THE UNIVERSITY
of LIVERPOOL

Topics in Classroom Discourse

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Richard Watson Todd

January 2003

Declaration

This work is original and has not been submitted previously in support of any degree, qualification or course.

Signature _____

Contents

List of Tables

List of Figures

Topics in Classroom Discourse

Richard Watson Todd

Abstract

This study aims to identify topics and follow topic progression in classroom discourse. A key aspect of the conceptual metafunction of language, topics are notoriously difficult to analyse, and previous research into topics has largely been inadequate since most studies have relied on intuition. Nevertheless, topics are a crucial consideration in the comprehension and production of discourse. This is perhaps especially true for classroom discourse where sequencing of content in eliciting is of vital importance. This study then attempts to conduct a rigorous analysis of topics in stretches of classroom eliciting.

Examining stretches of eliciting from an English language support course at a Thai university, the data was transcribed and divided into T-units, reference was resolved, and ellipsis was identified and supplied. Six methods of analysis likely to provide insights into topics were then applied: an analysis based on Sinclair and Coulthard (1975), another based on Hoey (1991), theme-rheme progression, given-new progression, and two topic-based analyses based on Watson Todd (1998), one focusing on logical relations and the other on associations. The results from these six approaches were then compared with each other, with a control analysis, and with the participants' perspectives on the discourse.

The findings suggest that three of the approaches (the analysis based on Hoey (1991) and the two topic-based analyses) are more productive for identifying topics than the other approaches, although the analysis based on Hoey (1991) has little to say concerning topic progression. Nevertheless, using a combination of approaches and viewing topics as expressed through concepts rather than propositions, it is possible to identify topics at different levels in a topic hierarchy and follow topic progression.

Acknowledgements

I would like to thank Professor Michael Hoey for his support throughout the conducting of this research and for his perceptive and critical feedback which has greatly improved this thesis. Thanks are also due to Dr. Sam Hartley for her help in sorting out connectionism for me and suggesting the use of extended trees. My mother also deserves gratitude for all her work in easing practical and administrative problems.

Most of all, I would like to express my esteem and love for my wife, without whose support and belief in me this research and much else would not have happened. This study is dedicated to her memory.

Chapter 1 Content Structure in the Classroom

1.1 Introduction

In 1998, before starting my doctorate, I was fortunate enough to have an article on classroom discourse published (Watson Todd, 1998). The topic of the article interested me greatly (and still does), and in deciding to study for a doctorate, I decided to follow up the research which had produced the article. To contextualise the present study, then, a brief explanation of the approach used, the findings, and the weaknesses of the article which stimulated this research is needed.

In the article, I focused purely on the conceptual (Widdowson, 1984) or ideational (Halliday, 1973) aspects of discourse, as I believed these to be under-reported in the literature. In investigating content in classroom discourse, I needed to be able to identify and follow the progression of topics, and I suggested several different ways in which this could be done.

To identify topics, there are several potentially productive approaches. We could look at nested levels of theme-rheme or given-new relations (Caron-Pargue and Caron, 1991), whereby frequently recurring themes could be indicative of topic; we could examine the density of links in lexical networks (de Beaugrande and Dressler, 1982; Hoey, 1991), where the most frequently linked items suggest topics; and we could attempt to identify the content schemata (Carrell and Eisterhold, 1983) involved in the discourse.

To follow progression of topics, we need to be able to distinguish between topic maintenance, topic shift and topic drift (see Crow, 1983). Topic maintenance, where a stretch of discourse concerns the same topic, can be taken as the default progression of topic. Topic shift, where there is a jump to a new topic, may be signalled by metadiscoursal markers (McCarthy, 1991) which in the classroom take the form of framing and focusing moves (Sinclair and Brazil, 1975; Sinclair and Coulthard, 1975). Topic drift, where succeeding topics meld into each other, might be identifiable in two ways. From a logical perspective, if we assume that propositions in discourse delimit a range of semantic space, subsequent propositions can be viewed as expanding the range of semantic space referred to by the discourse, and this expansion is equivalent to topic drift (Hurtig, 1977; van Dijk, 1977). Alternatively, we could follow theme-rheme or given-new progression (Daneš, 1974; Lautamatti, 1978), and focus on how new predominant

themes or given information are introduced into the discourse. (More details of these approaches are given in chapters 4 and 5.)

