRAST/CONCESSION
However, at above 450K, the phonon signal.......
Although the differences were small, the data....
In contrast, most of the MP-based fibres........

CONDITION - REAL
In the case of $x=0.95$, changeover..............
When sections were stained with TB, the cells...................

CAUSE - PURPOSE
In order to gain an insight into the band edges, the data........

LOCATION IN TIME - RWE
After longer aging times, these changes....

ADDITION - EMPHATIC
In addition, the relationship............

CAUSE - REASON/RESULT
Thus, the effect........

STATEMENT of IMPORTANT RESULTS

+

COMMENTS on RESULTS

generalisation
explanation
comparison

Pattern of Results (R) and Comments (C)

Alternating:
R1 + R2 + R3 + C
Sequential:
R1 + C1
R2 + C2
R3 + C3

REAL WORLD DOMAIN
The capacities of all five amines............
The value of Tn........
ESR and optical measurements...........
A large increase........
The temperature and pressure ranges........
The purity of p36.......
These results..........
The rate of C_2 formation..............
The v_2 absorption.......
These latter data.......
There was interlocking..
It is found that........

H&O DOMAIN
It is clear that........
The most striking feature of the gel spectrum...............
It should be obvious that...................
The most surprising observation.............
It is apparent that.....

PARTICIPANT DOMAIN
We.....................
Drijver and co-workers..

DISCOURSE DOMAIN
A similar explanation...

Figure 13.
RESULTS
Theme-Move map

CONTEXT FRAMES	DISCUSSION MOVES	SUBJECTS
LOCATION IN SPACE - RWE In pure NiFe, the chemical order.........	STATEMENT of MOST IMPORTANT FINDINGS [MOVE D1]	REAL WORLD DOMAIN The phase transition in $NaNO_3$................... Evaporated SiO_x thin films................... The results of this study...................
CONDITION - REAL When an epitaxial film is grown, dislocations..		
EXTERNAL VALIDATION In previous CdTe studies, a corresponding conclusion..............		REAL WORLD DOMAIN The result of such a reduction............... The absence of the X phase................... Such narrow spectral lines...................
CONTRAST/CONCESSION On the other hand, Kimura[15]..............	COMPARISON with PREVIOUS RESEARCH [MOVE D2a]	
However, it appears that production..............		PARTICIPANT DOMAIN Lynden-Bell (1989)...... We..................... Previous workers........
CAUSE - REASON/RESULT Since n is an integral product, this..........	+	
CONDITION - HYPOTHETICAL If the assumption is correct, these measurements...........	EXPLANATIONS and SPECULATIONS for RESULTS [MOVE D2b]	H&O DOMAIN The most probable cause. It is likely that....... There is some limited evidence that.......... It would appear that.... It is interesting to note that..............
ADDITION - EMPHATIC Moreover, this second reaction...............		DISCOURSE DOMAIN It can be argued that...
CONTRAST/CONCESSION Despite the shortcomings of eq.1, the existence of two contributors..... Though the LIF method was not precise, the experiments............. However, a complete investigation..........	[LIMITATIONS of PRESENT RESEARCH] + [RECOMMENDATIONS for FUTURE RESEARCH] [MOVE D3]	REAL WORLD DOMAIN Experiments to substantiate this....... The low value.......... The weak cordinating power...................
		PARTICIPANT DOMAIN We.....................
CAUSE - REASON/RESULT Therefore, a full understanding..........		DISCOURSE DOMAIN No explanation..........

Figure 14.
DISCUSSION
Theme-Move map

3.10.3 RA section Theme-Move maps

By means of the preceding analysis and description, it is possible to suggest an approximate mapping of the linguistic component of bottom-up Theme onto the top-down rhetorical Moves in the four major sections of the scientific RA. These Theme-Move maps (Figures 11-14 above) represent one aspect of generic characterisation and indicate how thematic content is sensitive to the overall conceptual macrostructure of the scientific RA and the rhetorical goals inherent at each stage of RA discourse.

It will be recalled that in the discourse-functional analysis of Subject roles, examples from the Real World domain predominated, even though there was noteworthy variation across the four RA sections in this one domain. In the four Theme-Move maps above, for a more accurate statistical appreciation of the relative presence and absence of thematic functions found in each RA section, reference should be made to the relevant tables (TABLE 5 for Context Frames and TABLE 8 for Subjects). However, as an approximate schematic indication, the examples given in the Theme-Move maps for each category indicate their more significant presence in the corpus. Likewise, the absence of particular marked and unmarked thematic functions indicates their lack of significance in signalling rhetorical moves and move boundaries.

In general, it can be seen from the four Theme-Move maps how Context Frames play a significant within-text structuring role in scientific RA discourse; move boundaries are clearly associated with the marked thematic choices relevant to the discrete rhetorical function of that local move. In addition, Subjects in the three non-Real World domains in particular, can be seen to contribute to the realisation of more global purposes as the discourse proceeds. They help to mark the changing balance throughout the scientific RA between, on the one hand, the *interpersonal* thematic strand of more

interactional Themes with greater writer visibility and, on the other, the *ideational* thematic strand of more topic-based Themes (see Figure 6 above).

It is also evident that Subjects which function as Real World Entities, Events and Processes in the vast majority of declarative clauses which make up scientific discourse, *stainless steels.., the mechanism..,* impose the most neutral unmarked thematic status. As such, the majority of these Themes play a much more indirect role in move dynamics, since they predominantly serve to set up move-characterising rhematic comment:

    ..stainless steels <u>have been widely used</u>..
    ..the role of type x collagen <u>remains to be defined</u>..
    ..the mechanism <u>is still not understood</u>..

The exception to this is perhaps found in the more rhetorically homogeneous Experimental section, which is itself distinguished by its relative lack of moves, and is described as light on Theme, heavy on Rheme (Swales 1990a). However, with modified Real World Subject phrases, *experiments to substantiate this.., no recent observations..,* such unmarked Themes also represent an important part of the RA writer's resources in signalling progress towards specific rhetorical goals.

3.11 CONCLUDING REMARKS ON THE ANALYSIS AND DESCRIPTION OF THEME

3.11.1 <u>The initial hypotheses</u>
An important objective in any descriptive study of discourse and genre is to indicate central tendencies in both the linguistic contexts and the social environments from which texts derive. This concluding section to Chapter Three addresses the various hypotheses which were presented earlier (section 3.3) in the light of the

central tendencies in textual organisation of the RA genre as revealed by means of corpus analysis.

[HYPOTHESIS #1]
Descriptions based on a dynamic model of language in use, such as that inherent in Halliday's social-semiotic perspective, 'must show how the discourse proceeds from one point to another and.. how the components of the discourse play their part in the achievement of some purpose' (Sinclair 1985a:15). The results presented above clearly confirm the general [HYPOTHESIS #1], that functional analysis highlights the changing options for, and constraints on, thematic choices across RA sections with developing rhetorical goals. Sinclair comments,

> 'printed material is presented in accordance with thousands of conventions and is measured to thousandths of an inch' (1985a:14).

Based on examination of a corpus of published scientific RAs, a highly standardised genre, the thematic options and constraints highlighted here can be taken as a strong reflection of the central linguistic and social tendencies of scientific RA discourse.

[HYPOTHESIS #1a]
The objective here was to investigate the potential that analysis of marked Context Frames has for understanding the propensities of within-text structuring of scientific RAs. Findings clearly show dynamic patterns which can be predicted on the basis of the rhetorical goals inherent in each section of RA discourse. The 'appropriate' textual flow of marked Themes is evidently perceived to contribute significantly to textual cohesion and coherence and thereby to the achievement of a RA writer's purpose and 'success' in publication. Davies's study of marked Theme choices across a variety of different genres indicates the 'quite direct relationship between certain selections of

marked theme and certain categories of texts' (1991:26),
and thus suggests the potential of broadening the analysis
presented here to other genres.

[HYPOTHESIS #1b]
The analysis of Subject as unmarked Theme presented here
clearly shows one aspect of the social-semiotic
organisation of the scientific RA genre. This hour-glass
structure is not simply a reflection of its conceptual
macrostructure, moving from the presentation of general to
particular information in the Introduction and *vice versa*
in the Discussion. It is strongly characterised by the
manipulation of Subject choices. Hence, it is the
progressive decrease and subsequent increase throughout RA
discourse in, on the one hand, topic-based Real World
Themes and, on the other, the three increasingly more
interactional non-Real World Themes which encode a major
element of the distinctiveness of this genre. Furthermore,
I suggest that the subtle balancing in scientific
discourse of more covertly interactional Themes with the
naturally more dominant topic-based Real World Themes
contributes to our understanding of the statement by
Bazerman & Paradis (1991) in the introduction to Chapter
Three: that our choices in writing structure our relations
with others and organise our perceptions of the world;
that through a better understanding of the genres of the
professions, we have greater insight into how discourse
communities constitute themselves and carry out their work
through texts.

[HYPOTHESIS #2]
Although they only represent an approximation of the
correlations between bottom-up and top-down levels, the
Theme-Move maps presented in section 3.10 of the four RA
sections clearly indicate the contribution of thematic
distribution to the characterisation of this genre and, in
turn, illustrate the constraints upon possible thematic
choices. In particular, with Sinclair's comments on the

integration of linguistic description in mind, comparison across the four Theme-Move maps permits a dynamic view of scientific discourse as interactive, directional and purposeful. This hypothesis is further substantiated by current research from other linguists; for example, working across a variety of written texts, Davies (1991) included extracts from eleven distinct genres in her analysis and was able to show how thematic resources may be exploited for the particular communicative purposes inherent in those distinct genres.

3.11.2 Pedagogical application of this work

Although attention has been paid to the description of central tendencies above, deviations from these norms do not necessarily imply 'poor' writing. Swales has commented that there may indeed be cases when writers 'ride rough-shod over many of their field's linguistic and rhetorical conventions' (1987a:123). However, recognition of genre-typical patterns allows valuable contrast and comparison with reference to the developmental skills of novice RA writers, in particular NNS researchers. For example, an interesting observation from a recent paper by Taylor & Chen (1991) concerned the writing of scientific papers by a group of Chinese researchers. The authors highlighted the fact that they paid relatively much less attention to summarising the literature in their fields of study compared to Anglo-American texts. It was suggested that one reason for this may be due to a culture-based unwillingness to implicitly criticise other researchers' work, to identify by name and then 'expose' gaps in research.

As indicated in the Introduction Theme-Move map (Figure 11) above, statements offering a critical evaluation of previous research, with their associated marked and unmarked thematic roles, express vital rhetorical functions of an effective Introduction. If novice NNS (or

of course NS) researchers are unable to express appropriately an evaluation of others' work, then clearly the opportunity for social interaction is diminished, as 'knowledge is gained in the process of dialogue not so much between the scientist and nature as between the contending views of one scientist and others' (Taylor & Chen 1991:331). If there is validity in such questions, they need to be addressed in helping raise the awareness of novice RA writers to the many between-the-lines issues in academic-professional writing and the sociology of scientific research.

With an increasing interest in genre-based curricula, these insights have potential benefits for English for Academic Purposes (EAP). Davies suggests that the major pedagogic implication arising from her work on thematic resources,

> 'would appear to be a case for developing student awareness of such resources, particularly as they occur in the different text and topic types students are required to read, and most particularly, to write' (1991:27).

Francis also comments that in focusing on elements of textual cohesion and coherence in writing, 'it seems possible to teach students about theme quite explicitly, and coherence can be improved as a result' (1990:86). With the practical application of the present work in mind, Hedge & Gosden (1991) have suggested that a focus on the ability to manipulate thematic components, based on an awareness of conventional generic patterns, is a promising means of assisting novice RA writers gain greater control of the writing and vital rewriting process, thereby helping raise students' 'learning ego'. In practical terms, Davies (1991) mentions the potential of EAP classroom activities such as re-ordering sections of texts

and providing students with Context Frame 'skeletons' of actual or constructed texts for completion.

To continue to explore the heuristic value of genre-based syllabuses, it is important to apply detailed analyses such as the one presented above to a variety of educational contexts and to producers and consumers of scientific research articles, namely, to both experienced and novice RA writers, to English NNSs as well as NSs. This will enable discourse analysts and EAP practitioners to not only draw useful distinctions between different genres, but to investigate the practical effects of consciousness-raising on the writing process and on written products.

With this conclusion of Part I of the thesis, <u>Description and Analysis</u> - <u>Focus on Genre/Product</u>, we now move on to Part II, comprising Chapters Four and Five, which focuses on the pedagogical application of this work and my attempts to apply insights gained from this research on Theme to a specific audience and workplace. The focus here is on a group of NNS novice researchers at a university in Tokyo who were attempting to 'successfully' write their first RAs in English, 'success' being judged here by acceptance for publication in international journals.

Chapter Four first reports on the results of a series of background surveys conducted in part to investigate the influence of thematic structure and other language-related aspects on the judgements by 'expert' journal editors of the merits of RAs written by NNSs. An overlapping survey is also conducted with the two other groups of participants in this task of getting NNSs' RAs successfully into print, namely, their 'experienced' research supervisors and the 'inexperienced' novices themselves. As this research moves from a more product-oriented perspective to one with a greater focus on the RA writing process, Chapter Five reports on a

second area of analysis which is concerned with corpora of both published NNS RAs and unpublished RA drafts written by novice NNSs. The thematic structure in a published corpus of 36 NNS RAs is compared with the parallel NS base norms established in Chapter Three; FIRST and FINAL unpublished RA drafts written by NNS novices are compared with each other in order to trace Theme-related aspects which were modified and manipulated in response to critique in an attempt to make the RA drafts more 'successful', that is, acceptable for publication.

PART II

APPLICATION - BACKGROUND SURVEYS AND ANALYSIS

CHAPTER 4 LANGUAGE-RELATED ASPECTS WHICH MAY AFFECT THE JUDGEMENT OF THE MERITS OF RESEARCH ARTICLES WRITTEN BY NNSs: THE INFLUENCE OF THEMATIC STRUCTURE

4.0 BACKGROUND TO THIS CHAPTER

As previously indicated, the specific motivation for the current research project arose as a result of teaching courses in English for Academic Purposes (EAP), in particular in Academic Writing, initially to NNS postgraduate science and engineering students at the English Language Unit, University of Liverpool and, subsequently, to Japanese doctoral students in the Graduate School of Science, Tokyo Institute of Technology. The general project background therefore is one where EAP support is offered to NNS novice researchers as part of the development of their *academic* communicative competence (Berkenkotter *et al.* 1991). More specifically, since the publication of English-language RAs is a graduation requirement for many Japanese doctoral science and engineering students, EAP support aims to assist their initiation and acculturation into the international discourse community through the many processes and stages involved in publishing in scientific journals. As part of this process, participants' perceptions of language-related aspects which may influence the judgement of the merits of NNSs' scientific RA drafts are therefore of particular interest.

Data and discussion in the preceding chapter on thematic distribution and its correlation with scientific RA discourse structure in the form of Theme-Move maps (section 3.10) led to a pedagogic interest in how NNS novice RA writers deal with thematic structure. In particular, we are interested in the influence of the 'appropriateness' of thematic choices at specific stages

of RA discourse on perceived 'success' (defined here by acceptance for publication), when NNS RAs come under review by editors of international English-language journals. The main objective of Parts II and III of this thesis then is to apply the insights gained from thematic analysis reported in Part I to a specific EAP audience and situation, namely NNS doctoral students writing their first RAs in English.

Some aspects of Parts II of this work are preparatory, that is, before we can investigate how to apply a genre-based approach within a systemic framework to supporting the RA writing activities of NNS novice researchers, it is necessary to have a better understanding of the nature of the RA writing task and of the *in situ* task environment for these NNS novices. To this end, Chapter Four reports on relevant sections of a series of surveys which was conducted as part of the background to the current research project. The rationale for using the research methodology of questionnaires is first outlined (4.1.1), followed by details of the major groups of participants in the international research reporting process whose perceptions are to be explored (4.1.2). Relevant sections of the first of three questionnaires reported in this chapter deal with NS editors (Questionnaire #1 - section 4.2), the second with NNS editors in Japan (Questionnaire #2 - section 4.3) and the third with both NNS doctoral students (Questionnaire #3) and their research supervisors (Questionnaire #3/s - section 4.4). Concluding remarks on implications arising from these three questionnaires are summarised in the final section (4.5).

4.1 RESEARCH METHODOLOGY

4.1.1 Research methodology - questionnaires

As indicated above, this chapter aims to explore participants' perceptions of language-related aspects which may influence the judgement of the merits of NNSs' scientific RA drafts, in particular the influence of the 'appropriateness' of thematic choices at specific stages of RA discourse on perceived 'success'. In this study, various questionnaires were used in an attempt to access participants' perceptions of these issues; this section therefore reviews the use of questionnaires as a research tool.

The use of questionnaires as a method of survey has a long and controversial history and, like all research methods, they are considered to have their relative strengths and weaknesses. They are generally seen as accessible, cheap and easy ways of obtaining information. However, there are obvious criticisms - test questions may be potentially leading or too vague and obscure, using unfamiliar or technical words (Belson 1981). Critics comment that the reliability of elicited information is open to question (Nesbit & Entwistle 1970). For example, in the present context, if we were to ask academic journal editors by means of questionnaires what they thought about the influence of X language factor when reviewing submitted RA drafts, how can we be sure that their replies reflected what they *actually* thought, or subsequently did, in their roles as editors? Nesbit & Entwistle (1970) thus caution that individuals' perceptions as revealed through survey replies are only to be taken as *indicators* of perceptions: as such, they may be seen as potentially valuable insights.

It has been suggested [Courtenay (1978); Belson (1981)] that the following points need consideration in the design of effective questionnaires: (i) it is necessary to

establish clear categories of questions which it would be relevant to test; knowledge of words, concepts and facts relevant to the understanding of survey questions must be carefully considered, as should the flow of questions and total length; (ii) questionnaire construction should facilitate data analysis; the theoretical constructs underpinning questionnaire design naturally reflect the wider research topics and underlying methodologies which are of central interest. Reflecting the 'mixed-form' research paradigm indicated earlier (section 1.7) which drives the present work, questionnaires themselves may be designed to yield both quantifiable statistical data and more qualitative anecdotal evidence; (iii) the use of pilot surveys to understand potential question failure is vital. As part of this process, Belson (1981) suggests the use of the 'question-testing method', whereby selected informants are asked questions about the wording of individual test questions after pilot questionnaire completion. This process provides valuable feedback for subsequent modifications of survey questions.

These above points were borne in mind in the design of the present questionnaires and discussion of method of survey is included where relevant at later stages in this chapter.

4.1.2 Participants

There were three major groups of participants in the process of international science research reporting whose opinions and perceptions about the task and the task environment of writing RAs in English as a second language (L2) were to be investigated:

(i) both English NS and NNS (in this case Japanese) editors and editorial board members of mainstream international English-language science journals which, as part of their regular acceptance process, conduct a

critical peer-review of submitted RAs. Johns (1990) describes these initiated members of the discourse community as the all-powerful, 'expert' readers.

(ii) NNS doctoral science students who were required or who wished to submit research articles in English to international journals. These RAs were, in the first instance, externally reviewed by the above editors in (i) who serve as gatekeepers of the academic community. For the present purposes, since the act of publication is assumed to be a valuable part of the RA writer's apprenticeship in developing awareness of the subtleties of the publication process, unpublished NNS novices are taken here as 'inexperienced' RA writers. From amongst this broad group, a core group of sixteen NNS doctoral students [four Grade (Year) 1; seven Grade 2; five Grade 3] from the Graduate School of Science, Tokyo Institute of Technology was selected to serve as informants for questionnaires and interviews. These students attended a weekly class in Academic Writing taught by myself during the academic session 1990/1991; they were chosen as core subjects, however, because at that time they were involved in the lengthy processes of writing their first academic research articles in English with the intention of publishing in international science journals. The requirement of their respective departments that they should publish RAs in English prior to graduation was undoubtedly a major factor in motivating students to join a class in Academic Writing.

(iii) the Japanese research supervisors of group (ii) above. In contrast to their students, as published academics, they may be seen as 'experienced' NNS RA writers. They may also serve as informants in (i), that is, as editors of scientific journals. Three NNS supervisors (additionally three NS) from each of the major science fields (see section 3.4.1), who were particularly competent NNSs of English, also played important roles as

technical experts and subject specialist informants throughout the present research project [see Selinker (1979); Huckin & Olsen (1984) on the use of informants in discourse analysis and Language for Specific Purposes]. For example, as indicated earlier (section 3.4.2), they assisted with minor coding difficulties of some Themes; more importantly, they also participated in the initial piloting of the surveys reported here in order to explore potential question failure.

Thus, with these three groups of participants, a broad spectrum of research and research reporting skills is represented from the group of (i) 'expert' editors to (iii) 'experienced' supervisors down to the most 'inexperienced' group of (ii) doctoral students. The next section introduces the first of the three questionnaires reported here, that is, with group (i) the 'expert' editors, and is prefaced with a brief overview of previous relevant research on academic writing surveys in the field of English for Academic Purposes.

4.2 QUESTIONNAIRE #1: NS EDITORS

4.2.1 Introduction

English is an important medium of research communication for countless English NNS researchers around the world. We may like to question the oft-quoted dominance of English as the *lingua franca* of research communication and in doing so appeal for a greater understanding of research reported in other languages. However, the perceived pressure of the medium is evident and many young NNS researchers may feel their chances of future academic success could be limited by ignoring English language medium reporting and hence written English proficiency.

For those EAP practitioners and researchers involved in supporting the research/L2 efforts of NNS novices, detailed linguistic analysis of components of the texts that science researchers read and aim to produce is a natural element of the wider RA writing task analysis which involves asking participants questions about the various demands this task makes upon them [Gosden (1991) reports more broadly on these aspects of task analysis and on the background to setting up courses in Research English]. Despite some criticisms concerning the methodology of survey questionnaires and the reliability of data [Johns (1981); Horowitz (1986a)], over the past decade or so, a number of surveys have addressed the question as to what kinds of academic writing tasks students are required to perform and how that writing is evaluated by their instructors [see Kroll (1979); Ostler (1980); Johns (1981); West & Byrd (1982); Bridgeman & Carlson (1984); Horowitz (1986a); Weir (1988); Canesco & Byrd (1989)]. A few studies have focused on novice NNS researchers [see Richards (1988); Shaw (1991); Casanave & Hubbard (1992); Jenkins *et al.* (1993)]. However, 'the writing requirements and problems of doctoral students have not been targeted in writing survey research' (Casanave & Hubbard 1992:33) and, in general, the teaching of Research English (RE) to novice NNS researchers has received relatively little attention (Swales 1985c, 1987b).

This neglect may be due in part to the task RE aims to support. In written research communication, writers are no longer attempting to satisfy the internal criteria of what Swales (1988) calls 'norm-developed' tasks. For example, with undergraduate written assignments, professors set criteria for examination and their students are expected to show familiarity with a defined body of knowledge. In contrast, writing in the academic-professional community is defined as 'norm-developing'. This implies persuasive reporting and

subtle negotiation through the review process with an external audience. In this way, the writing and publication of research papers constantly define the academic community's more dynamic norms.

It is presumed that when editors and reviewers consider a scientific RA draft for publication, 'it's the quality of research that counts'. As indicated in a study by Myers (1985), the lengthy and complex process of publishing academic papers is formidable enough for native speakers (NSs) of English and so it can be presumed that writing in a second language adds an extra burden to the task of international research communication. For NS and NNS researchers alike, the editors and reviewers of academic journals represent the audience whose criteria all RA writers are both initially and ultimately attempting to satisfy. Therefore, as gatekeepers to academic publication (and, thereby, to whatever follows on professionally from that), it would appear worthwhile to investigate both general and specific language-related aspects which may most seriously distract the attention of editors and reviewers away from appreciation of the scientific merits of papers written by NNSs of English; if through experience Research English teacher-researchers have strong intuitions about the nature of such language-related aspects, then it is important to systematically confirm these. These insights would be of benefit to both RE practitioners and NNS researchers, particularly apprentice researchers in universities and research institutes learning the skills of a new genre. With the extra burden of L2 proficiency, the overall objective of RE support is simply to play a role in assisting NNS researchers compete *on an equal research basis*.

[HYPOTHESIS #3]
It can be hypothesised that, given the nature of the initial RA review process, the language-/discourse-related aspects that may seriously disguise the real worth of a

submitted NNS RA draft are those related to thematic structure and progression as basic components of textual cohesion and coherence. As indicated from the hypotheses addressed in Chapter Three, both marked and unmarked thematic choices have considerable text-structuring and genre-characterising potential. A conventionally (RA-specific) 'appropriate' usage of both marked and unmarked thematic choices, as suggested by the results of the corpus analysis in Chapter Three, will play a more or less *positive*, at worst *neutral* role in first reviews of an RA draft by editors and reviewers (that is, comprehension of the science will not be significantly impaired). Conversely, a lack of proficiency in dealing with conventionally 'appropriate' thematic choices will imply a *negatively* influential factor, since this is expected to affect the overall cohesion and coherence of the RA, and hence for busy editors and reviewers, the perceived quality and current acceptability of the submitted paper.

4.2.2 NS editors and their journals

The criteria for selection of editors and journals were similar to those reported for the main corpus (see section 3.4.1): (i) academic journals are considered to be mainstream international publications in physical and life sciences and are roughly divided between the broad disciplines of physics, chemistry and biological sciences. With the nature of interdisciplinary research fields, this included areas such as physical chemistry, chemical physics, electronic chemistry, biochemistry, biotechnology, chemical engineering; (ii) journals are edited in the U.K., the U.S. or Canada; (iii) they publish regular length papers and only accept articles written in English; (iv) all submissions undergo a critical peer-review process.

4.2.3 <u>Method of survey</u>

A two-part pilot questionnaire was designed and distributed to the NNS and NS specialist informants mentioned earlier (section 4.1.2). Particular attention was paid to the wording of the ten language aspects to be investigated (see Figure 15 below). The aim of Part 1 of Questionnaire #1 was to investigate NS editors' perceptions of the influence of thematic structure and progression as basic components of textual cohesion and coherence; however, the outcome of the small pilot study conducted with subject specialist informants suggested that technical and non-technical usages of the terms *topic* and *theme* were likely to cause some difficulties. Following consultation, for the purposes of the present questionnaire to editors, the more general term *topic* was chosen (see aspect #2 in Figure 15) in preference to the specific technical meaning here of *Theme* which realises Halliday's textual metafunction. Nevertheless, in spite of different possible interpretations of terminology, it was suggested by informants that from *topic* similar functions to those of *Subject/unmarked Theme* would be understood by editors. Likewise, the *linking of sentences* (see #3 in Figure 15) was generally taken to imply reference to the important sentence boundary functions of *Context Frames* as cohesive text-structuring devices. A further modification to the questionnaire following piloting was that a three-point evaluation scale was chosen in preference to the originally piloted five-point scale, since it was suggested by informants that this might facilitate busy editors' completion and return of this mail-shot questionnaire.

Following the pilot, a modified two-part questionnaire was distributed by mail to editors with a covering letter outlining the aims of the current project. For the purposes of the questionnaire, in the instructions, "NNSs" were simply defined as "non-native speakers of English", whether as second or foreign language. Part 1 comprised a

closed, multiple-choice section yielding quantifiable statistical data; Part 2 consisted of a series of 10 open-ended questions which provided more anecdotal qualitative data about the problems NNSs face in attempting to publish RAs. As an example, a copy of Part 1 of the questionnaire as completed by one individual editor is reproduced below in Figure 15, with replies typewritten for clarity. [Since we are most interested here in the influence of language-related, specifically Theme-related aspects, this section draws attention only to replies from Part 1 of the questionnaire - Gosden (1992a) reports on wider issues concerned with L2 research writing].

4.2.4 Questionnaire #1 - Part 1: rationale

The 10 language-related aspects of Part 1 fall into three major areas of inquiry, from bottom-up, 'lower' to top-down, 'higher' order in nature: (i) *sentence-, clause-and word-level concerns*, such as mechanical accuracy (#5) and lexis (#1), which were of basic interest in Weir's (1988) academic writing survey; (ii) *discourse-, text-level concerns* with a focus on topic (Subject/Theme), the linking and development of sentences/propositions (Context Frames) and larger units of the RA (#2/#3/#4), as well as aspects of style/register (#6); (iii) the 10 aspects also included a wider *sociopragmatic dimension* specific to this highly refined research/discourse activity. The latter is derived from a number of social studies of science [see Latour & Woolgar (1979); Bazerman (1981, 1984, 1988); Knorr-Cetina (1981); Gilbert & Mulkay (1984)] and the work of Myers (1985, 1990) who sees the act of review and revision of an academic paper as 'the negotiation of the status that the scientific community will assign to the text's knowledge claim.' (1985:593). Moreover, Myers suggests claims and negotiations are the social processes in which science is constructed. This sociopragmatic dimension is naturally realised through linguistic resources, for example, an RA writer's choices

Questionnaire #1 - Part 1: It is anticipated that quality of research, rather than quality of research reporting, has prime importance. However, with reference to personal experience, what degree of influence might these following 10 aspects have when considering publication of a paper submitted by a NNS researcher? Please use a scale of 1-3, where 3 = GREAT influence, 2 = SOME influence and 1 = NO influence.

1. use of a wide range of vocabulary: 2
    _____

2. development of the topic from sentence to sentence   3
    in a coherent way: _____

3. logical and clear linking of sentences for the reader: 2
    _____

4. organisation of the different sections of a paper in   2
    a clear and logical way: _____

5. use of grammatically correct sentences: 3
    _____

6. writing in the style of academic written English   2
    and not everyday spoken English: _____

7. appreciation of the level of claim that can   2
    justifiably be made for their research: _____

8. ability to manipulate skilfully the language   3
    used in making this claim: _____

9. appreciation of the status of their work in the wider
    academic community and negotiation of this status in   2
    subsequent correspondence with editors: _____

10. ability to manipulate the language which reflects   1
    awareness of this status and its negotiation: _____

Figure 15. A completed copy of Questionnaire #1 - Part 1

when making a claim about research results. A claim that is either too hedged or unrealistically bold will probably require time-consuming negotiation and revision. Therefore, both the level of claim (a research activity) and its subsequent expression (a discourse activity) need to be carefully appreciated. Aspects (#7/#8/#9/#10) aimed to investigate the perceived validity of these points.

4.2.5 Returns

A total of 299 questionnaires (U.S. 146; Canada 36; U.K. 117) was sent to editors and associates of 166 journals; 154 replies (51.5% return) were received (U.S. 49% return, Canada 31%, U.K. 61%). Of the 154, 116 were fully completed for Part 1 and data and commentary are based on this number. The generally good return rate [compared with other broader academic writing surveys: Horowitz (1986a) ca. 5% - Casanave & Hubbard (1992) ca. 15% - Jenkins *et al.* (1993) ca. 31%] may be a reflection of editors' positive concerns for the matters under investigation.

4.2.6 Questionnaire #1 - Part 1: findings and comments

Editors were asked to judge the degree of *influence* that the 10 aspects might have when considering publication of NNS submissions. These were to be judged on a scale of 3 = *great,* 2 = *some* and 1 = *no influence.* Results are presented below in the form of a rank order of the 10 aspects according to mean (TABLE 10) and a graph comparing these mean values (Figure 16). These are followed by comments.

It is clear that "submissions from non-native speakers of English" is a broad category for editors to comment on. Language-related problems inevitably vary according to numerous factors, for example, L1 origin. Moreover, subjective opinions about matters of influence are also

likely to vary quite widely according to editors' personal experience. Nevertheless, my intention here is to draw some useful generalisations for the learning and teaching of research writing skills. Therefore, it is essential to highlight the aspects that appear to have the most negative influence on the review and revision processes, in particular those which may seriously distract attention from judgement of a paper's essential merits.

TABLE 10

Questionnaire #1 - Part 1: Rank Order of NS Editors' (N=116) Perceptions of *influence* - with original question number (#), Mean score {} and Standard Deviation [].

1 (#3)	logical and clear linking of sentences for the reader	{1.97} [.66]
2 (#2)	development of the topic from sentence to sentence in a coherent way	{1.96} [.69]
3 (#5)	use of grammatically correct sentences	{1.79} [.67]
4 (#8)	ability to manipulate skilfully the language used in making this claim	{1.78} [.69]
5 (#7)	appreciation of the level of claim that can justifiably be made for their research	{1.76} [.74]
6 (#4)	organisation of the different sections of a paper in a clear and logical way	{1.74} [.73]
7 (#9)	appreciation of the status of their work within the wider academic community and negotiation of this status in subsequent correspondence with editors	{1.55} [.58]
8(#10)	ability to manipulate the language which reflects awareness of this status	{1.52} [.55]
9 (#6)	writing in the style of academic written English and not everyday spoken English	{1.45} [.55]
10(#1)	use of a wide range of vocabulary	{1.44} [.50]

Figure 16. Questionnaire #1 - Part 1: comparison of mean values

1 = no: 2 = some: 3 = great influence

Values: 1 → 1.44, 2 → 1.96, 3 → 1.97, 4 → 1.74, 5 → 1.79, 6 → 1.45, 7 → 1.76, 8 → 1.78, 9 → 1.55, 10 → 1.52
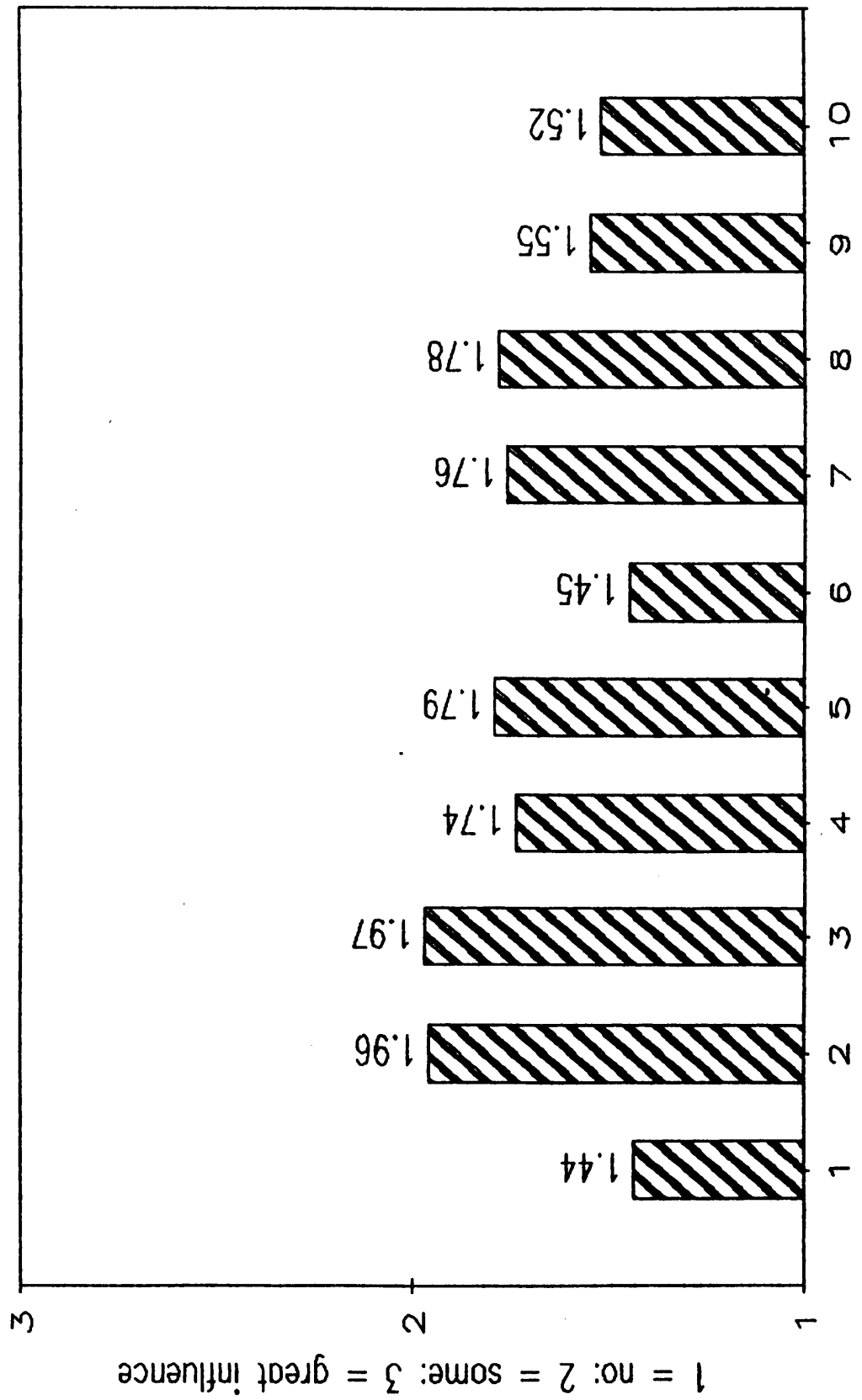
The first and second ranked aspects involve the cohesive linking (#3){1.97} of sentences/propositions and coherent topic progression (#2){1.96}. This reflects the fact that the most meaningful editorial help can generally only be given when the development of a writer's argument is logical and clear. (Of course, it should be noted here that the highest mean {1.97} still indicates perception of slightly below *some* influence). In the review process, and consequently in the writing process itself, attention to the dynamic threads of discourse (#2/#3) takes priority over simple grammatical accuracy (#5), ranked third {1.79}, although it is of course simplistic to minimise the need for mechanical accuracy in realising both textual cohesion and coherence. Ranked fourth {1.78} and fifth {1.76} are researchers' rhetorical expression (#8) and appreciation (#7) of their claims, giving some weight to the influence of these more subtle sociopragmatic aspects. The global management of the different sections of a paper (#4) appeared next {1.74}. Aspects reflecting a researcher's status within the academic community (#9){1.55} and (#10){1.52}, matters of style/register (#6){1.45} and lexical variety (#1){1.44} have the lowest mean values.

Figure 17 below shows the frequency distribution of the three degrees of *influence*. Although replies appear to be quite widely spread, the relative importance of (#2) and (#3) is confirmed in the fact that these two aspects were given the highest counts of *some* and *great*, and the least *no influence*. In response to [HYPOTHESIS #3], that those aspects related to thematic structure and progression would be seen as most influential in the review process, it can be confirmed that this is indeed the case.

Figure 17. Questionnaire #1 - Part 1: Frequency distribution of the three degrees of influence

1: no
2: some
3: great
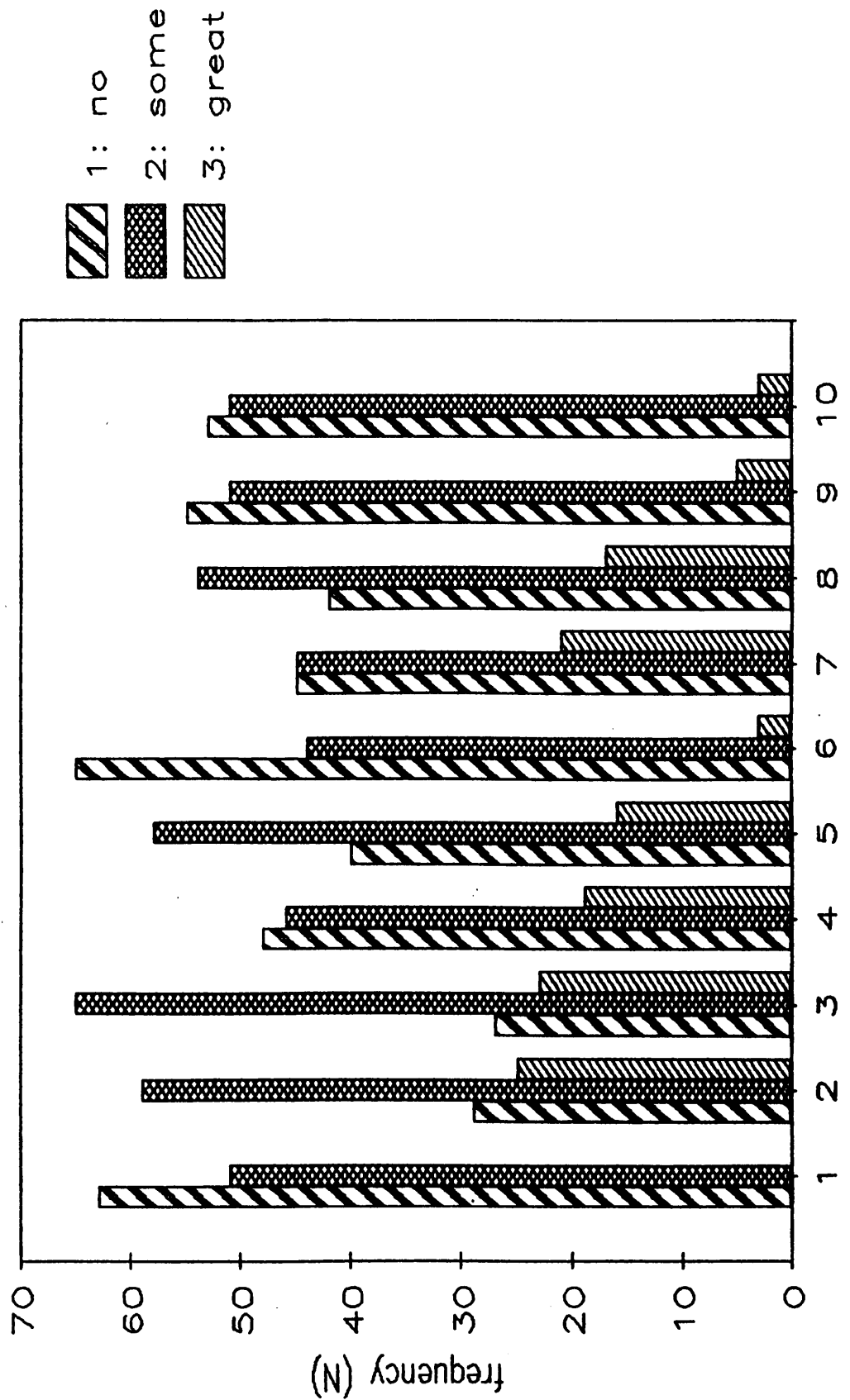
4.3 QUESTIONNAIRE #2: NNS EDITORS

4.3.1 <u>Introduction</u>
Commensurate with the high level of research in progress in Japanese universities and research institutes, there is a thriving market in Japan-published English-language scientific journals which are well-known in their respective fields and of international stature. The present research project focuses on the attempts of Japanese novices to get their RAs published and, quite naturally, local Japan-based publications often represent target destinations for novices' submissions. Since these journals and their published products merit attention, a modified version of Questionnaire #1 above (see Figure 15) was also used with a group of NNS (in this case, Japanese L1) editors and editorial board members. In addition to asking the same question related to the perceived *influence* of the same 10 aspects, the opportunity was taken with this group to enquire as to the perceived relative degrees of *difficulty* these same 10 aspects cause Japanese L1 novice writers of RAs in L2 English. This may provide an interesting comparison between perceptions of these two variables.

[HYPOTHESIS #4a]
As gatekeepers of international English-language science journals, NNS editors are 'expert', published academics; therefore, there are unlikely to be significant differences between the way these two groups (that is, NS and NNS editors) perceive the *influence* of the ten language-related aspects on NNS submissions, although there may be different levels of understanding of L2-related difficulties and problems. Furthermore, if there are no significant differences in data, this will indicate the "universal", non-culture-specific nature of international scientific reporting in English.

[HYPOTHESIS #4b]

NNS editors have both experience of the many 'between-the-lines' issues involved in getting an RA into print and knowledge of English-language education in Japan. Therefore, there are likely to be discrepancies in NNS editors' perceptions of the relative *difficulty* of the 10 language-related aspects when compared with their actual *influence* on the review process. There will thus be a weak correlation between aspects which may be perceived as *difficult* for novice Japanese RA writers, and whether these same aspects are in fact of particular *influence*.

4.3.2 NNS editors and their journals

The criteria for selection of journals were as follows: (i) for the field of journals, see above (section 4.2.2); (ii) they have Japanese editors and editorial board members and journals are primarily edited in Japan; (iii) they only accept papers written in English; (iv) all submissions undergo a peer-review process.

4.3.3 Method of survey

Section 4.2.3 above outlined the method of survey adopted with NS editors and a similar procedure was followed with NNS (Japanese) editors. As Questionnaire #2 added the second factor of *difficulty* to that of *influence*, a pilot of this questionnaire was conducted with NNS specialist informants to test for question failure; it appeared that judgement of *difficulty* was relatively straightforward, based on participants' first-hand knowledge and experience of their research students' English language/science education.

4.3.4 Returns

A total of 96 questionnaires was sent to Japanese editors, associate editors and editorial board members of

62 journals; 57 replies (59.4% return) were received. Of these, 45 were fully completed for Part 1 and data and commentary are based on this number. The return rate compares favourably with the NS editors' returns (51.5%).

4.3.5 Questionnaire #2 - Part 1: rationale

Section 4.2.4 above indicated the rationale for the 10 selected language aspects in Questionnaire #1 - Part 1 distributed to NS editors and these are identical in Questionnaire #2 for NNS (Japanese) editors. The rationale for inclusion of the additional factor of *difficulty* to that of *influence* is related to the investigation of [HYPOTHESIS #4b], namely that there are important contrasts between these two sets of judgements.

4.3.6 Questionnaire #2 - Part 1: findings and comments

Results are presented in the form of rank orders of (a) *influence* (TABLE 11) and (b) *difficulty* (TABLE 12) according to mean scores, followed by comments.

(a) *influence*: The three aspects ranked highest involve the cohesive linking of sentences/propositions (#3){2.38}, coherent topic progression (#2){2.24}, and the clear global organisation of a paper (#4){2.22}. As with the NS editors' data, these judgements reflect the fact that editorial help will most readily be given when the development of a writer's argument is already well set-out, that is, writers cannot leave it up to others to interpret what it is they want to say. Ranked fourth (#10){2.20} and fifth (#9){2.18} are aspects concerned with Japanese researchers' perceived status within the wider academic community. The subtle aspects involving researchers' appreciation (#7){2.11} of their claims and their subsequent rhetorical expression (#8){1.98} are ranked six and eight, respectively. In between, simple mechanical accuracy (#5){2.0} is given a relatively low

ranking at seven. Least influential are matters of style/register (#6){1.67} and lexical variety (#1){1.62}.

TABLE 11

Questionnaire #2 - Part 1: Rank Order of NNS Editors' (N=45)
Perceptions of *influence* - with original question number (#),
Mean score {} and Standard Deviation [].

1 (#3) logical and clear linking of sentences for the reader	{2.38} [.68]
2 (#2) development of the topic from sentence to sentence in a coherent way	{2.24} [.68]
3 (#4) organisation of the different sections of a paper in a clear and logical way	{2.22} [.63]
4(#10) ability to manipulate the language which reflects awareness of this status	{2.20} [.50]
5 (#9) appreciation of the status of their work in the wider academic community and negotiation of this status in subsequent correspondence with editors	{2.18} [.58]
6 (#7) appreciation of the level of claim that can justifiably be made for their research	{2.11} [.69]
7 (#5) use of grammatically correct sentences	{2.00} [.58]
8 (#8) ability to manipulate skilfully the language used in making this claim	{1.98} [.66]
9 (#6) writing in the style of academic written English and not everyday spoken English	{1.67} [.67]
10(#1) use of a wide range of vocabulary	{1.62} [.65]

(b) *difficulty*: ranked first (#8){2.40} is the crux of a paper - the precise expression of what is being claimed about one's research. Next, is the ability to mobilise a wide range of vocabulary (#1){2.36} - often considered by NNSs to be a 'NS-like' facility they may lack. Ranked third is the manipulation of language which may reflect a researcher's awareness of his/her status in the community (#10){2.16}. The first-ranked aspect in *influence* above,

namely the clear and logical linking of sentences (#3){2.09}, is ranked only fourth in *difficulty*. The next two are research-oriented factors - writers' appreciation of their claims (#9){2.07} and of their community status (#7){2.0}. Simple mechanical accuracy (#5){1.96} is ranked seven; eight and nine return to discourse-oriented factors - coherent topic development (#2){1.93} and global textual organisation (#4){1.73}. Bottom ranking is given to matters of style/register (#6){1.64}, that is, NNS editors generally think that the ability to write an RA in a formal academic style is the least of their difficulties.

TABLE 12

Questionnaire #2 - Part 1: Rank Order of NNS Editors' (N=45)
Perceptions of *difficulty* - with original question number (#),
Mean score {} and Standard Deviation [].

1 (#8) ability to manipulate skilfully the language used in making this claim	{2.40} [.65]	
2 (#1) use of a wide range of vocabulary	{2.36} [.60]	
3(#10) ability to manipulate the language which reflects awareness of this status	{2.16} [.60]	
4 (#3) logical and clear linking of sentences for the reader	{2.09} [.74]	
5 (#9) appreciation of the status of their work in the wider academic community and negotiation of this status in subsequent correspondence with editors	{2.07} [.69]	
6 (#7) appreciation of the level of claim that can justifiably be made for their research	{2.00} [.64]	
7 (#5) use of grammatically correct sentences	{1.96} [.60]	
8 (#2) development of the topic from sentence to sentence in a coherent way	{1.93} [.62]	
9 (#4) organisation of the different sections of a paper in a clear and logical way	{1.73} [.62]	
10(#6) writing in the style of academic written English and not everyday spoken English	{1.64} [.64]	

It is quite natural to suggest that the task of writing academic papers in a second language creates anxiety. If NNSs find specific aspects of the writing task difficult, it is evident that they are also going to be worried about their competence to deal effectively with them. In emphasis of this, when Clark & Ivanič (1991) asked a NNS group of EAP students to make a visual representation (using labelled cards on a large sheet of paper) of the L2 academic writing process as they saw it, EXPERIENCING PANIC, PAIN AND ANGUISH was the central component – all sixteen other components radiated out from this.

However, instead of experiencing a general feeling of anguish because these factors are 'difficult' as part of the L2 writing process, it would clearly be more constructive for NNS RA writers to be concerned about factors which will have a significant influence on the review process. This suggestion is an example of how reader-based awareness can be raised through participation in the academic community's activities. For apprentice researchers, this process helps to raise awareness of factors which initially appear to be outside the individual writer's control and which 'define, shape and ultimately judge a piece of writing' (Horowitz 1986a:446). In order to investigate factors of which apprentice RA writers may have to become more aware, components of these questionnaire findings are now compared and contrasted.

4.3.6.1 *Influence vs. difficulty*
The Pearson correlation coefficient of $r = -0.081$ indicates almost exactly no statistical correlation ($p<0.05$) between the means of NNS editors' perceptions of the two variables. This gives very strong support to [HYPOTHESIS #4b], that there were likely to be discrepancies in NNS editors' perceptions of *difficulty* of

Figure 18. Questionnaire #2 - Part 1: Comparison of NNS editors' (N=45) perceptions of influence and difficulty
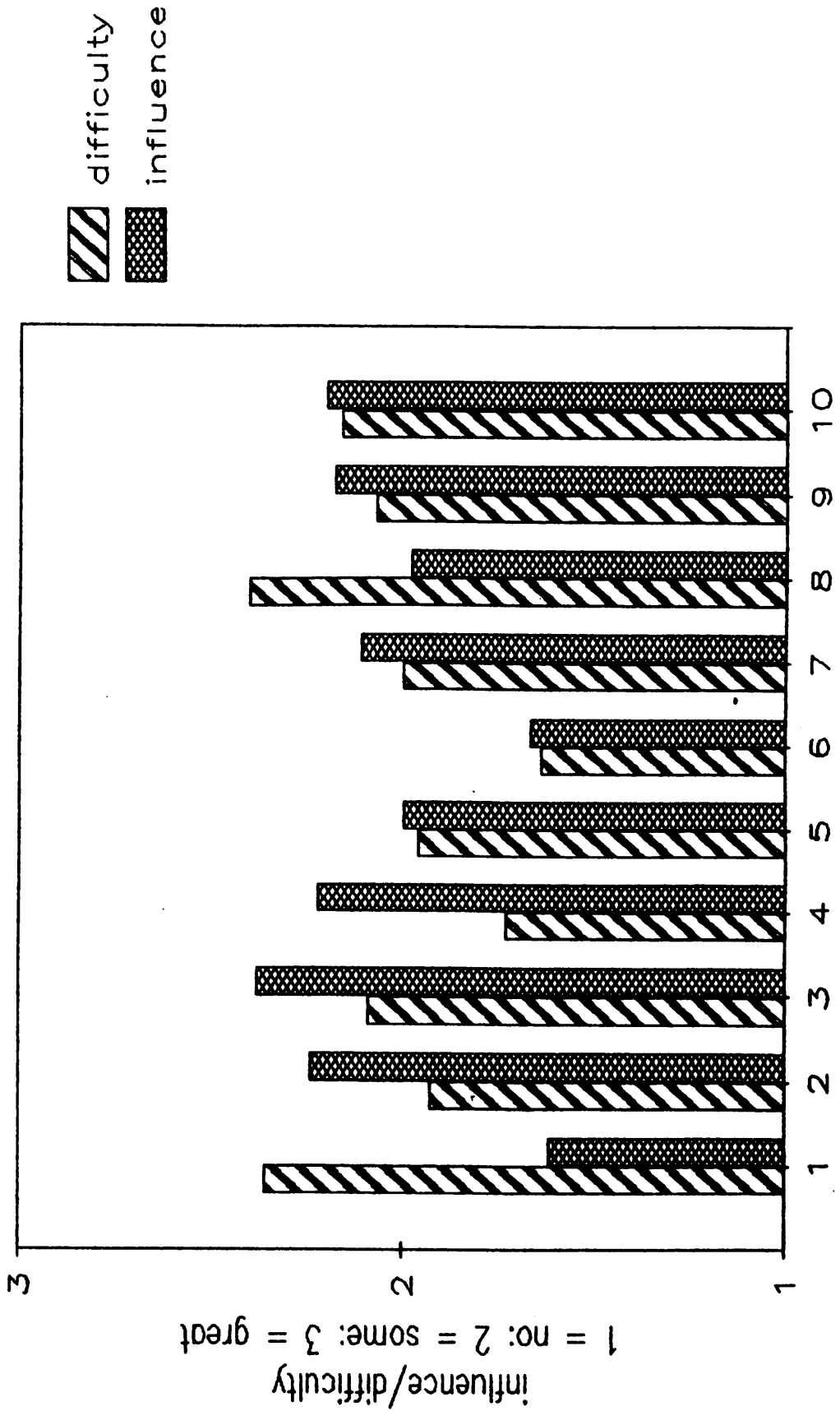
the 10 language-related aspects when compared with their actual *influence* on the review process. Figure 18 above shows that there are indeed evident contrasts.

For example, the top three ranked aspects in *influence* do not coincide with those of *difficulty*. In fact, the second and third aspects of *influence* (#2/#4) (concerning coherent topic development and global textual management) are only ranked eighth and ninth in *difficulty*. As another example, editors consider the ability to use a wide range of vocabulary (#1) as problematic - it is ranked second in *difficulty*. According to editors, however, the *influence* of (#1) in the review process is ranked tenth. These facts emphasise the need for NNS RA writers to distinguish factors which matter in terms of influence from those which are simply concerns and thereby pay them appropriate degrees of attention at different stages of writing and rewriting RA drafts.

4.3.6.2 <u>NNS editors vs. NS editors</u>

Comparisons between replies from NS editors (section 4.2.6) and NNS editors (section 4.3.6) can be drawn. Comparing the rank orders for *influence* statistically, the Spearman correlation coefficient is just significant at r_s = 0.624 ($p<0.05$), showing only weak evidence therefore of some measure of agreement and support for [HYPOTHESIS #4a]. However, editors agreed on the top and bottom two ranked aspects, (#3)/(#2) and (#6)/(#1), respectively. As can be seen from Figure 19 below, the most obvious contrast between the two sets of data is that, for all 10 aspects, the mean score for NNSs was higher than for NSs. Moreover, the range of mean scores was 1.62-2.38 for NNSs compared with 1.44-1.97 for NSs. In other words, NNSs felt that *all* aspects were more influential and this may again reflect the inherent anxiety of second language writing.

Figure 19. Comparison of NS and NNS editors' perceptions of influence

influence
1 = no; 2 = some; 3 = great

NS
NNS

181
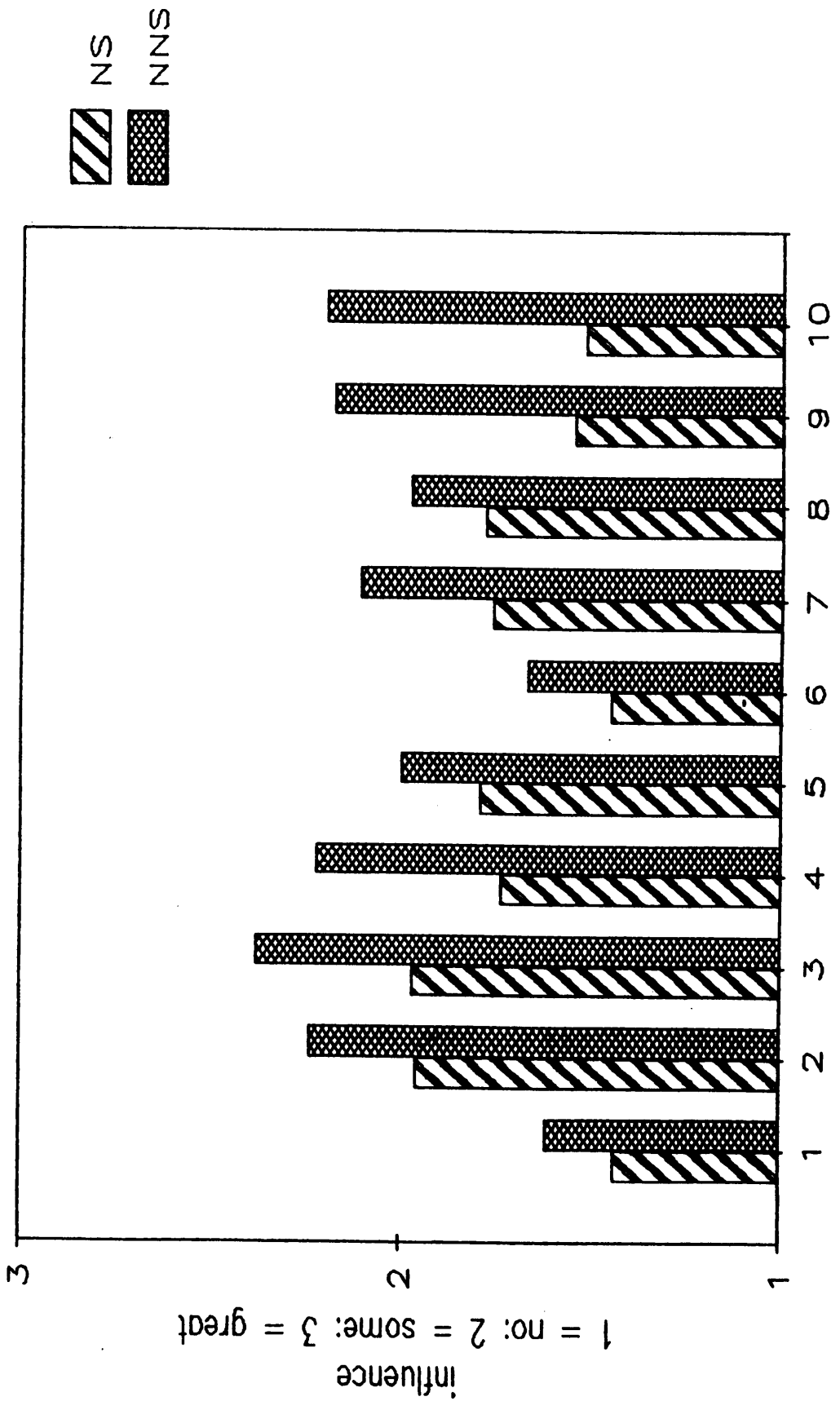
If we ignore absolute scores and compare relative rank orders, we can say that: (i) higher rankings were given by NNS editors to questions of status in the community (#9)/(#10); (ii) NNSs placed greater influence on global textual organisation (#4) than did NS editors; (iii) NSs emphasised more the ability to deal with claims (#7)/(#8), fundamental to the appreciation of what an RA purports to add to the existing body of knowledge; (iv) NSs saw simple mechanical accuracy (#5) as having a greater bearing on the review process. Overall, the broad implication from these comments is that NNS editors perceived more the influence of discourse-external factors. Behind the reasons for the latter may lie rather sensitive questions regarding possible bias against NNS submissions and thus a strong desire for fair representation on an equal research basis [Part 2 of the survey of NS editors reported in Gosden (1992a) provides a fuller discussion of this issue]. Conversely, for NS editors, there is greater focus on discourse-internal realisations of external factors. Following Myers (1985) noted earlier, the appreciation and manipulation of the language of claims may well have a sociopragmatic dimension, but clarity of their expression is most critical.

With questionnaires #1 and #2, we have now addressed the perceptions of the NS and NNS 'expert' groups of participants in this process of international research communication, in particular, in relation to language-related factors which may affect the judgement of the merits of submitted RA drafts. The next section reports on relevant parts of a similar survey of the perceptions of the 'experienced' and 'inexperienced' groups of participants.

4.4 QUESTIONNAIRES #3 AND #3/s

4.4.1 Introduction

A modified version of questionnaires #1 & #2 used with 'expert' NS and NNS editors was also used with the two other groups of participants [see section 4.1.2(ii)/(iii)], namely, 'experienced' supervisors (N=38) and 'inexperienced' doctoral students (N=40). As members of the Graduate School of Science, Tokyo Institute of Technology, they specialised in similar scientific fields to those mentioned above (see section 4.2.2) and the novice researchers were doctoral candidates at different stages of their graduate research (Grades/Years 1-4). Supervisors were either immediate mentors or advisory professors in the same department and can be seen to represent an influential intermediate 'norm-developing' audience for their students' RA drafts.

4.4.2 Method of survey and returns

In comparison with the earlier questionnaires which were mailed, copies of questionnaires #3 and #3/s with a covering letter were distributed by hand to the immediate research supervisors and research laboratory co-supervisors (N=38) of the core group of sixteen doctoral students [section 4.1.2(ii)]. The purpose of the survey was explained during a brief interview arranged with each supervisor and the completed questionnaire was collected by hand or delivered by campus mail. With the core group of sixteen doctoral students, questionnaires were completed and collected during the very first class of an Academic Writing course, prior to any discussion of the issues raised in the questionnaire; in addition, fellow graduate students in the core group's laboratories were visited for the purposes of completing a broader survey (N=40). Thus, 100% of those questionnaires distributed were returned.

Instructions to students
As a graduate student, how much (a) *difficulty* and
(b) *concern* [or worry] for their importance [or
influence] do you have with the following 13 areas
when preparing an academic paper in English for
publication.

Instructions to supervisors
As a supervisor of graduate students, please give
your personal opinion about (a) how *difficult*
graduate students find the following 13 areas when
preparing academic papers in English for
publication and (b) how *concerned* you are, as their
supervisor, for the relative importance and
influence of these 13 areas.

Please score as follows:
5 = enormous *difficulty* & *concern*
4 = a lot of *difficulty* & *concern*
3 = some *difficulty* & *concern*
2 = only a little *difficulty* & *concern*
1 = no *difficulty* & *concern* at all

(1)   writing grammatically correct sentences
(2)   using a variety of grammatical structures
(3)   using a wide and varied range of vocabulary
(4)   using appropriate vocabulary
(5)   spelling correctly
(6)   the scientific content of the paper
(7)   doing justice to the research when writing up
(8)   developing the topic in a coherent way
(9)   using correct punctuation
(10) linking sentences logically and clearly
(11) organising the different sections of the
      complete paper in a logical and clear manner
(12) preventing sentences   from becoming
      unnecessarily complex to read
(13) writing in the style of academic written
      English and not spoken English

Figure 20. A copy of Questionnaires #3 and #3/s

4.4.3 Questionnaires #3 and #3/s - rationale

Due to this particular audience and with EAP classroom activities and purposes in mind, the modifications to Questionnaires #3 & #3/s from those reported above primarily involved a greater focus on sentence-/discourse-level aspects and less on the sociopragmatic dimension investigated with journal editors. In addition, since doctoral students generally have little knowledge or experience in considering how influential specific language-related aspects may or may not be on the RA review process, the factor of *influence* was approached through their perceived "*concern* [or worry] for their importance [or influence]". A copy of the questionnaire used is reproduced above in Figure 20. [In contrast to the editorial surveys above, here the originally piloted five-point scale was used for this questionnaire, since the method of survey allowed for a longer time for completion and 100% return].

[HYPOTHESIS #5]
With their different levels of experience, in terms of both research and in writing RAs, there are likely to be significant differences in the perceptions of *influence* and *difficulty* of language-related aspects between novice NNS research students and their supervisors.

4.4.4 Questionnaires #3 and #3/s - findings and comments

It is interesting to note the correlation between *difficulty* and *concern* in this questionnaire. For doctoral students (TABLE 13 and Figure 21 below), this is significant (r = .834, $p<0.05$), that is, they are generally concerned about what they find difficult and *vice versa*.

A close correlation between these two factors (r = .828, $p<0.05$) was also the case for a group of M.Sc. students, who can be seen as a group of yet more 'inexperienced'

185

Figure 21. Questionnaire #3: Doctoral students' (N=40) perceptions of difficulty and concern

5.0 = enormous difficulty/concern
1.0 = no difficulty/concern at all

difficulty
concern
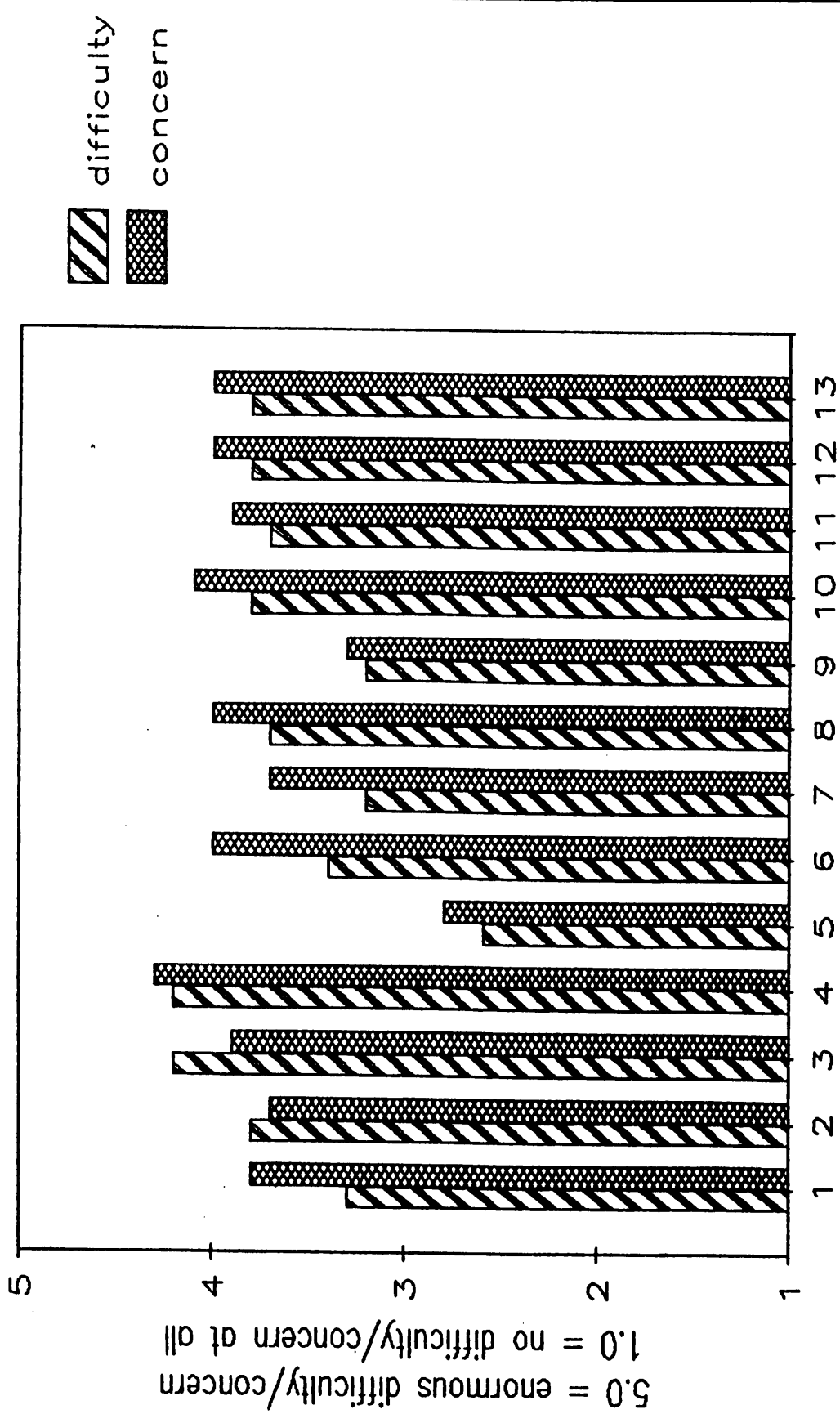
researchers, which was additionally surveyed for Questionnaire #3 [raw data are not reported here]. Interestingly, for M.Sc. students, *difficulty* preceded *concern* in 9 of 13 cases. No M.Sc. candidate had at the time of completion of the questionnaire written a paper in English and this may be the reason for the dominance of supposed *difficulty*. Doctoral students, however, perceived more *concern* than *difficulty* in 11 out of 13 cases. By this stage of their research experience, doctoral students had started to appreciate at least the global importance and relevance of English and the role that their own L2 competence may play in their academic-professional life.

TABLE 13

Questionnaire #3 - Rank Order of Doctoral Students' (N=40) Perceptions of *difficulty* and *concern*

(a) *Difficulty*				(b) *Concern*			
#	Mean[1]	rank	SD[2]	#	Mean[1]	rank	SD
1	3.3	10	.83	1	3.8	9	1.00
2	3.8	3=	.86	2	3.7	10=	.99
3	4.2	1=	.85	3	3.9	7=	.85
4	4.2	1=	.72	4	4.3	1	.82
5	2.6	13	.90	5	2.8	13	1.04
6	3.4	9	1.17	6	4.0	3=	1.10
7	3.2	11=	1.23	7	3.7	10=	1.08
8	3.7	7=	1.08	8	4.0	3=	.88
9	3.2	11=	.99	9	3.3	12	.88
10	3.8	3=	.88	10	4.1	2	.76
11	3.7	7=	1.02	11	3.9	7=	1.07
12	3.8	3=	1.07	12	4.0	3=	1.02
13	3.8	3=	.92	13	4.0	3=	.99

[1] range: 5 = enormous *difficulty* and *concern*; 4 = a lot; 3 = some; 2 = only a little; 1 = none at all
[2] Standard Deviation

The above comments on student groups contrast with those on data from supervisors. From TABLE 14 and Figure 22 below, although it can be seen that *concern* precedes

Figure 22. Questionnaire #3/s: Supervisors' (N=38) perceptions of difficulty and concern

difficulty

concern

5.0 = enormous difficulty/concern
1.0 = no difficulty/concern at all
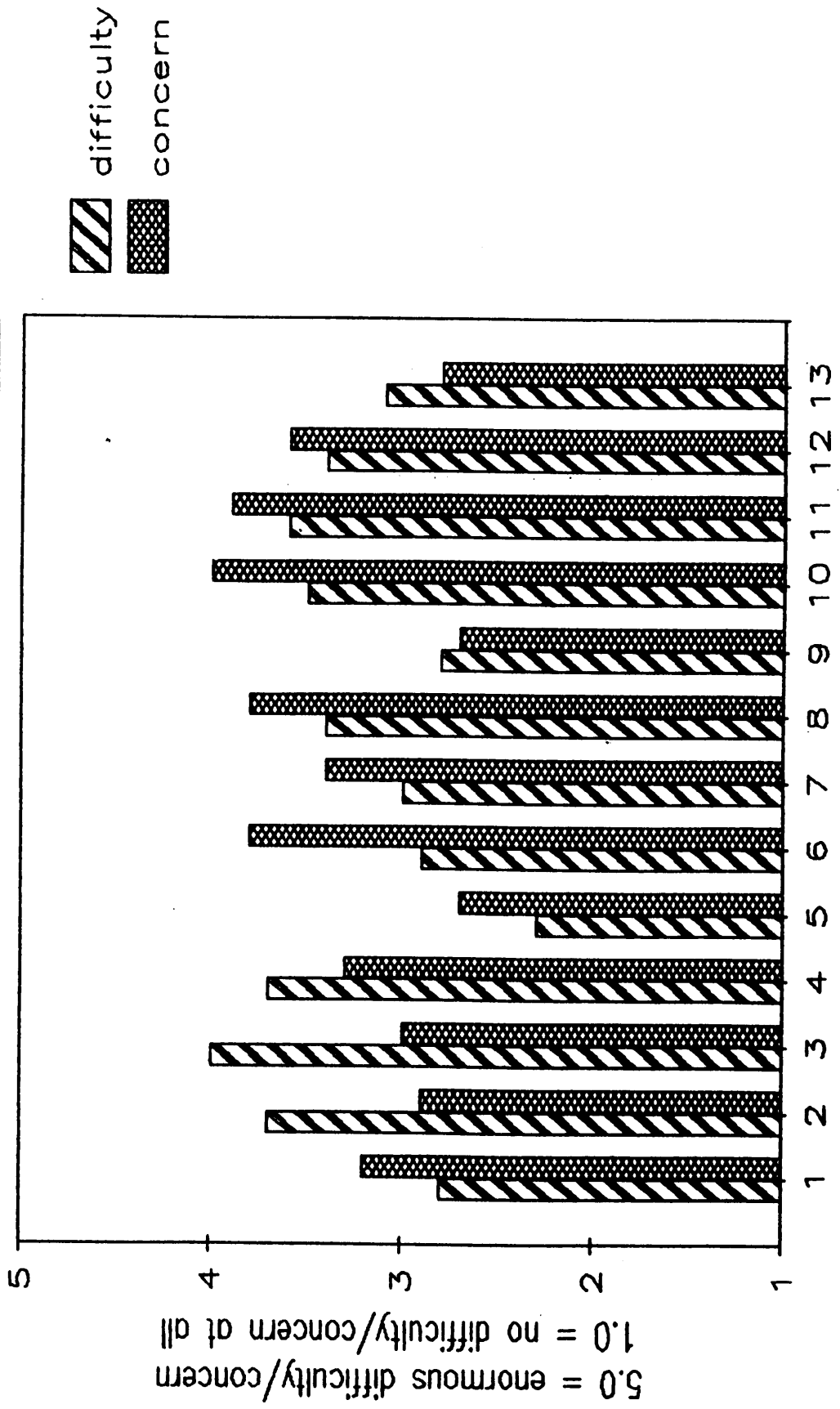
difficulty in 8 out of 13 cases, there is no statistical significance (r = .319, p<0.05) between the two factors. This value of r, when compared with those from NNS editors (-.081) and doctoral students (.834), does indeed imply a continuum in the perceptions of the 'norm-developing' audiences from novice through to experienced RA writers towards editorial acceptance and publication.

TABLE 14

Questionnaire #3/s - Rank Order of Supervisors' (N=38)
Perceptions of *difficulty* and *concern*

(a) *Difficulty*				(b) *Concern*			
#	Mean	rank	SD	#	Mean	rank	SD
1	2.8	11=	.83	1	3.2	8	1.05
2	3.7	2=	.90	2	2.9	10	.90
3	4.0	1	.73	3	3.0	9	.82
4	3.7	2=	.74	4	3.3	7	.85
5	2.3	13	.79	5	2.7	12=	1.09
6	2.9	10	.95	6	3.8	3=	1.20
7	3.0	9	.91	7	3.4	6	1.11
8	3.4	6=	.86	8	3.8	3=	1.00
9	2.8	11=	.78	9	2.7	12=	.81
10	3.5	5	1.01	10	4.0	1	.87
11	3.6	4	.92	11	3.9	2	.88
12	3.4	6=	1.08	12	3.6	5	.86
13	3.1	8	1.02	13	2.8	11	1.05

Let us now return to the initial [HYPOTHESIS #5] which stated that there were likely to be significant differences in the perceptions of *influence/concern* and *difficulty* between doctoral students and their research supervisors. TABLES 15/16 and Figures 23/24 present a summary of the top ranked aspects of *difficulty* and *concern*, respectively. [For the purpose of simplified ranking here, Standard Deviation values are used from the tables where there are equal Mean scores].

The statistical correlation between this homogeneous L1
Japanese group of supervisors' and doctoral students'
perceptions of *difficulty* is significant (r = .929,
p<0.05), with all eight aspects appearing in both groups'
lists. They are in agreement, therefore, that the general
word-/phrase-level factor of vocabulary, in particular,
the ability to appropriately use a wide and varied range
of vocabulary, is considered most problematic.

TABLE 15

Questionnaires #3 and #3/s: Comparison of the Top Eight
Ranked Aspects of *difficulty*.

	SUPERVISORS	DOCTORAL STUDENTS
1	wide range of vocabulary(#3)	appropriate vocabulary(#4)
2	appropriate vocabulary(#4)	wide range of vocabulary(#3)
3	grammatical structures(#2)	grammatical structures(#2)
4	organising RA sections(#11)	clear & logical linking(#10)
5	clear & logical linking(#10)	academic written style(#13)
6	coherent topic development(#8)	long complex sentences(#12)
7	long complex sentences(#12)	organising RA sections(#11)
8	academic written style(#13)	coherent topic development(#8)

Casanave & Hubbard (1992) conducted a survey of science
and technology graduate faculty (cf. supervisors) at a
U.S. university about the writing problems (cf.
difficulty) of NNS (and NS) graduates. The first three
ranked factors (of 12) in their survey were:
   (1) Correctness of punctuation/spelling
   (2=) Accuracy of grammar
   (2=) Appropriateness of grammar

In addition, in the context of overseas postgraduate
students at British universities, Weir (1988) conducted a
similar survey of both student and faculty opinions. Here,
academic staff perceptions of writing difficulties were:

Figure 23. Comparison of doctoral students' (N=40) and supervisors' (N=38) perceptions of difficulty

5.0 = enormous difficulty
1.0 = no difficulty at all

students
supervisors
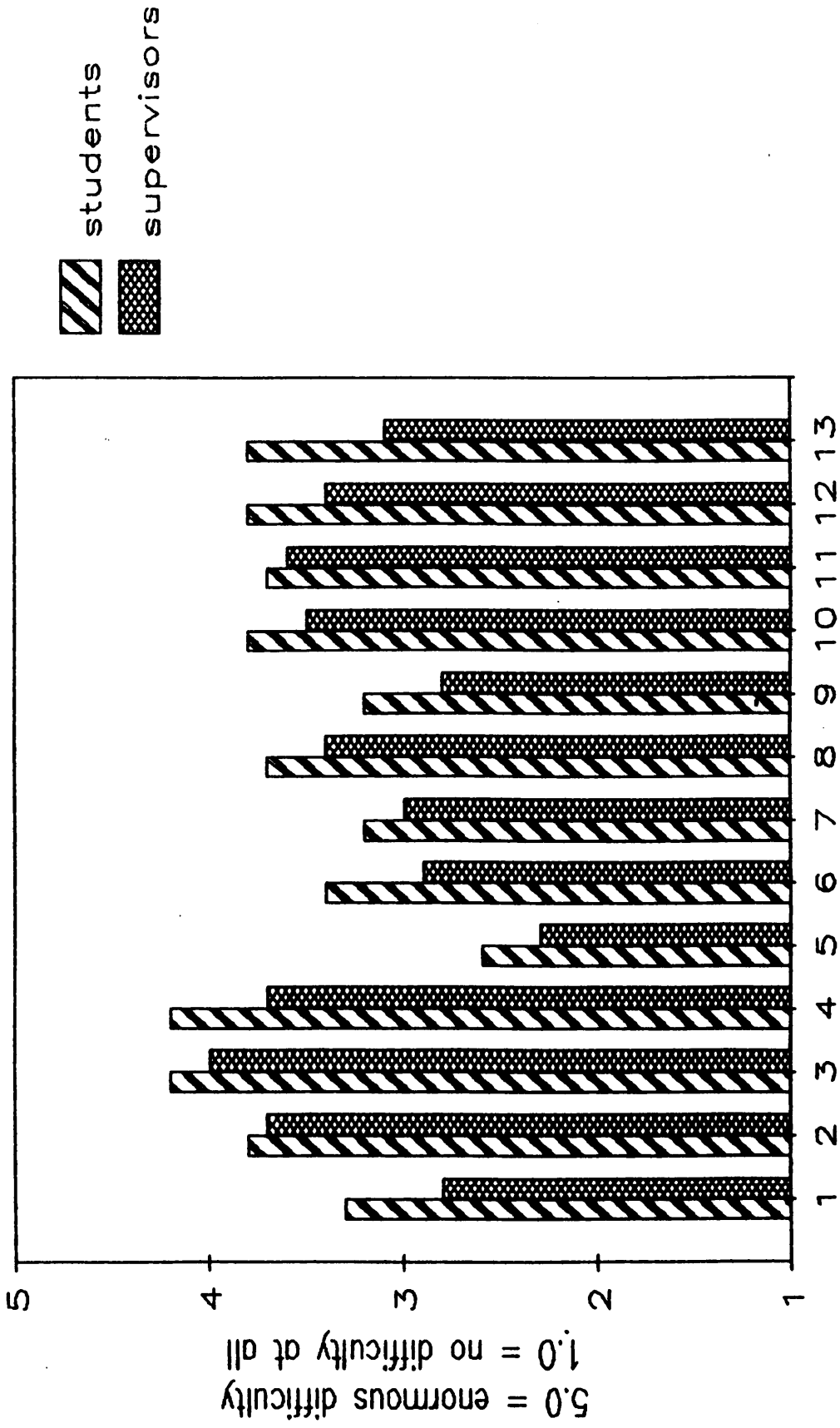
(1) Writing grammatically correct sentences
        (2) Using appropriate grammatical structures
        (3) Expressing what you want to say clearly

In comparison with the present data (supervisors'
perceptions of *difficulty* - see TABLE 15 above), there is
much more emphasis on grammatical difficulties in these
two other surveys. Size and appropriateness of vocabulary,
considered all-important in the present survey, were two
of the twelve factors offered in the surveys by both (i)
Weir and (ii) Casanave & Hubbard; however, size was ranked
(i) #5 and (ii) #6, respectively; appropriateness of
vocabulary in Weir's study was #7 and curiously appears
not to have been scored at all for NNSs (but for NSs) in
the Casanave & Hubbard survey. Although it is clearly not
possible to accurately compare these distinct populations
of NNSs, the #1 'writing problem' factor of correct
punctuation and spelling reported by Casanave & Hubbard is
quite surprising, particularly with the availability of
computer spellcheckers.

In comparison, the three factors considered most
difficult by overseas students in Weir's study conform
closely to the present data in TABLE 15 above:
        (1) Using a wide and varied range of vocabulary
        (2) Using a variety of grammatical structures
        (3) Using appropriate vocabulary

In contrast to *difficulty*, the ranking of factors which
influenced (cf. *concern*) the grades that faculty members
in Casanave & Hubbard's study gave first-year doctoral
students on written assignments (clearly a more
'norm-developed' task than 'norm-developing' RA writing)
indicated that discourse-level criteria (development of
ideas, adequate treatment of topic) were ranked high in
importance, whereas word- and sentence-level criteria
(accuracy of grammar, spelling/punctuation, size of
vocabulary) ranked low. In Weir's study, important

assessment criteria were also more global (subject content, clarity of expression).

These findings of other research are also reflected in the present data from supervisors' perceptions of *concern* which primarily reflect greater awareness of the aspects which may most negatively influence the RA review process and which, therefore, should be of more genuine *concern* in the writing and rewriting process. With supervisors' greater experience, it is not surprising that their data more closely reflect those of the academic journal gatekeepers (see sections 4.2 and 4.3).

TABLE 16

Questionnaires #3 and #3/s: Comparison of the Top Eight Ranked Aspects of *concern*.

	SUPERVISORS	DOCTORAL STUDENTS
1	clear & logical linking(#10)	appropriate vocabulary(#4)
2	organising RA sections(#11)	clear & logical linking(#10)
3	coherent topic development(#8)	coherent topic development(#8)
4	the scientific content(#6)	academic written style(#13)
5	long complex sentences(#12)	long complex sentences(#12)
6	doing justice to research(#7)	the scientific content(#6)
7	appropriate vocabulary(#4)	wide range of vocabulary(#3)
8	grammatical correctness(#1)	organising RA sections(#11)

For supervisors, *concern* is overwhelmingly discourse-level, particularly questions of clear and logical thematic development and overall RA management. Although doctoral students partly reflect supervisors' concerns, there is no statistical significance, (r = .607, $p<0.05$) between these data. From TABLE 16 (and Figure 24 below) it can be seen that word-level aspects of vocabulary figure more strongly under students' concerns.

Figure 24. Comparison of doctoral students' (N=40) and supervisors' (N=38) perceptions of concern

5.0 = enormous concern
1.0 = no concern at all

students
supervisors

194
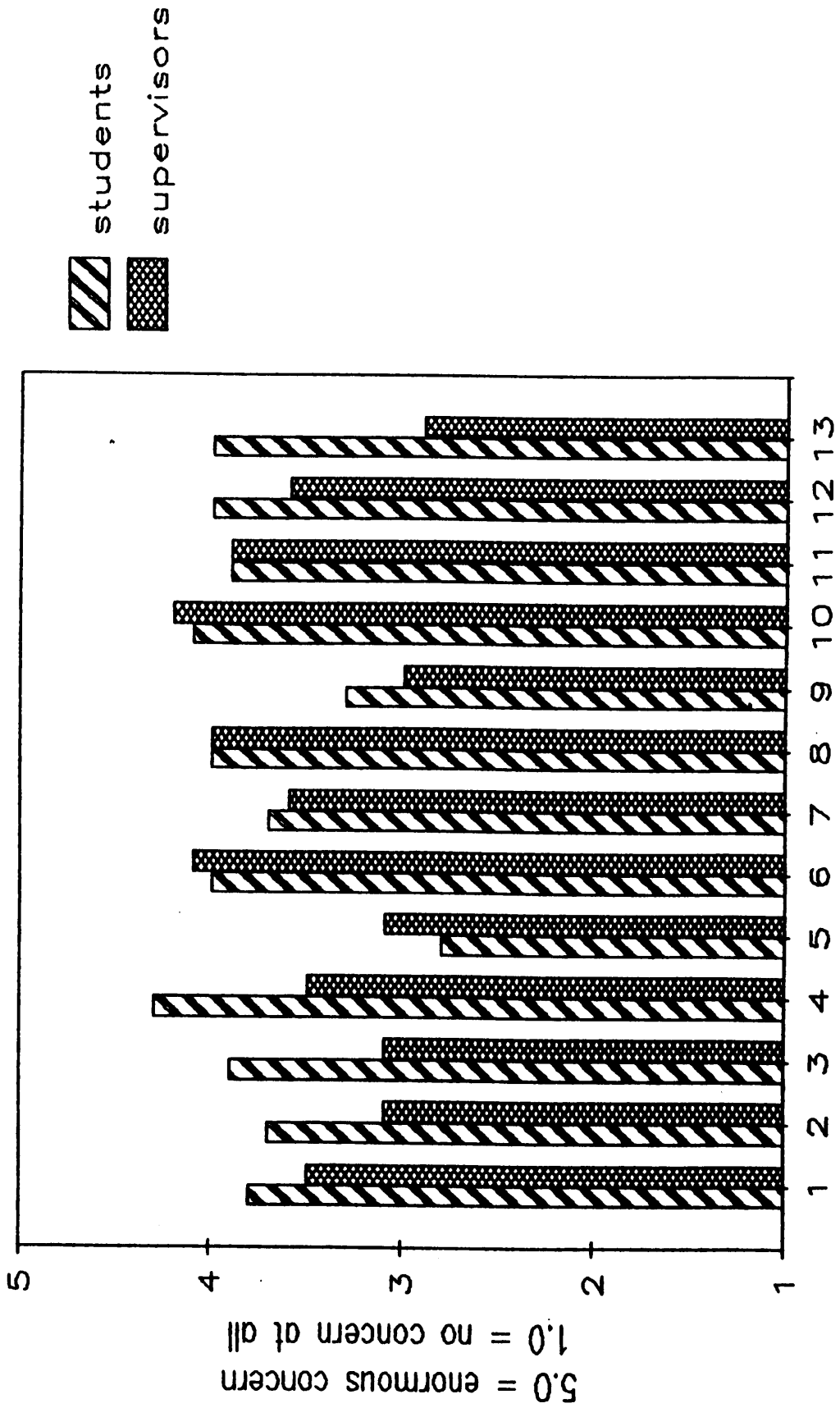
Thus, [HYPOTHESIS #5], which stated that there were likely to be significant differences in perceptions of the *influence* and *difficulty* of language-related aspects between novice NNS research students and their supervisors, appears to be confirmed for the variable of *concern/influence*; however, since these two groups share the same L1 insights into a similar L2 burden, this hypothesis is not valid for *difficulty*.

4.5 CONCLUDING REMARKS ON THE QUESTIONNAIRES

The surveys presented in Chapter Four confirm the important role of raising awareness about the respective influences on the RA writing task and the social environment in which the task is performed. Comparison between questionnaires specifically raises the following question: can the relevance of factors connected with the task of research writing, for example, awareness of aspects which can positively and adversely influence the review process, be seriously under- or overestimated? If so, this would be particularly to the disadvantage of inexperienced NNS researchers as novice RA writers. These questionnaires suggest that the answer to that question is indeed in the affirmative, which itself suggests the need for further guidelines to assist all NNSs who wish to fully participate in international research communication and thereby compete *on an equal research basis*.

Of the points raised in Gosden (1992a) concerning major pedagogical implications arising from Questionnaire #1, of relevance here is editors' focus on components of Theme and their contribution to perceived textual cohesion and coherence. It is, of course, hardly surprising that editors stress the need for the clear and coherent development of the topic of a research article. As part of this creation of a clear and logical text, the management

of thematic choices appropriate to this genre is seen as critical at the review stage of RAs, giving strong support to [HYPOTHESIS #3] above. Thus, thematic choices merit careful attention at all levels and stages of the RA writing process, for both novice and experienced researchers.

Analysis and description of Theme in Chapter Three were based on a corpus of published NS RAs. It would be a natural progression for the current research, therefore, to examine further the written products of the different levels and stages of RA writing experience in order to investigate research questions arising from work so far. For example, Japanese research supervisors and doctoral students tell me that the English-language *Japanese Journal of Applied Physics* is a well-respected international journal, widely read by physicists throughout the world in fast-moving fields such as semi-conductor technology. Published texts therein - overwhelmingly written by Japanese researchers - naturally represent significant 'models' for many Japanese novice applied physicists' own writing. Thus, the following question arises: compared to the NS base norms identified in Chapter Three, do published NNS RAs employ thematic resources across RA sections in similar ways? Additionally, for Japanese novices attempting their first RAs in English, what stages do thematic selections go through between RA drafts in approaching these conventional norms as an integral part of the processes of review and 'successful' acceptance for publication? How do any such thematic changes reflect the social-constructionist nature of research writing? These questions are examined in the next chapter.

CHAPTER 5   THEMATIC ANALYSIS AND DESCRIPTION OF PUBLISHED
            NNS RESEARCH ARTICLES AND UNPUBLISHED (FIRST &
            FINAL) RA DRAFTS WRITTEN BY NNS NOVICES

5.0 BACKGROUND TO THIS CHAPTER

As indicated earlier, Part II of this thesis which
comprises Chapters Four and Five represents the transition
from a research focus on genre-as-product to one which
increasingly integrates attention to genre-as-process; in
this way, the work moves from a theoretical focus on
issues related to the scientific RA genre and Theme to the
pedagogical application of insights gained from the
earlier functional description and analysis reported in
Chapter Three. In order to investigate, and indeed
confirm, the important influence of 'appropriate' thematic
control to the judgement of the merits of submitted
scientific RAs, and thereby to their perceived 'success',
a series of background surveys was reported in Chapter
Four.

As indicated in the concluding remarks to Chapter Four, we
now move on to a brief parallel quantitative-statistical
analysis of Theme in other relevant corpora, namely,
published RAs written by 'experienced' NNS researchers
(section 5.1) and 'inexperienced' NNS novices' unpublished
FIRST and FINAL RA drafts (section 5.2). In order to
integrate background data on the RA genre-as-process,
subsection 5.2.2 reports on the RA writing procedures and
strategies of this core group of NNS novices
(Questionnaire #4). Emphasis is placed here on those
aspects of RA redrafting procedures and strategies which
may be broadly related to the textual development of RA
drafts reported above, in particular those which relate to
recognised features of mature and immature writing and
which may be traced from FIRST to FINAL RA drafts. This
section 5.2 serves to introduce a qualitative

process-oriented commentary (section 5.3) on NNS novices tackling the complex redrafting of their first scientific RAs with a microanalysis of three FIRST and FINAL RA subsections.

5.1 PUBLISHED NNS RESEARCH ARTICLES

5.1.1 Introduction

A corpus of published NNS RAs was analysed according to the thematic discourse functions outlined in Chapter Three. In parallel with the selection criteria and method for the main NS corpus (see section 3.4), 36 RAs were selected from refereed English-language academic journals published in Japan, written by Japanese researchers from institutions in Japan [see APPENDIX D for full references]. Six articles were selected from two journals in each of three disciplines. Six articles - as against three in the NS corpus - were chosen due to the relatively more limited number of English-language journals published in Japan (see section 3.4.1 for other selection criteria). Of the six selected journals (see TABLE 17), four are published monthly and two bimonthly.

TABLE 17
NNS Corpus Journals

Journal of Physical Society of Japan
Japanese Journal of Applied Physics
Bulletin of Chemical Society of Japan
Chemical Pharmaceutical Bulletin
Cell Structure and Function
Japanese Journal of Physiology

When wishing to compare features of published NS and NNS scientific writing, it is of course impossible to judge the varying degrees to which NNS (and, of course, NS) submissions have received editorial assistance in

'improving' the language of the final product. Indeed, it is naturally desirable that advice and language support is sought from NS and more competent NNS colleagues, as journal editors often advise or even insist upon prior to submission, as indicated in Gosden (1992a). It was ascertained from agencies in Japan, professionally involved in scientific publishing in English, that any language-related editorial assistance generally tends to be on the level of cleaning up minor errors and generally does not include the kind of extensive rewriting by NSs which would much more seriously affect the discourse-level concerns under investigation here.

This next section briefly highlights results from comparison of published NNS RAs with the data on Theme from the NS corpus and addresses the following hypothesis:

The RA is generally seen as a most "universal" genre. For example, in their study of Chinese researchers, Taylor & Chen reported that 'there is no "Chinese way" of writing science that is attributable to features of the Chinese language system itself' (1991:330). Shaw's informants, from a variety of cultural and L1 backgrounds, also regarded science as 'a universal enterprise' (1991:199).

[HYPOTHESIS #6]
In recognition of the perceived "universal" nature of academic-professional scientific reporting, there will be few differences in usage of binary thematic selections (Context Frames + Subjects) and their distribution across RA sections between published NS and NNS scientists.

5.1.2 <u>Method of analysis</u>
The method of analysis adopted here exactly replicates that used with the NS corpus described earlier (for details see sections 3.4.1 and 3.4.2).

5.1.3 Findings and Comments

These are briefly given in parallel to those reported for the NS corpus (see sections 3.7 and 3.8); throughout, for ease of comparison, the relevant base NS corpus data are repeated in *italics*. Statistical correlations (*r*) are given where relevant.

5.1.3.1 SIEs - marked vs. unmarked Themes

TABLE 18 below shows the total number of SIEs in the 36 NNS texts where SIE = grammatical subject (GS), SIE = context frame (CF) and where SIE = non-GS/CF [see APPENDIX E for full data]. The homogeneity of the thematic component when comparing the two sets of texts points strongly to the process of scientific writing as rigidly product-constrained, that is, confirming its accepted stable nature as determined by the international discourse community.

TABLE 18
Total of Sentence-Initial Elements in the 36 NNS Texts
- *comparison with the NS corpus*

	NNS corpus	*NS corpus*
SIE = GS	2684 (66.4%)	*2929 (67.2%)*
SIE = CF	1349 (33.4%)	*1408 (32.3%)*
SIE = non-GS/CF	8 (0.2%)	*21 (0.5%)*
Total SIEs	4041 (100%)	*4358 (100%)*
[SD][1]	[8.4]	*[8.2]*

[1] [SD] Standard Deviation of Context Frames

The fact that approximately one-third of all SIEs are CFs in both sets of RAs indicates that such choices are based on writers' experience as consumers and producers of RAs and their awareness of the required rhetorical structuring of discourse. In effect, their goals as RA writers are strongly predetermined and thus a major challenge for novices is to become aware of those goals and their linguistic realisations.

5.1.3.2 GSs vs. CFs by RA section

In investigating the semantic functions of CFs and inherent generic contraints on their distribution throughout different stages of discourse, the relative percentages of GSs and CFs in each RA section are of particular interest and these are given in TABLE 19.

TABLE 19
Distribution of Subjects and Context Frames by RA Section
- *comparison with the NS corpus*

	NNS corpus				NS corpus			
SIE	= GS	CF	non-GS/CF	[SD][1]	GS	CF	non-GS/CF	[SD]
Int	63.9%	36.1%	0.0%	[15.3]	*66.1%*	*32.9%*	*1.0%*	*[12.8]*
Exp	81.5%	18.5%	0.0%	[10.4]	*80.7%*	*19.2%*	*0.1%*	*[10.8]*
Res	66.6%	33.1%	0.3%	[12.7]	*65.8%*	*33.8%*	*0.4%*	*[11.0]*
Dis	58.1%	41.6%	0.3%	[13.2]	*60.5%*	*39.0%*	*0.5%*	*[12.3]*

[1] [SD] Standard Deviation of Context Frames

There is again close similarity between writers' usage of CFs across RAs pointing to perceptions of clearly defined rhetorical goals to be achieved in each RA section. The NNS texts show a slightly greater usage of CFs than the NS texts in the two more multifunctional sections, Introduction and Discussion, with the reverse case in the middle two sections. The more rhetorically varied Introduction and the cyclical nature of Discussion commentary determine the higher frequency of sentence-boundary topic shifting, and hence topic shift signalling. This greater usage of CFs may indicate a writer's attempts to more overtly create a more cohesive text, although discourse markers can of course be overused as suggested by Johns (1984) or inappropriately used. [This point is naturally of considerable interest and will be taken up in some detail (see section 5.3) when dealing with NNS novices' RA drafts].

5.1.3.3 Context Frames by function

To compare the use of all CFs according to the semantic functions listed earlier (see section 3.5), TABLE 20 indicates their relative average percentages distributed throughout the whole RA [see APPENDICES F1-F3 for full data on CFs].