In my article, I attempted to combine all these approaches to produce a conceptual map of the discourse through which topics could be followed. The map was constructed by identifying keywords in the discourse based on frequency and representing them in a hyponymic tree, where relationships between keywords were identified by the logical relation of implication. The progression through keywords was then added to the conceptual map allowing topics to be identified and topic shift and topic drift to be followed (see more details in chapter 5).

A number of tentative conclusions arose from this analysis. Firstly, the ratio of movements between keywords in the conceptual map in proportion to units in the discourse may affect learners' comprehension. Secondly, the distances between keywords in semantic space as shown in the conceptual map are related to the different ways in which topics may progress. Thirdly, the patterns identified may be related to the need for clear sequencing and logic in discourse as advocated in much teacher training (e.g. Cole and Chan, 1987; Jantz, 1989; Kennedy, 1996; Watson Todd, 1997b).

In the article, I attempted to achieve a number of goals. Firstly, the article attempts to marry several disparate approaches into one analysis. Secondly, it tries to diagrammatically explain topic drift. Thirdly, it posits certain influences on learners' comprehension which have not been previously considered. Fourthly, I attempted to move towards the 'why' of discourse organisation rather than purely concentrating on 'what' or 'how'.

This is an ambitious amount of ground to cover in a single article, and I may be biased in my perceptions of it. Nevertheless, the fact that the article is on the reading lists of at least two Masters programmes at respected British universities suggests that it is not without merit. Despite this support for my views concerning the article, I realise that it has weaknesses, and the need to overcome these is what stimulated me to choose the topic of this present study.

The most obvious weakness is that the article involves the analysis of a single extract of classroom discourse from the perspective of a single method of analysis. If the method of analysis is to be of wider value, it must be applicable to other stretches of discourse, and it needs to be triangulated

both with other methods of analysis and with insider perspectives on the discourse. Other weaknesses in the analysis are specific to the approach and are discussed in more detail in chapter 5. They also, however, need to be addressed.

The purpose of this study, then, is to extend the analysis conducted in Watson Todd (1998), but on a much larger scale involving comparisons with other methods of analysis and with some improvements on the methodology. The two key focuses in Watson Todd (1998) are investigating topics and examining classroom discourse. In this chapter, I will focus on the latter, while topics will be discussed in chapter 2.

Examining classroom discourse is important and useful for two main reasons. Firstly, understanding how participants communicate in classrooms is important for understanding and developing educational practice in general. Secondly, extensive previous research into classroom discourse and the control over the discourse exercised by the teacher mean that language use in classrooms is particularly amenable to analysis. We cannot, however, investigate the entirety of classroom discourse, even within a single classroom, since it is a complicated phenomenon involving a multitude of factors. We therefore need to choose a particular aspect of classroom discourse to investigate. In focusing on a conceptual analysis of classroom discourse, the analysis is likely to be most productive if we concentrate on those parts of classroom discourse where the conceptual metafunction of language plays a particularly important role in communication. One such area, as we shall see, is eliciting. In this study, I therefore intend to examine topics and topic progression in eliciting, basing much of the analysis on the procedures in Watson Todd (1998).

In this chapter, I will provide some of the background for the analysis by discussing the nature of classroom discourse, previous conceptually oriented analyses of classroom discourse, the importance of conceptual or content sequencing in classroom discourse, and justifications for concentrating the analysis on eliciting.

1.2 Classroom learning

The language classroom has been defined as "the gathering, for a given period of time, of two or more persons (one of whom generally assumes the role of instructor) for the purposes of language learning" (van Lier, 1988: 47). As Allwright and Bailey (1991) point out, this is a very broad definition

encompassing groupwork as well as more traditional language teaching contexts. In this research, I will concentrate on the more traditional contexts where one person has been socially assigned the role of teacher.