It can be seen from TABLE 20 that for both sets of published RAs, only four functions (Contrast/concession, Location in Real World Time and Space, Cause-reason/result) account for 57.1% and *57.4%* (NNS and *NS* texts, respectively) of all CFs. There is some slight variation between NNS and NS usage, the largest difference being in internal Validation (6.2% vs. *3.3%*), but the overall significant correlation indicates remarkable agreement in how CFs are used and this is also reflected across RA sections.

TABLE 20
Distribution of Context Frames by Function throughout the RA
- *comparison with the NS corpus*

			NNS corpus	*NS corpus*
1a.	Location in Time	- RWE	12.4%	*11.2%*
1b.		- DE	1.5%	*2.8%*
2a.	Location in Space	- RWE	12.3%	*14.9%*
2b.		- DE	5.4%	*4.2%*
3a.	Addition - appositive		0.9%	*1.1%*
3b.	Addition - emphasis		8.4%	*8.4%*
4.	Contrast/concession		17.7%	*18.3%*
5a.	Cause - reason/result		14.7%	*13.0%*
5b.	Cause - purpose		4.4%	*5.3%*
6.	Means		1.8%	*2.0%*
7a.	Condition - real		9.6%	*8.4%*
7b.		- hypothetical	1.8%	*2.6%*
8a.	Validation - external		2.6%	*3.6%*
8b.		- internal	6.2%	*3.3%*
9.	Viewpoint		0.3%	*0.9%*
			100%	*100%*

correlation with
the *NS corpus* $r = .968$

202

Since the data in TABLE 20 are independent of rhetorical context, of greater significance are indications of local Theme choices as reflections of global macrostructure - these data are presented in the next section.

5.1.3.4 Distribution of CFs by RA section

TABLE 21 shows the distribution of Context Frames by RA section. Comments below highlight noteworthy differences between the NNS and NS corpora.

TABLE 21
Distribution of Context Frames by RA Section
- *comparison with the NS corpus*

	NNS corpus	*NS corpus*
Introduction	11.6%	*15.4%*
Experimental	11.2%	*12.6%*
Results	30.8%	*31.7%*
Discussion	42.7%	*40.3%*
	100%	*100%*

The data again indicate more notable differences between Introduction and Discussion sections. The <u>Discussion</u> section accounts for 42.7% (vs. *40.3%*) of all CFs [Results and Discussion taken together account for 73.5% and *72.0%*, respectively]; this indicates the greatest need here for the manipulation of thematic structure. From these data, it appears that the NNS RA writers may feel the need to make their texts work rhetorically harder in presenting their idealised hypothetico-deductive case in the critical Discussion section, since sequences of CFs indicating Contrast/concession + Cause-reason/result account for over one third of the total in this section. In emphasis of this, the highest totals of a single functional category of CF used by both NNSs and NSs occur in this section, namely Contrast/concession - 120 examples for NNSs and *108* for NSs. This is closely followed by Cause-reason/result

with 109 and *78* respectively. Cycles of consecutive
sentence-initial CFs evidently signal these critical
rhetorical sequences, the maximum number in one NNS RA
being 13 (*7* in the NS corpus). The similarity here between
the NNS and *NS* corpora is again striking with two (or
more) consecutive sentence-initial CFs making up 59% (*57%*)
of all CFs; threes 32% (*32%*) and fours 17% (*17%*).

5.1.3.5 CFs by function and type

From not only the discourse analytical but also from an
operational pedagogic point of view, the current work has
a strong interest in developmental phenomena in L2 RA
writing skills. So, although we are primarily interested
in a broad semantic view of thematic components, in order
to indicate teachable CF usage, a picture of CFs according
to both functional and grammatical paradigms may be
desirable. The overall raw data for the three grammatical
categories of Context Frames (see section 3.5.1 above) are
first indicated in TABLE 22.

TABLE 22
Distribution of Context Frames by Type
- *comparison with the NS corpus*

	NNS corpus	*NS corpus*
CF Type 1	40.0%	*37.5%*
CF Type 2	35.7%	*38.0%*
CF Type 3	24.3%	*24.5%*
	100%	*100%*

From TABLE 22, it can be seen that there is indeed a
greater usage of 'minimal' Type 1 CFs, as may have been
anticipated; but it is curious that this is not at the
expense of identical CF functions expressed by means of
Type 3 CFs. To examine these data in greater detail, TABLE
23 below presents a summary with percentages of CF Types
by function, followed by comments on noteworthy points of
interest.

TABLE 23
CF Usage according to Functional and Grammatical Paradigms
- *comparison with the NS corpus*

	CF Type 1	CF Type 2	CF Type 3
Location in Time-RWE	10.4% - *7.0%*	19.1% - *17.9%*	5.8% - *7.2%*
- DE	3.9% - *6.4%*	*[1] - *	* - *1.2%*
Location in Space - RWE	* - *1.1%*	33.8% - *37.6%*	* - *
- DE	1.5% - *	13.5% - *10.8%*	* - *
Addition - appositive	2.2% - *3.0%*	* - *	* - *
- emphatic	17.6% - *18.9%*	3.1% - *3.4%*	* - *
Contrast/concession	34.4% - *31.1%*	3.1% - *3.4%*	11.3% - *21.1%*
Cause - reason/result	21.1% - *22.2%*	4.0% - *2.2%*	20.5% - *15.7%*
- purpose	* - *	3.5% - *4.7%*	12.6% - *13.9%*
Means	* - *	1.2% - *	5.2% - *6.4%*
Condition - real	5.4% - *1.7%*	11.0% - *11.8%*	14.4% - *13.3%*
- hypothetical	* - *1.3%*	* - *	7.0% - *8.1%*
Validation - external	1.7% - *2.3%*	4.6% - *5.6%*	1.2% - *2.6%*
- internal	* - *1.7%*	3.1% - *1.1%*	21.1% - *9.3%*
Viewpoint	* - *2.3%*	* - *	* - *
	100% *100%*	100% *100%*	100% *100%*
correlations with the *NS corpus* r =	.981	.989	.819

[1] *denotes value under 1.0%*

Here Type 3 CFs mirror the functional usage of Type 1 CFs, that is, in the expression of cycles of Contrast/concession *although* + Cause-reason/result *since*, and specifically in indicating Cause/purpose *in order to,* which are by far the most common realisations for both NSs and NNSs. However the NNS texts clearly make considerably less use of these syntactically more complex Type 3 CFs in the Contrast/concession function (11.3% vs. *21.1%*), also slightly less for Cause-purpose (12.6% vs. *13.9%*) but the reverse is the case with Cause-reason/result (20.5% vs.

15.7%). As noted above, a curious anomaly is that one of the largest differences in all the data is in the usage of Type 3 CFs to express internal Validation (21.1% vs. 9.3%). This functional category features many prefabricated phrases such as *As shown in Figure 1*, etc. but it appears to demonstrate NNS writers' greater attempts to validate research findings by reference to visual data and graphic evidence, accompanied by textual description and rationale for data. However, if such usage were partly in lieu of commentary, it implies that graphs and tables are expected to stand alone as sufficient evidence. This in turn suggests that some NNS writers may expect readers to provide more textual coherence for themselves, Hinds's (1987) point that Japanese, for example, is a considerably more listener/ reader-responsible language than the target language English [see comments on question #2 (section 5.2.2.2) for further discussion of this point].

From consideration of Context Frames as marked thematic choices, we now move on to the analysis of unmarked Subject roles in the NNS corpus.

5.1.3.6 Distribution of Subject roles

By using the same functional domains and categories established for unmarked Themes (see section 3.6), TABLE 24 indicates the results of Subject analysis in the 36 NNS research articles.

Again, it must be noted that similarities in Subject role usage between the NNS and NS corpora are remarkable, particularly when considering the data from the dominant Real World domain versus the other three. Of interest is also the appearance in the NNS corpus of a new functional category, namely, Real World Participant. This resulted from the following kind of example; *the subjects underwent*

the experiment... in the Methods section of a bioscientific paper.

TABLE 24
Distribution of Subject Roles by Domain
- *comparison with the NS corpus*

Subject Role Domain	Mean % NNS corpus	Mean % *NS corpus*
Participant domain		
Discourse Participant	4.6%	*5.7%*
Participant Viewpoint	0.4%	*0.6%*
Interactive Participant	2.3%	*2.9%*
Real World Participant	0.2%	*1*
Total this section	7.5%	*9.2%*
Discourse domain		
Discourse Event/Process	0.7%	*1.0%*
Interactive Discourse Entity	0.7%	*1.6%*
Macro Discourse Entity	0.2%	*0.2%*
Micro Discourse Entity	3.8%	*2.1%*
Empty Discourse Theme	1.2%	*1.1%*
Total this section	6.6%	*6.0%*
H&O domain		
Hypothesised Entity	0.6%	*0.6%*
Objectivised Viewpoint	1.8%	*2.0%*
Hypothesised Viewpoint	0.6%	*0.6%*
Empty H&O Theme	3.4%	*4.4%*
Total this section	6.4%	*7.6%*
Real World domain		
Mental Process	0.4%	*1.2%*
Real World Entity	62.3%	*56.1%*
Real World Event/Process	14.9%	*17.9%*
Empty Real World Theme	1.9%	*2.0%*
Total this section	79.5%	*77.2%*

correlations	Introduction	r = .984
with the	Experimental	r = .999
NS corpus	Results	r = .991
	Discussion	r = .995

[1] *there were no examples of this domain category in the NS corpus*

5.1.4 Concluding remarks on published NNS RAs

In strong support of the statement of [HYPOTHESIS #6], that there would be few differences in the usage of unmarked and marked Themes and their distribution across RA sections between 'experienced' NS and NNS scientists, the above data emphasise the point made by Swales with reference to the selection of sources for text analysis: 'insistence on native-speaker purity has little to recommend it in the contemporary world' (1987a:124) and this statement appears particularly relevant for the world of international scientific cooperation. For novice researchers attempting to learn their L2 research reporting skills 'on the job' without formal instruction, NNS RAs evidently serve as equally authentic 'models' of thematic development in RA writing as do NS products.

Having already examined the written products of 'experienced' NS and NNS RA writers, the next section (5.2) concludes a tripartite quantitative analysis of RA products with an investigation of FIRST & FINAL RA drafts written by 'inexperienced' NNS novice researchers attempting their first research papers in English. However, by tracing the textual history of individual RAs, the focus of interest moves away from a heavily product-oriented investigation of thematic structure in research writing to one of greater process orientation and pedagogic application. Following the anthropological tradition established in the sociology of science [Latour & Woolgar (1979); Knorr-Cetina (1981); Gilbert & Mulkay (1984)], the approach adopted in tracing the 'manufacture of knowledge' from a first rough draft in English to its published version integrates attention to both product and process data, allowing a focus on both quantitative and qualitative aspects of thematic selection as an integral part of the redrafting process. In particular, thematic analysis across drafts is expected to give insights into the development of pragmatic competence, 'the knowledge and skills that are necessary for membership in a society

or community' (Mehan 1980:131); thereby, we may observe
aspects of the initial acculturation of a group of novice
NNS researchers into a well-defined international
discourse community.

5.2 UNPUBLISHED (FIRST & FINAL) RA DRAFTS WRITTEN BY NNS NOVICES

5.2.1 Introduction

Of the core group of sixteen doctoral students who were
the informants for the present work (see section 4.1.2),
ten had submitted FINAL drafts of their first papers to
international English-language scientific journals (as in
TABLES 1 and 17 above) by the time the project of
monitoring their RA writing was completed. In fact, in
eight cases the RAs had been published, were in press or
had been accepted for publication [see APPENDIX G]. Thus,
informants had participated fully, not only in the
processes of revising pre-submission drafts, but had also
entered into the vital post-review stages of editorial
correspondence with resultant negotiation and revision.

Of the ten subjects, four were from the department of
Physics (Applied), with three each from Chemistry and
Biological Sciences (Cell Biology), thus representing an
approximate balance of disciplines compared to the main 36
(12x12x12) NS and NNS corpora. However, deviating somewhat
from selection criteria of the main corpora (see section
3.4.1), only six of the ten RAs followed the internal
structure previously examined, that is, with four separate
sections (IERD). Due to specific format requirements by
journals, the remaining four RAs were written with
combined Results/Discussion sections, with a brief
Conclusion added (IER/DC). In addition, one of the RAs
included a separate subsection for theoretical discussion
- for the purpose of analysis, this was excluded. The

discourse-functional analysis of marked and unmarked Themes in FIRST and FINAL drafts in English was carried out according to previous methods (see section 3.4.2).

Of particular interest, then, is how 'appropriately' these ten novice RA writers dealt with the text-structuring potential of marked Themes and the discourse community-interactive potential of unmarked Subjects. From analysis of published RAs above, we already have clear evidence that there are strong similarities in the usage and distribution of thematic functions in RAs produced by 'experienced', published NS and NNS writers. Therefore, rather than being an issue of NS versus NNS, the issue of 'appropriacy' here is closely concerned with aspects of writing procedures and strategies, as well as language usage, recognised as being associated with 'inexperienced', immature writers as against 'experienced', mature writers. Furthermore, the development of these aspects from relative immaturity to recognised maturity between discrete stages of writing (in this case, between drafts of scientific RAs) is of considerable pedagogic interest.

Before we can formally explore the issue of 'appropriacy' of thematic selections in FIRST & FINAL RA drafts by means of hypotheses, we need to highlight some of the relevant apsects of language usage and the RA writing process which research has suggested are generally recognised features of mature and immature writing. These issues are now discussed in relation to the current target group of L2 RA writers.

5.2.2 <u>An investigation of recognised aspects of mature and immature writing: NNS novices' scientific RA writing procedures, strategies and language usage</u>
As indicated above, the main objective here is to reveal recognised aspects of mature and immature writing in

relation to both language usage and the RA writing procedures and strategies of the core group of sixteen novice NNS researchers. To this end, previous research in both L1 and L2 fields is reviewed and presented in the framework of analysis of verbal reports generated from individual structured interviews conducted with the sixteen informants. The rationale for this research methodology is briefly reviewed below (section 5.2.2.1).

Interview questions (Questionnaire #4) were selected on the basis of relevant research on and surveys of L2 writing and tailored to the particular demands of the current EAP setting of English NNS (L1 Japanese) novice researchers attempting their first scientific RAs in English. As with previous questionnaires used here (see section 4.1.1), questions were pre-tested for potential failure on NNS subject specialist informants (see section 4.1.2). The interviews focused on six major question areas:

#1) What are the main stages involved in your writing [the first draft of] a research article to be published in English?
#2) If you write first in Japanese, what method do you use to then translate into English?
#3) In general, how many RA drafts [printed out for comment] do you have to complete for a 'good' piece of writing, ready for submission to an international English-language scientific journal?
#4) When you complete different drafts in English, what do you try to change between drafts?
#5) What is the role of your research supervisor (a) in assisting with the writing of your papers in English and (b) when is this role most important?
#6) Have you ever been formally taught about academic writing style in English? Do you read or consult books on scientific writing [either in English or Japanese] and 'how to do it'?

Based on these six questions (as indicated in the section on findings, relevant supplementary questions were added if considered appropriate), interviews were recorded and transcribed. Subsequently, information from protocols was synthesised systematically in order to create categories of response. Comments presented below (section 5.2.2.2) on those responses are integrated with "quotes" from students.

5.2.2.1 Research methodology - verbal reports

Ericsson & Simon (1985, 1987) comment that verbal reports, have been used in the field of psychology,

> 'for many widely varying purposes and have been gathered and interpreted according to quite different methodologies. In the recent resurgence of their use, they have been employed within an information-processing framework, chiefly in the study of problem solving' (1987:24).

In L2 research, Cohen (1987a) and Faerch & Kasper (1987) comment on the growing interest in using learners' reports of their own intuitions and insights as a complement to classroom observation. The analysis of such verbal reports may open to observation the manner in which L2 learners express metalinguistic knowledge about the forms and processes of their interlanguage and their conscious learning and communication strategies. To distinguish between different types of verbal report data, Cohen (1987a:84) employs the terms *self-revelation*: a learner's think-aloud stream of consciousness disclosure of thought processes while information is being attended to; *self-observation*: the inspection of specific language behaviour, either while the information is in short-term memory, that is, introspectively, or after the event, that is, retrospectively; *self-report*: learners' descriptions of what they do, characterised by generalised statements about learning behaviour.

However, Grotjahn (1987) cautions that the methodology of verbal reports as a research tool has been the subject of fierce controversy throughout its history and the main criticisms are concerned with the reliability of the different types of data elicited. Ericsson & Simon (1984) maintain that retrospective reports are incomplete, inaccurate and affected by researcher bias. Flower & Hayes suggest that introspective analysis of what people did while writing is 'notoriously inaccurate and likely to be influenced by their notions of what they should have done' (1981b:368) and for this reason they favour *self-revelation*, think-aloud research data. Cohen, however, points out that it is often difficult to establish whether learners are actually 'thinking-aloud', without analysis, or whether they are 'observing' that behaviour; it is, therefore, likely that any given verbal report may contain different types of data, with the attendent potential drawbacks indicated above.

A further critical objection to the validity of verbal reports in both L1 and L2 research is that much of language learning takes place at an unconscious level and is, therefore, inaccessible to observation or verbalisation. Moreover, the data obtained may be affected by the mode and formality of elicitation. On the one hand, some researchers comment that the validity of verbal report data must be considered high, 'when minimal intervention on the part of the experimenter takes place and no pressure to verbalize is exerted in any way' (Krings 1987:173). Conversely, Cohen comments that 'faulty data may result from an inadequate understanding on the part of the respondents as to how they are supposed to report' (1987b:38) and suggests that subjects may need to be trained in how to provide the desired form of data with specific instructions.

What can be agreed upon is that verbal reports about learning behaviour and conscious strategies are limited to

those who actually speak up; thus, self-expression is likely to be more difficult if the reporting is in subjects' L2, since 'we are, in effect, asking for a description of language processing at the level of performing one's language competence' (Cohen 1987b:36). In recognition of this, Holmes & Ramos (1991), suspect the possibility of cognitive overload on students who are asked to produce think-aloud protocols in L2 whilst attending to other L2 tasks; they found that group observation conducted in students' L1 instead provided a more effective method for elicting data. [Of course, this procedure requires researchers and specialist informants with bilingual competence as well as a homogeneous L1 group of subjects - the latter was at hand for the present research, the former was not].

Based on initial experiences with the sixteen NNS doctoral students in Academic Writing class and group/individual activities, it was anticipated that questions which required subjects to verbalise awareness of their L2 writing procedures and strategies would be problematic and indeed this proved to be the case with many subjects; responses were frequently hesitant, particularly to those questions concerned with their writing and rewriting processes. However, Ericsson & Simon indicate that 'incompleteness of reports may make some information unavailable, but it does not invalidate the information that is present' (1980:243). Nevertheless, there are clearly potential limitations to verbal reports when subjects are unfamiliar with the required metalanguage and particularly the situational role. Indeed, quite a number of the sixteen subjects found the one-to-one interview situation and the novel contents of the questionnaire surprising (and even amusing!), since they had never thought about such matters, nor had they ever considered their possible relevance for their own research writing activities. For these NNS novice scientists, there is evidently an enormous gap in their L2 communicative

potential in a relatively familiar situation (for example, talking about their work at a poster presentation at an international science conference), compared to the linguistic performance required of this highly unusual interview about personal research writing habits of which they may or may not be cognisant.

Despite the argument about what it is that protocols may or may not open to observation, numerous studies which have employed this methodology indicate that the rich data resulting from verbal reports can be a useful research tool under certain conditions and with certain limitations. Any deficiencies in methodology help to emphasise the view that 'L2 research needs an eclectic approach and should eschew rigid methodological positions' (Ellis 1989:363). Furthermore, in focusing on the writing processes of NNS novice RA writers, difficulties with L2 verbal reporting emphasise the importance of gathering other types of process data, notably those generated from post-process analysis of RA drafts and other RA writing-type activities (see sections 5.3 and 6.1).

5.2.2.2 <u>Questionnaire #4 - findings and comments</u>

#1) What are the main stages involved in your writing [the first draft of] a research article to be published in English?

Four procedures were reported by students. They either:
(a) write the full paper (that is, as complete as possible) in their L1 Japanese, and then translate this directly into English. At each stage of L2 writing, full sentences are formed and a completed #1 draft aimed for;
(b) or write an outline of the paper in Japanese notes ("only main flow of ideas") which they subsequently

translate into English. At this stage of L2 writing, full sentences are formed and the completed #1 draft aimed for;

(c) or write directly English notes which are then worked up into full sentences and the completed #1 draft;

(d) or write the #1 draft of the full paper directly in English without any of the intermediate stages described above.

Of the 16 students, 13 employed either procedure (a) [6 students] or (b) [7 students] - of this small group then, over 80% wrote first in Japanese. One student used procedure (c) with the remaining two (d). The main reason given for adopting procedures (a) and (b), that is, writing first in Japanese, was that it was considered too difficult for them at that stage in their apprenticeship, both as novice science researchers and as 'inexperienced' RA writers, to do otherwise - they were simply "not skilled enough".

In response to a supplementary question here on their perceptions of 'what makes for skilled RA writing', informants strongly indicated that to be a skilled L2 writer, they thought it necessary to be a skilled L1 writer. When asked whether they thought they were indeed skilled writers in their mother tongue Japanese, the feeling was something akin to 'naturally, it's my L1', implying any problems as novice RA writers lay more directly in simple L2 proficiency rather than in learning the L2 rhetorical skills required of a new genre. However, this remark generally contradicts research findings which suggest that 'a lack of competence in writing in [L2] English results more from the lack of composing competence than from the lack of linguistic competence' (Krapels 1990:49).

Recent studies [for example, Cumming (1989)] have raised interesting questions concerning the nature of the

interaction between writing expertise and L2 proficiency and how relevant it is to distinguish between these two. Concerning the sixteen NNS novices who were informants in the present study, it was ascertained that none had spent extended periods of time in an English-speaking environment. They had all had compulsory General English language education up to the 2nd year undergraduate level, traditionally using grammar-translation methodology. As indicated in question #6 below, none had had any previous training, in Japanese or in English, in "how to write scientific papers". Beyond these brief details, there was no attempt to objectively measure any degree of homogeneity amongst this group in terms of their English language proficiency. In respect of writing expertise, however, all novice researchers (and, in particular, these sixteen NNSs) can be considered to be at a similar stage in their academic-professional careers.

However, despite these findings on the nature of the interaction between writing expertise and L2 proficiency, it would be simplistic to underplay the basic influence of L2 proficiency on 'the extra time, effort, and patience required to get NNS researchers' papers published' (Gosden 1992a:135). The maturation of L2 writing skills is inevitably demonstrated in the development of linguistic control over the new genre and its manifest socio-rhetorical goals.

To return to responses to question #1, some informants commented that they would already have copious notes in Japanese from laboratory meetings and experiments, 'the "core" around which the paper was to be constructed' (Knorr-Cetina 1981:129); students may first try to present these experimental results at a Japanese-language conference/forum. Thus, it would be natural to start writing sections of or full RAs first in Japanese as an expansion of these preliminary research reports. It is of particular interest that the more rhetorically complex the

particular RA section is considered to be (that is, Introduction compared to Experimental), the more likely a full draft is first made of this section in Japanese. This translation stage may well add to the burden of producing an effective Introduction.

In her study of the RA writing processes of Spanish scientists, St.John (1987) reported similar writing procedures; in particular, she noted the initial writing in L1 of RA sections considered the most problematic. In comparison, in Shaw's (1991) study of the writing processes of a group of 22 English NNS research students studying in the U.K., 17 of them (77%) did not first use their L1 at all when writing up their research; this suggests the positive influence of the L2 target environment and, conversely, the extra burden for *in situ* NNS novices, such as those interviewed here, who do not share the many aspects of that potential advantage.

One writing strategy that the current group did not report here but which has previously been observed [Selzer (1983); St.John (1987)] is the 'jigsaw' approach to writing and revising texts, where researchers sit down with all the relevant (English-language) RAs they have read and lift, combine and add expressions from these papers. Swales cautions that '[EAP] instructors ought to be able to offer an alternative to such match and mix techniques' (1990b:206), but clearly we are talking here about an acceptable strategy of plagiarising language rather than scientific content and this appears to be a common technique for many NNSs. [This particular issue will be taken up in greater detail in #4 below].

#2) If you write first in Japanese, what method do you use to then translate into English?

    (a) "phrase-by-phrase translation"   n=9

```
              "translate groups of 2/3 sentences"        n=3
              "just translate into short notes"          n=1
    (b)    no translation                                n=3
```

The 13 students who reported writing in Japanese as a first stage in #1(a/b) above recorded a number of translation problems or suggested strategies to overcome potential translation difficulties:

- "I think translation of Japanese manuscript in English too difficult and I can't... I don't look at Japanese manuscript but I look at the results and data";

- "the logic [in Japanese] is different, I can't simply translate from Japanese version to English version. I make 2/3 sentences and put in together and then in English, these translate to another 2/3 sentences in English, something like that";

- "I translate phrase by phrase, then change word order, I try to write like English order as possible";

- "English start with subject and then comes verb, but in Japanese, verb comes last, so I think it's a very big problem";

- "at the writing of my first paper, I translated word for word and I failed, very strange English";

For those who in question #1 above specifically reported procedure (b), that is, using only Japanese notes:

- "I see the connections only in English, I don't include them in Japanese".

As a means of eliminating some perceived translation difficulties, the notion of Anglo-Japanese scientific writing was reported by several students who followed procedure (b):

- "the Japanese is not the usual Japanese, it's very similar to the English Japanese.. that means some passive sentences, many passive sentences".

At the stage of writing L1 notes then, incorporation of discourse features of the target L2 paper was perceived to facilitate the translation stage. Anglo-Japanese also included a greater use of overt transition markers than might normally occur in a Japanese-language scientific text. With the support of some Japanese research, Hinds (1987) comments that,

'English readers.. expect, and require, landmarks along the way. Transition statements are very important. It is the writer's task to provide appropriate transition statements so that the reader can piece together the thread of the writer's logic which binds the composition. In Japanese.. the landmarks may be absent or.. these transition devices may be more subtle and require a more active role for the reader' (1987:146).

The intermediary strategy of writing in Anglo-Japanese reported here clearly shows raised awareness of the writing process as shaped by the demands of the intended audience and the need to have greater control over the written product. Moreover, if this intermediary strategy itself gradually becomes redundant with greater experience of writing RAs, it suggests how the 'model' target procedures and strategies of 'experienced' RA writers are themselves approached in stages of development.

#3) In general, how many RA drafts [printed out for comment] do you have to complete for a 'good' piece of research writing, ready for submission to an international English-language scientific journal?

"3 or 4"   n=7
"5 or 6"   n=6

For all science researchers, stages of writing partly
reflect the many processes inherent in scientific research
and its reporting. By following the traditional scientific
research model of conjecture and refutation, 'students can
see the critical importance of continuous feedback from
literature, experiments and other data in revising and
testing hypotheses' (O'Brien O'Keefe & Journet 1985:351).
However, in addition to the scientific reality of the
'laboratory process', there is the rhetorical reality of
the 'story of the paper' (Knorr-Cetina 1981). As
illustrated by Myers (1985, 1988), the stages of writing
drafts of a scientific RA are complex, with the 'long
process of rhetorical construction leading to the drafting
of the first full version [and] an equally long process of
rhetorical reconstruction leading to the published paper'
(Swales 1990a:121).

For NNS RA writers, there is of course the linguistic
dimension of improving and polishing L2 texts as a result
of critical feedback on both content and language from
whatever suitable sources are available: research
supervisors, co-workers, journal editors and reviewers,
visiting foreign researchers, correspondence with outside
institutions, EAP teachers etc. [The nature of actual
language revisons made between drafts will be discussed
under #4]. It is certainly possible that the extra burden
of L2 proficiency may increase the total number of RA
drafts; in the course of my recent work with NNS
researchers (both with the current novices but also with
experienced academics), every referee's report I saw
automatically commented on the requirement to improve the
standard of English, irrespective of the nature of
suggested emendations. In the RA review process, this
burden is confirmed as an influential problem from the
point of view of NS journal editors and reviewers - again,
it is a question of extra time, effort and patience.

#4) When you complete different drafts in English, what do you try to change between drafts? (NB: in questions #4 and #5 subjects were free to give more than one response and this is reflected in total numbers).

    (a) After the FIRST draft:
        "correct grammar - sentence structure" n=11
        "logical linking" n= 6
        "spelling" n= 2
    (b) After SUBSEQUENT drafts:
        "improve vocabulary" n= 8
        "check careless mistakes" n= 5
        "check style" n= 1
        "check balance of sections" n= 1
    (c) "the same checks all the way through" n= 8

It may be supposed that the procedures and strategies adopted by many NNS novices may well prolong the RA redrafting process and consequently the total number of drafts indicated in #3 above:

- "At first I write very roughly, don't check spelling or details of grammar and at that stage most important thing is what to write.. at second version I check.. difficult due to inexperience".

This compares with the majority (77%) of Shaw's (1991) informants' initial 'rough' or 'bad' English. It is clear that for novice NNS researchers there is a basic dichotomy. Through their supervisors (like their students, L1 Japanese) and their own reading of RAs, they may be aware of 'model' RA products, and also possibly optimal procedures and strategies that should be their objective as 'good' RA writers. Some students commented that they kept what were considered to be 'good' writing models of key papers in their fields: "my supervisor told me this was a good paper". [The role of NNS novices' research supervisors in the L2 writing process is

discussed under #5 below]. However, given the dual
constraints of their inexperience as both apprentice
science researchers and as novice RA writers who need to
fully concentrate on the science and the mechanics of L2
writing, they may only be able to achieve a partial focus
on these 'ideal' procedures when writing up their research
results:

- "a higher technique is needed but I can't do it well yet".

However, the comment from one informant above about his
'poor' writing strategy being "due to inexperience" may
rather indicate a false NNS perception of the writing
procedures and strategies of NSs who are thought not to
imitate his 'poor' procedure, but who skilfully practise
'higher techniques'.

In the process literature (see section 1.3), major
research in the field of L1 writing [see, for example,
Flower & Hayes (1977, 1980, 1981a/b, 1983)] has focused on
highlighting the different procedures and strategies used
by skilled and less skilled writers as a result of the
three processes of *planning* (generating ideas and setting
goals), *translating* (writing) and *reviewing* (evaluating
and revising). For example, in a business setting, Flower
(1989) showed that 70% of material generated at the
planning stage by inexperienced writers were statements of
content (that is, what they knew); in contrast, 69% of
material generated by experienced writers were statements
of rhetorical planning (for example, goals, audience,
implications). Thus, they suggest that less skilled
writers are more concerned about displaying their
knowledge, whilst more skilled writers use rhetorical
strategies to 'connect' with the reader, thereby
distinguishing between writer-based and reader-based
prose.

In terms of the *reviewing* process, which is of parallel
relevance here to NNS novices' attempts to redraft
scientific RA sections in response to feedback and
criticism, Zamel (1983) found that the more skilled L2
writers in her study delayed consideration of lexical and
grammatical problems until the final stages of their
writing; in contrast, the least skilled writer lost track
of her ideas whilst writing because she was determined not
to commit errors and, therefore, paid premature attention
to them. Faigley & Witte also reported a preferred
top-to-bottom approach amongst experienced L1 writers,
noting that they,

> 'clean[ed] up their manuscripts after they had satisfactorily
> dealt with their subjects. By this point inexperienced students
> had largely quit revising' (1981:409).

Beach (1976) and Sommers (1980) observed that immature L1
writers assumed their intended meaning was present in a
first draft and that, at the reviewing stage, there was
really no need to add more explanation and detail or
rearrange ideas to make their paper better. Similarly,
working as an EAP specialist with a group of NNS
researchers who were in the process of writing scientific
research papers in English, St.John observed resistance
to anything but changes at a local level, since,

> 'suggestions of changes at the rhetorical level to clarify the
> significance of information were seen as unnecessary. Changes
> related to grammar are more easily accepted than any relating to
> discourse or rhetoric' (1987:119).

In contrast to these more immature revising strategies,
then, research indicates that for experienced writers the
first draft helped define the territory and further
revision and redrafting helped them to continue to create
meaning and find the line of the argument.

If we look at replies to question #4 in detail, we see that eight students make the same checks all the way through; thus 50% of this sample of NNS novice RA writers suggest that the process of redrafting does not itself imply the refinement of textual revisions in RA sections from more macro, discourse-level changes, for example, manipulating thematic and information structure, to more micro, word-/phrase-level changes, for example, checking spelling. Thus, it appears that revision is generally seen by these writers as simple mechanical editing, rather than the opportunity to continue to create meaning. On the other hand, it appears that others in this group, manifesting 'better' revising strategies, only anticipate producing a satifactory, grammatically correct and logical draft at the second attempt, with subsequent drafts aimed at producing "NS-like" variety of expression and appropriate stylistic choices, as well as tidying up "careless mistakes".

In relation to this latter point, all informants here, as did the majority of Shaw's (1991), expected to widen their active vocabulary by lifting from published NS RAs "very smart expressions and idioms" which students know are "at least correct". One student commented,

- "I choose famous authors and addresses... I underline some good points and use the vocabulary, patterns, for example 'proceeded cleanly', 'there has been much interest', or Figure captions. I underline such points and use them when I'm writing".

The above commentary on responses to question #4 has generally highlighted aspects of mature and immature procedures and strategies in relation to L2 writing. To focus now more specifically on the aspect of language usage in relation to the present concern with NNS novices' FIRST & FINAL RA drafts, a report by Sharples (1988) suggests features which differentiate mature and immature [L1] writing at the three different levels of word/phrase,

225

sentence and section/text. A variety of sentence-level features are of particular interest to the present investigation of marked thematic development between RA drafts:

FOUND IN IMMATURE WRITING
- Coordination of sentences by and; but; then.

FOUND IN MATURE WRITING
- A wider use of cohesive devices
- Multiple levels of subordination

Sharples cautions that 'the presence of a "mature" feature is not a guarantee of quality, nor can the features simply be summed to give a single "maturity" score. They must be read as signposts that point, more or less accurately, toward writing produced at an early or late stage of cognitive development' (1988:7). In a study of cohesive devices in writing, Connor (1984) observed the lack of variety of cohesive devices employed in L2 learners' writing compared to 'good' NSs' writing and comments that this use of cohesion is developmental in nature.

In the light of these findings, when later considering the textual revisions made between NNS novices' unpublished FIRST and later FINAL RA drafts, the central research hypotheses are:

[HYPOTHESIS #7a]
In recognition of their relative inexperience as both RA writers and as science researchers, there will be notable differences in the usage and distribution of marked and unmarked thematic selections in the mini-corpus (N=10) of novices' unpublished FIRST drafts when compared to the genre-specific norms suggested by the main published NS corpus.

[HYPOTHESIS #7b]

It is expected that FIRST drafts will show evidence of recognised features of immature writing which are redrafted towards recognised more mature features. In particular, in relation to marked Theme choices, FIRST drafts are expected to exhibit a greater degree of simple coordination of structures by 'minimal' Context Frames <u>and</u>, <u>but</u>, <u>then</u>; in contrast, FINAL draft sentences will show a greater range of contextualising Frames and subordinate structures. Thus, comparison between thematic selections in FIRST & FINAL drafts will show a progressive movement towards conventional usage as represented by published NS models, as a result of the 'norm-developing' process of external review, negotiation and revision.

These specific hypotheses relating to language usage will be addressed in sections 5.2.3 and 5.2.4 below.

#5) What is the role of your research supervisor (a) in assisting with the writing of your papers in English and (b) when is this role most important?

(a)	"checks the English"	n=9
	"grammatical construction"	n=4
	"logical organisation"	n=3
	"Spoken English"	n=1
(b)	"first draft most important"	n=7
	"final stage"	n=1
	"the Introduction"	n=1
	"all the way through"	n=1

The most popular response "checks the English" is an example of the tendency towards generalisation when subjects may lack the metalanguage to be able to describe and analyse their writing in more detail - a problem mentioned at the beginning under methodology when informants are being asked to report in L2. It may also be

a reflection of the fact that, although these NNS novices anticipate that their English will be checked by their NNS research supervisors, matters of scientific content remain a priority at all stages of consultation. However, again based on personal experience working with NNS researchers, it can be suggested that the subtleties of expression in L2 employed in order to reflect the subtleties of scientific debate are perceived to a lesser degree, certainly by 'inexperienced' novice NNS researchers, but also as well by their more 'experienced' supervisors.

#6) Have you ever been formally taught about academic writing style in English? Do you read or consult books on scientific writing [either in English or Japanese] and 'how to do it'?

All 16 students answered "no" to the first part of the question and "yes" to the second part. Whereas the majority of Shaw's (1991) informants noted standard monolingual or bilingual dictionaries as their main reference source, here the most common type of material used was a Japanese-English reference text for science students. Some manuals specialise in particular subject areas, for example, chemical engineering, and give detailed Japanese-English translations of technical terms as well as hints on style and structure. Only a few students used English-language manuals which had been recommended by their supervisors, intended for English NSs [for example, Day (1979); Trelease (1982)] or specifically for NNSs [for example, Huckin & Olsen (1983)]. One student commented proudly that his laboratory library had 26 reference books on 'how to write academic papers in English' but how such help is generally used is illustrated by the comment:

- "I have not read all of the parts because I just look when I make mistake".

For these novice NNS RA writers, such manuals appear to be the most readily available resource in the workplace and it is particularly striking here that there is a history of a complete absence of direct face-to-face tuition in English for Academic Purposes, at this or at any previous level. In view of this, it is not surprising that the apprenticeship as a research student is considered the most powerful pedagogic relationship [Myers (1988); Spack (1988)]. When *in situ* tuition in the form of EAP courses is available, therefore, it is clear that a dual approach from within both language and subject departments is vital for effective support.

5.2.2.3 Concluding remarks on recognised features of mature and immature writing: NNS novices' RA writing procedures, strategies and language usage

Comments throughout this questionnaire, primarily on their L2 RA writing procedures and strategies, indicate that many of this group of NNS novice researchers are well in the process of refining their knowledge about aspects of interactive research communication. This process occurs as they mature as researchers and as RA writers by means of participation in the established mechanisms of communication of the discourse community. Comments demonstrate that they have already developed insights about the many standardised conventions of the written RA product through their reading and they will improve their RA writing skills and learn the broader 'rules of the academic game' through increasingly active participation. The data presented here may well be quite representative of many of the procedures and strategies adopted by NNS novice science researchers around the world learning their research reporting skills in the workplace, particularly those without access to formal instruction in English for Academic Purposes.

This investigation into aspects of L2 RA writing procedures, strategies and language usage gives rise to the interesting question concerning the interaction between writing expertise and L2 proficiency and how relevant it is to distinguish between these two. The initiation into the international academic community by means of publications is common (but not, of course, identical) for all researchers around the world, irrespective of L1 origin. It can be suggested, therefore, that developing expertise in the form of awareness of the social-constructionist nature (see section 1.3) of the 'hard', 'norm-developing' processes (Swales 1988, 1990a) of RA drafting, feedback, negotiation and redrafting, and of the procedures and strategies required to achieve these, is of primary importance to 'success' in research publication.

As indicated in the introduction to perspectives on composition theory and research, prominent models of the writing process [notably Flower & Hayes (1981)] may not be considered particularly useful in the genre-based context of English for Academic Purposes (see section 1.3 for a review of major criticisms). In particular, certain kinds of writing, such as the scientific texts which are the focus of interest in the present work and which have particular demands and constraints, are ignored in the cognitive process model and literature. Clark & Ivanič report their attempts to develop their own model of the academic writing process in an L2 EAP context, one which may more adequately differentiate the unique nature of the extended processes of subsequent negotiation and revision in response to the crucial activities of criticism and feedback:

'We felt that there are more components to the process than such models [Flower & Hayes (1981)] represent, and we felt intuitively that the process itself is more varied and

complicated than any diagram we had seen in the literature'
(1991:171).

Thus, in their academic writing-oriented process model,
Clark & Ivanič (1991) emphasised more strongly subtle
affective and social aspects of the writing process, such
as, *clarifying your commitment to your ideas* and *deciding
how to take responsibility: whether to mask or
declare your own position*. These subtleties are
indispensible in the negotiation of knowledge claims in
correspondence with journal editors in response to
critical reviews and naturally lead to a focus on the
linguistic realisations of such claims (for example, the
use of hedging devices).

Despite the limitations mentioned above (see section
5.2.2.1) of the methodological approach adopted here, it
can be suggested that the potential value of these verbal
report data to the design of EAP writing courses is
considerable:

> `As we find out more about the processes that learners use, we
> are better equipped to test hypotheses about strategies that we
> would predict are likely to produce the greatest success for
> given types of learners' (Cohen 1987a:92).

Verbal data can thus provide important insights to enhance
learners' attention to language input and the written task
environment, which, in turn, help to demystify academic
discourse (Spack 1988) and provide a strong rationale for
a genre-based approach to the teaching of writing. The
methodology also provides a solid student-centred
framework for the exploration of teaching and research
tools which enable the application in the EAP classroom of
the range of insights highlighted here about the RA
writing process; of interest are redrafting procedures and
strategies used to help create more 'successful' RAs,
particularly those concerning the 'appropriate'

manipulation of Theme. A detailed exploration of such a teaching/research tool is the subject of Chapter Six.

The final sections of this chapter return to the context introduced in section 5.2.1 above. They present both a quantitative and qualitative investigation of NNS novices' FIRST and FINAL RA drafts and focus on how 'appropriately' they deal with thematic structure, in particular, with the text-structuring potential of marked Themes and the interactive potential of unmarked Subjects as highlighted in Chapter Three.

5.2.3 Analysis of NNS novices' FIRST & FINAL RA drafts - findings and comments

In principle, the method of analysis adopted here also replicates that used with the published NS (see sections 3.4.1 and 3.4.2) and NNS (see section 5.1.2) corpora described above. However, in contrast to certain selection criteria for the 36 published RA corpora, here four of the ten unpublished sets of FIRST and FINAL RA drafts had a combined Results & Discussion section; separate data are given where relevant. The limited size of this mini-corpus of unpublished RA drafts (N=10) implies that quantitative comparison remains speculative.

Findings and comments are briefly given in parallel to those reported for the NS and NNS corpora above, but data focus on changes from FIRST to FINAL RA drafts. Relevant data from the base *NS corpus* are repeated here for convenience of comparison and statistical correlations are given. It should be recalled that since there were striking similarities in reported data between the 'experienced' published NS (sections 3.7 & 3.8) and NNS (section 5.1) RA data, of greater emphasis here are distinctions between these RA products and the relatively more immature writing of 'inexperienced' NNS novices - issues which address [HYPOTHESES #7a/b] (section 5.2.2.2).

5.2.3.1 SIEs - marked vs. unmarked Themes

For an initial overview, TABLE 25 indicates the distribution of SIEs in the mini-corpus (N=10) and the breakdown of GSs and CFs.

TABLE 25
Number of SIEs in NNS novices' (N=10) FIRST & FINAL RA Drafts
- *Comparison with the NS corpus*

Total N	FIRST > (863)	FINAL (1029)	NS corpus
SIE = GS	60.4%	61.6%	*(67.2%)*
SIE = CF	39.4%	38.3%	*(32.3%)*
SIE = non-GS/CF	0.4%	0.1%	*(0.5%)*
Total SIEs	100%	100%	*(100%)*
[SD][1]	[10.9]	[7.2]	*[8.2]*

[1] Standard Deviation of Context Frames

Of note is the overall higher usage of CFs in novices' RAs compared with the main NS corpus and the direction of decrease of CFs in FINAL texts, that is, slightly towards this norm. The lower Standard Deviation value in students' FINAL texts indicates a narrower range of CF usage.

5.2.3.2 GSs vs. CFs by RA section

TABLE 26 breaks down the above information by RA section in order to elucidate where CFs were used more and less frequently compared to NS norms. Due to the smaller sample of draft RAs (N=10), percentages are rounded to 1% throughout.

These data indicate that in novices' Introductions, nearly one-half of SIEs are CFs, as against the one-third of the NS corpus. In fact, the FINAL draft showed an increase to 49%. Standard Deviations indicate a much wider variation of usage amongst novices' Introductions [25] (versus NSs [*15.3*]) compared to other major sections.

TABLE 26

Distribution of Subjects and CFs by RA Section and Changes between
FIRST & FINAL RA Drafts - *Comparison with the NS corpus*

	FIRST>FINAL CHANGE			FIRST>FINAL CHANGE			*NS corpus*	
RA section	Int	Int		Exp	Exp		Int	Exp
N texts	N=10	N=10		N=10	N=10		N=36	N=36
SIE = GS	53%	51%	-	82%	86%	+	*66.1%*	*80.7%*
SIE = CF	47%	49%	+	18%	14%	-	*32.9%*	*19.2%*
SIE = non[1]	0%	0%	nc[2]	0%	0%	nc	*1.0%*	*0.1%*
[SD][3]	[25]	[25]		[12]	[13]		*[15.3]*	*[10.4]*

	FIRST>FINAL CHANGE			FIRST>FINAL CHANGE			*NS corpus*	
RA section	Res	Res		Dis	Dis		Res	Dis
N texts	N=6	N=6		N=6	N=6		N=36	N=36
SIE = GS	57%	62%	+	55%	56%	+	*65.8%*	*60.5%*
SIE = CF	42%	38%	-	44%	44%	nc	*33.8%*	*39.0%*
SIE = non[1]	1%	0%	-	1%	0%	-	*0.4%*	*0.5%*
[SD]	[11]	[8]		[15]	[10]		*[12.7]*	*[13.2]*

	FIRST>FINAL CHANGE			FIRST>FINAL CHANGE			*NS corpus*[4]
RA section	R/D	R/D		Con	Con		
N texts	N=4	N=4		N=4	N=4		
SIE = GS	52%	55%	+	83%	74%	-	
SIE = CF	48%	45%	-	17%	26%	+	
SIE = non[1]	0%	0%	nc[1]	0%	0%	nc	
[SD][2]	[15]	[15]		[25]	[17]		

[1] *non GS/CF*
[2] *nc = no change*
[3] *Standard Deviation of Context Frames*
[4] *No comparison possible here due to separate Results and Discussion sections in the NS corpus*

Results and then Discussion sections indicate the next clearest difference; it is interesting to note that in

234

Results, final drafting reduced the percentage of CFs. Standard Deviation values indicate much closer agreement over CF usage in these sections. As regards combined Res/Dis sections, CF usage is roughly equivalent to that in the Discussion section, indicating the nature of the combined section as cycles of Result + Comment/Discussion.

5.2.3.3 Context Frames by function

TABLE 27 examines the discourse-functional roles of CFs, an indication of their function as a means of textual development.

TABLE 27
Distribution of CFs by Function in FIRST & FINAL RA Drafts
- *Comparison with the NS corpus*

	FIRST>FINAL		CHANGE	*NS corpus*
1a. Location in Time - RWE	13%	12%	–	*11.2%*
1b. - DE	1%	2%	+	*2.8%*
2a. Location in Space - RWE	13%	13%	nc	*14.9%*
2b. - DE	3%	1%	–	*4.2%*
3a. Addition - appositive	1%	1%	nc	*1.1%*
3b. - emphatic	12%	10%	–	*8.4%*
4. Contrast/concession	16%	16%	nc	*18.3%*
5a. Cause - reason/result	17%	20%	+	*13.0%*
5b. - purpose	5%	8%	+	*5.3%*
6. Means	1%	2%	+	*2.0%*
7a. Condition - real	11%	8%	–	*8.4%*
7b. - hypothetical	2%	1%	–	*2.6%*
8a. Validation - external	2%	2%	nc	*3.6%*
8b. - internal	2%	3%	+	*3.3%*
9. Viewpoint	1%	1%	nc	*0.9%*
	100%	100%		*100%*
correlation with the *NS corpus* $r =$	.946	.923		

As between the published NS and NNS corpora, there is a striking correlation here which on the surface indicates an understanding of the rhetorical 'appropriacy' of respective CFs and how they contribute to the within-text structuring of RAs. A notable difference in usage between

'experienced' and 'inexperienced' RA writers' data lies in the increase of Cause-reason/result CFs to 20% in the FINAL drafts (vs. *13.0%*). [It should be pointed out, of course, that the correlation with the NS corpus actually decreases FIRST to FINAL].

5.2.3.4 CFs by function and type

Johns (1984) has commented that NNS may overuse 'minimal' CFs (such as additive *for example, and, or, in addition*). The separation of CFs by function and type may indicate whether this perception is valid and will thus address [HYPOTHESIS #7b]. TABLE 28 first describes the grammatical paradigm only, followed by both grammatical and functional paradigms in TABLE 29.

TABLE 28
Distribution of CF Types in FIRST & FINAL RA Drafts
- *Comparison with the NS corpus*

	FIRST	FINAL	CHANGE	*NS corpus*
CF Type 1	47%	44%	–	*37.5%*
CF Type 2	32%	31%	–	*38.0%*
CF Type 3	21%	25%	+	*24.5%*
	100%	100%		*100%*

As can be seen from TABLE 28, this group of ten novice RA writers did indeed use a greater percentage of CF Type 1 'minimals' - 47%>44%, as compared with *37.5%* for the main NS corpus. Swales comments that NNSs may 'signpost unerringly - even by overincorporating metadiscourse items... to compensate for the anticipated semantic and registral uncertainties' (1990b:204). However, NNS corpus data from TABLE 23 above (see section 5.1.3.5) suggest that overincorporation of CFs may be more clearly a feature of immature, 'inexperienced' versus mature, 'experienced' RA writing.

TABLE 29

CF Usage according to Functional and Grammatical Paradigms in FIRST & FINAL RA Drafts - *Comparison with the NS corpus*

	FIRST>FINAL CHANGE			FIRST>FINAL CHANGE			NS corpus	
	CF TYPE 1			CF TYPE 2			CF1	CF2
Location in Time-RWE	13%	8%	−	19%	20%	+	7.0%	17.9%
−DE	2%	3%	+	0%	0%	nc	6.4%	*[1]
Location in Space-RWE	0%	0%	nc	39%	43%	+	1.1%	37.6%
−DE	2%	1%	−	7%	2%	−	*	10.8%
Addition − appositive	2%	3%	+	0%	0%	nc	3.0%	*
− emphatic	21%	17%	−	3%	7%	+	19.0%	3.4%
Contrast/concession	30%	31%	+	2%	2%	nc	31.1%	3.9%
Cause − reason/result	23%	33%	+	6%	3%	−	22.1%	4.7%
− purpose	0%	0%	nc	1%	2%	+	*	2.2%
Means	1%	1%	nc	0%	2%	+	*	*
Condition − real	3%	1%	−	13%	7%	−	1.7%	11.8%
− hypo	1%	1%	nc	1%	1%	nc	1.3%	*
Validation − external	0%	1%	+	6%	5%	−	2.3%	5.6%
− internal	0%	0%	nc	3%	4%	+	1.7%	1.1%
Viewpoint	2%	0%	−	0%	2%	+	2.3%	*
	100%	100%		100%	100%		100%	100%

correlations with
the *NS corpus*   r = .972 .964   .990 .952

	FIRST>FINAL		CHANGE	NS corpus
	CF TYPE 3			CF3
Location in Time − RWE	7%	8%	+	7.2%
− DE	0%	1%	+	1.2%
Location in Space − RWE	0%	0%	nc	*
− DE	0%	0%	nc	*
Addition − appositive	1%	0%	−	*
− emphatic	6%	1%	−	*
Contrast/concession	7%	8%	+	21.1%
Cause − reason/result	19%	19%	nc	13.9%
− purpose	24%	30%	+	15.7%
Means	4%	4%	nc	6.4%
Condition − real	23%	21%	−	13.3%
− hypothetical	6%	3%	−	8.1%
Validation − external	0%	0%	nc	2.6%
− internal	3%	5%	+	9.3%
Viewpoint	0%	0%	nc	*
	100%	100%		100%

correlation with
the *NS corpus*   r = .727 .757

[1] * *denotes value under 1.0%*

237

If CF Type 1 minimals are 'overused' by inexperienced writers, it may be assumed that it is the more complex CF Type 3 which are consequently 'underused' as they frequently serve similar CF functions. However, the data in TABLE 28 show that the FINAL drafts use almost the same average proportion of these Type 3 CFs (25%) compared to the NS corpus (*24.5%*). Nevertherless, the correlation between NS and novices' FINAL drafts usage is certainly less significant ($r=.757$) than between CF Types 1 and 2. It is the adverbial/prepostional group of CF Type 2 that shows the greatest difference in overall percentages - 32%>31% as against *38.0%*. However, with the corpus size of drafts RAs (N=10), it is unwise to speculate on any significance here. In addressing [HYPOTHESIS #7b] above, that recognised features of immature writing are redrafted towards recognised more mature features, comments will therefore mostly concentrate on CF Types 1 and 3:

CF Type 1: The usage of the CF function of Cause-reason/result shows an interesting contrast - an increase from 23% in FIRST drafts to 33% in FINAL versions, (compared to *22.1%* for the NS corpus). *Therefore* and *thus* mainly accounted for this increase in Results and Discussion sections, indicating writers' attempts to more clearly emphasise their rationale for research actions and speculation about findings (see APPENDIX H for full CF Type 1 data from FIRST & FINAL drafts; see APPENDICES C1 and F1 for comparison with main NS and NNS corpora).

Concerning the usage of minimal coordinators *and*, *but*, *then*, concordances [using *MicroConcord* (Scott & Johns 1992), Oxford English Software] in TABLE 30a (FIRST), TABLE 30b (FINAL) and a summary (TABLE 30c), indicate their virtual disappearance between FIRST & FINAL drafts.

TABLE 30a

Corpus (N=10) of FIRST Drafts: Concordances of CFs Type 1

and (N=13); but (N=10); then (N=14)

cases. And possible factor for polyamine's effects on calf [2]
0.3 mM. And second, they lost their activation effects in t [2]
rature. And the potential of the specimen was kept constant [5]
strain. And the applied plastic strain values were varied i [5]
aining. And current decreased after the deformation was sto [5]
 other. And according to the result, the relation between t [5]
mation. And this result clarifies the fatigue crack propoga [5]
d ways. And the dependence of Hc on the total thickness dep [7]
reased. And when the thickness layer of Fe was constant, th [7]
r line. And the dependance was clearer in a) than b). Then [7]
g.7-b). And in Fig.7-c), the steps began to disappear. [7]
 first. And this columnar growth might be saturated when th [7]
change. And other enzyme inhibitor such as several glycosid [10]

tained. But S/N ratio was not enough to drive the wavenumbe [1]
rs (J). But the plots of transition wavenumbers in Table II [1]
ar one. But CO is not well known better than N_2 about their [1]
ne(24). But still there is not any information about sperma [2]
midine. But they were rather inhibitory at 0.3mM, producing [2]
chains. But at low temperatures, the weak interchain antife [6]
rystal. But in reality, a small birefringence due to the mi [6]
educed. But n^{ac} of $CsMnI_3$ have the high temperature tail of [6]
period. But the values became larger as the film period inc [7]
hnique. But the lattice relaxation proceeds further. These [9]

tively. Then, triethylenetetramine which was the same as sp [2]
mation. Then, the fatigue crack propogation tests were carr [5]
$KNiCL_3$. Then the DM interactions which are the origin of th [6]
 moved. Then the configuration of the interactions becomes [6]
for 2h. Then these three condition type specimens were comp [7]
d mica. Then we measured the step width in some films with [7]
f them. Then we made some multilayer films which had two Fe [7]
han b). Then as for one film heated up by two way described [7]
urface. Then the grains became larger as they grew. Saying [7]
Fig.10. Then the region which had same crystallographic ori [7]
 films. Then these properties may be measured another multi [7]
stress. Then the mismatch stress is relieved by the disloca [9]
give 3. Then 1,2-isopropylidene group of 3 was hydrolysed b [10]
overed. Then two step conversion was carried out, epoxidati [10]

[1-10] superscript numbers on the right refer to student[1], student[2] etc.

TABLE 30b

Corpus (N=10) of FINAL Drafts: Concordances of CFs Type 1

and (N=0); but (N=1); then (N=5)

---

t 300K. But n^{ac}(T)'s of CsMnI$_3$ and CsMnBr$_3$ contain high-tem [6]

states. Then the 1682-47983 configurations were constructed [1]
thanol. Then the precipitated DNA was dissolved in 10 ul of [2]
stress. Then the reversed plastic zone size is proportional [5]
atures. Then, in this study birefringence measurements ha [6]
c-axis. Then it may well be said that the symmetry of the c [6]

TABLE 30c

Corpora of main NS, NNS RAs and Novices' FIRST and FINAL Drafts: CFs

Type 1 and; but; then - Summary and Comparison

	NS corpus (N=36) [APPENDIX C1]	NNS corpus (N=36) [APPENDIX F1]	FIRST > FINAL (N=10) [APPENDIX H]
And	0	1	13 > 0
But	1	7	10 > 1
Then	3[a]	32[b]	14[c] > 5[d]

---

[a] Condition-real: N=3
[b] Condition-real: N=22; Location in Time-RWE: N=10
[c] Condition-real: N=2; Location in Time-RWE: N=8;
   Cause-reason/result: N=4
[d] Condition-real: N=1; Location in Time-RWE: N=2;
   Cause-reason/result: N=2

Although the changes between FIRST & FINAL drafts confirm [HYPOTHESIS #7b], that recognised features of immature writing are redrafted towards recognised more mature features, TABLE 30a clearly indicates that the usage of minimals *and, then, but* is highly characteristic of the individual NNS novice's relative level of RA writing (im)maturity. On the one hand, of the ten students, three did not employ these coordinators at all; on the other, student[5] and student[7] account for 10 out of 13 examples of *and.*

CF Type 2: FINAL drafts indicate a much lower usage of strongly interactive Location in Space-Discourse Entity CFs (*in this paper, from Figure 2*) than the NS corpus (2% vs. *10.8%*), although curiously the FIRST drafts showed 7%.

CF Type 3: The usage of Contrast/concession CFs was much lower in novices' drafts (7%>8%) than in the NS corpus (*21.1%*) – this 'underuse' does not appear to be compensated for by means of a greater usage of CF Type 1 minimals (30%>31% as against *31.1%*). Since contrast is a powerful means of textual development, a lack of overt CF signals may represent a weakness in thematic progression. (Of course, marked Themes are only one means of development through contrast). Another distinctive pattern with CF Type 3 is the higher usage of real Condition CFs (*when..,*) in novices' drafts (23%>21%) compared to the *13.3%* of NS RAs. In view of [HYPOTHESIS #7b] and an interest in evidence of modifications from coordinate to subordinate structures, concordances for the most notable CF Type 3 FIRST>FINAL draft change (24%>30%), Cause-purpose [*(in order) to..*], are given in TABLES 31a/b below. It is interesting to note the increased use of the full formal *In order to..* in preference to the short form *To...*

TABLE 31a

Corpus (N=10) of FIRST Drafts: Concordances of CFs Type 3

Cause-purpose (N=16)

ically.	To understand eucaryotic topoisomerase I reaction a 2
amines.	To elucidate whether the size of polyamines or the 2
ffects.	To investigate the interaction of these unusual pol 2
action.	To see whether polyamines stabilze tthe intermediat 2
ig.2b).	To determine the propogation velocity of CGB, we an 4
times.	To analyze quantitatively how the breakdown of a CG 4
plate.	To remove the effect of the solution surface flucta 5
bolism.	To study the DOI formation reaction, it is essentia 10
tivity.	To apply this technique, DOI is required in quantit 10

ations.	In order to remove salts and polyamine, samples wer 2
id DNA.	In order to investigate the effect of the GC conten 2

241

e(2,3). In order to investigate the role of the mitotic app [3]
s used. In order to investigate mitotic chromosomes more pr [3]
icolin. In order to investigate the relationship between th [3]
TN)/TN. In order to search the best fit of the data to (17) [6]
ements. In order to examine the effect of heat treatment on [7]

TABLE 31b

Corpus (N=10) of FINAL Drafts: Concordances of CFs Type 3

Cause-purpose (N=28)

---

state. To confirm the experimental assignment and to deduc [1]
action. To elucidate the mechanism for the formation of the [2]
sheet. To determine the propogation velocity of CGB, we an [4]
ng CGs. To analyze quantitatively how the breakdown of a CG [4]
ecimen. To remove the effects of the solution surface fluct [5]
$+n^{ac}m$. To extract $n^{ac}m(T)$ from the observed $n^{ac}m$, the latt [6]
$^{ac}m(T)$. To see the low temperature behavior of the lattice [6]

a lens. In order to monitor the scanning of the dye laser, [1]
hwhile. In order to understand the interaction of polyamine [2]
ations. In order to remove salts and polyamine, the mixture [2]
action. In order to investigate this possibility, the fully [2]
midine. In order to understand their physiological roles in [2]
e(2,3). In order to investigate the role of the mitotic app [3]
oplasm. In order to visualize chromosomes, the histones wer [3]
s used. In order to investigate mitotic chromosomes more pr [3]
icolin. In order to investigate the relationship between ch [3]
l shot. In order to analyze CGB, the sheet was divided into [4]
zation. In order to calculate the propogation velocity of C [4]
UCTION. In order to determine the process of CGB in a small [4]
ucture. In order to achieve those aims, appropriate estimat [6]
Tn)/Tn. In order to obtain an excellent fit of the data to [6]
plane. In order to check the above anisotrophy, an Fe thin [7]
ported. In order to elucidate the origin of the steps, we m [7]
layers. In order to confirm this hypothesis, we fabricated [7]
reases. In order to investigate the nature of such loops in [7]
H loop. In order to explain the relationship between magnet [7]
mation. In order to construct an *in vitro* system for any bi [10]
hnique. In order to develop a convenient analytical method, [10]

In general, compared to concordances of Type 1 CFs above, the usage of these Cause-purpose CFs appears to be less idiosyncratic. We will return to the relevance of these above concordances for [HYPOTHESIS #7b] in the more

qualitative analysis of NNS novices' FIRST & FINAL drafts
presented in section 5.3 below.

5.2.3.5 Distribution of Subject roles
Finally, we turn to unmarked Themes, that is, Subjects and
their discourse-functional roles throughout RA sections.
Compared to the text-organising potential of CF patterns,
the interest here lies in the dynamic patterning of
interactional Themes throughout RA discourse. It may be
suspected that NNS novices may not appreciate and hence
employ the full range of subtle interactive thematic
choices presented through non-Real World Themes (see
section 3.6 above). TABLE 32 below compares Subject roles
by domain between novice FIRST>FINAL drafts and the NS
corpus.

The correlations between this mini-corpus of 10 RAs and
the NS corpus continue to be significant (p<0.05), even
though FIRST>FINAL drafting did not actually improve this
already strong correlation. The total for the three
non-Real World domains for novices' drafts is 18.0%>19.0%
compared to *22.8%* for the NS corpus (20.5% for the main
NNS corpus). In particular, the data show a relative
'underuse' of the more interactional Participant
(6.9%>6.5% vs. *9.2%*) and H&O (4.7%>5.5% vs. *7.6%*) domains.
Davies (1988a, 1991) considers that Objectivised
Viewpoints (OV), as given statements of fact, and
'invisible Subjects' [here, Empty H&O Theme], as a subtly
hidden means of expressing evaluative comment may present
particular problems for novices in their reading and
writing of RAs. There is some evidence of an attempt
through FIRST>FINAL redrafting to increase the potential
of evaluative comment through Objectivised (0.8%>1.8% vs.
2.0%) and Hypothesised (0.1%>0.5% vs. *0.6%*) Viewpoints
(HV).

TABLE 32
Distribution of Subject Roles by Domain in FIRST and FINAL RA Drafts
- *comparison with the NS corpus*

Subject Role Domain	Mean % FIRST	>	Mean % FINAL		Mean % *NS corpus*
Participant domain					
Discourse Participant	5.0%		4.1%	–	*5.7%*
Participant Viewpoint	0.6%		0.6%	nc	*0.6%*
Interactive Participant	1.3%		1.8%	+	*2.9%*
Total this section	6.9%		6.5%	–	*9.2%*
Discourse domain					
Discourse Event/Process	0.2%		0.5%	+	*1.0%*
Macro Discourse Entity	0.1%		0.2%	+	*0.2%*
Micro Discourse Entity	4.1%		4.6%	+	*2.1%*
Interact. Discourse Entity	0.7%		0.7%	nc	*1.6%*
Empty Discourse Theme	1.3%		1.0%	–	*1.1%*
Total this section	6.4%		7.0%	+	*6.0%*
H&O domain					
Objectivised Viewpoint	0.8%		1.8%	+	*2.0%*
Hypothesised Viewpoint	0.1%		0.5%	+	*0.6%*
Hypothesised Entity	0.0%		0.5%	+	*0.6%*
Empty H&O Theme	3.8%		2.7%	–	*4.4%*
Total this section	4.7%		5.5%	+	*7.6%*
Real World domain					
Real World Entity	66.3%		64.8%	–	*56.1%*
Real World Event/Process	12.7%		12.8%	+	*17.9%*
Mental Process	0.3%		0.4%	+	*1.2%*
Empty Real World Theme	2.7%		3.0%	+	*2.0%*
Total this section	82.0%		81.0%	–	*77.2%*
	100%		100%		100%
correlations with the *NS corpus* $r =$	.991		.991		

However, again as with *and, but, then* data above, concordances in TABLES 33a/b of examples of Objectivised (OV) and Hypothesised Viewpoint (HV) indicate the predominance of usage by individual novices[6,9,10]. Perhaps curiously, Empty H&O Themes themselves decline (3.8%>2.7% vs. *4.4%*) between FIRST>FINAL drafts (see APPENDIX J for data).

TABLE 33a

Corpus (N=10) of FIRST Drafts: Concordances of H&O domain

OV (N=7) and HV (N=1)

OV ion. The *reason* why RbMnBr₃ takes the IC spin structure has no	6
OV far. The *key to the problem* probably reside the fact that RbMn	6
OV OK. This *fact* was not known when the structure of room-tempera	6
OV ure. The *major feature* of this technique involves the utilizat	9
OV nt. This *problem* must be solved by using the low-temperature m	9
OV ure. The *main feature* is that the degree of the lattice relaxa	9
OV her hand *no remarkable lattice relaxation* was observed at vari	9
HV ses. And *possible factor* for polyamine's effects on calf thymu	2

TABLE 33b

Corpus (N=10) of FINAL Drafts: Concordances of H&O domain

OV (N=19) and HV (N=5)

OV Table 2. *No significant difference* was found when the substrat	2
OV ed. This *fact* indicates that the conformational changes to the	3
OV dy, this *fact* was confirmed in living cells using light contro	4
OV nt. This *fact* had not been known when Visser et al. determined	6
OV Br₃. The *reason* why RbMnBr₃ takes the IC spin structure has no	6
OV ved. The *reason* is that the breaking of the structural symmetr	6
OV bic one. *One of the reasons* is the optical biaxiality indicate	6
OV own. The *key to the problem* may well be that RbMnB^{r}₃ undergoes	6
OV ure. The *problem* concerns how the exchange interactions on a h	6
OV ure. The *major feature* of this technique involves the utilizat	9
OV ure. The *main feature* is that the degree of the lattice relaxa	9
OV er. This *marked difference* in the lattice relaxation between M	9
OV and. The *large difference* between MBE and HR-CVD cannot be int	9
OV oreover, *no marked changes* were observed in the structure of t	9
OV °C. This *evidence* leads us to the conclusion that the stable p	9
OV ore, the *most suitable and reliable method* to assay the reacti	10
OV inosose. *One suitable analytical method* seems to be a GC-MS se	10
OV duction. *Some of those intriguing compounds* may be prepared wi	10
OV purpose. *One drawback* of the Ferrier reaction is that the hydr	10
HV ate. The *possible constituent electronic configurations* of the	1
HV eaction. *One of the feasible mechanisms* for the distinctive ef	2
HV itself. *Some candidates* for the process may be enzymatic reac	4
HV Cl₃, the *most probable structure* of RbMnBr₃ below T$_{s2}$ has the	6
HV any. The *latter possibility* cannot be applied for the biochemi	10

5.2.4 <u>Concluding remarks on unpublished (FIRST & FINAL) RA drafts written by NNS novices - Part I</u>

Clearly, data presented here have to be put into the perspective of the 'bigger picture': the potential of interactive unmarked Themes may be underappreciated, but this is not to say that this particular set of NNS novices' drafts will be necessarily 'unsuccessful' because of this. Although labelled as 'inexperienced' novices writing their first papers in English, concordanced data indicate the idiosyncratic variability amongst this group in terms of the level of awareness of the RA product and the wider task environment, including the negotiation of scientific content and its expression with the 'expert' journal editors who rate the value of their pre-publication drafts. Many of this group may well be 'advanced' novice researchers with excellent research to report; as mentioned above (see section 5.2.1), eight of the students had had their first papers published by the end of the monitoring period, six in Japanese English-language journals and two in U.S. journals [see APPENDIX G for references].

It could be speculated that submission to a Japan-based English-language journal, whose editors and reviewers may be more sympathetic to common L2 difficulties, represented an 'easier' first publication option for novices [there are certainly suspicions of potential bias by NS editors against Japanese submissions, as noted in Gosden (1992a)]. Thus, the fully negative impact of L2-related difficulties on the NS judgement of the merits of an RA may be somewhat cushioned. [A close reading of published RAs in refereed Japanese English-language journals may reveal a number of language 'errors' (for example, subject-verb agreement) and 'typos']. Nevertheless, since we have suggested here a dynamic graded scale from 'inexperienced' to 'experienced' to 'expert', it is natural that a number of these NNS novices have already started to become relatively more 'experienced' RA writers [by definition

here of course, once published, novices become relatively more 'experienced'] and active researchers through initiation into international research publication.

All statistical correlations in the above TABLES proved, to a greater or lesser degree, to be significant. This appears to dispute therefore the statement in [HYPOTHESIS #7a] that there would be notable differences between 'inexperienced' and 'experienced' RA corpora. However, although a quantitative, product-oriented perspective may certainly indicate areas of interest in text analysis that merit further investigation, it seems clear that quantitative data shed little light on the qualitative, process-oriented changes in individual novices' RA writing, in particular, on the exact nature of changes in thematic choices that took place with successive drafting of RAs. This is particularly evident since correlations from FIRST to FINAL drafts did not move in the expected direction towards NS norms (to any significant degree at least, since correlations were already strong in many cases).

Consequently, to elucidate the 'appropriacy' of thematic selections according to our [HYPOTHESIS #7b], the next section will concentrate on a closer textual microanalysis of a set of FIRST & FINAL RA section drafts written by three of the ten novices whose drafts were included above. It is of interest that, in their analysis of scientists' discourse, Gilbert & Mulkay observed that a comparatively detailed examination of a small number of papers, 'can demonstrate most forcibly the capacity of particular scientists to produce radically different versions of given actions' (1984:40), indicating the interplay of scientific and rhetorical realities.

247

5.3 CHANGES IN THEMATIC SELECTIONS AND PATTERNS OF THEMATIC PROGRESSION BETWEEN UNPUBLISHED (FIRST & FINAL) RA DRAFTS WRITTEN BY NNS NOVICES

5.3.1 Introduction

Three sets of RA section drafts are presented for the purpose of analysis and description of thematic changes and patterns of thematic progression as they evolved between the FIRST full RA draft available in English and the FINAL draft submitted or accepted for publication in an international English-language journal. As previous work (Chapters Two and Three) has highlighted aspects of the relationship (see Figure 3 above) between the extralinguistic context of culture (RA genre) and the linguistic plane of the lexicogrammar (Theme), it is important to look at RA sections of texts which have been conceived as a whole, rather than examine just isolated sentences from sections. Therefore, one complete Introduction (Example #1) and one complete sub-section of longer Discussion (Example #2) and combined Results & Discussion (Example #3) sections are presented. The Experimental section is not included here due to the relatively much greater textual stability of this RA component, that is, there are relatively few changes between drafts. Experimental [Methods] sections have been characterised as checklists (Knorr-Cetina 1981), heavy on Rheme but light on Theme (Swales 1990a). This is, of course, of interest in itself but not for the purposes of the following analysis.

As mentioned above, a quantitative, product-oriented approach to discourse and text analysis may indicate areas of potential interest for research, but shed little light on the nature of the many qualitative changes which evidently contribute to the perceived 'success' of the RA. In particular, it sheds little light on the dynamic nature of changes towards genre-specific 'appropriate' thematic selections that occur with successive redrafting of RAs in

response to external critique. As Knorr-Cetina observes about the complex processes of research writing and rewriting:

'The final version of the paper is not only the product of its authors, but of several other scientists as well, whose critical comments have been taken into account. The process of rewriting the first version is one of negotiation among authors and critics. The dynamics of this process is interesting in itself, since there is no smooth transition from one version to the next via the incorporation of comments and criticisms. Comments may be solicited, but not received, or unsolicited but received, or received one way or another and not taken into account, or received several times in different versions and resented etc.' (1981:104).

In order to introduce a greater focus here on the dynamic, 'norm-developing' nature of research writing and redrafting processes, there is the need for a more qualitative, exploratory-interpretative research paradigm [Grotjahn (1987); see section 1.7] to underpin the present analysis and description of the process of redrafting key components of these texts. In the present case, this paradigm is based on the dual roles that the EAP (English for Academic Purposes) teacher-practitioner adopts as both participant in and observer of the processes involved in getting NNS novices' RA drafts accepted for publication in English-language scientific journals. The underlying research perspective of a participant-observer viewpoint is briefly reviewed in the next section.

5.3.2 Research methodology - participant observation
As a method of social research, one of the ideal strengths of participant observation is considered to be that the researcher becomes a participant in a naturally occurring social activity, without the introduction of artificiality into social observation and investigation (Jary & Jary

1991). In settings where the social action under investigation may be generally characterised as covert, as is often the case amongst the academic tribes, the discovery-based approach of participant observation is seen as a particularly valuable research method, as evidenced by the classic ethnographic studies of scientists' discourse and laboratory life by Latour & Woolgar (1979), Knorr-Cetina (1981) and Gilbert & Mulkay (1984). In this approach, data are collected informally in the course of a researcher's interactions in normal social life whereby the accurate recording of data and systematically focused interviews of key informants are normally an essential feature of the approach.

As indicated above, the pedagogic setting in this study is the writing of first RAs in English by novice NNS researchers, in this case, doctoral students at a science and technology university in Japan whose major fields of research are in the broad areas of the physical and life sciences (see section 4.1.2). Having set up EAP courses (Gosden 1991), classes in Academic Writing Skills were available to these novice NNS researchers. After completion of a two-semester course (approximately 40 hours, over one academic year), I continued to work closely with novices and in consultation with their research supervisors throughout the many stages involved in the initial drafting and subsequent revision of RAs in English. Thus, the EAP practitioner takes on the dual teaching-/research-oriented roles as both participant in and observer of the many stages involved in the L2 research writing process.

In a recent survey (Gosden 1992a), the majority (74%) of NS journal editors (N=136) indicated that there is a danger for NNS researchers that the value and quality of their research may be disguised by the quality of its reporting. Consequently, the main purpose of EAP-oriented language/research support is simply to assist NNSs compete

on an equal research basis in L2 English. As part of this process, there is naturally the potential for the participant-observer EAP practitioner to play quite an intrusive role, in terms of assisting in rewriting RA drafts and polishing texts prior to submission to journals. However, an important long-term aim in teaching preparatory courses in Academic Writing is clearly to train NNS novices to become increasingly more independent RA writers rather than to train them to see EAP teachers and other available NSs in a short-term role as proofreaders. On the other hand, one-to-one tutoring and consultations with research students about their RA drafts are naturally a desired (from the point of view of the NNS with the burden of having to compete in an L2) and desirable (from the NS's point of view in helping NNSs compete on an equal research basis) part of the role as EAP teacher in providing L2/research support.

Nevertheless, for the present exploratory purposes of a description of textual revisions between FIRST and FINAL RA drafts, it appeared important to discount initially the direct effects of the EAP classroom and one-to-one tutoring, since this may represent the introduction of a kind of artificiality into social observation and investigation. Consequently, analysis here is based on NNS novices' RA drafts in English which had been submitted or accepted for publication prior to their participation in the Academic Writing course.

As indicated above, the following analysis and description of three novices' FIRST & FINAL RA drafts presents data and commentary from a participant observer 'viewpoint. Here, comments on sentence-by-sentence changes are made on the supposed or evident rationale for FIRST through to FINAL draft changes. The participant-observer approach to this analysis of redrafting modifications is a reflection of interpretative commentary based on the following accumulated sources: (i) post-process interviews recorded

with (a) doctoral students and (b) their immediate research supervisor about revisions at each stage of redrafting; (ii) correspondence from journal editors and reviewers following initial submission; (iii) notes on drafts from colleagues or co-supervisors; (iv) the various English-language RA drafts (three or four printed drafts in total).

It is intended that the participant-observer approach to text analysis, used in greatest detail in the longer RESULTS & DISCUSSION section (Example #3), should complement the predominantly more traditional functional orientation to analysis and description of thematic progression employed in the shorter INTRODUCTION (Example #1) and DISCUSSION (Example #2) sections.

5.3.3 Example #1 - INTRODUCTION

In Example #1, the two complete Introduction drafts are first presented, followed by TABLE 34 which lists details of thematic selections, their discourse functions and their derived Theme-Rheme progressions in the FIRST draft. This thematic flow is depicted in Figure 25 below, which also maps thematic structure onto rhetorical moves (see Figure 11 above for Introduction moves). To complete Example #1, similar detailed analysis is presented for the FINAL draft (TABLE 35 and Figure 26) and the section concludes with comments on changes between drafts. For ease of reference, Subject Themes are in *italics* and Context Frames are underlined; sentences are numbered for FIRST draft in round brackets (1) and FINAL draft sentences thus [1].

Example #1 - INTRODUCTION: FIRST draft

(T)   Orientation relationships between TiAl-phase and Ti3Al-phase
      in Ti-40at%Al alloy

(1) *As-cast ingots of gamma titanium aluminides with compositions
less than 50at%Al* usually exhibit lamellar microstructures which
consist of gamma TiAl with a crystal structure of L10 and alpha2
Ti3Al with a crystal structure of DO19. (2) *Lamellar microstucture* is
important because of its potential for ductility at room
temperature[1,2] (3) *Structure of interfacial boundaries which
influence glides of atomic layer[3], antiphase domains and twins in
TiAl* were investigated[4-7]. (4) *It is possible to* distinguish
orientation relationships[8]. (5) *Yang et al.* investigated
orientation relationships between the gamma and the alpha2 and
distinguished four orientation variants of the gamma phase from the
selected area diffraction pattern[9]. (6) <u>However</u> *we* found six
orientation variants of the gamma phase. (7) *This paper* shows that
the gamma phase has six orientation variants for the alpha2 matrix.

Example #1 - INTRODUCTION: FINAL draft

(T) Distribution of variants in gamma plates in Ti-40at%Al alloy

[1] *As-cast ingots of gamma titanium aluminides with compositions
less than 50at%Al* usually exhibit lamellar microstructures which
consist of gamma TiAl with a crystal structure of L10 and alpha2
Ti3Al with a crystal structure of DO19. [2] *Lamellar microstucture* is
important because of its potential for ductility at room
temperature(1,2). [3] *Structure of interfacial boundaries which
influence glides of atomic layer[3], antiphase domains and twins in
TiAl* have been investigated(4-6). [4] *It is possible to* distinguish
orientation relationships between the gamma and the alpha2(7), e.g.
Yang et al. investigated these and distinguished four orientation
variants of the gamma phase from the selected area diffraction
patterns(8). [5] <u>However, in our study</u> *we* found six orientation
variants of the gamma phase. [6] *This paper* shows that it is possible
to distinguish six variants in the gamma phase and the gamma phase
has six orientation variants for the alpha2 matrix. [7] <u>Furthermore</u>
we observed the distribution of variants in gamma plates.

253

TABLE 34
Example #1 - INTRODUCTION: FIRST Draft
Themes: how they derive from previous clauses and their discourse functions

Themes (N=13)		Derivation	Discourse functions
(T)	Orientation relationships between TiAl phase and Ti3Al phase in Ti-40at%Al alloy		
T^1 *a*	As-cast ingots of gamma titanium aluminides with compositions less than 50at%Al	it <- (T)	: Real World Entity
t^1 =*b*	which {LM}	ir <- R^1	: relative
T^2 *a*	Lamellar microstucture (LM)	pr <- R^1	: Real World Entity
t^2 x*b*	because of {its potential}	ir <- r^2	: adverbial
T^3 *a*	Structure of interfacial boundaries	it <- T^2	: Real World Entity
t^3 =*b*	which {structure}	it <- T^3	: relative
T^4 *a*	It {is possible}	NEW THEME	: Empty Real World Theme
T^5 1	Yang et al.	pr <- R^3	: Interactive Participant
$t_1{}^5$ +2[and]	(they)	it <- T^5	: Interactive Participant
T^6 1	<u>However,</u> we	it <- t^5	: <u>Contrast/concession</u> + Discourse Participant
T^7 *a*	This paper	it <- T^6	: Macro Discourse Entity
t^7 x*b*	[that] the gamma phase	pr <- R^4	: Real World Entity

(T)	RA TITLE (FIRST draft)	it	from immediately preceding Theme
T^1, T^2,	sentence-initial Themes	pt	from any preceding Theme
t^1, t^2,	dependent clause themes	ir	from immediately preceding Rheme
$t_1{}^1$, $t_2{}^1$	continuing clause themes	pr	from any preceding Rheme
R^1, R^2,	main clause Rhemes	=	expansion: elaboration
r^1, r^2,	dependent clause rhemes	+	expansion: extension
a; b,y,d	dominant; dependent	x	expansion: enhancement
1; 2,3	initiating; continuing		

```
  (T)              Orientation relationships between TiAl-phase
   ¦                    and Ti3Al-phase in Ti-40at%Al alloy.
   ¦
   ¦
(1)  T¹ ------------> R¹                         ┌──────────────┐
     (ingots    (lamellar microstructures-LMs)   │    KEY:      │
      <50at%Al)       *                           │              │
                      t¹ -----------> r¹          │  T   R   R   │
                      (which)       (γ-TiAl)      │  ¦   *   :   │
___MOVE 1___          ¦                           │  ¦   *   :   │
                      ¦                           │  T   T   R   │
(2)                   T² -----------> R²          └──────────────┘
                      (LMs)         (important)
                      ¦                  *
                      ¦                  t² -------> r²
                      ¦              (because of)  (ductility
                      ¦                             potential)
                      ¦
(3)                   T³ -----------> R³
                  (IB structure)  (investigations)
                      ¦                  *
                      t³ ----> r₁³        *
                  (which) (influence) *                (T)
                                      *                 ^
                                      *                 ¦
                                      *                 ¦
(4)  T⁴ ---------------------------------------------------> R⁴
     (It {is possible})                *          (distinguish
                                       *           orientation
                                       *           relationships-ORs)
                                       *                :
                                       *                :
                                       *                :
(5)                                    T⁵ --------------> R⁵
                                   (Yang et al)       (ORs
                                       ¦            investigation)
                                       ¦                :
                                       t₁⁵ ------------> r₁⁵
                                   ([they])         (distinguish
                                       ¦             4 γ-phase
                                       ¦             variants)
___MOVE 2___                           ¦                :
                                       ¦                :
(6)                                    T⁶ -------------> R⁶
                                   (However, we)    (6 γ-phase
                                       ¦             variants)
___MOVE 3___                           ¦                *
                                       ¦                *
                                       ¦                *        R⁶(T)
(7)                                    T⁷ ---------> R⁷  *         ^
                                   (This paper)  (shows)  *        :
                                                         *   t⁷ --> r⁷
                                                        (γ-phase) (six
___MOVE 4___                                                      variants)
```

Figure 25. Example #1 - INTRODUCTION: FIRST draft
Thematic flow of information

TABLE 35
Example #1 - INTRODUCTION: FINAL Draft
Themes: how they derive from previous clauses and their discourse functions

Themes (N=15)	Derivation	Discourse functions
[T] Distribution of variants in gamma plates in Ti-40at%Al alloy		
T^1 aAs-cast ingots of gamma titanium aluminides with compositions less than 50at%Al	it <- [T]	: Real World Entity
t^1 =b which {LM}	ir <- R^1	: relative
T^2 a Lamellar microstucture (LM)	pr <- R^1	: Real World Entity
t^2 xb because of {its potential}	ir <- R^2	: adverbial
T^3 a Structure of interfacial boundaries	it <- T^2	: Real World Entity
t^3 =b which {structure}	it <- T^3	: relative
T^4 1 It {is possible}	NEW THEME	: Empty Real World Theme
t_1^4 +2Yang et al.	pr <- R^3	: Interactive Participant
t_2^4 +3[and] (they)	it <- t_1^4	: Interactive Participant
T^5 1 However, we	it <- t_2^4	: Contrast/concession + Discourse Participant
T^6 1a This paper	it <- T^5	: Macro Discourse Entity
t^6 x1b[that] it {is possible}]	pt <- T^4	: Empty Real World Theme
t_1^6 +2[and] the gamma phase	pr <- r^6	: Real World Theme
T^7 1 Furthermore, we	pt < T^6	: Addition - emphatic + Discourse Participant

[T] RA TITLE (FINAL draft)
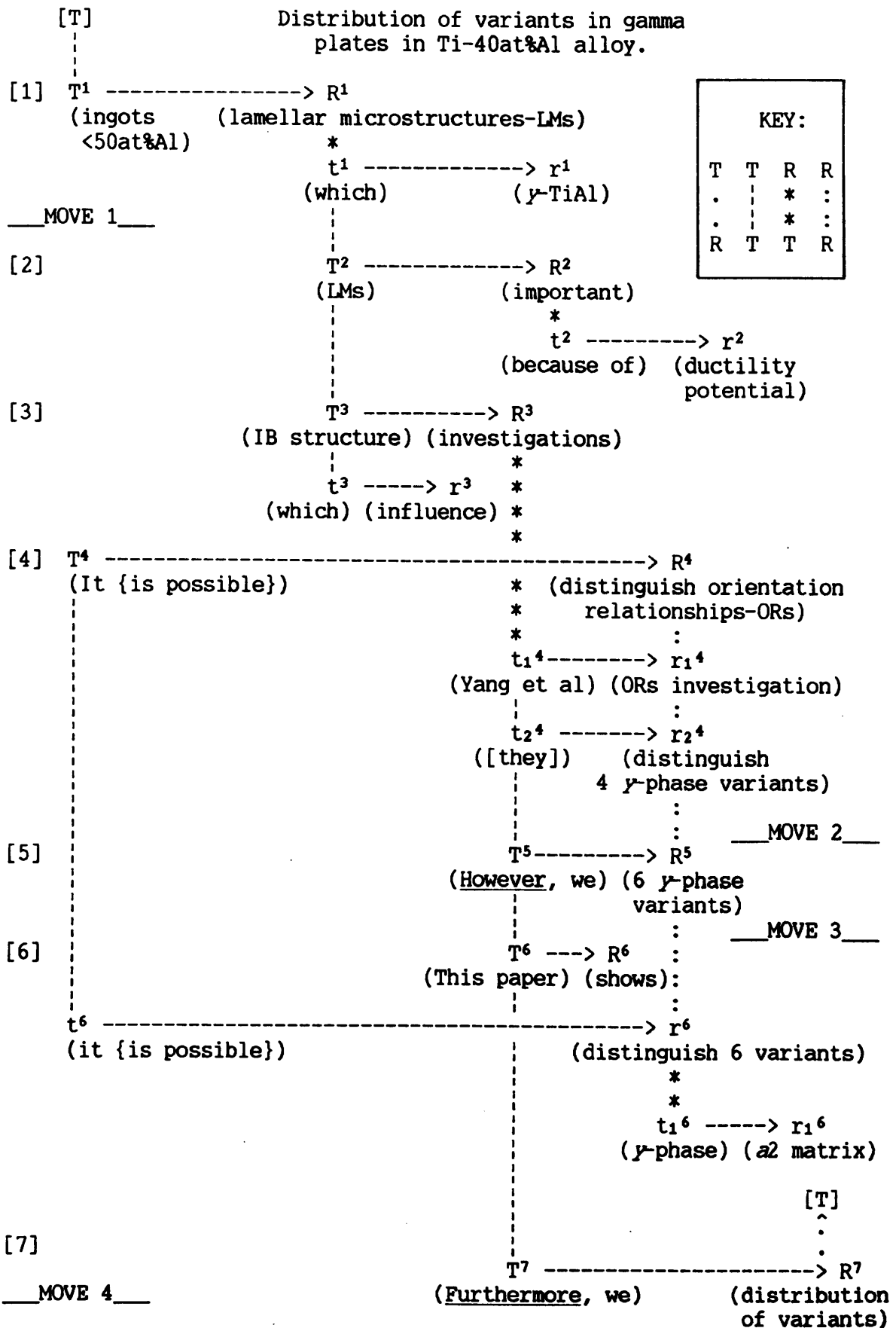
[T]
Distribution of variants in gamma
plates in Ti-40at%Al alloy.

[1] T¹ ----------------> R¹
     (ingots   (lamellar microstructures-LMs)
      <50at%Al)   *
                    t¹ -------------> r¹
                    (which)   (γ-TiAl)

___MOVE 1___

[2]   T² -------------> R²
                    (LMs)   (important)
                                         *
                              t² ----------> r²
                              (because of) (ductility
                                             potential)

[3]   T³ ----------> R³
                 (IB structure) (investigations)
                                     *
                    t³ -----> r³   *
                 (which) (influence) *
                                     *

[4] T⁴ ---> R⁴
     (It {is possible})   * (distinguish orientation
                                      *   relationships-ORs)
                                      *   :
                                   t₁⁴--------> r₁⁴
                                   (Yang et al) (ORs investigation)
                                      :   :
                                   t₂⁴ -------> r₂⁴
                                   ([they])   (distinguish
                                      :   4 γ-phase variants)
                                      :   :
                                      :   : ___MOVE 2___
[5]   T⁵---------> R⁵
                                   (However, we) (6 γ-phase
                                      :   variants)
                                      :   : ___MOVE 3___
[6]   T⁶ ---> R⁶   :
                                   (This paper) (shows):
                                      :   :
     t⁶ ---> r⁶
     (it {is possible})   :   (distinguish 6 variants)
                                     :   *
                                     :   *
                                     :   t₁⁶ -----> r₁⁶
                                     :   (γ-phase) (a2 matrix)

                                     :   [T]
                                     :   ^
                                     :   .
[7]   :   .
                                   T⁷ --------------------> R⁷
___MOVE 4___   (Furthermore, we)   (distribution
                                                         of variants)

Since the method of analysis outlined in Chapter Three (see section 3.4.2) concentrated on only sentence-initial Themes (T) in the NS corpus, these are again the main focus of interest in thematic analysis. However, it will be seen in the above TABLES and Figures that dependent clause Themes (t) are also coded. Some other researchers have included non-main clause thematic data in their analyses; for example, given the condensed nature of his Abstracts data, Gibson (1992) indicates that hypotactically related clauses contained relevant thematic material. Thus, secondary Themes are included in analyses here in order to reveal the contribution they may make to the textual flow and structuring of information in the RA drafts.

The more significant changes between these two drafts of Example #1 occur in the second part of the Introduction, the first three sentences being identical. However, the effects of a rewording of the title Theme [T] are interesting. The title is changed from a focus on 'orientation relationships' to 'distribution variants' and the need for greater coherence between the title and Move Four, a statement of present research objectives, is reflected in the addition of FINAL sentence [7]. In any discussion of the relationship between Theme and components of textual cohesion and coherence, it is important to correlate the linking of successive clauses and sentences with distinct patterns of theme-rheme progression. Fries (1983) lists three such major patterns identified earlier by Daneš (1974) and these are shown in Figure 27 below.

Fries uses these patterns to demonstrate his point that 'thematic progression correlates with the structure of a text' and 'with the levels of relevance of each sentence within a passage' (1983:121). Thus, in expository prose, it would be expected that each sentence should follow

logically from what has gone before, expanding on the
previous argument. So, as illustrated by the 'simple
linear progression' of Pattern (1), the Theme of one
clause picks up on preceding rhematic information.
Conversely, Fries further explains that, in narratives,
since events may involve parallel sequences of clauses
with common characters and settings, 'constant topic'
Pattern (2) will predominate. Successive Themes may also
be derived from a 'hypertheme', such as may be represented
by a title (T) or subheading (sT) in an RA section; this is
shown in Pattern (3).


```
(1)   T¹ -----> R¹
                ¦
                ¦
              T² (=R¹) -------> R²
                                ¦
                                ¦
                              T³ (=R²) -------> R³
(2)   T¹ -----> R¹

      T¹ -----> R²

      T¹ -----> R³

(3)                    T
              T¹ -----> R¹
                    T² -----> R²
                          T³ -----> R³
```

Figure 27. Major patterns of Theme-Rheme progression
            (Daneš 1974)

How does Example #1 relate to these patterns? It can be
seen from TABLE 34 that, in the FIRST 7 main clauses,
there are (including (T) title) 13 Themes. On examination
of how these derive from previous clauses, 5 relate to
previous rhematic information (ir=2; pr=3), 6 relate to
immediately preceding Theme, with 1 new unrelated Theme *It
is possible..* . (NB: 12 derivations are possible with 13
Themes). These correspond to 4 major thematic clusters

259

which contribute to the realisation of introductory moves (see Figure 11):

Move 1   : <50at%Al γ-phase TiAl   - (T) T^1 (R^1) t^1
Move 2   : lamellar microstructures   - T^2 (R^2) t^2 T^3 t^3 T^4 (T)
Move 2   : Yang et al.   - (R^3) T^5 $t_1{}^5$
Moves 3/4 : we/this paper/"our work"   - T^6 T^7 t^7 (R^6) (T)

Thus it appears that, within each discreet Introduction move, the predominant thematic progression more closely resembles Pattern (2), described as a pattern of parallel thematic notions, although this is by no means a neat fit.

Looking now at the FINAL text for comparison (TABLE 35 and Figure 26), the major effect of redrafting has been to add two thematic extensions which relate to preceding Themes: [t^6] *it is possible...* which parallels [T^4], part of Move 2, and [T^7] <u>Furthermore</u>, *we* which extends the notion in T^6 and, in turn, T^5 the initiator of conjoined Moves 3/4. Thus, in comparison with the FIRST draft, there is an attempt to create greater textual cohesion and coherence *across* rhetorical moves and this effect can be seen more clearly in Figure 26. The predominant pattern of thematic progression here appears to more closely resemble a combination of Patterns (2) and (3), where superordinate Themes help create related clusters of Themes extending across the text. Francis (1990) comments that in text analysis there is, in fact, often no neat way of characterising patterns of thematic progression and this is the case with Example #1.

Other minor modifications in (Context Frame +) Subject thematic selections between drafts may also be noted. SIE (6) in the FIRST draft has become [5] in the FINAL draft:

(T^6)   <u>Contrast/concession</u> + Discourse Participant (DP)
        However, we found...
becomes

[T⁵]   *However, in our study, we* found..

The writer expands the Context Frame in the FINAL draft to include a secondary context in our study which also adds to the strong participant flavour of this Move 3/4 transition. Its addition strengthens the contrast implicit between the review of previous findings and the preliminary results indicated here. [With the quantitative method of Theme analysis employed here (see section 3.4.2), this secondary context is not recorded and this therefore emphasises the importance of the present qualitative approach].   This contrast is further emphasised in the FINAL draft by the addition of an entirely new sentence [7] to the end of the original Introduction, contextualised by furthermore and the continuation of the strongly interactional Move 4 Topic *we*. The added presence of [7] confirms that, in scientific RAs, a preliminary statement of principle research findings is not infrequently a part of the Introduction [see Swales & Najjar (1987) who reported that 45% of their sample included such statements].

5.3.4 Example #2 - DISCUSSION

In parallel with Example #1 above, here two complete Discussion subsection drafts are first presented, followed by TABLES 36 & 37 and Figures 28 & 29 which list and illustrate thematic selections, discourse functions and derived progressions in FIRST & FINAL drafts. As above, the section concludes with comments on changes between drafts.

Example #2 - DISCUSSION: FIRST draft

(T)   Quantitative analysis of the process and propogation of
      cortical granule breakdown in echinoderm eggs
(sT) Process of a single CGB:

(1) *The process of a single CGB* was investigated by using a high-speed video system at the speed of 200 fields/sec. (2) *It was found that* a CG swelled up by ca. 20% in diameter and then broke down. (3) *Our data* were disagree with the report of Endo that a CG swelled two times as largely as the former one[7]. (4) *It is guessed that* he might confuse the swelling of a CG before breakdown with the swelling of CG inclusion after the breakdown, since we observed the phenomenon that CG inclusion was swelling up significantly after CGB. (5) *It could not be determined* when the fusion pore opened during CGB in this study. (6) However, *it is certain that* a CG would open first after the fusion of the CG and egg plasma membrane and the CG swells up.

Example #2 - DISCUSSION: FINAL draft

[T]   Quantitative analysis of the process and propogation of
       cortical granule breakdown in sea urchin eggs
[sT] Process of single CGB:

[1] *The process of a single CGB* was investigated by using a high-speed video system at the speed of 200 fields/sec. [2] *It was found* that a CG swelled up by ca. 20% in diameter and then broke down. [3] *Our data* did not coincide with the findings of Endo that a CG swelled in diameter by a factor of two during CGB[7]. [4] Since we observed the phenomenon that CG inclusions swelled up significantly after CGB, *it is suggested that* there was confusion between the swelling of a CG before breakdown and the swelling of CG inclusions after the breakdown. [5] *The time of fusion pore opening* could not be determined in this study. [6] However, *it seems likely that* a CG would open first after the fusion of the CG and egg plasma membrane and the CG swells up.