Perhaps the most important aspect of van Lier's definition is that the purpose of the language classroom is language learning. Despite decades of research into language learning, it is still unclear how the purpose of the language classroom can best be achieved - indeed, it is even unclear whether classroom learning is actually of any real benefit (Allwright, 1989). Attempts to conduct global comparisons between teaching methodologies have been fraught with problems and are ultimately unsatisfactory (see Allwright, 1988), so classroom research has generally become more descriptive and focused at more specific levels. Without a descriptive framework to build on, explanations of classroom events and learning cannot be formulated and thus effective ways of language learning/teaching cannot be identified (see Brumfit, 1995). The research in this thesis, then, is primarily descriptive in that it attempts to construct an analytic method for describing classroom language. It also, however, goes beyond this in that, having set up the descriptive framework, it attempts to see whether this framework can explain some aspects of language learning. To build such a descriptive framework and to use it to explain aspects of language learning is simply not possible if we try to tackle the whole of the teaching/learning process. We must therefore choose an area, ideally previously under-researched, to focus on and this chapter aims to justify the choice of area for research in this work.

Another consequence of the lack of information concerning classroom learning is that there is no certain foundation on which to base any research. Instead, we have to posit basic premises about the classroom to provide a foundation. In this research, I will assume that classrooms are social events where the prime purpose of all participants is to construct meaning and knowledge. This is achieved through interaction between participants, and in the traditional classroom much of this interaction is between the teacher (as designated by a wider social role) on the one hand, and the learners on the other. Teacher-learner interaction is largely talk (Lemke, 1989) or the social use of language, and thus, based on these premises, research into the classroom should focus on the classroom discursive practices which aim towards the construction of meaning and knowledge (Moita Lopes, 1995).

1.3 Discourse analysis and the classroom

Discourse analysis is a problematic term in that different authors use it to mean different things. For example, it is sometimes contrasted with text analysis, the latter referring to the analysis of written language while discourse analysis is reserved for spoken language. On the other hand, it can also be contrasted with conversational analysis, which takes an ethnographic approach to spoken language (see Stubbs, 1986 for a fuller discussion). Alternatively, discourse, and by implication discourse analysis, may be variously defined as equivalent to Saussure's concept of *parole*, or as the object of analysis of suprasentential linguistics, or as the communicative functions of language (Holec, 1985). It is therefore important to define discourse before we continue.

In this research I will take a broad view of discourse as encompassing both spoken and written language (and non-verbal language where appropriate), although the focus will be largely on the spoken language used in teacher-learner interactions. I will also consider discourse to be any coherent suprasentential stretch of language.

However, the discourse that I will investigate, while broad in definition, is restricted to the classroom. In taking this situational approach (Berger and Bradac, 1982), the generalisability of the findings to situations other than the classroom will, of course, be much weakened. However, I hope to show that discourse in the classroom is both worthy of investigation in itself and particularly amenable to analysis.

Discourse analysis has been applied to language classrooms in two main ways (Gibbons, 1994). Firstly, it informs the subject matter to be taught. That is, discourse analysis can be used to analyse the target language and the findings used to produce pedagogic materials (see Widdowson, 1979, 1984 for examples). Secondly, discourse analysis can be applied to the language of the classroom, i.e. to the discourse of teacher-learner interactions. This research is concerned with the second of these applications.

Research into the discourse of classrooms has been a very popular and, to some extent, fruitful area, indeed so popular that terms like *classroom discourse* and *classroom research* are widely used, whereas, say, *barroom discourse* and *courtroom research* are restricted to a few articles at most (van Lier, 1989). Given this prevalence of classroom discourse analysis, why is more research needed? One reason, as we have seen, is that descriptions of

what happens in classrooms are still inadequate. This is partly because some areas of classroom discourse have been nearly totally ignored in the research literature.

1.3.1 Aspects of classroom discourse

Language, and particularly discourse, can be classified according to its purpose. The most influential such classification is that of Halliday, who categorised the purposes of language into three functions: the ideational, the interpersonal, and the textual (1970, 1973). The ideational function is when "language serves for the expression of 'content'" (Halliday, 1970: 143); the interpersonal function concerns the establishment and maintenance of social relations; and the textual function involves the links within language such as cohesion. This model has been criticised by Widdowson (1984) who argues that the textual function and the interpersonal function are both "features of the communicative function of language" (p. 71). Thus Widdowson classifies the purposes of language into two functions only: the conceptual, analogous to Halliday's ideational function, and the communicative. These two functions also correspond to Brown and Yule's (1983a) transactional and interactional categories and to Lyons' (1977) descriptive/social-expressive classification.