TABLE 36
Example #2 - DISCUSSION: FIRST Draft
Themes: how they derive from previous clauses and their discourse functions

Themes (N=16)	Derivation	Discourse functions
(T) Quantitative analysis of the process and propogation of cortical granule breakdown (CGB) in echinoderm eggs		
(sT) Process of a single CGB:		
T^1 *a* The process of a single CGB	it <- (sT):	Real World Process
t^1 x*b* by {using}	ir <- R^1 :	adverbial
T^2 1*a* It {was found}	it <- t^1 :	Empty Real World Theme
t^2 1x*b*[that] a corticle granule	pt <- T^1 :	Real World Entity
t_1^2 x2[and then] (it)	it <- t^2 :	Real World Entity
T^3 *a* Our data	NEW THEME :	Participant Viewpoint
t^3 x*b* [that] a CG	pt <- t_1^2 :	Real World Entity
T^4 *a* It {is guessed}	NEW THEME :	Empty H&O Viewpoint
t^4 x*b*[that] he	pr <- R^3 :	Interactive Participant
t^{4-1} x*y*[since] we	pt <- T^3 :	Discourse Participant
t^{4-2} x*d*[that] CG inclusion	pt <- t^3 :	Real World Process
T^5 *a* It {could not be determined}	pt <- T^2 :	Empty Real World Theme
t^5 x*b*[when] the CG fusion pore	pt <- t^{4-2}:	Real World Entity
T^6 1*a* However, it {is uncertain}	pt <- T^5 :	<u>Contrast/concession</u> + Empty H&O Viewpoint
t^6 1x*b*[that] a CG	pt <- t^5 :	Real World Entity

(sT) RA TITLE (FIRST draft)
t^{4-1}, t^{4-2} dependent clause sequence

```
(T)        Quantitative analysis of the process and propogation of
              cortical granule breakdown (CGB) in echinoderm eggs.

              (sT)     Process of a single CGB:

(1)         T¹ --------------------------------> R¹
            (CGB process)                        (investigation)
                                                        *
                                                 t¹ ------ r¹
                                                 (by {using}) (video)

                                                 T²->R²
(2)         t² --------------------------> r₁²   (It {was found})
            (a CG)                (swelling x20%)

            t₁² --------------------------> r₂²
__MOVE D1___ (it)                  (breakdown)

(3)  T³ -------------------> R³
     (Our              (Endo's
     data)             report)
                          *
            t³ --------------------------> r³
            (a CG)       *      (swelling x2)    __MOVE D2a__
                         *
(4)    T⁴->R⁴            *
       (It {is          *
       guessed})        t⁴ ----> r⁴
                        (he) (after/before
                              CG confusion)
    t⁴⁻¹ --------------------------> r⁴⁻¹
    (we)                    (our "after"
                            observation)

            t⁴⁻² --------------------------> r⁴⁻²
__MOVE D2b__ (CG inclusion)        (swelling)
                                                 T⁵->R⁵
(5)                                              (It {could not
                                                 be determined})

            t⁵--------------------------> r⁵
            (CG fusion pore)     (CGB opening)    __MOVE D3___

     KEY:                                        T⁶->R⁶
(6)                                              (However, it
     T   R   R                                   {is certain}
     |   *   :
     |   *   :
     T   T   R   t⁶ --------------------------> r⁶
                 (a CG)             (opening)
                                                 __MOVE D2b__
```
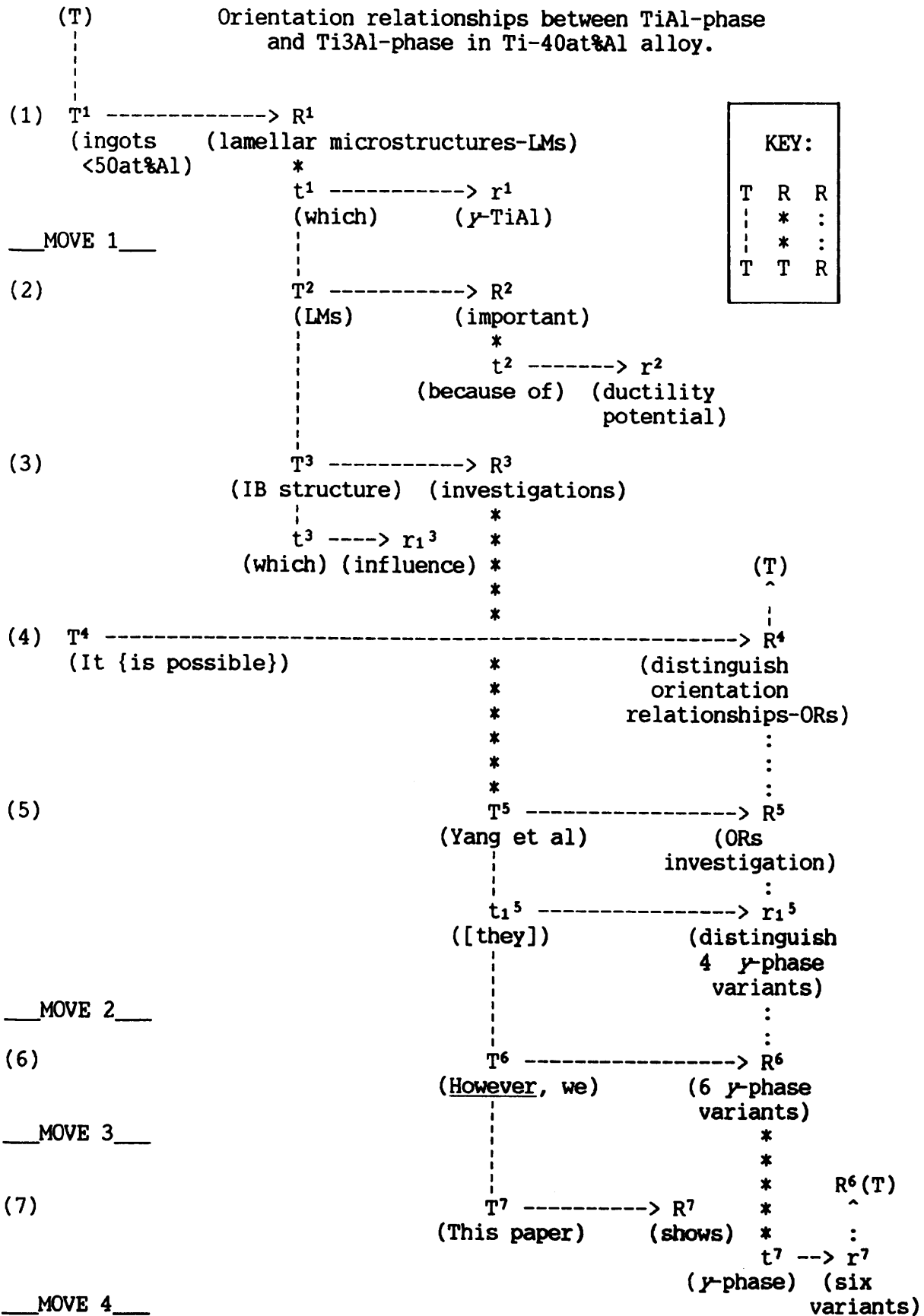
Figure 28. Example #2 - DISCUSSION: FIRST draft
Thematic flow of information

TABLE 37
Example #2 - DISCUSSION: FINAL Draft
Themes: how they derive from previous clauses and their discourse functions

		Themes (N=13)	Derivation	Discourse functions
[T]		Quantitative analysis of the process and propogation of cortical granule breakdown (CGB) in sea urchin eggs		
[sT]		Process of a single CGB:		
T^1	*a*	The process of a single CGB	it <- [sT]	: Real World Process
t^1	x*b*	by {using}	ir <- R^1	: adverbial
T^2	1*a*	It {was found}	it <- t^1	: Empty Real World Theme
t^2	1x*b*[that] a corticle granule		pt <- T^1	: Real World Entity
t_1^2	x2	[and then] (it)	it <- t^2	: Real World Entity
T^3	*a*	Our data	NEW THEME	: Participant Viewpoint
t^3	x*b*	[that] a CG	it <- t_1^2	: Real World Entity
T^4	x*b*	Since we observed the phenomenon that CG inclusions swelled up significantly after CGB, it {is suggested}	pt <- T^3	: Cause - reason/result + Empty H&O Viewpoint
t^4	*a*	[that] there {was confusion}	pr <- R^3	: Empty H&O Viewpoint
T^5	1	The time of fusion pore opening	ir <- r^4	: Real World Entity
T^6	*a*	However, it {seems likely}	ir <- R^5	: Contrast/concession + Empty H&O Viewpoint
t^6	x*b*	[that] a CG	pt <- t^3	: Empty Real World Theme

[sT] RA TITLE (FINAL draft)

265

```
[T]        Quantitative analysis of the process and propogation of
              cortical granule breakdown (CGB) in sea urchin eggs.

                      [sT]   Process of a single CGB:
                        |
[1]                    T¹ -----------------------------> R¹
                       (CGB process)                 (investigation)
                        |                                   *
                        |                                  t¹ ------ r¹
                        |                             (by {using}) (video)
                        |                      T¹           |
[2]                     |                      ^       T²->R²
                        |                      |      (It {was found})
                       t² ---------------------> r²       .
                       (a CG)              (swelling x20%) .
                        |                            :     .
___MOVE D1___          t₁² --------------------->r₁²      .
                       (it)                 (breakdown)   .
                        |                            :     .
[3]  T³ ------------------> R³                        :     .
     (Our               |    (Endo's                 :     .
      data)             |     findings)              :     .
      |                 |        *                    :     .
      |                t³ --------------------------> r³    .
      |                (a CG)  *                 (swelling x2) . ___MOVE D2a__
[4]  T⁴->R⁴             |      *                      :     .
     (Since we         |      *                      :     .
      observed..,      |     t⁴ --------------> r⁴    .
      it {is           |   (there {was}      (CG swelling .
      suggested})      |    confusion)      before/after) .
                       |                          *     .
___MOVE D2b___         |                          *     .
                       |                         T⁵ -----> R⁵
[5]                    |                    (CG opening  (could not be
                       |                        time)    determined)
___MOVE D3___          |                          .          *
                       |                          .          *
[6]                    |                          .        T⁶->R⁶
                       |                          .       (However, it
                       |                          .        {seems likely}
                       |                          .
                      t⁶ --------------------> r⁶
                      (a CG)              (opening)

___MOVE D2b___
                                            _____
                                           |   KEY:         |
                                           |                |
                                           | T   T   R   R  |
                                           | .   |   *   :  |
                                           | .   |   *   :  |
                                           | R   T   T   R  |
                                           |_____|


              Figure 29. Example #2 - DISCUSSION: FINAL draft
                      Thematic flow of information
```

266

In both Example #2 Discussion drafts, sentence-initial Themes (1-3)[1-3] remain unchanged, as they did in Example #1, suggesting that the realisation of the more factual opening moves in Introductions and Discussions presents fewer problems. Labelled Move D1, (1)[1] makes reference to the main purpose of the work (an optional move element) and main findings (see Figure 14 above). (3)[3] realises Move D2a, a comparison with the previous research results of Endo. The next two sentences (4)[4] corresponding to Move D2b - Speculation on Previous Results - and (5)[5] Move D3 - Limitations of Present Research - are the focus of major thematic manipulation in redrafting and these are discussed in more detail below. Sentence (6)[6] returns to speculative comment on results [Move D2b], rather than final concluding comments, perhaps not unusual as we are dealing with an initial Discussion subsection here.

As is illustrated by Figures 28 and 29 above, both drafts predominantly reflect the thematic progression of Pattern (2) above. This can be seen in the linear vertical structures of the central Theme->Theme (11 out of 15 Themes in FIRST; 7 of 12 in FINAL) and corresponding Rheme->Rheme lines. In the FIRST draft, the writer's predominant usage of relatively more 'marked' sentence-initial thematic constructions, for example, the Empty Themes *It was found* (T^2), *It is guessed* (T^4), *It could not be determined* (T^5), _However, it is certain_ (T^6), means that subordinate Themes carry the parallel scientific Real World development of corticle granules (t) and the process of their breakdown (r); thus the Empty Themes (particularly the two H&O Viewpoint) carry most of the rhetorical weight of the text. However, in the redrafting process, this relative thematic 'emptiness', notably in sentences (4)[4] and (5)[5], has been modified.

(T^4)   Empty H&O Theme
                 It is guessed...

becomes

Cause-reason/result + Empty H&O Theme

[T⁴]   *Since we observed..., it is suggested...*

Both the motivation for and the corollary of this
manipulation is a change in the perceived flow of
Given-New information in the text. Fries defines the terms
thus: 'New information is information which is presented
as "newsworthy" - as worthy of the listener's attention,
while Given information is presented as information which
is recoverable in some way' (1992:2). Halliday indicates
that 'what is treated as recoverable may be so because it
has been mentioned before; but that is not the only
possibility. It may be something that is in the situation'
[or something presented]'as Given for rhetorical purposes'
(1985a:277). As with other functional terms, these glosses
may be difficult to apply in practice. Fries (1992) points
out that the evident problem in analysis of information
structure is that the functions of Given and New are
generally only signalled in spoken language; however,
written discourse is clearly able to present Given and New
information.

Glatt (1982) comments that the writer assumes the burden
of responsibility for the Given-New contract since it is a
special, implicit agreement between 'conversants' - and
one means of achieving this is through the correlation of
the tone group with clause structure. Halliday comments
that 'the unmarked position for the New is at the end of
the information unit' (1985a:276) resulting in 'end-focus'
(Quirk *et al.* 1985:1357). 'It is often said that writers
tend to use this association of placement at the end of
the clause with importance to make their written language
flow naturally' (Fries 1992:2). Other things being equal,
then, we generally expect Given information in Theme
position with New in Rheme and research suggests that this
matching of writer/reader (speaker/listener) expectations
helps communication to occur more easily than when there

is a mismatch (Haviland & Clark 1974; Clark & Clark 1977;
Clark & Haviland 1977). Francis (1990) suggests that one
reason why texts produced by inexperienced L2 writers may
appear disconnected and disjointed is because there is no
clear information structure; moreover, 'incomplete
understanding of the meanings of both theme and focus of
new information often leads to unintended emphases,
making it difficult.. to understand the point being made'
(1990:86).

To return to Example #2, the manipulation of thematic
choices between drafts may therefore imply a deliberate
refocusing of information in the text. In FIRST (4), the
information flow appears to be marked New-Given and in
redrafting the writer evidently realised the need to more
clearly contextualise his explicit criticism of a fellow
researcher's study in (4). Hence in the FINAL draft, the
subordinate clause *Since we observed..* is fronted in order
to emphasise the justification for the writer's
suspicions. The reason *Since..* is now taken as Given,
partly for rhetorical purposes and partly in consideration
of the recoverability of this information after statements
[2] and [3]. The suggested *confusion* (t^4->r^4) thus becomes
the information which is to be presented as 'newsworthy'.

With the conventions of accepted norms of research
critique in mind, the directly face-threatening wording of
(4) in the FIRST draft *he might confuse..* is subsequently
softened and objectivised with *it is suggested that there
was confusion...* In addition, *suggesting* is clearly more
acceptable academic evaluation than *guessing*. In their
sociological analysis of scientists' discourse, Gilbert &
Mulkay demonstrate that scientists move between two
'interpretative repertoires, or linguistic registers'
(1984:39), the contingent repertoire of informal
scientific discourse similar to conference discussions and
interviews, and the formalised empiricist repertoire of
RAs. The more informal *he might confuse* appears to be an

example of contingent evaluation since in an empiricist account, 'each scientist's actions and beliefs, no matter how inconsistent they appear to be with those of other researchers, are presented as those of any competent scientist' (Gilbert & Mulkay 1984:56). As an indication of how the process of FIRST>FINAL redrafting approaches published RA empiricist norms, neither of the eliminated FIRST draft Themes *he...* and *it is guessed...* appeared once in the 72 NS/NNS RAs analysed in the main corpora, that is, in over 8300 sentence-initial Themes [but see example #3 FINAL draft[5] below].

Turning now to sentence (5), the information flow in the FIRST draft also appears marked New-Given. Empty Themes are convenient devices for 'providing some kind of dummy theme which will enable the originator to indicate the 'new' status of a whole clause, including its subject' (Quirk *et al.* 1985:1402). Previous discussion in (4) infers that the information about the procedural issue of *time* is recoverable from context, that is, Given; this FIRST draft sentence (5) therefore suffers from what Bloor & Bloor (1992) call the problem of Empty Rheme, particularly ending with *in this study.* It is indeed the fact that the researchers were unable in their study to *determine* the CGB time that is New (=worthy of attention), if not exactly New (=unfamiliar or rather unpredictable, since T^5 relates back in a way to T^2). So in the unmarked case, it is this information which should carry, in Prague terms, the highest degree of communicative dynamism and thus end-focus.

(T^5, t^5) *It could not be determined* when *the fusion pore*
              ---------------> NEW <--------------------------
              opened during CGB in the study.
              ------------------------ GIVEN
becomes

[T⁵]   *The time of fusion pore opening* could not
            GIVEN ------------------------------------
            be determined in this study.
            ---> NEW <----------- GIVEN

Moreover, in the FINAL draft, the fronting of the causal
CF in the preceding sentence now creates a
rhematic>thematic (r⁴>T⁵) link as part of a Pattern (1)
linear progression of expository information.

A minor modification of sentence (6) highlights a
different kind of thematic manipulation, but it is a most
critical one embedded in the social-constructionist nature
of research reporting:

(T⁶)   Empty H&O Theme
              it is certain that...
becomes
[T⁶]   Empty H&O Theme
              it seems likely that...

The crux of a scientific RA's potential contribution lies
in the expression of what it is claiming to add to the
present state of knowledge. In anticipation of the
community's validation of the current research findings,
the crucial claim is decisively hedged in the FINAL draft
of [6]. As indicated by the survey of NS journal editors
above (see section 4.2), the originally more confident
claim may well have created time-consuming delays in
correspondence and debate (if editors disagreed with the
claim, that is, and the writers appeared to suspect that
they would). Swales and Najjar confirm that writing an RA
Introduction is 'strongly modulated by perceptions of the
anticipated reactions of peer-colleagues' (1987:175) and
this is perhaps more so in the matter of judging the
highest-possible level of claim in Discussion sections. On
the other hand, this new, hedged position opens up future
research space, that is, a search for stronger evidence,

and we can expect the writer and co-workers to fill these gaps at some point with further publications.

5.3.5 Example #3 - RESULTS & DISCUSSION

In this third section, comments on sentence-by-sentence changes below continue to be made from the viewpoint of the supposed or evident rationale for changes from the base FIRST draft to the FINAL version. Textual changes naturally include the cleaning up of 'careless mistakes' but of particular interest are those which arise from and thus reflect the 'harder', 'norm-developing' writing processes (Swales 1988, 1990a) which anticipate and countenance the reactions of the intended readership in the light of criticism and feedback.

However, in this Results & Discussion subsection, the TABLES and Figures which illustrated thematic progression in Examples #1 and #2 are not reproduced. Instead, as part of the participant-observer commentary which follows the extracts, specific background observations of interest made by the informant's supervisor are integrated in [square brackets] and cross-references of relevance to other L2 writing/EAP/systemic-functional literature are made in {these brackets}.

As above, sentences are numbered for convenience of discussion, with those from the FIRST draft indicated by round brackets (13) - on the left side of the commentary - and those from the FINAL draft in square brackets [13] - on the right. Since we are particularly interested in thematic control, marked Themes as coded throughout are underlined with unmarked Subjects in *italics*. The complete RA subsection is presented in four main extracts, again for convenience of discussion.

(RA Title) Structure and magnetic properties of Fe/Ag multilayer
            films prepared by DC sputtering
(subTitle) 3-1 Magnetic properties

(1) *The magnetic properties of the multilayer films with a thickness
ratio tFe/tAg of about 1/10* were measured by VSM. (2) *A typical loop*
is shown in Fig.2a). (3) *The loop shape* indicates that the films
prepared by this apparatus have a uniaxial magnetic anisotropy in the
plane. (4) *We* guess that this uniaxial anisotropy might be introduced
during sputter deposition in this apparatus, because the incident
direction of sputtered atoms was not perpendicular to the substrate
plane. (5) *A magnetic structure of Fe thin film (about 20nm) observed
by Lorentz microscopy with TEM (200kV)* is shown in Fig.2b). (6)
Ripple fringes were observed and magnetization direction was found to
be parallel to the average easy axis which was measured by VSM.

RESULTS & DISCUSSION: FINAL draft [1] - [5]

[RA Title] Hysteresis loops and microstructures of Fe/Ag multilayer
            films
[subTitle] 3.1 Magnetic properties of Fe/Ag multilayer films

    [1] *A typical m-H loop measured by VSM for a bilayer film
having a thickness ratio of about 1/9 ((109/900)₁)* is shown in
Fig.2(a). [2] *The loop shape* indicates that the film has a uniaxial
magnetic anisotropy in the film plane. [3] <u>In order to check the
above anisotrophy</u>, *an Fe thin film about 20nm thick, prepared in the
same apparatus*, was observed by Lorentz microscopy, as shown in
Fig.2(b). [4] *Ripple fringes* are observed and the magnetization
direction is found to be parallel to the average easy axis measured
by VSM. [5] <u>Therefore</u>, *we* guess that this uniaxial anisotropy might
be introduced during sputtering deposition in this apparatus, because
the incident direction of sputtered atoms was not perpendicular to
the substrate surface.

FIRST   DRAFT   >   >   >   >   >   >   >   FINAL   DRAFT
(sentence #)   [sentence #]

(sT) subsection Title: specify in greater detail as in title.   [sT]

(1) This repeats information just mentioned in Experimental –
     it's not a Result – delete.

(2) This is a Statement of Location of Results {see Figure
     13} so it should now come first with the less essential
     details in (1) embedded: *having the thickness ratio...*   [1]

(4) *We guess..* speculation before adequate description of
     Result which comes later in (5/6) – delay this comment;
     with the reverse order of sentences (5+6+4), this
     speculation can be more clearly contextualised with our
     reason for the speculation with *therefore.*   [5]

(5) Why did we observe this *thin film 20nm thick* sample? – to
     check the *anisotropy* mentioned in (3); so contextualise
     now with *In order to check the above anisotropy.* [The
     correlation between the sentences is not clear, she only
     wrote the results not the <u>purpose</u> – the readers cannot
     understand the results if you don't explain the purpose
     of the study].   [3]

{Research on expository writing in Japanese draws an
interesting parallel here. Hinds noted a common style
which he described as 'delayed introduction of
purpose'(1990:98). Hinds analysed texts translated first
literally from Japanese, then translated for an
English-speaking audience. In the first case,

   'the purpose of the article is not seen until the final
   paragraph. On the surface at least, this an indication of

an inductive style of writing.. . The translation for English readers has a clear statement of purpose as its initial sentence, and the following sentences develop or expand on this statement of purpose' (1990:91).

This novice RA writer may lack appreciation of the fact that scientific reporting is wholly deductive in style and, following Swales (1990b), that RA readers have discoursal expectations about the expression of an RA writer's underlying purpose. In addition, Knorr-Cetina (1981) makes the important point that in writing up scientific research there is commonly a reversal of the reality of the "laboratory process" and the "story of the paper": 'the impression of a problem-pushed solution which has been researched, rather than encountered by chance, is created in the text through the hierarchical organisation of arguments through which the solution appears *derived* rather than original. However, this reversal is not the effect of misrepresentation, but part of the literary strategy of the text' (1981:101). The increase of Cause-purpose Context Frames (see section 3.5.2) such as *in order to* from FIRST to FINAL drafts here (see concordances in TABLES 31a/b) contributes to the creation of this rhetorical rather than scientific reality}.

(6) The reporting tense in sentence (3) is present, so change *was found* to *is found* to emphasise current findings.   [4]

> Cycle #1 of Result + Comment {see Figure 13} is complete, so the next cycle can now be marked with a new paragraph.

{In her detailed analysis of FIRST to FINAL RA draft changes, Knorr-Cetina (1981) noted three major strategies of modification: deleting particular statements made in the original version; reshuffling the original statements; changing the modality of certain assertions. FINAL paragraph #1 [1-5] certainly confirms the first two (see [18] for modality), although, in general, the

275

reshuffling of statements on a macro-level is relatively limited in this text. The <u>addition</u> of statements would clearly appear to be a major fourth modification in the present FINAL draft: moreover, additional statements in [10][17] [25][26][32] all have important global text-organising properties}.

RESULTS & DISCUSSION: FIRST draft (7) - (11)

---

(7) <u>When the thickness of Fe layers were from 6.5nm to 45nm</u>, *some small steps* were observed in the M-H loops for the average easy axis, as shown in Fig.3. (8) <u>From many M-H loop measurement for Fe/Ag films</u>, *it was found* that the number of steps were equal to the number of Fe layers sandwiched in the multilayer film. (9) <u>When the thickness ratio was 1/1</u>, *similar steps* were also observed. (10) <u>In addition</u>, *similar steps* were observed in the films deposited on glass slide and mica. (11) <u>Then</u> *we* measured these "step width" in some films with various periods.

RESULTS & DISCUSSION: FINAL draft [6] - [11]

---

[6] <u>When the thickness of Fe layers were from 6.5nm to 45nm</u>, *some small steps* were observed on the M-H loops measured for the average easy axis, as shown in Fig.3. [7] <u>From many M-H loop-measurement of Fe/Ag films</u>, *it was found* that the number of steps was equal to that of Fe layers sandwiched in the multilayer film. [8] <u>In addition, when the thickness ratio was 1/1</u>, *similar steps* were observed. [9] *Similar steps* were also observed for the films deposited on glass slides and mica. [10] <u>Therefore</u>, *it was concluded* that the appearance of such steps was independent of the thickness ratio and the substrate species, for Fe-layer thicknesses within the examined range. [11] <u>To our knowledge</u>, *there has to date been no detailed analysis* of such steps on the M-H loops, although similar steplike loops have been reported.[6)]

---

(9) The themes of (9) and (10) *similar steps* are identical; so *In addition* doesn't make sense between these two sentences; it should be between (8) and (9) which together represent the expansion of [7].   [8]

The #2 Result has now been given [6-9], but as yet there is no accompanying #2 Comment, that is, explanation or

speculation {see Figure 14}. [At first a lot of Japanese
students only ever write Abstract, then Experimental and
Results, without Discussion, sections in English]. As in
[5] above, Comment #2 needs to be introduced with clearly
signalled reason *Therefore, it was concluded...*   [10]

(11) Also, the reason for the measurement of *steps* in (11)
     needs to be clearly stated; so, indicate the comparative
     gap in research {see Figure 14} *there has to date been no
     detailed analysis*; the negative tone of this can be
     better contextualised with *to our knowledge.* [We did a
     KEYWORD computer citation search but the reference list
     is nevertheless short: it's a relatively new field; if we
     don't include *to our knowledge* this effort is not clear].   [11]

     The original use of *Then* without the addition of
     sentences [10][11], makes for a weak connection of ideas.
     [Japanese students usually use words such as *then*; this
     word does not mean exactly the connection between the two
     sentences; in Japanese language we usually use these
     words].

{Many minimal adjuncts, such as *Then,*   have   several
meanings: in this case, reason, condition, time (Quirk *et
al.* 1985:634-636). Research   has   suggested   that L2
students may have a restricted knowledge of the diversity
of such linking words in English (Bacha & Hanania 1980)}.

RESULTS & DISCUSSION: FIRST draft (12) - (17)

---

(12) *Step width* is the difference of Hc for the neighbor. (13) *The
results* was shown in Fig.4. (14) When deposition was started from Fe
layer on substrate, *the step widths* were almost the same as of those
of the film with the same film period. (15) But *the values* became
larger as the film period increased. (16) After all, if the Fe layers
were stacked periodically with the same thickness, *the steps* were
observed with the same width. (17) When deposition was started from
Ag layer, *the step widths* were also similar to one loop and the
values were nearly equal in all film periods.

---

[12] <u>In order to elucidate the origin of the steps,</u> *we* measured the step widths for several films having various film periods with a constant thickness ratio of 1/10. [13] <u>Here</u> *the step width* means the difference in Hc between two adjacent steps on the M-H loop, as shown in Fig.3. [14] *The results* are shown in Fig.4 for films whose total thickness is about 1000nm: the number of Fe layers decreases with increasing the film period. [15] <u>When deposition was started from an Fe layer on substrate,</u> *the resulting films* showed steps with almost the same step widths for a given film period, although the step widths increased with increasing film period. [16] <u>When deposition was started from a Ag layer,</u> *the step widths* were nearly the same, independent of the film period. [17] *This different dependence* will be discussed in 3.3 in relation to structural properties.

As the gap in research (*no detailed analysis*) has now been indicated [11], this can be linked with the present purpose: *In order to elucidate.* [12]

(12) We have just mentioned *step widths* in [12] without confirming the present definition for readers - add *Here* as context. [13]

(13) As in [4], use the present tense to refer to Fig. 4; *the result was* becomes *results are.* [14]

We've just directed readers to Fig.4, but we haven't actually stated the Result; briefly include as part of sentence [14].

(14) This represents an abrupt contrast: simple <u>But</u> + short statement. (15) directly follows from [15], so (14) and (15) can be combined with a more 'mature' subordinate clause *Although.* [15]

(16) This Result belongs to subsection 3-2 - delete from 3-1. The topic of (16) and (17) is going to be discussed in greater detail in a later subsection, so signal this explicitly *will be discussed in 3.3.* [17]

---

(18) <u>From these results, especially from the point that the number of steps were equal to the number of Fe layers</u>, *we* suppose that each Fe layer had a different Hc, and that the steps were interpreted by a summention of them. (19) <u>Then</u> *we* made some multilayer films which had two Fe layers with various thicknesses. (20) *The construction of the film* is drawn schematicaly in Fig.5. (21) *The M-H loop observed for this film* is shown at the right part of this figure. (22) <u>Considering the value of magnetization (Ms)</u>, *the loop* is separated to two loops shown in figure. (23) <u>From this result</u>, *we* found that each Fe layer had different coercive forces (Hc) and the steps were interpreted by a summention of them. (24) <u>On the other hand</u>, *we* devided these loops with step widths into some loops by the above mentioned way. (25) <u>And</u> *the dependence of Hc on the total thickness deposited under the Fe layer concerned* were investigated as shown in Fig.6. (26) *The values of Hc* increased when the total thickness under the Fe layer increased. (27) <u>And when the thickness of Fe layer was constant</u>, *the values of Hc* were on similar line. (28) <u>And</u> *the dependence* was clearer in a) than b). (29) <u>Then as for one film heated up by two way described in the experimental procedure</u>, *M-H loops* were measured. (30) <u>For as-prepared film</u>, *some steps* were observed but at the high applied magnteic field, the divisions of steps were not clear as shown in Fig.7-a). (31) <u>After light heat treatment</u>, *the steps* were observed clearly as shown in Fig.7-b). (32) <u>And in Fig.7-c)</u>, *the steps* began to disappear.

[18] <u>From these results, especially from the finding that the number of steps is equal to that of Fe layers</u>, *we* hypothesize that each Fe layer had a different Hc, and that the M-H loops can be interpreted as a summation of the M-H loops which correspond to the sandwiched Fe layers. [19] <u>In order to confirm this hypothesis</u>, *we* fabricated some multilayer films which containing two Fe layers with various thicknesses: [19a] *their construction* schematically drawn in the left part of Fig.5. [20] *The M-H loop measured for this film* is inserted in the right part of the figure. [21] <u>Considering that the value of magnetization (Ms) is proportional to the thickness of the Fe layer</u>, *we* can separate the measured M-H loop into two loops as shown in the figure. [22] <u>Thus, on the basis of this analysis</u>, *we* divided all the M-H loops into several constituent subloops. [23] <u>As a result</u>, *it was found* that Hc depended on the total thickness of layers inserted between the substrate and the Fe layer concerned, as demonstrated in Fig.6. [24] <u>When the thickness of the Fe layers is constant</u>, *the values of Hc* show a linear dependence on the total thickness of the underlying layers. [25] *This trend* is observed more clearly in Fig.6(a) than in Fig.6(b). [26] <u>Consequently, from the analysis of these results</u>, two conclusions can be reached concerning the coercive force of each Fe layer sandwiched in the Fe/Ag multilayer film: [26a] *the first* is that, when the thickness of each Fe layer in a film is equal, the Hc increases as the total thickness of layers inserted between the Fe layer and the substrate increases.

[27] *The second* is that, when the total thickness of layers inserted between the Fe layer and the substrate is constant, the Hc of the Fe layer increases as the thickness of the Fe layer increases.

[28] <u>In order to investigate the nature of such loops in more detail</u>, *a film* was measured after various heattreatment, and its loops were compared with each other. [29] <u>As shown in Fig.7(a)</u>, <u>for as-prepared film</u>, *the division of steps* is not clear, and only 4 steps are easily recognized. [30] <u>After annealing at 373K for 24h</u>, *5 steps* are clearly seen as shown in Fig.7(b). [31] <u>However</u> <u>after annealing at 473K for 2h</u>, *several steps* disappeared, as shown in Fig.7(c). [32] *The cause for this* will be discussed in the next section in reference to the results of microstructure investigations.

---

(18) The #3 Result cycle is now completed and the #3 Comment section {see Figure 13} can thus be marked with a new paragraph. Change more spoken English form *suppose* to *hypothesize*; it has been suggested that the degree of confidence in our interpretation needs hedging, so modify *were interpreted* to *can be interpreted.*   [18]

(19) As in [12], minimal *then* may not be helpful; create a stronger link of our purpose with *hypothesize* in previous [18] - *In order to confirm this hypothesis..*   [19]

(20) The pointer to Fig.5 can be appended to [19] instead of its own separate short sentence.   [19a]

(21) The research process wasn't *observation* but *measurement.*   [20]

(22) We mention the *value of magnetization (Ms)* but we should clarify this; moreover, there is an important difference between the contexts *Considering the value* (i.e. now discussing) and *Considering that the value* (i.e. on this condition or for this reason); we mean the latter. Change *the loop is separated* to *we can separate the loop* as the former does not emphasise our own research contribution sufficiently - a switch to the higher writer visibility of *we* as Discourse Participant {see section 3.6.2} is appropriate.   [21]

(23) This is a premature Comment on a major Conclusion {see

Figure 13}, so should be delayed until all relevant
Results have been mentioned as part of this cycle. Again,
as in [5] [10], signal the reason for our conclusions,
this time (for variety) *Consequently*, see [26].

(24) The original context of contrast *On the other hand* does
not now make much sense following on from [21]; we
should clearly signal the <u>consequence</u> of this analysis
- <u>Thus</u>...*we divided*.. [22]

(25) The connector *And* is multifunctional like *Then* in [12] &
[19] {see Quirk *et al.* (1985:930-932); Eiler (1986:58)}.
And gives little indication of the Statement of Results
{see Figure 13} to come, so the connection should be more
explicit as in previous cases [5][10][21]: *As a result.*
[Sentence [22] contains the <u>purpose</u> of the study so after
that, comes Result; in the usual case, we explain
experimental results with such words <u>as a result</u>]. In
contrast to [21] *it was found that* is now preferable to
we found that, since the Empty Real World Theme {see
section 3.6.2} signals a research Result rather than a
higher profile statement of our own research action/
technique. [23]

{Halliday & Hasan comment that, since the *and* relation may
be felt by mature NSs to be structural and not cohesive,
'this is why we feel a little uncomfortable at
finding a sentence in written English beginning with *And*,
and why we tend not to consider that a [NS] child's
composition [cf. a novice NNS writer's FIRST RA draft]
having *and* as its dominant sentence linker can really be
said to form a cohesive whole' (1976:233). Moreover, in
the English-language RA genre, *And* would be highly marked
sentence-initial, but [the equivalent in Japanese is OK
sentence-initial]. In her corpus of L2 writing, Johns
(1980) found that the most frequent errors involved those
elements which students probably learnt first, for
example, conjunctions *and*, *but*; Eiler (1986) comments that

the extensive presence of *and* reinforces the oral nature
of discourse and Bloor & Bloor (1992) note that
inexperienced writers may incorporate features of
intonation and stress into their mental perception of
writing. Pettinari (1985) comments on the overproduction
of certain linguistic forms in her NNS data and
hypothesises that as NNSs learn the relevant discourse
organisation (in this case, of surgical reports in
English), they "fill in" the discourse functions with more
appropriate linguistic forms. In connection with this
hypothesis, see concordances of sentence-initial *And*, *But*,
Then in TABLES 30 a/b above which indicate their virtual
disappearance between novices' FIRST and FINAL RA drafts}.

(26) The major Conclusions and implications of this subsection
     can be clearly grouped together and then they can be
     signalled more effectively.   [26a]

(27) The same weak connector *And* - delete it's not additional. [24]

(28) As in [23][24], delete minimal *And*; (consecutive *And And*
     And - immature and 'poor style'); signal to the readers
     what the graphs collectively show i.e. *this trend is*
     observed.   [25]

     Having decided to group major Conclusions in [26] [27],
     this section can now be clearly signalled: *Consequently,*
     from the analysis of these results, two conclusions can
     be reached...   [26]
     As part of this major reiteration of Conclusions, (27)
     is now more clearly restated and signalled: *the second*
     is that...   [27]

(29) As with previous redraftings [3][12][19], the motivation
     for the research action is not clear with *Then.* The
     intention was to investigate *the nature of such loops in*
     more detail, hence the new context: *In order to.* This new
     cycle of Results {see Figure 13} is indicated by a new

(30) Fronting the pointer to Fig.7(a) *As shown in..* emphasises
the negative nature of this Result:*the division of steps
is not clear*; moreover, *some steps* should clearly be the
more precise *4 steps.*   [29]

(31) Again *some steps* is clearly inadequate; this is a case of
reference to a Figure without accompanying comment and
therefore a weak explanation - change to *5 steps.*   [30]

(32) The experimental treatments in (30) and (31) were not the
same - this is what makes the difference to the number of
steps (4 > 5) seen in the graphs. The original marked
theme:*And, in Fig.7-c)* (*and* yet again) doesn't at all
reflect this contrast in treatments, so it should be
strongly highlighted: *However, after annealing at 473K
for 2h..*   [31]

The cause of Results stated in [29] and [30] need not be
explained under this section as they are concerned with
structural features the subsection heading of 3.2. The
link: *The cause for this will be discussed in the next
section* provides cohesive forward momentum from this
first Results & Discussion subsection to the next.   [32]

5.3.6 Concluding remarks on unpublished (FIRST & FINAL) RA drafts written by NNS novices - Part II

The qualitative analysis of three sets of FIRST and FINAL
RA sections above clearly shows that a critical part of
the process of 'successfully' redrafting these extracts
entailed the manipulation of thematic and information
structure, within the bounds of generic and rhetorical
constraints which operate at different stages of the RA.
In addition, it is clear that such manipulations and
certain seemingly minor stylistic modifications are not
just the cleaning up of 'errors', but are an integral part
of the development of an RA under the interactive

influence of external sources - consultations and negotiations with colleagues, supervisors and journal editors - as revealed from a participant-observer viewpoint. In emphasis of the defining role of this interactive influence, Knorr-Cetina comments that the process of negotiation which precedes publication,

'illustrates the fact that the content of a published paper is not merely the result of an author adhering to the conventions of scientific writing. ..*technical critique* and *social control* are inseparably intertwined' (1981:106).

For EAP practitioners and researchers, it is clearly important to understand how such interaction is 'appropriately' manifested in linguistic terms so that awareness of such issues, and attendant procedures and strategies, can be raised as part of the development of novice NNS writers' academic communicative competence.

Our initial [HYPOTHESIS #7b] above suggested that NNS novices' FIRST drafts would show evidence of recognised features of immature writing which by means of redrafting processes would be emended towards recognised more mature features. In particular, in relation to marked Theme choices, FIRST drafts were expected to exhibit a greater degree of simple coordination of structures by multifunctional devices *and*, *but*, *then*; in contrast, FINAL draft sentences would show a greater range of cohesive devices and subordinate structures. Both the quantitative data, including concordances and statistical comparison with the base NS corpus (see section 5.2), and the qualitative data, in particular, the participant-observer commentary on Example #3: RESULTS & DISCUSSION above, present strong evidence which confirms [HYPOTHESIS #7b]. The data indicate how the redrafting process does indeed reflect elements of this transformation of a once relatively immature, unpublishable piece of writing into a relatively more mature version, now acceptable to the

'expert' readers who function as the gatekeepers to the academic community.

The unifying thread throughout the present work is Halliday's textual metafunction of language, Theme; this focus has been justified on the grounds of its text-structuring and genre-characterising potential. In relating to concepts of textual cohesion and coherence, however, it is natural that any discussion of thematic structure should make reference to the component of Rheme, as in some of the above comments, tables and figures related to Examples #1 and #2. It is particularly relevant to note that throughout the various drafts of the ten NNS novices' RA sections, rhematic content often remained largely unchanged. Since the most 'newsworthy' information of scientific interest to be reported carries the greatest communicative dynamism, it evidently becomes more quickly 'settled' in the unmarked conflated New/Rheme position. In turn, the manipulation of the Theme component, and to some extent information focus, become crucial resources for writers to create 'successful' RAs.

A genre-based writing syllabus, and the learner-centred nature of the methodology employed in the EAP classroom, aim to integrate attention to both the RA writing process and the written product. An important role here for the EAP practitioner is the exploration of teaching and research tools which enable the effective application of a wide range of insights about 'mature' redrafting procedures, strategies and language usage which help to create more 'successful' RAs. A detailed exploration of such a teaching/research tool as a heuristic for raising NNS novices' awareness about thematic structure and its manipulation as an integral part of the RA redrafting process is the subject of the final Part III of this thesis.

PART III

APPLICATION - FOCUS ON GENRE/PROCESS

CHAPTER 6   INVESTIGATION OF THE SCIENTIFIC RESEARCH ARTICLE WRITING PROCESS: PROPOSITIONAL CLUSTERS AS A HEURISTIC FOR RAISING NNS NOVICES' AWARENESS ABOUT THE MANIPULATION OF THEMATIC STRUCTURE IN REDRAFTING RA SECTIONS

6.0 BACKGROUND TO THIS CHAPTER

The previous section (5.3) highlighted aspects of NNS novices' redrafting of FIRST to FINAL RA drafts in response to feedback and criticism. By means of process-oriented data, particular attention was paid to novices' thematic control and manipulation and their evident contribution to the perceived 'success' of the finally published RA. Part III of this thesis continues to apply these insights on 'successful' RA writing, with the research setting having moved more clearly into the EAP (English for Academic Purposes) classroom.

As a result of the accumulation of both theoretical and applied insights into tracing thematic structure and control in RA writing, this chapter explores the use of the teaching/research tool of Propositional Clusters as a heuristic for raising NNS novices' awareness about the manipulation of Theme in the process of redrafting RA sections. Here, further process data are gathered from Academic Writing classroom exercises by means of observation of RA writing-type activities and analysis of several RA section drafts, with students attempting to comment on their redraftings. These process-oriented data are supported by a brief comparative statistical analysis with the findings from the base NS corpus reported in Chapter Three. Concluding remarks summarise the major findings from this research in relation to earlier reported work and pedagogical insights gained for practical application in the EAP classroom.

6.1 INTRODUCTION

In the previous chapter, analysis of thematic changes between novice NNS RA writers' FIRST and FINAL drafts suggests the following question: since the 'appropriacy' of thematic selections clearly plays an important role in creating 'successful' RAs, as judged by acceptance for publication in international English-language journals, how can awareness of the full manipulative potential of thematic components be raised amongst novice RA writers as part of L2 and research support? In other words, there is a need to explore effective techniques and EAP classroom exercises which function as research tools for teacher-researchers and as heuristic procedures for students. According to Flower & Hayes, 'a heuristic is an alternative to trial and error. It is simply the codification of a useful technique or cognitive skill. It can operate as a discovery procedure or a way of getting to a goal' (1977:450). This section focuses on one potentially valuable heuristic, namely the use of Propositional Clusters (PCs).

The use of PCs aims to integrate attention to elements of both bottom-up linguistic knowledge (Theme) and awareness of top-down rhetorical constraints (Moves) which guide the RA writing and redrafting processes (see section 3.10 for RA Theme-Move maps). Moreover, with the attention of novice NNS RA writers heavily focused on the strong pedagogic and heuristic framework of the written product, the use of PCs attempts to shift the focus of consciousness-raising (C-R) activities to the recursive processes of drafting and redrafting RAs in the light of critical language- and content-related feedback. The general approach to how PCs exercises are executed in Academic Writing classes thus aims to reproduce certain elements of the 'norm-developing' social-constructionist environment from which all research writing derives.

Since 'successful' RA writing has been defined here by means of external criteria, that is, acceptance for publication, for the purposes of judging the relative 'success' of written products arising from EAP classroom exercises, other criteria are evidently necessary. For this reason, subject specialist informants (see section 4.1.2) who had editorial experience were asked to rate the acceptability of draft RA extracts produced from PCs exercises.

The following sections first outline background research on propositions as a basis for text production, the form of PCs and the rationale behind their use as part of C-R activities with the same group of sixteen NNS novices who were interviewed above (see section 5.2.2.2); subsequently, the procedures adopted in their application in the EAP classroom are described.

6.1.1 <u>Research methodology - propositions as a basis for text production</u>

The term 'propositional analysis' derives from Kintsch's description of the system of mental operations that underlie the processes occurring in text comprehension and production, research into the latter being primarily concerned with the generation of recall and summarisation protocols (Kintsch 1974, 1977, 1985; Kintsch *et al.* 1975, 1977; Kintsch & Kozminsky 1977). Kintsch & van Dijk explain the central relevance of propositions to these processes:

'We assume that the surface structure of a discourse is interpreted as a set of propositions. This set is ordered by various semantic relations among the propositions. Some of these relations are explicitly expressed in the structure of the discourse; others are inferred during the process of interpretation with the help of various kinds of context-specific or general knowledge' (1978:365).

The semantic structure of a discourse is characterised at the levels of microstructure, the local level of the individual propositions and their relations, and the macrostructure, the global level which characterises the discourse as a whole and which is described in terms of proposition sequences. Thus, in consideration of thematic progression and development throughout a discourse, the propositions of a text,

'must be connected relative to what is intuitively called a *topic of discourse*, that is, the theme of the discourse or a fragment thereof. Relating propositions in a local manner is not sufficient. There must be a global constraint that establishes a meaningful whole, characterized in terms of a discourse topic' (Kintsch & van Dijk 1978:365).

A discourse is coherent only if its respective sentences and propositions are connected, and if these propositions are organised globally at the macrostructure level. In relation to the present work, these descriptions naturally correlate with one component of the microstructure, thematic structure, and with the global organisation of rhetorical moves which were integrated to create RA Theme-Move mappings (see section 3.10). It is the role of Propositional Clusters in developing novice NNS writers' awareness of the interaction of these micro- and macrostructural components of textual coherence and cohesion in research article writing that I wish to explore in this section.

The current pedagogic interest in PCs originates from Rutherford's (1987) work on the application of C-R methodology to aspects of L2 grammar. However, since the present work is guided by a systemic-functional framework, it should be noted at the outset that there are potential conflicts between this framework and Rutherford's own approach to the analysis of form and meaning. Most notably, his discussion of the Theme/Rheme distinction

does not generally conform to traditional Hallidayan viewpoints - his interpretation of 'initial rheme' (1987:77) is clearly inconsistent with the Hallidayan notion of Theme as 'point of departure', however its scope may be limited (see section 3.2). In connection with this, there is also confusion in Rutherford's account about the 'combining' or 'separating' approaches (see Fries 1983) taken in respect to the mapping of Theme and Rheme onto Given and New.

However, despite these inconsistencies, it is Rutherford's use of PCs exercises to raise students' consciousness about the writing process which is of interest here, since C-R is concerned with,

> 'developing the learner's powers of judgement and discrimination in respect of the semantic and discoursal demands on the grammatical structures of the target language, emphasizing in this reflective mode how well the learner understands the relationships between form and meaning' [Candlin (1987):preface to Rutherford (1987)].

Moreover, this reflective mode of C-R activities may also produce verbal protocols from subjects, although comments above (see section 5.2.2.1) suggest that the extent of self-report about L2 processes may be extremely limited for a variety of reasons.

6.1.2 The form of Propositional Clusters and their use for teaching and research purposes

A cluster consists of a non-finite verb + associated 'arguments' (noun phrases) arranged in a random order to the right of the verb. As Rutherford (1987) comments, this randomness is intended to avoid the bias of pre-selected word-order, since word-order choices and the consequent grammaticalisation attendant upon such choices are what we want students to make. Thus, the expansion of a PC into

complete clause(s)/sentence(s) is intended to shape
awareness of the effects upon surface form of the three
interrelated systems of discourse, semantics and syntax.
Let us look at an example of a PC from Rutherford
(1987:168) below:

> *solve - professor - problem - computer*

Possible expansions of this cluster into a sentence are:

[1a] *The professor solved the problem with a computer.*

[1b] *The professor's problem was solved with a computer.*

[2] *A computer solved the problem for the professor.*

[3a] *The problem was solved by the professor with a computer.*

[3b] *The problem's solution was found by the professor with
(the aid of) a computer.*

[4] *The solution to the problem was found by the professor with
(the aid of) a computer.*

In each case, different thematic choices evidently trigger
certain decisions about the formation of the rest of the
sentence. As a classroom writing exercise, this task may
help students to practise the purely mechanical
manipulation of propositions arising from thematic
choices. However, Rutherford emphasises that the
manipulation of isolated PCs as illustrated above, without
'the control of a schema which constrains their
operations' (Kintsch & van Dijk 1978:366), entirely
disregards the motivation behind competing thematic
selections in running text, which of course lies beyond
sentence boundaries. The work of Kintsch *et al.* and the
findings presented here in Chapter Three strongly indicate
that a matching of local and global structures is
necessary for the production of coherent texts as
exemplars of a specific genre.

From a pedagogic viewpoint then, it is necessary to create
contexts for C-R activities which serve 'to sensitize the
learner to the importance of discourse as the prime
influence upon the linear ordering of information chunks'
(Rutherford 1987:99). Given the clear discourse context of

drafting RA sections, choices will not be seen to be simply arbitrary as above, but dependent on thematic/rhematic development in the framework of generic constraints, such as Introduction moves and their inherent discourse functions. Thus, the process of drafting RA extracts made up of expanded PCs should not result in a random collection of unrelated sentences but needs to be guided by the many external forces, such as intended audience, which shape and define texts.