Applying Widdowson's model to classroom discourse, we could focus our analysis on either the conceptual or the communicative functions of language. Whichever we choose, it should be stressed that the focus is in the analysis, not in the language. Most language use serves both conceptual and communicative functions simultaneously and both play vital roles in communication. Choosing to focus on one, then, is a matter of selecting a method of analysis.

In discourse analysis generally, and especially in classroom discourse analysis, most researchers have chosen to investigate the communicative function of language. The two most influential studies of classroom discourse (Mehan, 1979; Sinclair and Coulthard, 1975) are attempts to describe the communicative function of classroom language. The extent of the influence of communicative approaches to classroom discourse analysis is such that terms like *discourse structure* and *lesson structure* are taken to mean communicative structure, usually in terms of Halliday's interpersonal function (e.g. Ellis, 1994; Gibbons, 1994; Marland and Osborne, 1990; Richards and Lockhart, 1994; Thompson, 1994; van Lier, 1988; Wong-Fillmore, 1985; D. Woods, 1996).

The lack of research using conceptual approaches to classroom discourse is in many ways surprising. Content, the focus of conceptual approaches, is after all a vital component of every interaction (van Lier, 1988), except possibly for phatic communication (though even here Gardner (1987) argues that content is important). Content, then, must be crucial in classroom discourse, which we are viewing as teacher-learner interaction, and warrants investigation.

Specific to the classroom, several authors have designed models to classify factors in classroom discourse, similar to Halliday's and Widdowson's models describing language in general. Sinclair and Brazil (1982) identify three purposes in classroom language: the subject matter of lessons, the organisation of lessons, and discipline; Johnson (1990) presents a model for analysing classroom language comprising three aspects, one of which is the pedagogical which includes the informative mode to cover conceptual purposes of language; Brumfit (1994) includes topic/content as one of six basic categories for the analysis of language teaching; and Allwright and Bailey (1991) argue that topic is one of five aspects of classroom interaction. All of these classifications include a category designed to account for Widdowson's conceptual function of language. To provide a full description of classroom discourse, then, models must consider the conceptual function of classroom language, or the content or topics used in the teaching/learning process, and so research in this area is sorely needed.

1.3.2 Previous conceptual analyses of classroom discourse

Given the importance of the conceptual function in classroom language, we should not be surprised that there has been some research into the content of classroom discourse. What is surprising is that serious attempts to investigate the content of classroom discourse are few and far between. In this section, I will briefly examine these attempts.

Although topic was included as one category of analysis in many of the schemes of Interaction Analysis prevalent in the sixties and seventies (see Allwright and Bailey, 1991; Long, 1983; Malamah-Thomas, 1987; McDonough, 1981), it was often treated as an afterthought. The main findings of these schemes concerned the communicative function of classroom language. With regard to content, the frequencies of the communicative aspects of language were mapped against the topic as one way of explaining how these communicative aspects arose. No systematic

attempts were made to describe or explain the conceptual function of language through these schemes since the researchers were more interested in the communicative function.

Still concentrating on the communicative function of language, but with a greater emphasis on topics and investigating the language classroom, is Allwright (1980). In this paper, Allwright examined classroom discourse in terms of turns, topics and tasks. For our purposes, it is the section entitled "A topic analysis of classroom turn-taking behavior" (p. 174) that is of interest. From this heading we can see that, as with the Interaction Analysis schemes, the conceptual function of language is being examined to explain a communicative aspect of interaction. However, the findings about topic are of interest in themselves. Allwright used four categories of topics in his analysis:

> "M instances of the target language intended primarily (if not exclusively) as 'models' ...
> I instances of communication concerned primarily (if not exclusively) with information ... about the target language and/or about instances of it ...
> P instances of communication concerned primarily (if not exclusively) with pedagogical/procedural matters.
> O any other ... use of language"
>
> (p. 174)