In the process of drafting texts, Rutherford suggests that syntactic adjustment of propositions following from grammatical choice in context makes possible the achievement of text cohesion as a natural rather than a contrived feature. In addition, with respect to the redrafting process, the manipulation of relationships between propositions and clusters aims to raise NNS novices' awareness about the nature of immature versus mature procedures, strategies and language usage in relation to the L2 RA writing task (see section 5.2.2).

With the current research based on a corpus of scientific RAs, PCs derived from scientific writing predominantly represent Real World topical Themes (see section 3.8), that is, the specialised technical vocabulary of which students have a high level of familiarity. Only a minor part of the demands of the exercise of PCs expansion lies in understanding technical language; it is, therefore, the 'appropriate' provision of absent text-structuring devices such as Context Frames (see section 3.5) which is of particular interest in the development of cohesive and coherent RA extracts, along with the 'appropriate' balancing of technical 'topic-based' Real World Themes and the more clearly 'interactional' non-Real World Themes (see section 3.6).

6.1.3 <u>Preparation for exercises using Propositional Clusters and EAP classroom procedures adopted for consciousness-raising activities</u>

The various steps involved in the preparation of PCs exercises and the classroom procedures adopted for their use in C-R activities with the sixteen doctoral students are outlined below. Comments on background issues raised at each step are integrated where relevant.

STEP #1) To investigate the potential of PCs as a pedagogic tool and a heuristic for Theme development and, since clusters are semantically conceived, it is evident that students must operate within an individually meaningful discourse context, namely, their own attempts at writing RA drafts. PCs must be tailored for each student and so RA drafts in English were collected some weeks prior to their intended use in the final sessions of the Academic Writing course. These drafts may still be rough notes of any section of the RA; however, particular attention was paid to the more discursive Introduction and Discussion sections, whose Themes according to previous data are evidently more varied in their discourse functions. For the present purposes of description of PCs, let us here refer to an example of the Introduction.

STEP #2) As original RA Introduction drafts were generally short (ca. 150-200 words: 6/8 SIEs), they were used in their entirety (longer Discussion subsections served as several separate PCs exercises). PCs were formulated and typed out, then separated into individual strips of paper. Rutherford's formulation of a PC (verb + noun phrases) is adapted for use with the greater number of arguments in complex sentences. An example of this adaptation is given below where the original sentence written in a novice's RA draft was: *Although several methods have proposed[2,3], these methods were limited to the tissue below surface.*

This sentence was split into two separate clusters:

```
propose - several methods - [2,3]
```

```
limit - below surface - methods - tissue
```

The reason for random manipulation of the order of the original propositions was threefold. First, to avoid the suggestion that PCs could simply be expanded by the linear reading from left to right of PCs, with students left to fill in articles/prepositions etc. In other words, thematic choices may have to be selected from a number of competing potential Themes embedded within the chain of arguments. Secondly, verbs in the original RA draft may be presented as nominalised forms, or vice versa. The intention here is for students to appreciate that a high degree of nominalisation is a characteristic of scientific writing and therefore the PCs exercise should develop their ability to manipulate such forms. Thirdly, to help counteract the possibility that students could simply remember word-for-word their original text. This possibility can never be entirely discounted; however, it was not intended that completed PCs products should in any way be measured against students' original piece of writing - the text may therefore be formulated in quite a different way.

STEP #3) At the start of the procedure in class, the first activity was for students to arrange the PCs, that is, the strips of paper (on average 6-10) according to their perceived schemata (see Figures 11-14).

STEP #4) A first draft expansion of PCs was attempted and this stage may also have involved changes to the first ordering of PCs strips. Students were requested not to

erase any emendations at any stage of the exercise so that evidence of changes would be tangible [in the Japanese education system, from elementary school right through to the graduate level, it is a habit for students to write in pencil so that mistakes can be instantly erased].

STEP #5) Since groups of students taking the Academic Writing Skills class came from the same laboratories, they now showed their #1 draft to a fellow student, generally well-informed about the research topic. Spontaneous discussion, which generally proceeded in Japanese, may focus on both scientific content and language emendations.

STEP #6) After reading the #1 draft, I may have commented on the text in my role as EAP teacher, attempting to elicit explanation from students through questions. Concerning the degree of form-focused intervention, no explicit mechanical correction was undertaken, but discourse-level concerns, for example, cohesive devices and the flow of information, were commented on in the style recommended by Clark & Ivanič (1991): 'I'm not sure I understand this, I don't quite get the connection between this sentence and this part'. However, much of this activity remained strongly content-focused for students; for example, if I were to say: 'I don't quite understand this bit' - ostensibly because of how propositions were linked together - the student would often start to explain the science by means of diagrams. From their experience of consciousness-raising (CR) activities within an EAP framework, Holmes & Ramos (1991) indicate that L2 learners do not, unaided, analyse their own problems very successfully and, moreover, have limited means of naming and discussing these problems. Lacking metalinguistic awareness and/or the ability to express such awareness in the L2, it may be the natural recourse for such students to focus on 'the science'. However, the dominance of content-focus over form-focus may be a potential obstruction to the cyclical sub-processes of

redrafting and correction of errors. For example, I found it curious that some very basic mechanical errors (misspelling of frequently used and very familiar technical words), as well as less surprising errors (S-V agreement), remained in the PCs drafts all the way through. In terms of the Flower & Hayes (1981b) cognitive process theory of writing, the monitor interface between the reviewing (evaluating and revising) process and the task environment (text produced so far) may be weak in inexperienced RA writers.

STEP #7) A #2 draft was completed within the same class time as the #1 draft. Students were told that they should at this stage aim to complete and finally polish the draft ready for presentation to an external source (for example, an outside researcher, a journal editor) as part of the review process of their research reporting.

STEP #8) At the next class one week later, a now-typed version of the #2 draft was handed back to students. Having gained some distance from their own work, students may revise their text at any level; this would become draft #3. They may further discuss their work with the EAP teacher or laboratory colleagues if they chose to do so. Rutherford stresses that,

> 'one of the principal exponents of the concept of grammatical C-R in L2 writing is therefore making the learner aware of what the nature of writing really is - to inculcate in him a disposition toward the emendation of written text as a *necessary*, even enjoyable, part of the writing endeavour' (1987:175).

However, this final stage in the PCs activity was entirely voluntary with no pressure to undertake any revision.

In the above description of classroom procedures, the use of students' own RA drafts as the focus of C-R activities

is seen as indispensible. This contrasts radically with an approach taken by Allwright *et al.* (1988) with their reformulation strategy where they themselves supplied students with their own selected propositional content on a given general topic. They comment that 'apart from the obvious advantage of ensuring comparability, we find that supplying the basic propositional content also helps students accept writing tasks that are not central to their subject-matter interests' (1988:237). This may be true with different EAP groups of students, for example, First Year undergraduates, but at the advanced level of Research English language support for doctoral students, there is little reason (or time) for students to accept tasks that are not central to their subject-matter interests. The more obvious the simulation of the target task, the greater is the risk of potential alienation from the very nature of the 'real' task. Moreover, Prabhu indicates that a lack of motivation in such teacher-designed tasks will not bring about in the learner the all-important 'preoccupation with meaning (or the message or content), since success in understanding things and success in saying what one wishes to say are both pleasurable activities for people - hence motivating' (1985:170).

To investigate the effect that externally-conceived PCs may have on a C-R activity, two trial PCs exercises were conducted prior to monitoring PCs data. One set of PCs was based on my own general title [The problems faced by Japanese graduate students in writing research papers in English] and the other was extracted from students' own RA drafts, as described above. An introductory lesson explained the aim of PCs exercises in supporting the 'real' task of writing RA drafts for publication, as well as the mechanical procedures involved in the completion of PCs exercises themselves.

From personal observation during the conduct of these trial PCs exercises, the earnest manner, keen interest and relative ease with which students tackled the PCs exercise based on their own propositional content and, on the other hand, the lack of those same attitudes during completion of the PCs exercise of my own invention, lead me to express serious reservations about the limited pedagogic value of the latter. The purely subjective evidence for this is partly based on Prabhu's (1987) level of 'surfacings', possible symptoms of involuntary language awareness with preoccupied deliberations, prolonged gazes and retrospective alterations throughout the course of task completion. Since the expansion of a network of PCs into a coherent and cohesive text depends on how interrelated links between propositions are first perceived and then realised, it is not surprising that the imagined links in the mind of one person may not correspond to those of another person, disregarding matters such as differing cultural expectations, language proficiency and content schemata. It was also very noticeable that, in the trial PCs exercise based on my title, the pattern of classroom dynamics was totally teacher-centred rather than learner-centred - in fact, discussion between class members at STEP #5 was entirely absent. A major conclusion from the work of Allwright *et al.* was that 'it seems that class discussions may be more powerful than reformulations in influencing non-native writers to modify their writing' (1988:252). The opportunity for students to discuss meaningful work with each other, thereby reflecting aspects of the 'norm-developing', social-constructionist nature of RA writing, is vital.

Another strong reason for individual, content-based PCs may lie in aspects of the cultural background of students. In reviewing work on L1/L2 writing groups, Nelson & Murphy comment on some of the difficulties of transferring

pedagogic methods used successfully with English L1 groups
to L2 groups:

'L2 students come from different cultural backgrounds that
reflect different attitudes toward working in groups... L2
groups also have different attitudes toward student-centred
activities such as peer-review of other students' compositions.
... In addition, L2 students often have differing notions
concerning the characteristics of "good" writing' (1992:173).

Rigorous peer-review is a sensitive issue in Japan, even
sometimes at the level of academic publications. In the
trial PCs exercises described above, it was certainly
clear that group-based discussion and critique of
students' texts on the common title [The problems faced by
Japanese graduate students in writing research papers in
English] was not fruitful. When asked to make suggestions
for 'improvements' to texts (at any level), even pairs of
students working together were very reluctant to imply
criticism of their partner's English language proficiency,
since, sharing a common title and similar opinions, there
could be very little debate amongst students about the
substance of the topic. The situation was quite different
with the PCs exercise based on students' own RA drafts, as
'correction' was approached through discussion of content
and students were motivated to communicate 'successfully'.

6.1.4 Findings and comments - Part I
In parallel with the presentation of findings in Chapter
Five, I will now briefly look at PCs exercises from both
qualitative (I) and quantitative (II) viewpoints. As with
the comparison of FIRST and FINAL RA drafts (see section
5.3), a qualitative perspective provides process data by
means of analysis of the intermediate drafts completed
prior to the finished product.

First of all, two examples - one INTRODUCTION and one
DISCUSSION - of expanded PCs written by one student are
presented. Each example includes three possible drafts, as
a result of the eight steps outlined above.

6.1.4.1 PCs Example #1 - INTRODUCTION

> RA TITLE: Chromosomal behaviour in Starfish (Asterina
> pectinifera): zygotes under the effect of aphidicolin,
> an inhibitor of DNA polymerase

> inhibit - DNA polymerase - [1] - aphidicolin

> undergo - A. pectinifera - zygotes - treatment - aphidicolin
> - cell division - [2]

> abandon - first cleavage - [3,4] - chromosomes

> investigate - mitotic apparatus - role - kinetochores
> - chromosomes - cell division

> explore - behaviour - chromosomes - value

> observe - mitosis - living cells - species - zygotes
> - unsuitability

> be - reason - poor transparency - cytoplasm - high optical density

Regarding the initial activity of ordering PCs strips
(STEP #2), it is clear that Swales's (1981) 4 Moves model
was extremely influential in its description of RA
text-organisation as this pattern was frequently referred
to by students in their justification for their chosen

300

ordering. Swales's work had been extensively covered as part of the Academic Writing course, students' first ever experience of formal instruction in writing RAs. It is therefore clear that subjects were strongly influenced by what was then perceived to be the preferred way to describe the structure of the Introduction section. Nevertheless, despite the obvious suggestion of contamination of data through 'pre-training', my purely subjective feeling from the way students responded is that reference to Swales's moves enabled them to neatly express intuitions they had hitherto been unable to verbalise, but which through the osmosis-like processes of their extensive RA reading they had undoubtedly possessed. Clark & Ivanič comment that one advantage of this kind of approach to consciousness-raising is that 'it generates an intuitive sort of metalanguage for all to draw on - metalanguage which has grown out of experience: a genuine need to "label"' (1991:181).

PCs Example #1 - INTRODUCTION #1 draft

---

(1) *Aphidicolin* inhibits DNA polymerase[1].

(2) By the treatment of aphidicolin, *zygotes of A. pectinifera* undergo cell division[2].

(3) However, at the first cleavage *chromosomes* abandon[3,4].

(4) Therefore, *it is necessary to* investigate the roles of kinetochores and chromosomes in making mitotic apparatus during cell division.

(5) However, at the observation of mitosis in living cells, *this species of zygotes* has unsuitability.

(6) *The reason* is their poor transparency of cytoplasm, which has high optical density.

(7) By using of new technique, *we* could observe chromosomes in living cells.

(8) In this paper, *we* explore behaviour of chromosomes and suggest the value of chromosome.

---

[1] *It is known that* aphidicolin inhibits DNA polymerase[1].

[2] <u>By the treatment of aphidicolin,</u> *zygotes of A. pectinifera* undergo cell divisions[2], although chromosomes are abandoned at the first cleavage[3,4].

[3] <u>Therefore,</u> *it is necessary to* investigate the roles of chromosomes and, especially, kinetochores in mitotic apparatus during cell divisions.

[4] <u>However,</u> *this species of zygotes* are not suitable for the observation of mitosis in living cells, owing to their poor transparency of cytoplasm, which has high optical density.

[5] <u>By using of new technique,</u> *the chromosomes* can be observed in living cells.

[6] <u>In this paper,</u> *behaviour of chromosomes* is explored and suggested the values of chromosomes and kinetochore.

The way this student has listed (and actually numbered) his sentences from the outset indicates a basic difficulty in his perception of how PCs may be combined to form a flowing paragraph/text. In PCs exercises, it was in fact not uncommon for first drafts to be formulated as separate sentences with sentence-combining and paragraphing only evident as part of the second drafting process. This may be a problem connected with the classroom task and how separated strips of PCs are perceived. On the other hand, in response to the questionnaire reported above (see section 5.2.2.2), the student whose PCs we are presently examining was amongst those who reported that they wrote sections of their first RA draft in the form of Japanese 'memos'. Therefore, this note-like approach to the PCs exercise may in fact be an integral part of the English #1 draft process for this particular student.

Thematic changes between drafts #1 and #2 are discussed below from the viewpoint of the #2 draft (with sentence number).

#2[1]   The predication of *aphidicolin* by *it is known that* now emphasises the recognised community validation of this statement. The writer now wishes to focus on the use of *aphidicolin* as a valid research tool, not the obvious fact concerning its effect. This refocusing of information changes the emphasis from what Myers (1990) terms the discourse of nature to the discourse of science.

#2[2]   This combines #1(2+3), where a CF Type 1 minimal <u>however</u> has been replaced with a CF Type 3 clause <u>although..</u>, and the former CF <u>at the first cleavage</u> in #1(2) is shifted to the end focus of New information in #2(2).

#2[3]   There is no thematic change from #1(4) but a slight adjustment of Rheme.

#2[4]   This combines #1(5+6). The suggested PC argument may have presented a too strongly marked choice in the topical Theme *reason*, functioning as Objectivised Viewpoint in #1(6). This became a dependent clause Context Frame Cause-reason/result CF <u>*owing to*</u> in #2(4). The information regarding *the observation of mitosis in living cells* is what is 'newsworthy' and has therefore been decontextualised from #1(5).

#2[5]   The topical Theme *chromosomes* has become research- rather than researcher-orientated; the Discourse Participant *we* possibly represents an overfocus on their personal contribution at this introductory stage of the RA.

#2[6]   As in the previous sentence, there is a further downgrading of Participant status *we* through the thematic selection of a Real World Theme *behaviour of chromosomes*.

{1} *It is known that* aphidicolin inhibits DNA polymerase[1]. {2} <u>By the treatment of aphidicolin,</u> *zygotes of A. pectinifera* do undergo cell divisions[2], although chromosomes are abandoned at the first cleavage[3,4]. {3} <u>Therefore,</u> *it is necessary to* investigate the roles of chromosomes and, especially, kinetochores in mitotic apparatus during cell division. {4} <u>Unfortunately,</u> *this species of zygotes* are not suitable for the observation of mitosis in living cells, owing to their poor transparency of cytoplasm and high optical density. {5} <u>However, by using of a new technique,</u> *the chromosomes* could be observed in living cells. {6} <u>In this paper,</u> *the behaviour of chromosomes* was explored and the values of chromosomes and kinetochore were suggested.

The PCs have become a paragraph in this #3 draft and thematic changes focus on the manipulation of CFs to create a tighter rhetorical structure in the text.

#3{2} The student's use of the emphatic operator 'do' is interesting. In considering the many problems that inexperienced writers have, Bloor & Bloor (1992) speculate that their mental perception of what they write may incorporate features of intonation and stress. This appears to be the case here with spoken emphatic 'do'. However, it is of course curious that it appears for the first time in this third draft and suggests the writer's attempts to more strongly argue for his own research space. [However, see DISCUSSION point #2[2] below, where the emphatic operator 'did' is removed, this time after the #1 draft].

#3{4} Personal Viewpoint CF <u>unfortunately,</u> a very marked choice when considering its frequency (cf. section 3.7) now replaces <u>however</u> in #2[4]. This may be because, in the next sentence #3{5}, the writer has

decided on the need for a stronger contextualisation of the Means CF by using of a new technique, which is instrumental in the identification of the new research technique being presented here. The additional primary CF Contrast/concession however in #3{5} would have left two consecutive sentence-initial examples of an identical CF. It can be speculated that this repetition may have caused the retrospective selection of unfortunately, rather than the writer's desire to more strongly indicate his personal Viewpoint.

#3{6} In this sentence, as in the previous one, the tense has been changed from present to past; the reporting tense is now in line with the conventions of announcing principal findings (see Swales & Najjar 1987) in the student's field. The awkwardness of the seemingly subjectless dependent clause in #2(6) ...and suggested the values... has been resolved by promoting values of chromosomes... to thematic position.

6.1.4.2 PCs Example #2 - DISCUSSION

RA TITLE: Chromosomal behaviour in Starfish (A pectinifera):
zygotes under the effect of aphidicolin, an inhibitor of DNA
polymerase
Subsection title: visualization of chromosomes by
fluorescently labeled whole histones.

visualize - present study - chromosome movement - polar
body formation - cleavage

visualize - means - microinjection - fluorescently labeled
histone - starfish oocyte

find - egg - treatment - aphidicolin - alignment - chromosomes -
equatorial plane - metaphase spindle - first cleavage

take place - cleavage - no movement - chromosomes

become - chromosomes - incompetent

take place - subsequent cleavages - achromosomal condition
- reports [2,3]

enter - nuclear proteins (e.g. histones) - nucleus - cytoplasm
- well-known [16-18]

show - this study - fluorescently labeled histones - entering
- germinal vesicle/nucleus - starfish oocytes/eggs

investigate - nuclear behavior - Drosophila embryos
- Minden *et al.* - use - fluorescently labeled histones

---

(1) <u>In the treatment of aphidicolin</u>, *the egg* was found to align chromosomes on the equatorial plane of metaphase spindle during the first cleavage.

(2) *The first cleavage* did take place without chromosome movement.

(3) <u>At this process</u> *chromosomes* became incompetent.

(4) <u>In the achromosomal condition</u> *subsequent cleavages* have taken place.

(5) *Some* report [2,3] that.

(6) *It is well-known* [16-18] that nuclear proteins (e.g. histones) enter nucleus from cytoplasm.

(7) *Minden et al.* have investigated nuclear behavior of Drosophila embryos by by use of fluorescently labeled histones.

(8) <u>To visualize chromosomes</u>, *the fluorescently labeled histones* were introduced into starfish oocyte by means of microinjection.

(9) *This study* showed entering of fluorescently labeled histones into germinal vesicle/nucleus in starfish oocytes/eggs.

(10) <u>Furthermore</u>, *chromosome movement* was visualized during both polar body formation and cleavage in the present study.

PCs Example #2 - DISCUSSION #2 draft

---

[1] <u>In the treatment of aphidicolin</u>, *the egg* was found to align chromosomes on the equatorial plane of metaphase spindle during the first cleavage. [2] *The first cleavage* took place with no movement of chromosomes, which became incompetent. [3] *Some reports* [2,3] show that in the achromosomal condition subsequent cleavages have taken place.

[4] *It is well-known [16-18]* that nuclear proteins (e.g. histones) enter nucleus from cytoplasm. [5] <u>For example</u>, *Minden et al.* have investigated the nuclear behavior of Drosophila embryos by using of fluoresently labeled histones. [6] <u>In order to visualize chromosomes</u>, *we* introduced the fluorescently labeled histones into starfish oocyte

by means of microinjection. [7] *Our study* showed that they entered germinal vesicle and nucleus in starfish oocytes and eggs. [8] Furthermore, in the present study, *chromosome movement* was visualized during both polar body formation and cleavage.

With the DISCUSSION section, this student continued to list PC expansions in his #1 draft, but his #2 draft was formed as two distinct paragraphs.

#2[2]   This neatly combines #1(2+3) with a loss of at this process, now an unnecessary context. The same comment applies here as in INTRODUCTION #3{2} above regarding the use of the emphatic operator 'did' in #1(2).

#2[3]   This combines #1(4+5) which clarifies that the statement in #1(4) is a report of previous research, rather than a possible statement of present findings.

#2[5]   With the number of references used in this particular paper, the for example prior to the Interactive Participant *Minden et al.* now contextualises a broader range of reviewed work.

#2[6]   Thematic changes here in the DISCUSSION are the exact reverse of those in the closing statements of the same student's INTRODUCTION section above. The now more clearly interactive Themes *we* in DISCUSSION #2[6], *our study* in #2[7] and the added context in the present study in #2[8] serve to more clearly define exactly what it is that is novel about the new investigative technique that the researchers are presenting in this paper and moreover they emphasise its perceived usefulness. This kind of thematic adjustment to both closing INTRODUCTION and DISCUSSION statements now helps to create a more 'appropriate' balance of topic-based and interactional Themes which conforms to the 'hour-glass' social organisation of RA discourse.

{1} <u>In the treatment of aphidicolin</u>, *the egg* was found to align chromosomes on the equatorial plane of metaphase spindle during the first cleavage. {2} *The first cleavage* took place with no movement of chromosomes, which became incompetent. {3} *Some reports [2,3]* show that in the achromosomal condition subsequent cleavages have taken place.

{4} *It is well-known* [16-18] that nuclear proteins (e.g. histones) enter nucleus from cytoplasm. {5} <u>For example</u>, *Minden et al.* have investigated the nuclear behavior of Drosophila embryos by using of fluoresently labeled histones. {6} <u>In order to visualize chromosomes</u>, *we* introduced the fluorescently labeled histones into starfish oocyte by means of microinjection. {7} *Our study* showed that they entered germinal vesicle in starfish oocytes and nucleus in eggs. {8} <u>Furthermore</u>, <u>in the present study</u>, *chromosome movement* was visualized during both polar body formation and cleavage.

There are no thematic changes at all between these two drafts, in fact only a slight adjustment to content in #3{7} was made.

6.1.5 <u>Concluding remarks</u> - <u>Part I</u>

Various levels of changes between PCs drafts, particularly those concerning Theme, are of considerable interest. In investigating the quality of such changes above, it is clear that the more significant manipulations between drafts take place in response to text-external influences and affect, for example, the flow of Given-New information in the text. However, in order to judge the 'appropriacy' of thematic selections, the key question as to whether thematic changes between drafts actually make for a more cohesive and coherent RA extract needs to be addressed. It would certainly appear that such changes are an integral part of the negotiated rewriting process which arises

through debate over scientific content and how new contributions are to be presented to the academic community. Entering into this level of scientific dialogue is generally expected to make an RA draft more 'successful', that is, more likely to be accepted for publication.

However, when we consider the present efforts to simulate the 'real' task through EAP classroom PCs exercises, it would only seem possible to gain a highly impressionistic answer to the key question as to whether thematic modifications between PCs drafts make for a more 'successful' piece of research reporting. Comments on the examples of expanded PCs from one individual student have highlighted aspects of his redrafting processes. The follow-up to this analysis involved the interview of his immediate research supervisor. This NNS academic had served on the editorial board of an English-language scientific publication and, as a reviewer of RAs submitted by NNSs, was well acquainted with the potential influence of L2-related infelicities at the review stage of RAs. The interview procedure with the supervisor was as follows:

The three drafts of the INTRODUCTION and DISCUSSION sections derived from PCs Example #2 were presented separately for comments. The extracts were now typed as single paragraphs and were not marked in any way to indicate their stages of drafting. The supervisor was asked the following two questions:

(1) Given the title, which extract is the clearest, that is, the best piece of research reporting?
(2) In your opinion as an RA reviewer, is this better draft now acceptable for publication as it is or does it require yet further revision? If yes, what needs attention?

The replies were as follows:

(1) Draft #3 was selected as the best piece of research reporting for both sections.

(2) The supervisor commented on the INTRODUCTION #3 draft, "it needs revision - some attention to historical aspects; it needs references". When asked to clarify the 'historical aspects', he indicated his concern that the research finding *although chromosomes are abandoned at the first cleavage* in #3{2} needed greater contextualisation, that is, readers would require deeper background on this significant qualification. From this it can be interpreted that the validity of this scientific information introduced by means of the marked Contrast/concession CF may not be taken as Given at present; thus the information and thematic structure would require further manipulation. Concerning the DISCUSSION, "it needs revision, I think. English polishment, especially the usage of article, for example, *starfish oocyte* needs *the*". Compared to the external discourse-community oriented concern indicated above for the INTRODUCTION draft, this last observation originates from quite the opposite level of concern - of the cleaning-up of minor errors.

As suggested by protocol analysis (see section 5.2.2) which investigated novices' RA writing procedures and strategies, the elicitation of self-report type data from NNS researchers may be extremely problematic. They may well consult bilingual technical writing manuals (see section 5.2.2.2 - #6) but they will generally have had no exposure to the discussion of procedures and strategies in the RA writing task. In emphasis of this difficulty, my comments on the supervisor's statements above are primarily interpretive, rather than a verbatim report of what he could actually express:

'What we call our data are really our own constructions of other people's constructions of what they... are up to' (Geertz 1973:9).

An internationally-published academic, the supervisor commented that, although he was very pleased to help me (and thereby his students) with my questions, this task of commenting on his judgements of others' L2 RA drafts "was more than painful" due to his inadequate range of expression in English for this particular exercise.

A minimal amount of verbal report data was also the output from his research student, and it must be noted that, despite some interesting protocols on RA writing procedures and strategies (see section 5.2.2.2), he was not an exceptional case amongst the group of sixteen NNS novices. My prompting for 'think-aloud' commentary in English during PCs activities was often met with anxious silences, sharp intakes of breath (a much-valued Japanese habit when being indirect), plenty of false starts in L1 and, in English, "I don't know exactly"; occasionally, this particular student would respond with surface comments: "I change this to this". He would happily show me the result of what he had changed on paper but we could only really gain insight into aspects of his RA redrafting processes by means of analysis of written drafts.

In contrast, it must be remembered that in STEP #5 of the procedure, discussions between students in their L1 about scientific content and their L2 writing were at times lively, recalling the earlier point by Holmes & Ramos (1991) about the potential advantages of L1 reporting. However, in the L2, at any rate verbally, the reasons for emendations remained out of reach. It seems clear that, if we want L2 students to be able to talk about their writing processes, 'it is likely that the development of a user-friendly metalanguage... is a necessary condition for improvement' (James & Garrett 1991:20). However, for the EAP practitioner providing *in situ* L2 support for NNS novices, time constraints may put serious practical limitations on EAP course objectives, including the degree of detail and amount of time that can be spent on

introducing user-friendly metalanguage (on a "need-to-label" basis) to enable NNS novices to talk about aspects of their L2 writing processes. This is naturally particularly valid when it is the submission of their latest RA to an academic journal which is uppermost in their minds.

In defence of ethnographic validity, Swales comments that,

> 'whatever the approach, it would seem that ethnographic accounts offer a narrative richness of context, a depth of insight into particularity that is of incontrovertible value in our attempts to understand modern professional life, insofar as it is enacted in communicative conduct' (1987a:126).

In the present case, attempts to understand initiation into modern academic-professional life, as enacted in aspects of communicative conduct within the scientific community, have needed to draw on indirect interpretation of the L2 writing processes and written products of the science culture. However, the fact that immensely valuable process data can still be gathered from above PCs procedures emphasises the usefulness of a functional, genre-based approach to the participant-observation of educational settings such as the EAP classroom and the science laboratory.

6.1.6 <u>Findings and comments</u> - <u>Part II</u>
In the presentation of previous findings throughout the present corpus-driven study, the quantitative perspective has preceded the qualitative. However, the major emphasis in this section on process in PCs exercises has been highlighted by the presentation of qualitative data first. To round off the present analysis, in this second part of PCs data the products from two PCs exercises (in total 64 drafts of expanded PCs - FIRST and FINAL drafts of one INTRODUCTION and one DISCUSSION section from the 16

doctoral students) are briefly examined from a quantitative perspective. As a result of having compared unpublished NNS data (section 5.2) with the main published NS corpus (sections 3.7 & 3.8), we can speculate on the following hypotheses which can be tested quantitatively:

[HYPOTHESIS #9]
As a general hypothesis, the exercise of expanding PCs will be 'authentic' in the replication of RA thematic structure in that PCs data on the variable distribution of marked and unmarked Themes across RA sections will correlate significantly with those previously determined by the 'base' NS corpus analysis.

[HYPOTHESIS #9a]
However, bearing in mind previous analysis of NNS novices' FIRST RA drafts (see section 5.2), there will be a tendency in PCs FIRST drafts for novices to 'overuse' minimal Type 1 CFs and 'underuse' Type 3 CFs, compared to previously established norms. However, the FINAL redrafting of PCs, partly in response to external feedback, will help to create a relatively more 'mature' piece of research writing and this will be reflected in the modified usage of Context Frames.   .

[HYPOTHESIS #9b]
The subtly evaluative, interactive potential of non-Real World Themes (for example, Empty H&O Theme) vs. predominant Real World Themes may not be fully realised in PCs FIRST drafts, but FINAL redrafting may increase their usage.

In order to address [HYPOTHESIS #9] and thus verify the 'authenticity' of PCs exercises, the brief findings reported here are limited to basic data, including correlations between sets of Introduction and Discussion PCs and the NS corpus norms. Fuller data are given in appendices where indicated.

6.1.6.1 SIEs - GSs vs. CFs

TABLE 38 shows the overall percentages of marked CFs vs. unmarked Subjects (GSs) used as SIEs in FIRST and FINAL PCs drafts. These data can be compared with those of the *NS corpus* (see TABLE 3).

TABLE 38
PCs Exercises: SIEs - GSs vs. CFs
- *comparison with the NS corpus*

	INTRODUCTION		DISCUSSION			*NS corpus*	
	FIRST > FINAL		FIRST > FINAL			*INT*	*DIS*
Total N	(90)	(84)	(103)	(93)			
SIE = GS	42%	42%	61%	53%		*66.1%*	*60.5%*
SIE = CF	58%	58%	39%	47%		*32.9%*	*39.0%*
	100%	100%	100%	100%		2	
SD[1]	[16]	[18]	[17]	[18]		[*15.3*]	[*13.2*]

[1] *Standard Deviation of Context Frames*
[2] *values of SIE non-GS/CF not included in totals*

As in the case of the RA drafts from 10 of these same students (see section 5.2), there is a noticeably higher usage of CFs in Introduction sections compared to the NS texts. It is worth noting that further drafting did not change the 58% in Introductions, but actually increased CF usage in Discussion sections from 39%>47%. In both sections, further redrafting involved sentence-combining.

6.1.6.2 CFs by function

It must be remembered that, although PCs supplied a number of propositions from which predominantly Real World topical Themes were to be selected, Context Frames were to be integrated by students themselves in an effort to create texture and they are, therefore, of particular interest. The full data of marked Themes by discourse function are given in APPENDIX K1 (Introduction) and K2 (Discussion). TABLE 39 gives the correlations with the appropriate sections of the main NS corpus.

315

TABLE 39
PCs Exercises: CFs by Function
- *correlations with the NS corpus*

	INTRODUCTION FIRST > FINAL		DISCUSSION FIRST > FINAL	
r =	.903	.915	.832	.872

As has been consistently seen before, correlations ($p<0.05$ throughout) between the norms of the NS corpus and other samples of RA writing are strong - in this case, there is greater agreement over CF usage in multifunctional Introductions than with Discussion sections; the latter may be due in part to the fact that novices' Discussion drafts, from which PCs were extracted, may have been intended as combined Results & Discussion sections. However, in both cases, further drafting moved the correlation slightly towards greater agreement with the NS corpus, although with the sample size involved, conclusions must remain speculative. It is important to note that the measure of correlations with the NS corpus is not taken here as proof of 'good', 'mature' or necessarily 'successful' writing *per se*. Given that any EAP-classroom based exercise can really only attempt to simulate, to a greater or lesser extent, components of the 'real' RA writing task, correlations are of interest here in order to establish the 'authenticity' of PCs exercises in being able to 'appropriately' replicate the thematic patterning of published, externally-validated RA products.

6.1.6.3 CFs by Type
In addressing [HYPOTHESIS #9a], did FIRST>FINAL drafting of PCs involve a move from predominant minimal CF Type 1 usage to more complex Type 3 CFs as was suggested in the corpus (N=10) of novices' FIRST>FINAL RA drafts (see TABLES 28-31b)? TABLE 40 below indicates this was valid for Introductions but not for Discussion sections, where

the usage of Type 1 CFs increased - a mixed response, therefore, to [HYPOTHESIS #9A]. It is evident that CF Type 3 usage is far below the NS corpus base norms, but again, PCs data may be too limited to make a reliable comparison.

TABLE 40
PCs Exercises: CFs by Type
- *comparison with the NS corpus*

	INTRODUCTION FIRST>FINAL	CHANGE	*NS corpus*	DISCUSSION FIRST>FINAL	CHANGE	*NS corpus*
TYPE 1	58% > 53%	-	*43%*	40% > 52%	+	*46%*
TYPE 2	38% > 37%	-	*39%*	47% > 39%	-	*28%*
TYPE 3	4% > 10%	+	*18%*	13% > 9%	-	*26%*

However, since CF Types reflect the grammatical and not the functional paradigm of marked Themes, it must be remembered that a lack of Type 3 CFs in NNS novices' drafts is not itself a discoursal 'error', but a possible indication of 'less mature' L2 writing. These limited data suggest therefore that, even though students may come to recognise RA writing 'as a fundamentally interactive enterprise' (Widdowson 1983:44) where drafts are revised in response to feedback and criticism, EAP classroom-based writing exercises may in the end only be able to maintain a weak interface with that enterprise.

For further details, APPENDICES L1 (Introduction) and L2 (Discussion) list Type 1 CFs used in the PCs mini-corpus and APPENDICES M1 (Introduction) and M2 (Discussion) the Type 3 CFs.

6.1.6.4 Distribution of Subject roles

Analysis reported (see section 3.8) of the discourse functions of unmarked Subjects indicated a clear pattern of Real World and more textually interactive non-Real

World Themes throughout RAs. Were similar patterns evident in expanded PCs? TABLE 41 below indicates a strong correlation with the NS corpus; APPENDICES N1 and N2 list data for Introduction and Discussion sections, respectively.

TABLE 41
PCs Exercises: Distribution of Subject Roles
- *correlations with the NS corpus*

	INTRODUCTION FIRST > FINAL		DISCUSSION FIRST > FINAL	
$r =$	.967	.983	.960	.953

The ratio of Real World to non-Real World Themes in the PCs drafts compared to the *NS corpus* was 35%>31% to *33.6%* for Introductions and 34%>36% to *34.4%* for Discussions. Therefore, given the limited data, it appears that there is no clear evidence, quantitatively at least, to substantiate [HYPOTHESIS #9b], that the more interactive potential of non-Real World Themes was not fully realised in PCs FIRST drafts and that FINAL redrafting would increase their usage. However, as seen from other process data above (see sections 5.3 and 6.1.4.1/2), this does not reflect the quality of thematic choices and the need for further revision of drafts.

From TABLE 41 above, it can be seen that, overall, FIRST>FINAL redrafting improved the significant correlation (p<0.05) in the Introduction section with the *NS corpus*, but not with the Discussion. As an indication of the usage of one subtly interactive structure which Davies (1988a) has highlighted as being problematic for novice academic writers to deal with, APPENDICES P1-P4 list concordances of Empty H&O Themes (*it is likely..* etc.) in FIRST>FINAL Introduction and Discussion drafts. According to *NS corpus* data, it would be expected that the Discussion (*13.6%*) of results and claims would imply a greater need for H&O Themes than in Introductions (*6.8%*)

and this was also the case with PCs drafts. It is
interesting to note that the FIRST>FINAL redrafting of
Discussion PCs decreased the usage of Empty H&O Themes
(8%>5%), and the H&O domain overall (20%>17%), and this
was compensated for by an increase in the other two
non-Real World domains, Participant (12%>15%) and
Discourse (2%>4%). The number (10%>10% vs *2.8%* - see
APPENDIX N2) of OV examples in Discussion PCs (headwords
difficulty, discrepancy, reason, evidence, drawback, fact)
is generally explained by their presence as one of the
original propositions drawn from novices' earlier RA
drafts. For example, the PC for one student,

> *take place - epitaxial growth - evidence - condition*

was rendered in the FIRST draft as,

> *One of the evidences is that epitaxial growth take*
> *place under this condition.*

but in the FINAL as,

> *Moreover, epitaxial growth took place under this*
> *condition. This evidence....*

creating a stronger sequence of Result + Comment (see
Figures 13/14 above). In other cases, nominalised forms
could be avoided (*difference -> differ, different*) or
students translated the discourse function (*reason -> due
to*). However, the usage of OV propositions does indicate
the technical difficulty of formulating PCs which should
where possible remain interactively 'neutral' and Real
World-based.

6.1.7 Concluding remarks - Part II
In commenting on the reasons for developing an academic
writing syllabus which involves attention to both top-down
and bottom-up strategies in parallel, Davies observes,

> 'gaining control over a particular genre, requires by
> definition an awareness of the top level formal and content
> schemata for structuring it. But an exclusive focus on top level
> features has limited potential; it may provide an outline and

section headings for a text but will not result in a full and comprehensible realisation of it. Conversely, a focus exclusively on form, at sentence level, will not equip students for the creation of coherent discourse. What is required is a match between top-down and bottom-up knowledge' (1988b:134).

The fact that the PCs data presented above correlate significantly with the main corpus of 36 NS texts, taken throughout the present work as the base norms of RA Theme selections against which to compare 'inexperienced' NNS RA writers' drafts, suggests that expanded PCs are generally authentic in their replication of the conventionalised thematic patterns understood to characterise this genre. This is not to say that there are no significant contrasts between corpora concerning developmental phenomena in RA writing; but it can be suggested that PCs provide a valuable heuristic and pedagogic framework within which both top-down and bottom-up linguistic information are attended to in creating the recognised thematic texture of the scientific RA.

However, it is clear from commentary on the quality of FIRST>FINAL process data from novices' unpublished RA drafts and PCs products that statistical correlations are only a starting point from which to proceed. In line with earlier comments on EAP classroom techniques by Clark & Ivanič (1991) and Allwright *et al.* (1988), the writing models and written products deriving from such activities are only as important as the opportunity provided for a forum of discussion about the many processes and stages involved in RA writing and in getting RA submissions accepted for publication. The participants in this forum are those whose interactions have been the focus of this work: 'inexperienced' NNS novices, their 'experienced' NNS research supervisors, 'expert' NS and NNS editors through survey findings and EAP teacher-researchers. Despite the significant limitations on the breadth of such discussions imposed by the L2 medium, novices' RA and PCs drafts

provide an important focal point for the development of academic communicative competence.

Through student-centred techniques, such as PCs exercises outlined above, it is important that attention to the development of the *individual* novice's RA writing processes is integrated with his/her intuitive reader-based knowledge of the many well-documented features of the written RA product. This matching of levels of both implicit knowledge, that is, the kind of knowledge which NNS doctoral students bring with them to an Academic Writing class, but about which they may not be able to speak, and explicit knowledge, that is, the structured input from a functionally-based Academic Writing course [for example, Davies (1987); Weissberg & Buker (1990)], can provide a powerful boost to the 'learning egos' (Hedge & Gosden 1991) of NNS novices. Importantly, this approach capitalises on the 'learning readiness' (James & Garrett 1991) of novice RA writers who experience an immediate need for a particular skill. In this way, on the threshold of their careers in international scientific research, this kind of L2 support may better prepare NNS novices in their future quest for 'success' and recognition *on an equal research basis*, as measured by the conventions of an increasingly competitive academic world.

CHAPTER 7   CONCLUDING REMARKS

7.0 INTRODUCTION

Working within a systemic-functional framework of analysis
and description, this thesis has presented a theoretical
and applied study entitled "A genre-based investigation of
Theme: product and process in scientific research articles
written by NNS novice researchers".

The theoretical analysis and description set out to
investigate the discourse-functional relationship between
Theme and generic elements of structure in a corpus of
research articles (RAs), written by NS (English
native-speaker) scientists and published in international
academic journals. Subsequent applied discourse analysis
and description focused on aspects of 'appropriate'
thematic control in the RA drafts of a group of NNS
(English non-native speaker) novice researchers attempting
their first RAs in English. Concluding remarks on findings
from the study reported here address the six major areas
of research outlined in Chapter One (see section 1.6);
these general research questions and hypotheses are
repeated below for convenience of reference.

GENERAL RESEARCH HYPOTHESIS #1: Halliday's notion of Theme
plays a significant role in the characterisation and
dynamic within-text structuring of one particular genre,
the scientific research article (RA).

GENERAL RESEARCH HYPOTHESIS #2: Such Themes can be
categorised according to their discourse-functional roles
at given stages of RA discourse and throughout the RA as a
whole.

GENERAL RESEARCH HYPOTHESIS #3: Thematic selections in
this genre are constrained by the changing rhetorical

purposes, signalled by means of moves, which operate throughout the different stages of RA discourse, here, taken as formally separated sections [Introduction, Experimental, Results, Discussion].

GENERAL RESEARCH HYPOTHESIS #4: The 'appropriate' usage (as highlighted through analysis of a corpus of NS RAs) of thematic selections across RA sections appears to contribute significantly to the perceived coherence and cohesion of a scientific research article. It is thus a significant measure of academic communicative competence and 'success', as defined and validated by acceptance of an RA for publication by an international English-language scientific journal.

GENERAL RESEARCH HYPOTHESIS #5: An examination of the RA writing procedures and strategies, and, in particular, of the progressive drafts of written products of NNS novice researchers in response to feedback and criticism illustrates the development of this crucial aspect of academic communicative competence.

GENERAL RESEARCH HYPOTHESIS #6: A pedagogic focus on the manipulation of genre-specific thematic choices in RA writing activities in the EAP classroom may raise NNS novices' awareness about the potential of thematic manipulation to help create 'successful' texts.

A summary of how the major findings of the present work have addressed these general research hypotheses is followed by comments on apparent limitations of this study with consequent suggestions for areas of relevant further research.

7.1 SUMMARY OF MAJOR FINDINGS

Findings which address GENERAL RESEARCH HYPOTHESES #1, #2
and #3 above relate to the major theoretical component of
the thesis title "A genre-based investigation of Theme:"
and are therefore grouped together.

An evaluation of current approaches to genre analysis in
the fields of linguistics and language education
highlighted the theoretical and methodological traditions
of English for Academic Purposes (EAP) and
systemic-functional linguistics which this study sought to
combine. In order to contextualise this "genre-based
investigation", Chapter Two outlined a systemic stratified
view of the social systems and structures of genre and
register which both determine and reflect choices in the
language systems and structures of the semantic and
lexicogrammatical strata (see Figure 3 above). Thereby,
the 'higher' order plane of genre as realised by
systemicists' *elements of structure*, and EAP analysts'
rhetorical *moves*, was related to the 'lower' order
lexicogrammatical structure of Theme. The central interest
herein lay in a detailed correlation of these levels of
extralinguistic (genre) and linguistic (Theme)
description.

Consequently, a discourse-functional categorisation of the
binary components of marked Theme (CONTEXT FRAMES) and
unmarked grammatical Subject (TOPICS) was devised and
described in Chapter Three. Based on the analysis of a
corpus of published NS scientific RAs (N=36),
discourse-functional analysis of Context Frames and
Subjects indicated the constraints acting on thematic
choices across the four formally labelled RA sections
(Introduction, Experimental, Results, Discussion)
according to their respective rhetorical goals as realised
in generic moves.

Analysis of marked thematic Context Frames by means of nine proposed discourse-functional categories clearly showed dynamic patterns across RA sections which could be predicted on the basis of the inherent 'method of development' (Fries 1983) of the scientific RA. Context Frames used here reflected the specific hypothetico-deductive framework of scientific reporting (Cause-reason/result, Cause-purpose; Condition-real, Condition-hypothetical), as well as employing more general methods of textual development (Contrast/concession; Addition-appositive, Addition-emphatic; Location in Real World Time and Space). Thus, it was suggested that a conventionally 'appropriate' flow of marked thematic patterns contributed significantly to the within-text structuring of the scientific RA and thereby to its perceived textual cohesion and coherence.

Analysis of unmarked Subjects by means of four proposed discourse-functional domains showed a different aspect of the social-semiotic organisation of the scientific RA. The progressive decrease and subsequent increase throughout RA discourse in, on the one hand, Real World Themes and, on the other, the three non-Real World domains (Participant, Discourse and Hypothesised & Objectivised Viewpoint), encoded a major element in the distinctiveness of this genre. This underlying conceptual macrostructure provides insights into how scientific discourse communities in the academic professions constitute themselves and how they carry out their work through texts.

Earlier discussion of the 'higher' order plane of register and its realisations (see section 2.3 - Figure 3) indicated that Field, Tenor and Mode choices tend to be reflected in the ideational, interpersonal and textual metafunctions and the major systems and structures of transitivity, mood and theme, respectively. Analysis of unmarked Subject Themes showed that 'appropriate' thematic

patterning helped writers to stage the progressively changing balance throughout scientific RAs of *topic-based* versus *interactional* discourse. Since the motivation for this dynamic balance can be seen as a reflection of the register variable of Tenor and hence the interpersonal metafunction, it confirms Berry's (1989b) proposal for a more flexible correspondence between register variables and Halliday's metafunctions than previously suggested in systemic literature.

In summary of the theoretical component (Part I) of this thesis, and with the intended aim of integrating the description of generic moves and the component of Theme, four separate RA section Theme-Move maps illustrated an approximate correlation between these top-down and bottom-up levels. These maps clearly indicated the contribution of thematic structure to the overall characterisation of the scientific RA genre. Further Theme analysis of a similar corpus of published NNS scientific research articles (N=36) highlighted significant correlations with the base NS corpus and thus suggested 'universally' recognised Theme-Move patterns within this standard product of the scientific discourse community.

Findings which address GENERAL RESEARCH HYPOTHESES #4, #5 and #6 above relate to the applied component of this work as reflected in the title "product and process in scientific research articles written by NNS novice researchers".

GENERAL RESEARCH HYPOTHESIS #4 suggested that 'appropriate' thematic control throughout the stages of scientific RA discourse was a significant measure of 'success', as defined here by acceptance for publication in an English-language scientific journal. In order to test this hypothesis, a series of surveys was conducted with the various participants involved in the process of

international research communication, namely, 'inexperienced' NNS novices, their more 'experienced' NNS research supervisors and the 'expert' readers of RAs, NS (and NNS) scientific journal editors. The results of the surveys reported in Chapter Four provided insights about the perceived degrees of *influence*, *difficulty* and *concern*: it appeared that these factors were clearly associated with 'appropriate' thematic control and related to the judgement of the merits of NNS researchers' RAs submitted for publication in an English-language scientific journal, and thereby to its 'success'.

GENERAL RESEARCH HYPOTHESIS #5 addressed the RA writing procedures and strategies of a group of NNS novices and suggested that an examination of progressive RA drafts would illustrate the development of thematic control in response to feedback and criticism. The analysis and description undertaken in Chapter Five of a group of NNS novices' FIRST and FINAL RA drafts indicated the stages of thematic development that RAs went through in order to be accepted for publication in an English-language scientific journal. The manipulation of thematic and information structure was clearly seen as a critical part of the redrafting process, with such modifications being an integral part of the development of an RA under the interactive influence of external sources.

In particular, FIRST to FINAL draft textual analysis of a number of recognised features of immature language usage indicated how the redrafting process reflected the transformation of a relatively immature, unpublishable RA draft into a relatively more mature version, acceptable for publication. Both qualitative and quantitative data elicited in these analyses provided important insights which appeared to enhance learners' attention to aspects of the RA writing process; thereby, the important role of raising awareness about the influence on the RA writing

task of the social environment in which the task is performed was confirmed.

The final GENERAL RESEARCH HYPOTHESIS #6 aimed to address the pedagogic application of insights gained from work so far. A major research question of interest to EAP practitioners is whether it is possible for RA writing activities and procedures in the EAP classroom to focus effectively on the potential of thematic manipulation to help create 'successful' texts. To this end, the development of a research/teaching tool - Propositional Clusters (PCs) - as a heuristic for raising NNS novice RA writers' awareness about 'appropriate' thematic control was described. The use of PCs principally aimed to integrate NNS RA writers' attention to elements of previously described bottom-up linguistic knowledge (Theme) in relation to the top-down rhetorical constraints (moves) which must guide RA drafting and redrafting processes. Data analysis and both internal and external feedback indicated that the use of individually student-centred PCs exercises appeared to help NNS novices gain greater control over the scientific RA genre. Although PCs were not directly compared with other methodologies, they can be seen as legitimate and potentially valuable EAP classroom writing activities which provide an important focal point for the development of aspects of NNS novice researchers' academic communicative competence.

7.2 LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Like the inherently cyclical nature and process of research undertakings, this study raises as many questions as it attempted to investigate. The major areas which merit further attention are concerned with the following:

Having set out a relatively transparent method of Theme analysis in the present work, based on optional marked Context Frames and obligatory unmarked Subjects, parallel comparative studies across a variety of different genres would be valuable additions to current knowledge in the field of educational linguistics on the relationship between Theme and genre. It is likely that some functional categories noted in the present work on Theme in scientific RAs would not be registered in the thematic analysis of other genres; likewise, additional discourse-functional categories would be required to cover the different rhetorical purposes inherent in other genres with their distinct generic elements of structure.

Concerning methodological limitations in the present work, marked Context Frames require further evaluation for potential differences with other scientific research process genres. Scientific discourse is characterised by significant variations according to the discipline - there may well be interesting differences in data that can be expected in the usage and distribution of Context Frames, for example: between Biological Sciences, Physics and Chemistry RAs (taken here as one corpus) and other 'softer' sciences; between RAs and Letters, Notes, Communications; between formal conference presentations and informal poster sessions; between biweekly versus quarterly publication of journals. Investigation of these questions will likely lead to the refinement of the present discourse-functional categorisation of thematic selections and suggested Theme-Move patterns.

In addition, analysis of unmarked Subjects here showed just one aspect of the social-semiotic organisation of the scientific RA with its distinctive pattern of relatively more and less interactional Themes throughout RA discourse - this pattern encoded a major element of the distinctiveness of this genre. In parallel with the present method of analysis, it would be interesting to

investigate by similar means the social-constructionist nature of discourse in a variety of other complete exemplars of genres.

In the applied part of this work with a group of NNS novice researchers, one obvious methodological weakness was the difficulty encountered in the elicitation of L2 verbal reports from informants. In considering future research based on the writing activities of such learners, this difficulty raises the question as to whether informants need to be trained in how to provide the desired form of data – a point mentioned earlier when discussing the research methodology of verbal reports. This may be a useful suggestion but the extra time required to do so may add to the burden of an already tight schedule for conducting EAP classroom- and laboratory-based research with NNS novices, more so with research supervisors and scientific journal editors. The delicate balance of asking busy academics for their time and assistance, whether through mail-shot questionnaires or when face-to-face, but not to overimpose on them, implies a limitation to the degree to which EAP researchers may reasonably involve participants in the process of investigating others' research writing activities.

Finally, the data in the study originated from three distinct groups of science researchers and RA writers: (i) 'experienced' NSs (ii) 'experienced' NNSs and (iii) 'inexperienced' novice NNSs. The missing fourth party here is, of course, a group of novice NS researchers. In continuing the ethnographic traditions of the sociology of science, but equipped with the analytic tools of the systemic-functional framework described here, it would be worthwhile to investigate the research writing processes and products of such a group in a manner parallel to the present study.

REFERENCES

Adams Smith,D. (1984). Medical discourse: aspects of author's comment. *The ESP Journal*, 3, 25-36.

Allwright,R., M.-P.Woodley & J.Allwright (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics*, 9/3, 236-256.

Bacha,N. & Hanania,E. (1980). Difficulty in learning and effectiveness of teaching transitional words: a study on Arabic-speaking university students. *TESOL Quarterly*, 14/2, 251-254.

Baldauf,R. & Jernudd,J. (1983a). Language of publications as a variable in scientific communication. *Australian Review of Applied Linguistics*, 6, 97-108.

Baldauf,R. & Jernudd,J, (1983b). Language use patterns in the fisheries periodical literature. *Scientometrics*, 5, 245-255.

Bazerman,C. (1984). Modern evolution of the experimental report in physics: spectroscopic articles in *Physical Review* 1893-1980. *Social Studies of Science*, 14/2, 163-196.

Bazerman,C. (1985). Physicists reading physics: schema-laden purposes and purpose-laden schema. *Written Communication*, 2, 3-24.

Bazerman,C. (1988). *Shaping Written Knowledge*. Madison,WI: University of Wisconsin Press.

Bazerman,C. & Paradis,J. (1991). Introduction. *Textual Dynamics and the Professions*, 3-10. Madison,WI: University of Wisconsin Press.

Beach,R. (1976). Self-evaluation strategies of extensive revisers and non-revisers. *College Composition and Communication*, 27, 160-164.

Belson,W. (1981). *The Design and Understanding of Survey Questions*. Aldershot: Gower.

Berkenkotter,C., Huckin,T. & J.Ackerman (1991). Social context and socially constructed texts. In C.Bazerman &

J.Paradis (eds.) *Textual Dynamics and the Professions*, 191-215. Madison,WI: University of Wisconsin Press.

Berry,M. (1975/1989). *An Introduction to Systemic Linguistics. I Structures and Systems*. University of Nottingham, Department of English Studies.

Berry,M. (1982). Review of Halliday, M.A.K., Language as a social semiotic. *Nottingham Linguistic Circular*, 11, 64-92.

Berry,M. (1987a). The functions of Place-Names. In T. Turville-Petre and M.Gelling (eds.) *Leeds Studies in English New Series XVIII*, 71-88. University of Leeds, School of English.

Berry,M. (1987b). Is teacher an unanalysed concept? In M.A.K. Halliday & R.Fawcett (eds.) *New Developments in Systemic Linguistics, Volume 1: Theory and Description*, 41-63. London: Pinter.

Berry,M. (1989a). Thematic options and success in writing. In C.Butler, R.Cardwell & J.Channel (eds.) *Language and Literature - Theory and Practice*, 62-80. University of Nottingham, Department of Linguistics.

Berry,M. (1989b). They're all out of step except our Johnny: a discussion of motivation (or lack of it) in systemic linguistics. *Occasional Papers in Systemic Linguistics*, 3, 5-67.

Berry,M. (1990). Thematic analysis and stylistic preferences. Paper presented at the XVIIth International Systemic Congress, University of Stirling, July 1990.

Berry,M. (to appear). Towards a study of generic choices and discourse demands. In *Discourse in Society: Systemic-functional Perspectives*. New Jersey: Ablex.

Bizzell,P. (1982). Cognition, convention and certainty: what we need to know about writing. *PRE/TEXT*, 3, 213-241.

Bloor,M. & Bloor,T. (1992). Given and new information in the thematic organization of text: an application to the teaching of writing. *Occasional Papers in Systemic Linguistics*, 6, 33-43.

Brandt,D. (1986). Text and context: how writers come to mean. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 93-107. Norwood,NJ: Ablex.

Bridgeman,B. & Carlson,S. (1984). Survey of academic writing tasks. *Written Communication*, 1, 247-280.

Brown,G. & Yule,G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Brown,R. & Herndl,C. (1986). An ethnographic study of corporate writing: job status as reflected in written text. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 11-28. Norwood,NJ: Ablex.

Butler, C. (1985). *Systemic Linguistics: Theory and Applications*. London: Batsford.

Butler, C. (1989). Systemic models: unity, diversity and change. *Word*, 40 (1/2), 1-35.

Candlin,C., Bruton,J. & J.Leather. (1976). Doctors in casualty: specialist course design from a database. *IRAL*, 14, 245-72.

Candlin,C. (1987). Preface. In W.Rutherford *Second Language Grammar: Learning and Teaching*. New York: Longman.

Canesco,G. & Byrd,P. (1989). Writing required in graduate courses in business administration. *TESOL Quarterly*, 23, 305-316.

Casanave,C. & Hubbard,P. (1992). The writing assignments and writing problems of doctoral students: faculty perceptions, pedagogical issues and needed research. *English for Specific Purposes*, 11/2, 33-49.

Chafe, W. (1976). Givenness, contrastiveness, definiteness and topics. In C.Li (ed.) *Subject and Topic*, 27-56. London: Academic Press.

Christie,F. (1986). Writing in schools: generic structures as ways of meaning. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 221-240. Norwood,NJ: Ablex.

Christie,F. (1987). Genres as choice. In I.Reid (ed.) *The Place of Genres in Learning: Current Debates*, 22-34. Victoria, Australia: Deakin University Press.

Clark,H. & Clark,E. (1977). *Psychology and Language*. New York: Harcourt Brace Jovanovich.

Clark,H. & Haviland,S. (1977). Comprehension and the given-new contract. In R.Freedle (ed.) *Discourse Production and Comprehension*. Norwood,NJ: Ablex.

Clark,R. & Ivanič,R. (1991). Consciousness-raising about the writing process. In C.James & P.Garrett (eds.) *Language Awareness in the Classroom*, 168-185. London: Longman.

Cohen,A. (1987a). Using verbal reports in research on language learning. In C.Faerch & G.Kasper (eds.) *Introspection in Second Language Research*, 82-95. Clevedon: Multilingual Matters.

Cohen,A. (1987b). Studying learner strategies: how we get the information. In A.Wenden & J.Rubin (eds.) *Learner Strategies in Language Learning*, 31-40. New York: Prentice Hall.

Connor,U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics*, 17/3, 301-316.

Connor,U. (1987). Research frontiers in writing analysis. *TESOL Quarterly*, 21/4, 677-696.

Courtenay,G. (1978). Questionnaire design. In G.Hoinville & R.Jowell (eds.) *Survey Research Practice*. London: Heinemann.

Couture,B. (1986). Effective ideation in written text: a functional approach to clarity and exigence. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 69-92. Norwood,NJ: Ablex.

Crookes,G. (1986). Towards a validated analysis of scientific text structure. *Applied Linguistics*, 7, 57-70.

Cumming,A. (1989). Writing expertise and second-language proficiency. *Language Learning*, 39/1, 81-141.

Daneš,F. (1970). An instance of Prague School methodology: functional analysis of utterance and text. In P.Garvin (ed.) *Method and Theory in Linguistics*. The Hague: Mouton.

Daneš,F. (1974). Functional Sentence Perspective and the organisation of the text. In F.Daneš (ed.) *Papers of Functional Sentence Perspective*, 106-128. The Hague: Mouton.

Davidse,K. (1987). M.A.K. Halliday's functional grammar and the Prague School. In R.Dirven & V.Fried (eds.) *Functionalism in Linguistics*, 39-79. Amsterdam: Benjamins.

Davies,F. (1987). *Academic Writing Course*. Liverpool University, UK: English Language Unit, mimeo.

Davies,F. (1988a). Reading between the lines: thematic choice as a device for presenting writer viewpoint in academic discourse. *Especialist 9*, 1/2, 173-200.

Davies,F. (1988b). Designing a writing syllabus in English for academic purposes: process and product. In P.Robinson (ed.) *Academic Writing: Process and Product*, 130-142. [ELT Documents 129]. Modern English Publications/British Council.

Davies,F. (1989). Developing competence in academic discourse: the role of language awareness. ELU Working Paper No.1, University of Liverpool, English Language Unit.

Davies,F. (1991). Writing across text-types and genres; the potential of marked Theme as a device for structuring text. University of Liverpool, English Language Unit.

Day,R. (1979). *How to Write and Publish a Scientific Paper*. Philadelphia: ISI.

Dhaif,H. (1985). Comments on Swales's paper. In R.Quirk & H.Widdowson (eds.) *English in the World: Teaching and Learning the Language and the Literatures*, 224-225. Cambridge: Cambridge University Press/British Council.

Downing,A. (1991). An alternative approach to Theme: a systemic-functional perspective. *Word*, 42, 119-144.

Eiler,M. (1986). Thematic distribution as a heuristic for written discourse function. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 49-68. Norwood,NJ: Ablex.

Ellis,J. (1987). The logical and textual functions. In M.A.K. Halliday & R.Fawcett (eds.) *New Developments in Systemic Linguistics: I Theory and Description*, 107-129. London: Pinter.

Ellis,R. (1989). Review. *Applied Linguistics*, 10/3, 360-364.

Enkvist,N. (1973). Theme dynamics and style: an experiment. *Studia Anglica Posnaniensia*, 5, 127-35.

Ericsson,K. & Simon,H. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.

Ericsson,K. & Simon,H. (1984). *Protocol Analysis*. Cambridge,MA: MIT Press/Bradford.

Ericsson,K. & Simon (1985). Protocol analysis. In T.van Dijk (ed.) *Handbook of Discourse Analysis, Volume 2: Dimensions of Discourse*, 259-268. London: Academic Press.

Ericsson,K. & Simon,H. (1987). Verbal reports on thinking. In C.Faerch & G.Kasper (eds.) *Introspection in Second Language Research*, 24-53. Clevedon: Multilingual Matters.

Faerch,C. & Kasper,G. (1987). From product to process – introspective methods in second language research. In C.Faerch & G.Kasper (eds.) *Introspection in Second Language Research*, 5-23. Clevedon: Multilingual Matters.

Faigley,L. & Witte,S. (1981). Analyzing revision. *College Composition and Communication*, 32, 400-414.

Feyerabend,P. (1975). *Against Method*. London: Verso.

Fine,J. (1992). Functions of probabilities on linguistic systems. *Occasional Papers in Systemic Linguistics*, 6, 9-18.

Firbas,J. (1974). Functional Sentence Perspective and the organisation of the text. In F.Daneš (ed.) *Papers in Functional Sentence Perspective*. Prague: Academia.

Firth,J. (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.

Flower,L. (1979). Writer-based prose: a cognitive basis for problems in writing. *College English*, 41, 19-38.

Flower,L. (1989). Rhetorical problem solving: cognition and professional writing. In M.Kogen (ed.) *Writing in Business Professions*. Urbana,IL: NCTE and the Association for Business Communication.

Flower,L. & Hayes,J. (1977). Problem-solving strategies and the writing process. *College English*, 39, 449-461.

Flower,L. & Hayes,J. (1980). The cognition of discovery: defining a rhetorical problem. *College Composition and Communication*, 31, 21-32.

Flower,L. & Hayes,J. (1981a). Plans that guide the composing process. In C.Frederiksen & J.Dominic (eds.) *Writing: Process, Development and Communication*, 39-58. Hillsdale,NJ: Lawrence Erlbaum.

Flower,L. & Hayes,J. (1981b). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.

Flower,L. & Hayes,J. (1983). The cognition of discovery: defining a rhetorical problem. *College Composition and Communication*, 31, 21-32.

Francis,G. (1989a). Thematic selection and distribution in written discourse. *Word*, 40, 201-222.

Francis,G. (1989b). Aspects of nominal-group lexical cohesion. *Interface*, 4/1, 27-53.

Francis,G. (1990). Theme in the daily press. *Occasional Papers in Systemic Linguistics*, 4, 51-87.

Fried,V. (1972). *The Prague School of Linguistics and Language Teaching*. Oxford: Oxford University Press.

Fries,P. (1983). On the status of Theme: arguments from discourse. In J.Petöfi & E.Sözer (eds.) *Micro and Macro Connexity of Texts*. Papers in Textlinguistics, 45, 116-152. Hamburg: Helmut Buske.

Fries,P. (1992). Information flow in written advertising. Unpublished manuscript. Central Michigan University.

Fries,P. (to appear). Patterns of information in initial
    position in English. In P.Fries & M.Gregory (eds.)
    Discourse in Society: Functional Perspectives.
    Norwood,NJ: Ablex.

Fries,P. & Francis,G. (1992). Exploring Theme: problems
    for research. *Occasional Papers in Systemic Linguistics,*
    6, 45-59.

Geertz,C. (1973). *The Interpretation of Cultures.* New
    York: Basic Books.

Geertz,C. (1983). *Local Knowledge: Further Essays in
    Interpretive Anthropology.* New York: Basic Books.

Gibson,T. (1992). Towards a discourse theory of abstracts
    and abstracting. University of Nottingham. Unpublished
    Ph.D. dissertation.

Gilbert,G. & Mulkay,M. (1984). *Opening Pandora's Box: a
    Sociological Analysis of Scientists' Discourse.*
    Cambridge: Cambridge University Press.

Givón,T. (1982). Topic continuity in discourse: an
    introduction. In T.Givón (ed.) *Topic Continuity in
    Discourse: a Quantitative Cross-language Study,* 1-41.
    Amsterdam: Benjamins.

Glatt,B. (1982). Defining thematic progressions and their
    relationship to reader comprehension. In M.Nystrand
    (ed.) *What Writers Know: The Language, Process and
    Structure of Written Discourse,* 87-103. New York:
    Academic Press.

Gosden,H. (1991). Teaching Research English to ESL
    researchers: questions of cost and benefit. *Journal of
    Asian Pacific Communication,* 2/1, 139-153.

Gosden,H. (1992a). Research writing and NNSs: from the
    editors. *Journal of Second Language Writing,* 1/2,
    123-139.

Gosden,H. (1992b). Discourse functions of marked Theme in
    scientific research articles. *English for Specific
    Purposes,* 11/3, 207-224.

Gosden,H. (1993). Discourse functions of Subject in
    scientific research articles. *Applied Linguistics.* 14/1,
    56-75.

Green,B. (1987). Gender, genre and pedagogy. In I.Reid (ed.) *The Place of Genres in Learning: Current Debates*, 83-90. Victoria, Australia: Deakin University Press.

Gregory,M. (1967). Aspects of varieties differentiation. *Journal of Linguistics*, 3, 177-198.

Gregory,M. (1987). Meta-functions: aspects of their development, status and use in systemic linguistics. In M.A.K.Halliday & R.Fawcett (eds.) *New Developments in Systemic Linguistics: I Theory and Description*, 94-106. London: Pinter.

Gregory,M. & Carroll,S. (1978). *Language and Situation: Language Varieties and their Social Contexts*. London: Routledge & Kegan Paul.

Grotjahn,R. (1987). On the methodological basis of introspective methods. In C.Faerch & G.Kasper (eds.) *Introspection in Second Language Research*, 54-81. Clevedon: Multilingual Matters.

Halliday,M.A.K. (1967). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, 3, 177-274.

Halliday,M.A.K. (1970a). Language structure and language function. In J.Lyons (ed.) *New Horizons in Linguistics*, 140-165. London: Penguin.

Halliday,M.A.K. (1970b). Functional diversity in language, as seen from a consideration of modality and mood in English. *Foundations of Language*, 6, 322-361. In G.Kress (ed.) (1976) *Halliday: System and Function in Language*, 189-213. Oxford: Oxford University Press.

Halliday,M.A.K. (1973). *Explorations in the Functions of Language.* London: Edward Arnold.

Halliday,M.A.K. (1977). Text as semantic choice in social contexts. In T.van Dijk & J.Petöfi (eds.) *Grammar and Descriptions*. Berlin: de Gruyter.

Halliday,M.A.K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning.* London: Edward Arnold.

Halliday,M.A.K. (1979). Modes of meaning and modes of expression: types of grammatical structure and their determination by different semantic functions. In

D.Allerton, E.Carney & D.Holdcroft (eds.) *Function and Context in Linguistic Analysis: Essays offered to William Haas*, 57-79. Cambridge: Cambridge University Press.

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

Halliday,M.A.K. (1985b). Dimensions of Discourse Analysis: Grammar. In T.van Dijk (ed.) *Handbook of Discourse Analysis, Volume 2: Dimensions of Discourse*, 29-56. London: Academic Press.

Halliday,M.A.K. (1988). On the language of physical science. In M.Ghadessy (ed.) *Registers of Written English: Situational Factors and Linguistic Features*, 162-178. London: Pinter.

Halliday,M.A.K. (1989a). *Spoken and Written Language*. Oxford: Oxford University Press.

Halliday,M.A.K. (1989b). Part A. In M.A.K.Halliday & R.Hasan *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Oxford: Oxford University Press.

Halliday,M.A.K., McIntosh,A. & P.Strevens (1964). *The Linguistic Sciences and Language Teaching*. London: Longman.

Halliday,M.A.K. & Hasan,R. (1976). *Cohesion in English*. London: Longman.

Halliday,M.A.K. & Fawcett,R. (1987). Introduction. In M.A.K.Halliday & R.Fawcett (eds.) *New Developments in Systemic Linguistics: I Theory and Description*, 1-13. London: Pinter.

Harre,R. (1981). The positivist-empiricist approach and its alternative. In P.Reason & J.Rowan (eds.) *Human Inquiry: a Sourcebook of New Paradigm Research*. New York: Wiley.

Harris,S. (1987). Court discourse as genre: some problems and issues. *Occasional Papers in Systemic Linguistics*, 2, 35-74.

Hasan,R. (1978). Text in the systemic-functional model. In W.Dressler (ed.) *Current Trends in Text Linguistics*, 228-246. Berlin: de Gruyter.

Hasan,R. (1984). Coherence and cohesive harmony. In J.Flood (ed.) *Understanding Reading Comprehension: Cognition, Language and the Structure of Prose*. Newark,DA: International Reading Association.

Hasan,R. (1987). The grammarian's dream: lexis as most delicate grammar. In M.A.K.Halliday & R.Fawcett (eds.) *New Developments in Systemic Linguistics: I Theory and Description*, 184-211. London: Pinter.

Hasan,R. (1989). Part B. In M.A.K.Halliday & R.Hasan *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Oxford: Oxford University Press.

Haviland,S. & Clark,H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13, 512-521.

Hedge,N. & Gosden,H. (1991). Language Awareness and EAP courses. In C.James & P.Garrett (eds.) *Language Awareness in the Classroom*, 186-197. London: Longman.

Heslot,J. (1982). Tense and other indexical markers in the typology of scientific texts in English. In J.Hoedt *et al.* (eds.) *Pragmatics and LSP*, 83-103. Copenhagen: School of Economics.

Hesse,M. (1980). *Revolutions and Reconstructions in the Philosophy of Science*. Bloomington: Indiana University Press.

Hill,S., Soppelsa,B. & G.West (1982). Teaching ESL students to read and write experimental-research papers. *TESOL Quarterly*, 16, 333-347.

Hinds,J. (1987). Reader versus writer responsibility: a new typology. In U.Connor & R.Kaplan (eds.) *Writing across Languages: Analysis of L₂ Text*, 141-152. Reading,MA: Addison-Wesley.

Hinds,J. (1990). Inductive, deductive, quasi-inductive: expository writing in Japanese, Korean, Chinese and

Thai. In U.Connor & A.Johns (eds.) *Coherence in Writing*, 89-109. Alexandria,VA: TESOL Inc.

Hjelmslev,L. (1961). *Prolegomena to a Theory of Language*. Madison,WI: University of Wisconsin Press.

Hoey,M. (1983). *On the Surface of Discourse*. London: George Allen & Unwin.

Holmes,J. & Ramos,R. (1991). Talking about learning: establishing a framework for discussing and changing learning processes. In C.James & P.Garrett (eds.) *Language Awareness in the Classroom*, 198-212. London: Longman.

Horowitz,D. (1986a). What professors actually require: academic tasks for the ESL classroom. *TESOL Quarterly*, 20/3, 445-492.

Horowitz,D. (1986b). Process not product: less than meets the eye. *TESOL Quarterly*, 20, 141-144.

Huckin,T. & Olsen,L. (1983). *English for Science and Technology: a Handbook for Non-native Speakers*. New York: McGraw-Hill.

Huckin,T., & Olsen,L. (1984). On the use of informants in LSP discourse analysis. In A.Pugh & J.Uljin (Eds.) *Reading for Professional Purposes: Studies and Practices in Native and Foreign Languages* (pp.120-129). London: Heinemann.

Huddleston,R. (1988). Review Article. *Journal of Linguistics*, Volume 24, 137-174.

Huddleston,R. (1991). Further remarks on Halliday's Functional Grammar - a reply to Matthiessen & Martin. *Occasional Papers in Systemic Linguistics*, 5, 75-129.

Huddleston,R. (1992). On Halliday's functional grammar: a reply to Martin and to Martin & Mathiessen. *Occasional Papers in Systemic Linguistics*, 6, 197-211.

Hudson,R. (1986). Review Article: Systemic Grammar. *Linguistics*, 24, 791-815.

Hunt,K. (1965). Grammatical structures written at three grade levels. Research Report No.3. Urbana,Ill: NCTE.

Hymes,D. (1974). *Foundations in Sociolinguistics: an Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.

James,C. & Garrett,P. (1991). The scope of Language Awareness. In C.James & P.Garrett (eds.) *Language Awareness in the Classroom*, 3-20. London: Longman.

Jary,D. & Jary,J. (1991). *Sociology*. New York: HarperCollins.

Jenkins,S., Jordan,M. & P.Weiland (1993). The role of writing in graduate engineering education: a survey of faculty beliefs and practices. *English for Specific Purposes*, 12/1, 51-67.

Johns,A. (1980). Preventing global discourse errors: problems and approaches to ESOL writing. *CATESOL Occasional Papers*, 6, 65-70.

Johns,A. (1981). Necessary English: a faculty survey. *TESOL Quarterly*, 15, 51-57.

Johns,A. (1984). Textual cohesion and the Chinese speaker of English. *Language Learning and Communication*, 3/1, 69-74.

Johns,A. (1986). Coherence and academic writing: some definitions and suggestions for teaching. *TESOL Quarterly*, 20/2, 247-265.

Johns,A. (1990). L1 composition theories: implications for developing theories of L2 composition. In B.Kroll (ed.) *Second Language Writing*, 24-36. Cambridge: Cambridge University Press.

Kaplan,R. (1983). Contrastive rhetorics: some implications for the writing process. In A.Freedman, I.Pringle & J.Yalden (eds.) *Learning to Write: First Language/Second Language*, 139-161. New York: Longman.

Kintsch,W. (1974). *The Representation of Meaning in Memory*. Hillsdale,NJ: Erlbaum.

Kintsch,W. (1977). On comprehending stories. In M.Just & P.Carpenter (eds.) *Cognitive Processes in Comprehension*, 33-62. Hillsdale,NJ: Erlbaum.

Kintsch,W. (1985). Text processing: a psychological model. In T.van Dijk (ed.) *Handbook of Discourse Analysis,*

Volume 2: Dimensions of Discourse, 231-243. London: Academic Press.

Kintsch,W. & Kozminsky,E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology*, 69, 491-499.

Kintsch,W. & van Dijk,T. (1978). Towards a model of text comprehension and production. *Psychological Review*, 85/5, 363-394.

Kintsch,W., E.Kozminsky, W.Streby, G.McCoon, & J.Keenan (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning Behavior*, 14, 196-214.

Kintsch,W., T.Mandel, & E.Kozminsky (1977). Summarizing scrambled stories. *Memory and Cognition*, 5, 547-522.

Kirk,J. & Miller,M. (1986). *Reliability and Validity in Qualitative Research*. Beverly Hills,CA: Sage.

Knorr-Cetina,K. (1981). *The Manufacture of Knowledge*. Oxford: Pergamon Press.

Krapels,A. (1990). An overview of second language writing process research. In B.Kroll (ed.) *Second Language Writing*, 37-56. Cambridge: Cambridge University Press.

Kress,G. (1982). *Learning to Write*. London: Routledge & Kegan Paul.

Kress,G. (1989). *Linguistic Processes in Sociocultural Practice*. Oxford: Oxford University Press.

Krings,H. (1987). The use of introspective data in translation. In C.Faerch & G.Kasper (eds.) *Introspection in Second Language Research*, 159-176. Clevedon: Multilingual Matters.

Kroll,B. (1979). A survey of the writing needs of foreign and American college freshmen. *ELT Journal*, 33, 219-226.

Kuhn,T. (1970). *The Structure of Scientific Revolutions*. 2nd. edition. Chicago: University of Chicago Press.

Latour,B. & Woolgar,S. (1979). *Laboratory Life: the Social Construction of Scientific Facts*. Beverly Hills,CA: Sage.

Lauttamatti,L. (1987). Observations on the development of the topic of simplified discourse In U.Connor & R.Kaplan

(eds.) *Writing across Languages: Analysis of L2 Text*, 87-114. Reading, MA: Addison-Wesley.

Lincoln,Y. & Guba,E. (1985). *Naturalistic Inquiry*. Beverly Hills,CA: Sage.

Lowe,I. (1987). Sentence initial elements in English and their discourse function. *Occasional Papers in Systemic Linguistics*, 2, 5-33.

Malinowski,B. (1923). The problem of meaning in primitive languages. Supplement to C.Ogden & I.Richards *The Meaning of Meaning*. London: Routledge & Kegan Paul.

Martin,J. (1985). *Factual Writing: Exploring and Challenging Social Reality*. Oxford: Oxford University Press.

Martin,J. (1986). Intervening in the process of writing development. In C.Painter & J.Martin (eds.) *Writing to Mean: Teaching Genres across the Curriculum*. Occasional Papers, 9, 11-43. Applied Linguistics Association of Australia.

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

Martin,J. (1992b). Theme, method of development and existentiality: the price of reply. *Occasional Papers in Systemic Linguistics*, 6, 147-183.

Martin,J. (1993). Genre and literacy: modelling context in educational linguistics. *Annual Review of Applied Linguistics (1993)*, 13, 141-172.

Martin,J & Rothery,J. (1986). What a functional approach to the writing task can show teachers about 'good writing'. In B.Couture (ed.) *Functional Approaches to Writing: Research Perspectives*, 241-265. Norwood,NJ: Ablex.

Martin,J., Christie,F & J.Rothery (1987). Social processes in education: a reply to Sawyer and Watson (and others). In I.Reid (ed.) *The Place of Genre in Learning: Current Debates*, 58-81. Victoria, Australia: Deakin University Press.

Martin,J. & Matthiessen,C. (1992). A brief note on Huddleston's reply to Matthiessen & Martin's response to

Huddleston's review of Halliday's Introduction to Functional Grammar. *Occasional Papers in Systemic Linguistics*, 6, 185-195.

Mathesius,V. (1961). In J.Vachek (ed.) *A Functional Analysis of Present Day English on a General Linguistic Basis*. Prague (1961). The Hague: Mouton (1975).

Matthiessen,C. (1992). Interpreting the textual metafunction. In M.Davies & L.Ravelli (eds.) *Advances in Systemic Linguistics: Recent Theory and Practice*, 37-81. London: Pinter.

Matthiessen,C. & Martin,J. (1991). A response to Huddleston's review of Halliday's Introduction to Functional Grammar. *Occasional Papers in Systemic Linguistics*, 5, 5-74.

McGregor,W. (1990). On the notion of Theme in semiotic grammar. Paper presented at Fourth International Conference on Functional Grammar, Copenhagen 1990.

Mehan,H. (1980). The competent student. *Anthropology and Education Quarterly*, 11/3, 131-151.

Miller,C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151-167.

Morgan,J. & Sellner,M. (1980). Discourse and linguistic theory. In R.Spiro, B.Bruce & W.Brewers (eds.) *Theoretical Issues in Reading Comprehension*, 165-200. Hillsdale,NJ: Lawrence Erlbaum Associates.

Morrow P. (1989) Conjunct Use. *English for Specific Purposes*, 8/3, 239-254.

Myers,G. (1985). Texts as knowledge claims: the social construction of two biology articles. *Social Studies of Science*, 15, 593-630.

Myers,G. (1988). The social construction of science and the teaching of English: an example of research. In P.Robinson (ed.) *Academic Writing: Process and Product*, 133-150. [ELT Documents 129]. Modern English Publications/British Council.

Myers,G. (1990). The social construction of popular science: the narrative of science and the narrative of

nature. In *Writing Biology*, 141-192. Madison,WI: University of Wisconsin Press.

Naerssen,M.van & Kaplan,R. (1987). Language and Science. In R.Kaplan (ed.) *Annual Review of Applied Linguistics (1986)*, 86-104.

Nelson,G. & Murphy,J. (1992). An L2 writing group: task and social dimensions. *Journal of Second Language Writing*, 1/3, 171-193.

Nesbit,J. & Entwistle,N. (1970). *Educational Research Methods*. London: University of London Press.

Nesbitt,C. & Plum,G. (1988). Probabilities in a systemic-functional grammar: the clause complex in English. In R.Fawcett & D.Young (eds.) *New Developments in Systemic Linguistics: II Theory and Application*, 6-38. London: Pinter.

Nwogu,K. (1990). *Discourse Variation in Medical Texts: Schema, Theme, and Cohesion in Professional and Journalistic Accounts*. Monographs in Systemic Linguistics, Volume 2. University of Nottingham.

O'Brien O'Keefe,K. & Journet,A. (1985). A hypothetico-deductive model for teaching the research paper. *Journal of Technical Writing and Communication*, 15/4, 339-352.

Ostler,S. (1980). A survey of academic needs for advanced ESL. *TESOL Quarterly*, 14, 489-502.

Patton,M. (1978). *Utilization-focused Evaluation*. Beverly Hills,CA: Sage.

Pettinari,C. (1985). Comparison of the production of surgical reports by native and non-native speaking surgeons. In J.Benson & W.Greaves (eds.) *Systemic perspectives on Discourse, Volume Two*, 187-203. Norwood,NJ: Ablex.

Plum,G. (1988). Textual and contextual conditioning in spoken English: a genre-based approach. Unpublished Ph.D. thesis. University of Sydney.

Popper,K. (1959). *The Logic of Scientific Discovery*. London: Hutchison.

Popper,K. (1976). *Unended Quest: an Intellectual Autobiography*. Glasgow: Fontana.

Potter,J. & Wetherell,M. (1987). *Discourse and Social Psychology*. London: Sage.

Prabhu,N. (1985). Coping with the unknown in language pedagogy. In R.Quirk & H.Widdowson (eds.) *English in the World*, 164-173. Cambridge: Cambridge University Press/British Council.

Prabhu,N. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Quirk, R., Greenbaum,S., Leech,G. & J.Svartvik (1985). *A Contemporary Grammar of the English Language*. London: Longman.

Reid,I. (1987). A generic frame for debates about genres. In I.Reid (ed.) *The Place of Genres in Learning: Current Debates*, 1-7. Victoria, Australia: Deakin University Press.

Reid,J. (1984a). The radical outliner and the radical brainstormer: a perspective on composing processes. *TESOL Quarterly*, 18, 529-533.

Reid,J. (1984b). Comments on Vivian Zamel's 'The composing processes of advanced ESL students: six case studies.' *TESOL Quarterly*, 18, 149-159.

Richards,R. (1988). Thesis/dissertation writing for EFL students: an ESP course design. *English for Specific Purposes*, 7/3, 171-180.

Rutherford,W. (1987). *Second Language Grammar: Learning and Teaching*. New York: Longman.

Selinker,L. (1979). On the use of informants in discourse analysis and "language for specialized purposes"'. *International Review of Applied Linguistics in Language Teaching*, 17/3, 189-215.

Selzer,J. (1983). The composing processes of an engineer. *College Composition and Communication*, 34, 178-187.

Sharples,M. (1988). The development of a cognitive theory of writing. ESRC Report C002300113. University of Liverpool.

Shaw,P. (1991). Science research students' composing processes. *English for Specific Purposes*, 10/3, 189-206.

Silva,T. (1990). Second language composition instruction: developments, issues and directions in ESL. In B.Kroll (ed.) *Second Language Writing: Research Insights for the Classroom*, 11-23. Cambridge: Cambridge University Press.

Silverman,D. (1985). *Qualitative Methodology and Sociology*. Aldershot: Gower.

Sinclair,J. (1985a). On the integration of linguistic description. In T.van Dijk (ed.) *Handbook of Discourse Analysis, Volume 2. Dimensions of Discourse*, 13-28. London: Academic Press.

Sinclair,J. (1985b). Selected Issues. In R.Quirk & H.Widdowson (eds.) *English in the World*, 248-254. Cambridge: Cambridge University Press/British Council.

Sinclair,J. (1989). The Divorce of Theory and Practice. Paper presented at JALT Conference, Ookayama, Japan. November 1989.

Sinclair,J. & Coulthard,R. (1975). *Towards an Analysis of Discourse: the English used by Teachers and Pupils*. Oxford: Oxford University Press.

Sommers,N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31, 378-388.

Spack,R. (1988). Initiating ESL students into the academic discourse community: how far should we go? *TESOL Quarterly*, 22/1, 29-51.

St.John,M. (1987). Writing processes of Spanish scientists publishing in English. *English for Specific Purposes*, 6/2, 113-120.

Stainton,C. (1992). Language awareness: genre awareness – a focused review of the literature. *Language Awareness*, 1/2, 109-121.

Stainton,C. (to appear). What is this thing called genre? *Nottingham Working Papers*.

Stanley,R. (1984). The recognition of macrostructure: a pilot study. *Reading in a Foreign Language*, 2, 156-168.

Swales,J. (1981). *Aspects of Article Introductions*. Aston ESP Research Reports #1. The University of Aston in Birmingham, Language Studies Unit.

Swales,J. (1983). Developing materials for writing scholarly introductions. In R.Jordan (ed.) *Case Studies in ELT*, 188-200. London: Collins.

Swales,J. (1985a). English as an international language of research. *RELC Journal*, 16/1, 1-7.

Swales J. (1985b). English language papers and authors' first language: preliminary explorations. *Scientometrics*, 8, 91-101.

Swales,J. (1985c). ESP - the Heart of the Matter or the End of the Affair?. In R.Quirk & H.Widdowson (eds.) *English in the World*, 212-223. Cambridge: Cambridge University Press/British Council.

Swales,J. (1986a). Citation analysis and discourse analysis. *Applied Linguistics*, 7, 39-56.

Swales,J. (1986b). A genre-based approach to language across the curriculum. In M.Tickoo (ed.) *Language across the Curriculum*, 10-22. SEAMEO RELC.

Swales,J. (1987a). Language and the professions: current developments and future prospects. *Annual Review of Applied Linguistics*, 7, 121-130.

Swales,J. (1987b). Utilizing the literatures in teaching the research paper. *TESOL Quarterly*, 21/1, 41-68.

Swales,J. (1988) Discourse communities, genres and English as an international language. *World Englishes*, 7/2, 211-220.

Swales,J. (1990a). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Swales,J. (1990b). Global coherence and local management. In U.Connor & A.Johns (eds.) *Coherence in Writing*, 189-207. Alexandria,VA: TESOL Inc.

Swales,J. (1992). Re-thinking genre: another look at discourse community effects. Paper presented at the Re-thinking Genre Colloquium, Ottawa, Canada. April 1992.

Swales,J. & Najjar,H. (1987). The writing of research article introductions. *Written Communication*, 4/2, 175-191.

Tarantino,M. (1991). English for science and technology: a quest for legitimacy. *English for Specific Purposes*, 10, 47-60.

Tarone,E., Dwyer,S., Gillette,S. & V.Icke. (1981). On the use of the passive in two astrophysics journal papers. *The ESP Journal*, 1, 123-140.

Taylor,G. & Chen,T. (1991). Linguistic, cultural and subcultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts. *Applied Linguistics*, 12/3, 319-336.

Thompson,S. (1985). Initial vs. final purpose clauses in written English discourse. *Text*, 5(1/2), 55-84.

Trelease,S. (1982). *How to Write Scientific and Technical Papers*. Cambridge,MA: MIT Press.

Vachek,J. (1972). The linguistic theory of the Prague School. In V.Fried (ed.) *The Prague School of Linguistics and Language Teaching*, 11-28. Oxford: Oxford University Press.

Ventola,E. (1983). Contrasting schematic structures in service encounters. *Applied Linguistics*, 4, 242-258.

Ventola,E. (1988). Text analysis in operation: a multilevel approach. In R.Fawcett & D.Young (eds.) *New Developments in Systemic Linguistics: II Theory and Application*, 52-77. London: Pinter.

Ventola,E. (1989). Problems of modelling and applied issues within the framework of genre. *Word*, 40 (1/2), 129-161.

Weir C. (1988). Academic writing - can we please all the people all the time?, In P. Robinson (ed.) *Academic Writing: Process and Product*, 17-34. [ELT Documents 129]. Modern English Publications British Council.

Weissberg,R. (1984). Given and New: paragraph development models for scientific English. *TESOL Quarterly*, 18, 485-500.

Weissberg,R. & Buker,S. (1990). *Writing up Research*. Englewood Cliffs,NJ: Prentice Hall Regents.

West,G. & Byrd,P. (1982). Technical writing required of graduate engineering students. *Journal of Technical Writing and Communication*, 12, 1-6.

Widdowson,H. (1983). New starts and different kinds of failure. In A.Freedman, I.Pringle & J.Yalden (eds.) *Learning to Write: First Language/Second Language*, 34-47. New York: Longman.

Winter,E. (1977). A clause-relational approach to English texts. *Instructional Science*, 6. Amsterdam: Elsevier.

Witte,S. & Cherry,R. (1986). Writing processes and written products in composition research. In C.Cooper & S.Greenbaum (eds.) *Studying Writing: Linguistic Approaches*, 112-153. Newbury Park: Sage.

Zamel,V. (1983). The composing processes of advanced ESL students: six case histories. *TESOL Quarterly*, 16, 195-209.

Journal Title/Volume/Year/Author(s) - issues/year

Journal of Physics: Condensed Matter - 51/year
1. Vol.2 (1990) pp. 2743-2750; Cranshaw,T.
2. Vol.2   pp. 5191-5198; Barnard,R.
3. Vol.2   pp. 5517-5527; Harris,M. *et al.*
Journal of Materials Science - 12/year
4. Vol.25 (1990) pp. 829-834; Dobb,M. *et al.*
5. Vol.25   pp. 1275-1284; Smith,J. *et al.*
6. Vol.25   pp. 3002-3007; Khan,G. *et al.*
Canadian Journal of Physics - 12/year
7. Vol.67 (1989) pp. 287-293; Cook,J. *et al.*
8. Vol.67   pp. 298-303; Rao,T. *et al.*
9. Vol.67   pp. 365-369; Paquin,L. *et al.*
Journal of Applied Physics - 24/year
10. Vol.67/6 (1990) pp. 2894-2898; Clevenger,L.
11. Vol.67/9   pp. 4093-4098; Chang,K. *et al.*
12. Vol.67/12   pp. 7478-7482; Caton,R. *et al.*
J.of Chem.Soc.: Faraday Transactions - 24/year
13. Vol.86/1 (1990) pp. 65-68;   Ajiboye,S. *et al.*
14. Vol.86/10   pp. 1751-1756; Frost,M. *et al.*
15. Vol.86/22   pp. 3803-3807; Burch,R. *et al.*
The Biochemical Journal - 24/year
16. Vol.265 (1990) pp. 45-50;   West,S. *et al.*
17. Vol.265   pp. 277-282; Winchester,B. *et al.*
18. Vol.268   pp. 387-392; Lappin,D. *et al.*
Journal of Catalysis - 12/year
19. Vol.124 (1990) pp. 73-85;   Bourzutschky,J.
20. Vol.124   pp. 97-108; Umansky,B. *et al.*
21. Vol.124   pp. 127-132; Nag,N. *et al.*
Journal of Chemical Physics - 24/year
22. Vol.92/3 (1990) pp. 1687-1695; Zhu,L. *et al.*
23. Vol.92/3   pp. 1790-1797; Farley,R. *et al.*
24. Vol.92/3   pp. 1796-1803; Marinelli,W.
Cellular Signalling - 6/year
25. Vol.2/3 (1990) pp. 265-276; Ali,S. *et al.*
26. Vol.2/3   pp. 277-284; Harrison,R.
27. Vol.2/3   pp. 305-310; Budd,J. *et al.*
Journal of Cell Science - 12/year
28. Vol.95 (1990) pp. 599-604; Smith,P. *et al.*
29. Vol.95   pp. 617-622; McCaig,C.
30. Vol.95   pp. 639-648; Thomas,J. *et al.*
Journal of Cell Biology - 12/year
31. Vol.110 (1990) pp. 27-34;   Gould,S. *et al.*
32. Vol.110   pp. 97-104; Baker,H. *et al.*
33. Vol.110   pp. 147-153; Bowen,B. *et al.*
Cell Motility and the Cytoskeleton - 12/year
34. Vol.12 (1989) pp. 113-122; Sweet,S. *et al.*
35. Vol.15 (1990) pp. 12-22;   Heintzelmann,M.
36. Vol.15   pp. 34-40;   Campos,R. *et al.*

Main corpus of 36 NS Research Articles
SIEs: *N* and % GS vs. CF by 4 RA sections
[average SIEs = 121.1 - max 202: min 77]

INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	

UK: Physics 1-3 = J.Phys:Condens.Matter

1. 21:14/7	9:9/0	*28:19/8	29:18/11	*87;60/26	*+1 non-GS/CF SIE
2. 13:7/6	7:5/2	27:13/14	30:13/17	77;38/39	
3. 17:14/3	11:7/4	10:9/1	*61:44/16	*99;74/24	*+1 non-GS/CF SIE

UK: Physics 4-6 = J.Material Science

4. 15:5/10	19:15/4	25:12/13	23:11/12	82;43/39	
5. 14:8/6	21:18/3	45:25/20	37:22/15	117;73/44	
6. 31:24/7	12:11/1	18:14/4	18:16/2	79;65/14	

Can: Physics 7-9 = Can.J.Phys

7. 14:8/6	18:16/2	33:27/6	*61:32/27	*126;83/41	*+2 non-GS/CF SIEs
8. 13:7/6	9:9/0	53:37/16	29:21/8	104;74/30	
9. 20:14/6	17:12/5	58:38/20	19:16/3	114;80/34	

US: Physics 10-12 = J.Appl.Phys

10. 18:11/7	30:25/5	30:25/5	39:25/14	117;86/31	
11. 36:31/5	15:13/2	62:40/22	38:19/19	151;103/48	
12. 13:11/2	45:34/11	29:21/8	50:24/26	137;90/47	

225	213	418	434	1290	(1-12) total *n* SIEs
18.7	17.7	34.8	36.2	107.5	(1-12) mean *n* SIEs

154/71	174/39	280/137	261/170	869/417	(1-12) total *n* GS/CF
0	0	1	3	4	(1-12) total *n* *non-GS/CF

68.4%	81.7%	67.0%	60.1%	67.4%	(1-12) mean % GS
31.6%	18.3%	32.8%	39.2%	32.3%	(1-12) mean % CF
0.0%	0.0%	*0.2%	*0.7%	*0.3%	(1-12) mean % *non-GS/CF

APPENDIX B (continued)
Main corpus of 36 NS Research Articles
SIEs: *N* and % GS vs. CF by 4 RA sections

INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	

UK: Chemistry 13-15 = J.Chem.Soc.Faraday Trans.

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
13.	10:6/4	21:21/0	40:28/12	25:20/5	96;75/21	
14.	14:5/9	40:34/6	13:9/4	53:34/19	120;82/38	
15.	19:12/7	16:16/0	22:18/4	*68:29/37	125;75/48	*+2 non-GS/CF SIEs

UK: Chemistry 16-18 = Biochemistry J.

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF
16.	16:11/5	21:19/2	39:21/18	17:9/8	93;60/33
17.	16:11/5	16:14/2	34:29/5	44:32/12	110;86/24
18.	13:13/0	36:29/7	24:14/10	32:28/4	105:84/21

US: Chemistry 19-21 = Journal of Catalysis

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
19.	19:14/5	14:12/2	*92:55/36	*60:39/20	*185;120/63	*+2 non-GS/CF SIEs
20.	32:21/11	*42:28/13	57:38/19	60:30/30	*191;117/73	*+1 non-GS/CF SIE
21.	14:8/6	15:15/0	19:17/2	29:21/8	77;61/16	

US: Chemistry 22-24 = J.Chem.Phys.

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
22.	31:23/8	26:20/6	49:35/14	*88:60/26	*194;138/54	*+2 non-GS/CF SIEs
23.	12:10/2	24:15/9	*31:15/15	71:40/31	*138;80/57	*+1 non-GS/CF SIE
24.	40:32/8	39:36/3	*82:60/21	41:19/22	*202;147/54	*+1 non-GS/CF SIE

INT	EXP	RES	DIS	TOTAL	
236	310	502	588	1636	(13-24) total *n* SIEs
19.7	25.8	41.8	49.0	136.3	(13-24) mean *n* SIEs

INT	EXP	RES	DIS	TOTAL	
166/70	259/50	339/160	361/222	1125/502	(13-24) total *n* GS/CF
0	1	3	5	9	(13-24) total *n* *non-GS/CF

INT	EXP	RES	DIS	TOTAL	
70.3%	83.6%	67.5%	61.4%	68.7%	(13-24) mean % GS
29.7%	16.1%	31.9%	37.8%	30.7%	(13-24) mean % CF
0.0%	*0.3%	*0.6%	*0.8%	*0.6%	(13-24) mean % *non-GS/CF

APPENDIX B (continued)
Main corpus of 36 NS Research Articles
SIEs: *N* and % GS vs. CF by 4 RA sections

INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
UK: Biology 25-27 = Cellular Signalling					
25. 22:15/7	22:17/5	37:23/14	34:16/18	115;71/44	
26.*14:8/5	40:32/8	32:19/13	24:13/11	*110;72/37	*+1 non-GS/CF SIE
27. 7:4/3	42:34/8	13:8/5	16:14/2	78;60/18	
UK: Biology 28-30 = J.of Cell Science					
28.*15:10/3	28:21/7	*29:21/7	44:30/14	*116;82/31	*+3 non-GS/CF SIEs
29. 13:9/4	15:12/3	17:14/3	65:44/21	110;79/31	
30. 12:6/6	53:46/7	45:27/18	39:19/20	149;98/51	
US: Biology 31-33 = J. of Cell Biology					
31. 12:10/2	50:41/9	49:37/12	45:27/18	156;115/41	
32. 15:4/11	27:20/7	59:27/32	42:23/19	143;74/69	
33.*22;15/6	15:12/3	36:24/12	25:14/11	*98;65/32	*+1 non-GS/CF SIE
US: Biology 34-36 = Cell Motility and the Cytoskeleton					
34. 23;13/10	42;28/14	25;11/14	50;27/23	140;79/61	
35. 29;16/13	30;22/8	34;18/16	23;12/11	116;68/48	
36.*18;8/7	35;26/9	22;19/3	26;19/7	*101;72/26	*+3 non-GS/CF SIEs

202	399	398	433	1432	(25-36) total *n* SIEs
16.8	33.2	33.2	36.1	119.3	(25-36) mean *n* SIEs

118/77	311/88	248/149	258/175	935/489	(25-36) total *n* GS/CF
7	0	1	0	8	(25-36) total *n* *non-GS/CF

58.4%	77.9%	62.3%	59.6%	65.3%	(25-36) mean % GS
38.1%	22.1%	37.4%	40.4%	34.1%	(25-36) mean % CF
*3.5%		*0.3%		*0.6%	(25-36) mean % *non-GS/CF

Main Corpus of 36 NS Research Articles
SIEs: *N* and % GS vs. CF by 4 RA sections
Overall Summary and Standard Deviations

INT	EXP	RES	DIS	TOTAL *N*	
663	922	1318	1455	4358	(1-36) total *n* SIEs
18.4	25.6	36.6	40.4	121.1	(1-36) mean *n* SIEs
438	744	867	880	2929	(1-36) total *n* GS
218	177	446	567	1408	(1-36) total *n* CF
7	1	5	8	21	(1-36) total *n* *non-GS/CF
66.1%	80.7%	65.8%	60.5%	67.2%	(1-36) mean % GS
32.9%	19.2%	33.8%	39.0%	32.3%	(1-36) mean % CF
*1.0%	*0.1%	*0.4%	*0.5%	*0.5%	(1-36) mean % *non-GS/CF
[15.3]	[10.4]	[12.7]	[13.2]	[8.2]	(1-36) [Standard Deviation of CFs]

Main Corpus of 36 NS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 1 N = 528/1408 = 37.5%

Functional Category	INT	EXP	RES	DIS	TOTAL
1a. **Location in Time - Real World Entity = 7.0%**					
Recently	9	–	–	4	13
Finally	–	4	–	–	4
Initially	1	–	1	1	3
Subsequently	1	1	–	1	3
First	–	1	–	1	2
Now	–	1	1	–	2
Second	–	–	–	1	1
Later	1	–	–	–	1
Next	–	1	–	–	1
Ultimately	–	–	1	–	1
Previously	–	–	1	–	1
Eventually	–	–	1	–	1
At this point	–	–	–	1	1
As yet	–	–	–	1	1
Sometimes	–	–	–	1	1
Simultaneously	–	–	1	–	1
Total	*12*	*8*	*6*	*11*	*37*
1b. **Location in Time - Discourse Entity = 6.4%**					
Finally	3	–	3	7	13
First	–	1	1	4	6
In brief	–	5	–	–	5
In summary	–	1	1	1	3
Second	–	–	1	2	3
In conclusion	–	–	–	2	2
So far	–	–	–	1	1
Initially	–	–	1	–	1
Total	*3*	*7*	*7*	*17*	*34*
2a. **Location in Space - Real World Entity = 1.1%**					
Here	–	–	–	5	5
There	–	–	1	–	1
Total			*1*	*5*	*6*
2b. **Location in Space - Discourse Entity = 0.2%**					
Here	–	–	1	–	1
Total			*1*		*1*
3a. **Addition - appositive = 3.0%**					
For example	4	1	3	5	13
That is	1	–	–	–	1
For instance	1	–	–	–	1
In other words	–	–	–	1	1
Total	*6*	*1*	*3*	*6*	*16*

	Functional Category	INT	EXP	RES	DIS	TOTAL
3b.	Addition - emphatic = 19.0%					
	In addition	7	1	9	9	26
	Moreover	3	–	3	11	17
	Furthermore	2	–	3	8	13
	Also	3	1	3	5	12
	Indeed	1	–	1	9	11
	Similarly	1	–	2	5	8
	Specifically	2	1	–	2	5
	In particular	2	–	–	1	3
	In this respect	–	–	1	1	2
	In this regard	–	–	–	1	1
	Essentially	–	1	–	–	1
	Equally	1	–	–	–	1
	Total	*22*	*4*	*22*	*52*	*100*
4.	Contrast/concession = 31.2%					
	However	19	2	34	51	106
	In/by contrast	2	–	11	9	22
	Nevertheless	3	–	3	5	11
	On the other hand	3	–	2	5	10
	Alternatively	–	–	–	4	4
	Instead	–	–	1	2	3
	Rather	1	–	1	–	2
	Otherwise	–	1	–	1	2
	But	–	–	–	1	1
	By comparison	–	–	1	–	1
	Even so	–	–	–	1	1
	Nor	1	–	–	–	1
	Total	*29*	*3*	*53*	*79*	*164*
5a.	Cause - reason/result = 22.1%					
	Thus	10	5	23	37	75
	Therefore	1	1	3	11	16
	As a consequence	3	–	1	5	9
	Hence	–	2	3	4	9
	As a result	1	–	–	2	3
	Accordingly	2	–	–	–	2
	For this reason	1	–	1	–	2
	So	–	–	–	1	1
	Total	*18*	*8*	*31*	*60*	*117*
5b.	Cause - purpose = 0.4%					
	To this end	2	–	–	–	2
	Total	*2*				*2*
6.	Means = 0.4%					
	In this way	–	–	–	2	2
	Total				*2*	*2*

Main Corpus of 36 NS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 1 N = 528/1408 = 37.5%

	Functional Category	INT	EXP	RES	DIS	TOTAL
7a.	**Condition - real = 1.7%**					
	In this case	–	–	1	3	4
	Then	–	–	–	3	3
	In practice	–	1	–	1	2
	Total		*1*	*1*	*7*	*9*
7b.	**Condition - hypothetical = 1.3%**					
	Perhaps	–	–	–	3	3
	In principle	–	–	1	–	1
	Conceivably	–	–	–	1	1
	Potentially	–	–	–	1	1
	As a rule	–	–	–	1	1
	Total			*1*	*6*	*7*
8a.	**Validation - external = 2.3%**					
	In general	1	–	4	3	8
	Usually	1	–	–	–	1
	Normally	–	1	–	–	1
	Structurally	–	–	–	1	1
	Qualitatively	–	–	1	–	1
	Total	*2*	*1*	*5*	*4*	*12*
8b.	**Validation - internal = 1.7%**					
	Clearly	–	–	1	3	4
	Certainly	–	–	–	1	1
	Evidently	–	–	1	–	1
	Obviously	–	–	–	1	1
	Strictly speaking	–	–	–	1	1
	Significantly	–	–	1	–	1
	Total			*3*	*6*	*9*
9.	**Viewpoint = 2.3%**					
	Unfortunately	–	1	1	4	6
	Interestingly	–	–	2	2	4
	Surprisingly	–	–	1	–	1
	Apparently	–	–	–	1	1
	Total		*1*	*4*	*7*	*12*

Main Corpus of 36 NS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 2 N = 535/1408 = 38.0%

Functional Category	INT	EXP	RES	DIS	TOTAL
1a. Location in Time - RWE = 17.9%					
After (the removal of IFNs)	–	30	7	5	42
Upon (exposure)	–	–	10	2	12
Following (release)	1	8	–	–	9
At (fertilisation)	2	4	2	–	8
Prior to (growth)	–	4	–	2	6
During (annealing for 1h)	2	–	3	–	5
Before (reactant addition)	–	3	1	1	5
By (embryonic day 12)	1	–	3	–	4
Over (the next few days)	2	–	–	–	2
In (the recovery period)	–	–	–	1	1
Since (then)	1	–	–	–	1
In the course of (time)	1	–	–	–	1
Total	*10*	*49*	*26*	*11*	*96*
1b. Location in Time - DE = 0.2%					
Throughout (the discussion)	–	–	1	–	1
Total			*1*		*1*
2a. Location in Space - RWE = 37.6%					
In (serial sections)	17	13	37	26	93
At (294K)	8	–	18	15	41
For (the Mn alloy)	1	7	13	16	37
From (the value of O_{225})	1	–	5	4	10
Within (the transition zone)	–	–	2	2	4
Below (the spectra)	–	–	3	–	3
On (a glass substrate)	–	1	1	–	2
Among (the carbon fibres)	–	1	1	–	2
With (zeolites)	1	–	–	1	2
Above (95K)	–	–	1	–	1
Outside (of the epithelium)	–	–	1	–	1
Near (540K)	–	–	–	1	1
(50 cm upstream)	–	1	–	–	1
Up to (the total dose explored)	–	–	–	1	1
Throughout (the domain)	–	–	1	–	1
Of (the MP based fibres)	–	–	–	1	1
Total	*28*	*23*	*83*	*67*	*201*
2b. Location in Space - DE = 10.8%					
In (this paper/figure)	18	–	19	8	45
From (Figure 2b)	–	–	7	1	8
For (the work described here)	–	2	2	–	4
On (the figure)	1	–	–	–	1
Total	*19*	*2*	*28*	*9*	*58*

3a. Addition - appositive [no examples]

	Functional Category	INT	EXP	RES	DIS	TOTAL
3b.	Addition – emphatic = 3.4%					
	In addition to (localisation)	–	1	3	2	6
	As for (the MV core proteins)	–	–	1	1	2
	Like (the 10°0 signals)	–	–	2	–	2
	By analogy (with the SH system)	–	–	–	2	2
	As with (observations of proteins)	–	–	–	1	1
	Similar to (silicidation)	1	–	–	–	1
	Concomitant with (this increase)	–	–	1	–	1
	In association with (the increase)	–	–	–	1	1
	As in the case of (discharge)	–	1	–	–	1
	Besides (Bi₄M)	–	–	1	–	1
	Total	*1*	*2*	*8*	*7*	*18*
4.	Contrast/concession = 3.9%					
	In contrast to (p36p33)	2	–	4	4	10
	Despite (this information)	4	–	–	1	5
	Unlike (the present data)	–	–	1	1	2
	In comparison to (In₀.₁₅Ga)	–	–	1	–	1
	Instead of (this surface pattern)	–	–	1	–	1
	Except for (a % of trajectories)	–	–	–	1	1
	With the exception of (copper)	–	–	1	–	1
	Total	*6*		*8*	*7*	*21*
5a.	Cause – reason/result = 2.2%					
	Because of (the similarities)	3	–	1	1	5
	In view of (these difficulties)	–	1	1	2	4
	Due to (chemical reaction)	–	1	–	–	1
	In light of (these results)	–	–	–	1	1
	In pursuit of (the concept)	–	–	–	1	1
	Total	*3*	*2*	*2*	*5*	*12*
5b.	Cause – purpose = 4.7%					
	For (evaluation of the gap)	2	13	5	4	24
	As a final check (on channel 1b)	–	–	–	1	1
	Total	*2*	*13*	*5*	*5*	*25*
6.	Means = 0.4%					
	By (a number of criteria)	–	–	1	–	1
	With (this construction)	–	1	–	–	1
	Total		*1*	*1*		*2*

	Functional Category	INT	EXP	RES	DIS	TOTAL
7a.	Condition - real = 11.8%					
	In the case of (x = 0.92)	3	3	8	17	31
	In the presence of (2m-GdmHCl)	–	–	8	1	9
	Under (standard conditions)	–	2	5	1	8
	With (the addition of forskolin)	–	2	2	1	5
	For (any redox system)	1	–	2	2	5
	In the absence of (these compounds)	1	–	1	1	3
	For the case of (copolymers)	–	–	1	–	1
	Without (data at higher doses)	–	–	–	1	1
	Total	*5*	*7*	*27*	*24*	*63*
7b.	Condition - hypothetical = 0.2%					
	With (such possibilities)	1		–	–	1
	Total	*1*				*1*
8a.	Validation - external = 5.6%					
	In (previous studies)	6	–	–	7	13
	On the basis of (these results)	2	–	–	1	3
	From (data of Itoh *et al.*)	1	–	–	3	4
	According to (Keuch[5])	–	–	–	2	2
	In agreement with (Noyce[23])	–	–	1	1	2
	Following (suggestions of Bell[3])	–	–	–	2	2
	As in (the experiments above)	–	–	1	–	1
	In support of (this)	–	–	–	1	1
	Consistent with (observations of x)	–	–	–	1	1
	In accordance with (x roles)	1	–	–	–	1
	Total	*10*		*2*	*18*	*30*
8b.	Validation - internal = 1.1%					
	In (refs.1 and 2)	–	–	–	2	2
	In accordance with (gene expression)	–	–	–	1	2
	From (X-ray studies)	–	–	–	1	1
	As in (experiments discussed above)	–	–	1	–	1
	On the basis of (these results)	–	–	1	–	1
	Total			*2*	*4*	*6*
9.	Viewpoint = 0.2%					
	In the broader view	–	–	–	1	1
	Total				*1*	*1*

Main Corpus of 36 NS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 3 N = 345/1408 = 24.5%

	Functional Category	INT	EXP	RES	DIS	TOTAL
1a.	Location in Time - RWE = 7.2%					
	As (the mismatch becomes larger)	1	–	4	5	10
	When (observing LIF from CH_3O)	–	2	2	–	4
	After (polishing to 100nm)	–	3	–	1	4
	While (studying the absorption edge)	1	–	–	1	2
	(Once initiated)	1	–	1	–	2
	Before (experiments were begun)	–	1	–	1	2
	In (determining the flow rate)	–	1	–	–	1
	Total	*3*	*7*	*7*	*8*	*25*
1b.	Location in Time - DE = 1.2%					
	Having concluded (that)	–	–	–	2	2
	(To start with one extreme)	–	–	–	1	1
	In (describing their work on NH_3)	–	–	–	1	1
	Total				*4*	*4*
2a.	Location in Space - RWE = 0.6%					
	Where (uptake occurred)	–	1	1	–	2
	Total		*1*	*1*		*2*
2b.	Location in Space - DE [no examples]					
3a.	Addition - appositive [no examples]					
3b.	Addition - emphatic = 0.6%					
	(As far as MP fibres are concerned)	–	–	1	–	1
	(Taken together)	1	–	–	–	1
	Total	*1*		*1*		*2*
4.	Contrast/concession = 21.1%					
	Although(there is controversy)	12	4	13	24	53
	While (the fraction was 59%)	1	1	7	9	18
	Whereas (the latter took up more slowly)	–	–	1	–	1
	Just as DMJ can be considered	–	–	–	1	1
	Total	*13*	*5*	*21*	*34*	*73*
5a.	Cause - reason/result = 15.7%					
	Since (there is anisotropy)	3	8	10	17	38
	As (there is no change)	2	–	1	4	7
	Because (t-PA binds fibrin)	1	2	1	2	6
	Considering (that)	–	–	–	2	2
	In view of the fact (that)	–	–	–	1	1
	Total	*6*	*10*	*12*	*26*	*54*
5b.	Cause - purpose = 13.9%					
	In order to (obtain HL cells)	5	14	19	9	47
	In an attempt (to)	1	–	–	–	1
	Total	*6*	*14*	*19*	*9*	*48*

	Functional Category	INT	EXP	RES	DIS	TOTAL
6.	Means = 6.4%					
	(Using ionophore induction)	1	4	4	6	15
	By (using Henry's law)	1	–	2	2	5
	In (analysing the dependence)	1	–	–	1	2
	Total	*3*	*4*	*6*	*9*	*22*
7a.	Condition – real = 13.3%					
	When (fk = 1 is applied)	3	1	16	14	34
	If (the polarity is reversed)	1	–	–	2	3
	Unless (otherwise stated)	–	2	–	1	3
	Given (a collision time of 4ms)	–	–	2	–	2
	Under the conditions where($a>b$)	1	–	–	–	1
	(Incorporated in x program)	–	–	1	–	1
	Depending on (the conditions)	–	–	–	1	1
	Bearing in mind (the miscibility gap)	–	–	–	1	1
	Total	*5*	*3*	*19*	*19*	*46*
7b.	Condition – hypothetical = 8.1%					
	If (C^{s+} works as CsOH)	–	–	1	21	22
	Assuming that (V = constant)	–	–	1	2	3
	On/under the assumption that (A=c)	1	–	–	1	2
	Whatever (the fate of the cells)	–	–	–	1	1
	Total	*1*		*2*	*25*	*28*
8a.	Validation – external = 2.6%					
	As (is well known)	–	–	2	3	5
	Based on (kinetic arguments)	–	–	–	2	2
	Referring to (data of Haber[5])	–	–	–	1	1
	As (noted earlier[13,19])	1	–	–	–	1
	Total	*1*		*2*	*6*	*9*
8b.	Validation – internal = 9.3%					
	As (shown in figure 1)	–	–	22	7	29
	As (expected)	–	–	2	–	2
	As (evidenced by particle distribution)	–	–	1	–	1
	Total			*25*	*7*	*32*
9.	Viewpoint [no examples]					

Journal Title/Volume/Year/Author(s) - issues/year

Journal of Physical Society of Japan - 12/year
37. Vol.58/1 (1989) pp. 269-278; Saito *et al.*
38. Vol.58/2   pp. 726-732; Watanabe,K.
39. Vol.58/3   pp. 930-939; Akishige,Y.
40. Vol.58/3   pp. 1080-1086; Miyake *et al.*
41. Vol.58/8   pp. 2867-2876; Hayakawa,T.
42. Vol.58/8   pp. 2994-3000; Itoh,M.

Japanese Journal of Applied Physics - 12/year
43. Vol.29/2 (1990) pp. 301-304; Uematsu *et al.*
44. Vol.29/2   pp. 308-312; Aso *et al.*
45. Vol.29/2   pp. 356-360; Kumeda *et al.*
46. Vol.29/3   pp. 564-568; Majima *et al.*
47. Vol.29/3   pp. 585-590; Itabashi *et al.*
48. Vol.29/6   pp. 1059-1065; Tanaka *et al.*

Bulletin of Chemical Society of Japan - 12/year
49. Vol.62/10 (1989) pp. 3075-3080; Yamashita,S.
50. Vol.62/10   pp. 3098-3102; Sanamasa,I.
51. Vol.63/1 (1990) pp. 102-107; Hamada *et al.*
52. Vol.63/1   pp. 192-198; Tanabe *et al.*
53. Vol.63/4   pp. 1221-1225; Aika *et al.*
54. Vol.63/6   pp. 1695-1699; Sakurai,M.

Chemical Pharmaceutical Bulletin - 12/year
55. Vol.38/3 (1990) pp. 765-768; Tsuji *et al.*
56. Vol.38/4   pp. 988-992; Ohno *et al.*
57. Vol.38/4   pp. 1035-1038; Miyazaki,K.
58. Vol.38/6   pp. 1648-1652; Koumoto *et al.*
59. Vol.38/6   pp. 1656-1659; Yatsuyanagi,J.
60. Vol.38/7   pp. 1810-1814; Shimabayashi,S.

Cell Structure and Function - 6/year
61. Vol.14/1 (1989) pp.   35-43; Kinoshita,Y.
62. Vol.14/1   pp.   61-74; Ono *et al.*
63. Vol.14/1   pp.   75-85; Yamamoto,N.
64. Vol.14/1   pp. 129-140; Okazaki *et al.*
65. Vol.14/4   pp. 459-471; Koyasu *et al.*
66. Vol.15/3   pp. 143-150; Ohkura *et al.*

Japanese Journal of Physiology - 6/year
67. Vol.39/2 (1989) pp. 229-240; Yanagidaira,Y.
68. Vol.39/3   pp. 371-383; Miyoshi,H.
69. Vol.39/4   pp. 549-558; Shimamoto,Y.
70. Vol.39/4   pp. 559-570; Karakida,T.
71. Vol.39/5   pp. 659-671; Maruta,K.
72. Vol.39/5   pp. 755-765; Takano,N.

Corpus of 36 NNS Research Articles
SIEs: *N* and % GS vs. CF by 4 RA sections
[average SIEs = 112.2 - max 180: min 69]

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	

Japan Physics 37-42 = J.of Physical Soc. of Japan

37.	9;6/3	18;13/5	30;21/9	78;34/44	135;74/61	
38.	11;7/4	15;10/5	*20;10/9	*48;14/33	*94;41/51	*+2 non-GS/CF SIEs
39.	15;11/4	19;15/4	*51;32/18	*58;29/28	*143;87/54	*+2 non-GS/CF SIEs
40.	13;8/5	14;13/1	41;26/15	71;35/36	139;82/57	
41.	11;8/3	19;16/3	36;19/17	114;54/60	180;97/83	
42.	24;12/12	20;17/3	45;35/10	*64;39/24	*153;103/49	*+1 non-GS SIE

Japan Physics 43-48 = Japanese J. of Applied Physics

43.	11;5/6	13;13/0	28;17/11	40;28/12	92;63/29	
44.	20;15/5	13;12/1	38;27/11	34;19/15	105;73/32	
45.	16;10/6	15;10/5	23;18/5	55;29/26	109;67/42	
46.	14;6/8	10;9/1	20;10/10	54;27/27	98;52/46	
47.	17;11/6	16;14/2	23;21/2	44;27/17	100;73/27	
48.	11;7/4	24;22/2	95;71/24	37;23/14	167;123/44	

INT	EXP	RES	DIS	TOTAL	
172	196	450	697	1515	(37-48) total *n* SIEs
14.3	16.3	37.5	58.1	126.2	(37-48) mean *n* SIEs

106/66	164/32	307/141	358/336	935/575	(37-48) total *n* GS/CF
0	0	2	3	5	(37-48) total *n* *non-GS/CF

61.6%	83.7%	68.2%	51.4%	61.7%	(37-48) mean % GS
38.4%	16.3%	31.3%	48.2%	38.0%	(37-48) mean % CF
0.0%	0.0%	*0.5%	*0.4%	*0.3%	(37-48) mean % *non-GS/CF

INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	

Japan Chemistry 49-54) = Bull.Chem.Soc. Japan

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
49.	8;6/2	17;15/2	40;22/18	45;30/15	110;73/37	
50.	12;9/3	11;9/2	14;11/3	*52;39/12	*89;68/20	*+1 non-GS/CF SIE
51.	6;3/3	19;18/1	49;34/15	15;10/5	89;65/24	
52.	10;8/2	21;17/4	46;22/24	44;20/24	121;67/54	
53.	13;6/7	26;23/3	41;28/13	38;22/16	118;79/39	
54.	7;3/4	6;6/0	12;7/5	44;28/16	69;44/25	

Japan Chemistry 55-60 = Chem.Pharm.Bull

	INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF
55.	11;9/2	38;31/7	35;30/5	18;13/5	102;83/19
56.	13;10/3	44;37/7	44;29/15	26;15/11	127;91/36
57.	6;3/3	32;26/6	30;25/5	32;18/14	100;72/28
58.	12;6/6	26;23/3	44;20/24	32;17/15	114;66/48
59.	8;5/3	28;26/2	27;18/9	21;17/4	84;66/18
60.	25;22/3	41;34/7	52;38/14	20;13/7	138;107/31

INT	EXP	RES	DIS	TOTAL	
131	309	434	387	1261	(49-60) total *n* SIEs
10.9	25.7	36.2	32.2	115.0	(49-60) mean *n* SIEs
90/41	265/44	284/150	242/144	881/379	(49-60) total *n* GS/CF
0	0	0	1	1	(49-60) total *n* *non-GS/CF
68.7%	85.8%	65.4%	62.5%	69.9%	(49-60) mean % GS
31.3%	14.2%	34.6%	37.2%	30.0%	(49-60) mean % CF
0.0%	0.0%	0.0%	*0.3%	*0.1%	(49-60) mean % *non-GS/CF

INT	EXP	RES	DIS	TOTAL *N* SIE;GS/CF	
Japan Biology 61-66 = Cell Structure and Function					
61. 10;5/5	13;13/0	*25;15/9	29;21/8	*77;54/22	*+1 non-GS/CF SIE
62. 17;9/8	37;27/10	44;31/13	21;14/7	119;81/38	
63. 12;8/4	32;26/6	24;15/9	25;17/8	93;66/27	
64. 17;13/4	40;33/7	32;22/10	49;34/15	138;102/36	
65. 11;6/5	17;16/1	53;40/13	30;23/7	111;85/26	
66. 19;14/5	40;35/5	56;33/23	22;10/12	137;92/45	
Japan Biology 67-72 = Japanese J. of Physiology					
67. 10;7/3	29;16/13	15;11/4	38;21/17	92;55/37	
68. 8;4/4	35;23/12	40;27/13	17;16/1	100;70/30	
69. 16;7/9	32;27/5	15;9/6	38;25/13	101;68/33	
70. 10;8/2	27;17/10	32;15/17	17;11/6	86;51/35	
71. 8;5/3	20;17/3	58;40/18	29;21/8	115;83/32	
72. 11;7/4	29;20/9	*23/17/5	33;17/16	*96;61/34	*+1 non-GS/CF SIE
149	351	417	348	1265	(61-72) total *n* SIEs
12.4	29.2	34.7	29.0	114.6	(61-72) mean *n* SIEs
93/56	270/81	275/140	230/118	868/395	(61-72) total *n* GS/CF
0	0	2	0	2	(61-72) total *n* *non-GS/CF
62.4%	76.9%	65.9%	66.1%	68.6%	(61-72) mean % GS
37.6%	23.1%	33.6%	33.9%	31.2%	(61-72) mean % CF
0.0%	0.0%	*0.5%	0.0%	*0.2%	(61-72) mean % *non-GS/CF

Corpus of 36 NNS Research Articles
SIEs: N and % GS vs. CF by 4 RA sections
Overall Summary and Standard Deviations

INT	EXP	RES	DIS	TOTAL N SIE;GS/CF	
452	856	1301	1432	4041	(37-72) total n SIEs
12.6	23.8	36.1	39.8	112.2	(37-72) mean n SIEs
289	699	866	830	2684	(37-72) total n GS
163	157	431	598	1349	(37-72) total n CF
0	0	4	4	8	(37-72) total n *non-GS/CF
63.9%	81.7%	66.6%	58.0%	66.4%	(37-72) mean % GS
36.1%	18.3%	33.1%	41.7%	33.4%	(37-72) mean % CF
0.0%	0.0%	0.3%	0.3%	0.2%	(37-72) mean % *non-GS/CF
[12.8]	[10.8]	[11.0]	[12.3]	[8.4]	(37-72) Standard Deviation of CFs

Corpus of 36 NNS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 1 N = 540/1349 = 40.0%

	Functional Category	INT	EXP	RES	DIS	TOTAL
1a.	**Location in Time - RWE = 10.4%**					
	Recently	11	–	1	5	17
	Then	1	8	1	–	10
	Next	–	1	3	2	6
	Thereafter	–	1	4	–	5
	Finally	–	3	1	–	4
	First	–	2	1	1	4
	At present	–	–	–	4	4
	Previously	1	–	–	–	1
	Subsequently	–	1	–	–	1
	Until now	–	–	–	1	1
	Second	–	–	–	1	1
	Third	–	–	–	1	1
	Sometimes	–	–	–	1	1
	Total	*13*	*16*	*11*	*16*	*56*
1b.	**Location in Time - DE = 3.9%**					
	In conclusion	–	–	–	6	6
	In summary	–	–	1	3	4
	Finally	–	–	–	4	4
	In brief	–	2	–	–	2
	First	–	–	–	1	1
	Earlier	–	–	1	–	1
	To begin with	–	–	–	1	1
	So far	–	–	–	1	1
	Basically	–	1	–	–	1
	Total		*3*	*2*	*16*	*21*
2a.	**Location in Space - RWE = 0.5%**					
	Here	–	–	2	–	2
	There	–	–	–	1	1
	Total			*2*	*1*	*3*
2b.	**Location in Space - DE = 1.5%**					
	Here	–	–	3	5	8
	Total			*3*	*5*	*8*
3a.	**Addition - appositive = 2.2%**					
	That is	–	–	1	5	6
	For instance	1	–	–	1	2
	In other words	–	–	1	1	2
	For example	–	–	1	–	1
	Namely	–	–	–	1	1
	Total	*1*		*3*	*8*	*12*

	Functional Category	INT	EXP	RES	DIS	TOTAL
3b.	Addition - emphatic = 17.6%					
	Furthermore	8	3	8	17	36
	In addition	5	–	7	7	19
	Moreover	3	–	5	5	13
	Also	–	–	1	3	4
	In fact	–	–	1	3	4
	Indeed	–	–	–	3	3
	In particular	1	–	–	2	3
	Actually	–	–	–	3	3
	Similarly	–	–	2	–	2
	In this connection	1	–	–	1	2
	Besides	–	–	1	–	1
	Of course	–	–	–	1	1
	And	–	–	–	1	1
	In this regard	–	–	–	1	1
	Analogously	1	–	–	–	1
	Concretely	–	–	–	1	1
	Total	*19*	*3*	*25*	*48*	*95*
4.	Contrast/concession = 34.4%					
	However	16	1	36	61	114
	On the other hand	2	1	20	15	38
	In/by contrast	1	–	8	4	13
	But	–	–	2	5	7
	Nevertheless	1	1	1	2	5
	Anyway	–	–	–	2	2
	On the contrary	–	–	–	–	2
	Alternatively	–	–	–	2	2
	In any event	–	–	–	1	1
	Yet	–	–	–	1	1
	Otherwise	–	–	–	1	1
	Total	*22*	*3*	*67*	*94*	*186*
5a.	Cause - reason/result = 21.1%					
	Thus	8	3	12	23	46
	Therefore	6	1	12	27	46
	Hence	–	–	1	6	7
	Accordingly	–	–	2	2	4
	As a consequence	–	–	–	4	4
	As a result	–	–	–	3	3
	So	–	–	2	1	3
	Thereby	–	–	–	1	1
	Total	*14*	*4*	*29*	*67*	*114*
5b.	Cause - purpose [no examples]					

	Functional Category	INT	EXP	RES	DIS	TOTAL
6.	Means = 0.4%					
	In this way	–	–	1	1	2
	Total			*1*	*1*	*2*
7a.	Condition - real = 5.4%					
	Then	1	–	4	17	22
	In this case	1	2	3	1	7
	Total	*2*	*2*	*7*	*18*	*29*
7b.	Condition - hypothetical = 0.2%					
	Possibly	–	–	1	–	1
	Total			*1*		*1*
8a.	Validation - external = 1.7%					
	In general	–	–	3	1	4
	Usually	–	1	–	1	2
	Histologically	–	–	–	1	1
	Typically	–	1	–	–	1
	On the whole	–	–	1	–	1
	Total		*2*	*4*	*3*	*9*
8b.	Validation - internal [no examples]					
9.	Viewpoint = 0.7%					
	Interestingly	–	–	1	1	2
	Intriguingly	–	–	1	–	1
	Unfortunately	–	–	–	1	1
	Total			*2*	*2*	*4*

373

	Functional Category	INT	EXP	RES	DIS	TOTAL
1a.	Location in Time - RWE = 19.1%					
	After (cooling treatment)	–	33	14	9	56
	During (saline injection)	2	2	3	1	8
	At (the end of the period)	–	3	3	2	8
	Upon (further cooling)	1	–	4	1	6
	In (the earlier stage)	–	–	4	1	5
	For (the first time)	–	–	–	2	2
	Prior to (each measurement)	–	2	–	–	2
	Following (incubation for 1h)	–	1	1	–	2
	On (day two)	–	1	–	–	1
	Over (a period of time)	–	1	–	–	1
	(40 minutes) later	–	1	–	–	1
	Total	*3*	*44*	*29*	*16*	*92*
1b.	Location in Time - DE [no examples]					
2a.	Location in Space - RWE = 33.8%					
	In (Raman spectroscopy)	14	13	29	35	91
	For (alcohol solutions)	1	4	5	14	24
	At (T_Tc)	1	2	11	4	18
	Among (these techniques)	2	–	5	2	9
	Above (T_Tc)	–	–	2	4	6
	From (the x-sections)	–	–	3	3	6
	Below (T_Tc)	2	–	–	1	3
	On (Ru/A.C.)	–	–	–	2	2
	To (the mixture)	–	2	–	–	2
	Between (the samples)	–	–	–	1	1
	Within (these cells)	1	–	–	–	1
	Total	*21*	*21*	*55*	*66*	*163*
2b.	Location in Space - DE = 13.5%					
	In (this paper/figure)	19	3	16	26	64
	From (the figure)	–	–	1	–	1
	Total	*19*	*3*	*17*	*26*	*65*
3a.	Addition - appositive [no examples]					
3b.	Addition - emphatic = 3.1%					
	As for (crystals of NaCl)	–	–	2	10	12
	Concerning (the constant E)	–	–	1	1	2
	Along with (the decrease in E')	–	–	1	–	1
	Total			*4*	*11*	*15*

APPENDIX F2 (continued)
Corpus of 36 NNS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 2 N = 482/1349 = 35.7%

	Functional Category	INT	EXP	RES	DIS	TOTAL
8a.	Validation – external = 4.6%					
	In (previous studies)	8	–	2	5	15
	According to (Szabo *et al.*)	–	1	1	2	4
	In support of (identification)	–	–	–	1	1
	As in (a previous study)	–	–	–	1	1
	On the basis of (structure)	–	–	–	1	1
	Total	*8*	*1*	*3*	*10*	*22*
8b.	Validation – internal = 3.1%					
	From (the results in Fig.1)		–	2	4	6
	On the basis of (structure)	–	–	1	3	4
	According to (Fig.4)	–	–	2	2	4
	In (the present results)	–	–	–	1	1
	Total			*5*	*10*	*15*
9.	Viewpoint [no examples]					

Corpus of 36 NNS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 3 N = 327/1349 = 24.3%

	Functional Category	INT	EXP	RES	DIS	TOTAL
1a.	Location in Time - RWE = 5.8%					
	After (remaining at location)	–	5	6	1	12
	As (the temperature increases)	–	–	4	1	5
	Prior to (applying the method)	–	1	–	–	1
	On (increasing energy)	–	–	–	1	1
	Total		*6*	*10*	*3*	*19*
1b.	Location in Time - DE [no examples]					
2a.	Location in Space - RWE [no examples]					
2b.	Location in Space - DE [no examples]					
3a.	Addition - appositive [no examples]					
3b.	Addition - emphatic = 0.9%					
	Besides (being effective)	–	–	–	1	1
	(As far as the trend is concerned)	–	–	1	–	1
	Considering (the role of hair)	–	–	–	1	1
	Total			*1*	*2*	*3*
4.	Contrast/concession = 11.3%					
	Although(there is controversy)	5	2	4	19	30
	Compared to (the resistivity)	–	–	3	1	4
	While (the fraction was 59%)	–	–	1	2	3
	Total	*5*	*2*	*8*	*22*	*37*
5a.	Cause - reason/result = 20.5%					
	Since (there is anisotropy)	7	6	7	19	39
	Because (*t*-PA binds fibrin)	3	1	2	7	13
	As (there is no change)	–	–	4	5	9
	(Differing from triamines)	–	–	2	1	3
	In (applying H-H formalism)	1	–	–	–	1
	In taking x into account)	–	–	–	1	1
	Considering (that)	–	–	–	1	1
	Total	*11*	*7*	*15*	*34*	*67*
5b.	Cause - purpose = 12.6%					
	In order to (obtain HL cells)	5	9	15	11	40
	For (distinguishing *a* and *b*)	–	–	–	1	1
	Total	*5*	*9*	*15*	*12*	*41*
6.	Means = 5.2%					
	Using (Fura-2)	2	–	3	7	12
	By (using Henry's law)	1	–	1	2	4
	In (applying H-H formalism)	–	–	1	–	1
	Total	*3*		*5*	*9*	*17*

Corpus of 36 NNS Research Articles
CFs by FUNCTIONAL CATEGORIES by 4 RA SECTIONS
CF TYPE 3 N = 327/1349 = 24.3%

	Functional Category	INT	EXP	RES	DIS	TOTAL
7a.	Condition - real = 14.4%					
	When (fk = 1 is applied)	–	5	20	12	37
	If (the polarity is reversed)	–	1	2	4	7
	Unless (otherwise stated)	–	1	–	–	1
	(Keeping this in mind)	–	–	–	1	1
	As long as (hw_s is considered)	–	–	–	1	1
	Total		*7*	*22*	*18*	*47*
7b.	Condition - hypothetical = 7.0%					
	If (C^{s+} works as CsOH)	1	1	2	15	19
	On the assumption that (A = c)	–	–	–	3	3
	Assuming that (V = constant)	–	–	–	1	1
	Total	*1*	*1*	*2*	*19*	*23*
8a.	Validation - external = 1.2%					
	As (reported previously[21,22])	2	1	–	1	4
	Total	*2*	*1*		*1*	*4*
8b.	Validation - internal = 21.1%					
	As (shown in figure 1)	–	4	38	12	54
	As (noted earlier)	–	1	3	9	13
	As far as (we are aware)	–	–	–	1	1
	As (expected)	–	–	1	–	1
	Total		*5*	*42*	*22*	*69*
9.	Viewpoint [no examples]					

1. Sekine,S. *et al.*
Chemical Physics Letters
Vol.180/3, 173-178. 1991.
2. Hamasaki,N. *et al.*
Manuscript submitted for publication.
3. Saiki,T. *et al.*
Experimental Cell Research
Vol.192, 380-388. 1991.
4. Kato,T. *et al.*
Journal of Physical Society of Japan
Vol.61/1, 275-285. 1992.
5. Shimojo,M. *et al.*
ISIJ International
Vol.31, 870-874. 1991.
6. Mohri,T. *et al.*
Cell Structure and Function
Vol.15, 309-315. 1990.
7. Haga,T. *et al.*
Japanese Journal of Applied Physics
Vol.30,12B, 180-184. 1991.
8. Yanabe,Y. *et al.*
Manuscript submitted for publication.
9. Gotoh,J. *et al.*
Japanese Journal of Applied Physics
Vol.29/10, 1767-1770. 1990.
10. Yamauchi,N. *et al.*
Journal of Antibiotics
Vol.45/5, 756-766. 1992.

APPENDIX H
Corpus (N=10) of NNS novices' FIRST and FINAL RA drafts
Comparison of usage of Context Frames – Type 1

FIRST (N=161)			FINAL (N=173)		
CF	N	%	CF	N	%
However	26	16%	However	38	22%
Therefore	24	15%	Therefore	30	17%
Then	14	9%	Thus	14	8%
And	13	8%	Furthermore	12	7%
But	10	6%	On the other hand	9	5%
Furthermore	8	5%	Then	5	3%
On the other hand	7	4%	In addition	5	3%
Thus	6	4%	Consequently	5	3%
First	6	4%	Moreover	4	2%
In addition	4	2%	In contrast	4	2%
Recently	3	2%	First	3	2%
Moreover	3	2%	In particular	3	2%
Finally	2	1%	In other words	3	2%
Second	2	1%	For example	3	2%
In this case	2	1%	Finally	2	1%
Unfortunately	2	1%	Generally	2	1%
Hence	2	1%	In this case	2	1%
For example	2	1%	Subsequently	2	1%
In contrast	2	1%	As a result	2	1%
Especially	2	1%	Recently	2	1%
So	2	1%	Indeed	2	1%
Besides	1	*	+Because	2	1%
In conclusion	1	*	Second	1	*
Interestingly	1	*	Thirdly	1	*
On the contrary	1	*	Indeed	1	*
In summary	1	*	In conclusion	1	*
That is to say	1	*	Interestingly	1	*
By the way	1	*	Similarly	1	*
In principle	1	*	Perhaps	1	*
Of course	1	*	Afterwards	1	*
Hereafter	1	*	Nevertheless	1	*
Actually	1	*	Now	1	*
Namely	1	*	But	1	*
Until now	1	*	Up to now	1	*
After all	1	*	Here	1	*
As a consequence	1	*	Accordingly	1	*
Here	1	*	In summary	1	*
In this manner	1	*	In this context	1	*
+While	1	*	Originally	1	*
Last	1	*	In this manner	1	*
			+While	1	*

indicates value below 1%   + incorrect usage as minimal CF 1

380

APPENDIX J
Corpus (N=10) of NNS novices' FIRST and FINAL drafts
Usage of Empty H&O Theme

FIRST Empty H&O Theme (N=29)	FINAL Empty H&O Theme (N=21)
It was necessary (4)	It was suggested (4)
It is possible [that] (3)	It is worthwhile (2)
It seems (2)	It was notable (2)
It was guessed (2)	It seems likely (2)
It is important (2)	It is essential (2)
It is attractive (1)	It was necessary (1)
It is expected	It is possible that
It would be worth while	It is important
It was worthwhile	It is clarified
It was notable	It may well be said
It is likely	It is to be noted
It was doubtful	It appears
It is certain	It seems appropriate
It is intriguing	It seemed difficult
It is suggested	
It is clarified	
It seems reasonable	
It seems likely	
It is essential	
It is easy to show	
+It is rather high	

+ *would appear to be incorrect usage*

Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
INTRODUCTION - CFs by function

		FIRST	FINAL	CHANGE	NS corpus
1a.	Location in Time - RWE	6%	6%	nc[1]	11.5%
1b.	- DE	0%	0%	nc	1.4%
2a.	Location in Space - RWE	10%	10%	nc	12.8%
2b.	- DE	10%	10%	nc	8.7%
3a.	Addition - appositive	2%	2%	nc	2.8%
3b.	- emphatic	11%	10%	-	11.0%
4.	Contrast/concession	28%	27%	-	22.0%
5a.	Cause - reason/result	21%	21%	nc	12.8%
5b.	- purpose	0%	4%	+	4.1%
6.	Means	6%	2%	-	1.4%
7a.	Condition - real	0%	0%	nc	4.6%
7b.	- hypothetical	0%	0%	nc	0.9%
8a.	Validation - external	6%	6%	nc	6.0%
8b.	- internal	0%	0%	nc	0.0%
9.	Viewpoint	0%	2%	+	0.0%
		100%	100%		100%

Correlation with NS corpus: r = .903   .915

---

[1] nc: no change

Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
DISCUSSION - CFs by function

		FIRST	FINAL	CHANGE	NS corpus
1a.	Location in Time - RWE	2%	2%	nc	5.3%
1b.	- DE	0%	2%	+	3.7%
2a.	Location in Space - RWE	28%	23%	−	12.7%
2b.	- DE	2%	2%	nc	1.6%
3a.	Addition - appositive	0%	2%	+	1.1%
3b.	- emphatic	8%	16%	+	10.4%
4.	Contrast/concession	20%	19%	−	21.2%
5a.	Cause - reason/result	23%	23%	nc	16.0%
5b.	- purpose	5%	2%	−	2.5%
6.	Means	0%	0%	nc	1.9%
7a.	Condition - real	2%	2%	nc	8.8%
7b.	- hypothetical	0%	0%	nc	5.5%
8a.	Validation - external	5%	5%	nc	4.9%
8b.	- internal	5%	2%	−	3.0%
9.	Viewpoint	0%	0%	nc	1.4%
		100%	100%		100%

Correlation with *NS corpus:* r = .832   .872

383

APPENDIX L1
Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
INTRODUCTION: Context Frames - Type 1

FIRST (N=30)		FINAL (N=26)	
Context Frame	N	Context Frame	N
However	8	However	7
Therefore	4	Therefore	4
On the other hand	4	On the other hand	2
In addition	2	In addition	2
Recently	2	Recently	2
Thus	2	Thus	1
Moreover	1	Moreover	1
Indeed	1	Indeed	1
For example	1	For example	1
Consequently	1	Consequently	1
Furthermore	1	As a result	1
Then	1	Then	1
+Because	1	Unfortunately	1
+Although	1	+Although	1

+ *incorrect usage as minimal CF Type 1*

APPENDIX L2
Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
DISCUSSION: Context Frames - Type 1

FIRST (N=16)		FINAL (N=23)	
Context Frame	N	Context Frame	N
However	3	Therefore	4
Therefore	3	However	3
Furthermore	3	Furthermore	3
But	2	Moreover	3
Thus	2	Thus	2
In contrast	1	On the other hand	2
On the other hand	1	In contrast	1
Then	1	In conclusion	1
		For example	1
		So	1
		But	1
		In addition	1

APPENDIX M1
Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
INTRODUCTION: Context Frames - Type 3

FIRST

(N=2)

FINAL

(N=5)

Context Frame	N
Although...............	1
Though................	1

Context Frame	N
Though..................	2
Though..................	1
In order to............	1
To.....................	1

APPENDIX M2
Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
DISCUSSION: Context Frames - Type 3

FIRST
(N=5)

FINAL
(N=4)

Context Frame	N
Although...............	1
In order to............	1
To.....................	1
As seen in Fig.1.......	1
As [reason]............	1

Context Frame	N
Although...............	1
In order to............	1
As [reason]............	1
As shown in Fig.1......	1

Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
INTRODUCTION - Subject roles by domain

	FIRST	FINAL	CHANGE	*NS corpus*
Participant domain				
Discourse Participant (DP)	11%	12%	+	*7.8%*
Participant Viewpoint (PV)	0%	0%	nc	*0.3%*
Interactive Participant (IP)	7%	7%	nc	*8.8%*
	18%	19%	+	*16.9%*
Discourse domain				
Discourse Event/Process (DEP)	2%	0%	–	*1.7%*
Interactive Dis. Entity (IDE)	5%	5%	nc	*5.6%*
Macro Discourse Entity (MaDE)	1%	1%	nc	*0.9%*
Micro Discourse Entity (MiDE)	0%	0%	nc	*0.3%*
Empty Discourse Theme	1%	1%	nc	*1.4%*
	9%	7%	–	*9.9%*
H&O domain				
Objectivised Viewpoint (OV)	3%	1%	–	*2.6%*
Hypothesised Viewpoint (HV)	0%	0%	nc	*0.2%*
Hypothesised Entity (HE)	0%	0%	nc	*0.5%*
Empty H&O Theme	5%	4%	–	*3.5%*
	8%	5%	–	*6.8%*
Real World domain				
Real World Entity (RWE)	33%	37%	+	*40.0%*
Real World Event/Pro. (RWEP)	28%	26%	–	*21.7%*
Mental Process (MP)	1%	2%	+	*1.7%*
Empty Real World Theme	3%	4%	+	*3.0%*
	65%	69%	+	*66.4%*
	100%	100%		*100%*

correlation
with *NS corpus:*   *r* =   .967   .983

APPENDIX N2
Corpus (N=16) of NNS novices' FIRST and FINAL PCs drafts
DISCUSSION - Subject roles by domain

	FIRST	FINAL	CHANGE	NS corpus
Participant domain				
Discourse Participant (DP)	9%	11%	+	8.6%
Participant Viewpoint (PV)	0%	1%	+	1.4%
Interactive Participant (IP)	3%	3%	nc	3.9%
	12%	15%	+	13.9%
Discourse domain				
Discourse Event/Process (DEP)	0%	0%	nc	1.6%
Interactive Dis. Entity (IDE)	0%	1%	+	1.7%
Macro Discourse Entity (MaDE)	0%	0%	nc	0.3%
Micro Discourse Entity (MiDE)	1%	2%	+	1.7%
Empty Discourse Theme	1%	1%	nc	1.6%
	2%	4%	+	6.9%
H&O domain				
Objectivised Viewpoint (OV)	10%	10%	nc	2.8%
Hypothesised Viewpoint (HV)	1%	1%	nc	1.6%
Hypothesised Entity (HE)	1%	1%	nc	1.6%
Empty H&O Theme	8%	5%	−	7.6%
	20%	17%	−	13.6%
Real World domain				
Real World Entity (RWE)	40%	38%	−	43.1%
Real World Event/Pro. (RWEP)	15%	16%	+	18.2%
Mental Process (MP)	4%	3%	−	2.4%
Empty Real World Theme	7%	7%	nc	1.9%
	66%	64%	−	65.6%
	100%	100%		100%

correlation
with *NS corpus:*   r =   .968   .966

387

APPENDIX P1
Corpus (N=16) of NNS novices' FIRST PCs drafts
INTRODUCTION - Concordances of Empty H&O Theme (N=4)

92.4%. There is a great difference in nuclear characteris [6]

ation, it has been emphasised that the role of pHi increa [10]
efore, it is necessary to investigate the roles of kineto [11]
olite. It is of great interest to study secondary metabol [12]

APPENDIX P2
Corpus (N=16) of NNS novices' FINAL PCs drafts
INTRODUCTION - Concordances of Empty H&O Theme (N=3)

---

ation, it has been emphasised that the role of pHi increa [10]
efore, it is necessary to investigate the roles of kineto [11]
olite. It is of great interest to study secondary metabol [12]

APPENDIX P3
Corpus (N=16) of NNS novices' FIRST PCs drafts
DISCUSSION - Concordances of Empty H&O Theme (N=9)

---

otein. It is assumed that sperm activates the G protein w [2]
egion. It also suggests that the assumption of rectilinea [3]
ation. It appears that isotherms of bna slightly decrease [6]
eeded. It is needed that the theoretical analysis using t [7]
(N+1). It appears that the predominant cause of this disc [8]
rmore, it is needed to control four hydroxyl group with e [12]
wever, it lies in the drawback that major product of dias [12]
ethod. It is difficult to measure the absolute value of d [14]
t. But it is necessary to develop the measurement system [14]

APPENDIX P4
Corpus (N=16) of NNS novices' FINAL PCs drafts
DISCUSSION - Concordances of Empty H&O Theme (N=5) .

---

otein. It is assumed that sperm activates the G protein w [2]
egion. It also suggests that the assumption of rectilinea [3]
(N+1). It appears that the oversimplification of the pert [8]
rmore, it is necessary to control remained four hydroxyl [12]
t. But it is necessary to develop the measurement system [14]