This classification is unsatisfactory in that category O, the most content-oriented category and presumably set up as a catch-all category, was by far the most frequent. This finding further emphasises the need to investigate the conceptual function of language in the classroom. In addition, contrasting relationships between turn-taking and, respectively, language-oriented topics (M and I in Allwright's analysis), procedural topics (P) and content-oriented topics (O) suggest that there may be differences in the ways language-oriented and content-oriented topics are manifested in classroom discourse. This, in turn, suggests that language classrooms could be a productive situation in which to investigate the conceptual functions of teacher-learner interaction.

Still greater stress was placed on topic by Hudak (1987). He was interested in the power relations between teachers and learners, and specifically in how teachers use their power to control topics in the classroom. By taking an

ethnographic case study approach in an American high school mass media class, he charts the formulation and maintenance of a topic by the teacher and challenges to the topic by students through a lesson. However, his ultimate concern is not with the conceptual notion of topic but with the communicative notion of power relations. Thus topic is only investigated as a means to understand the end goal of power in the classroom. A study similar to Hudak's and more relevant to the situation examined in this thesis is Moita Lopes (1995) who conducted an ethnographic microanalysis of a Portuguese reading class in Brazil. Again, however, he is more interested in teacher power than in the lesson topics themselves.

The first serious conceptually-focused examination of topics in the classroom is Lemke (1989). Investigating science classrooms, Lemke argues that classrooms combine activity structures and thematic systems. Activity structures serve the communicative function of language, so that a classroom lesson and a textbook chapter covering the same content have little in common concerning activity structures. Thematic systems, or "systems of relations among ... themes" (p. 10) (it should be noted that, although Lemke does not define *theme* in this work, it appears to be referring to discourse topic), can show how meaning is constructed in terms of content, and so describe the conceptual function of language. In conducting a thematic analysis, Lemke identifies themes, draws them into a diagram and connects the themes by lines to show relationships. This provides valuable data concerning how topics are developed within a lesson, but suffers two major drawbacks. Lemke relies on intuition in identifying themes and their relationships, and there is no attempt to show that what Lemke identifies as themes are what the discourse participants, i.e. the teacher and the learners, consider to be the themes of the lesson. Lemke even admits at one point that readers would probably not have noticed one key relationship in the thematic system he constructs. Lemke's work, then, is important in suggesting an overall framework for analysing the content of classrooms (identifying themes and their relations, and drawing these into diagrams) but weak in terms of the method of analysis.

Building on Lemke's work in describing the progression of content in lessons but with a more rigorous analysis is my work on English for Science and Technology (EST) classrooms (Watson Todd, 1997a, 1998). The first of these is a preliminary article exploring the possibilities of using logical relations such as implication to identify the relations between intuitively identified themes. The arguments concerning this will be presented in detail

in chapter 5. For the moment, it is sufficient to conclude that the analysis is promising. However, this article still ignores the problems of identifying themes.

In the second article (Watson Todd, 1998), I attempted to provide a more rigorous approach both to theme or topic identification and to the identification of relationships between topics. Topics, it is suggested, can be identified based on theme-rheme analysis, lexical networks and schemata. In addition to the logical relations used in the earlier article, metadiscoursal markers and the ways in which topics typically progress in conversation may, it is argued, facilitate the identification of relations between themes. While hinting at valuable insights into classroom discourse and suggesting key characteristics of clearly sequenced classroom discourse, the article analyses a single transcript only and lacks triangulation to gain the perspectives of the discourse participants.

To summarise, although the method of analysis in Watson Todd (1998) appears productive and valuable, further work is needed, both to iron out problems in the analysis and to collect a reasonable quantity of triangulated data from which it is possible to draw stronger conclusions. Findings from such work could begin to fill the large gap in the research literature concerning the conceptual function of classroom discourse.

1.3.3 Characteristics of classroom discourse

I hope to have shown that one reason for investigating the conceptual function of language in the classroom is the lack of previous research into this area. Another reason is that classrooms are particularly amenable to discourse analysis.

Classroom discourse differs from conversational discourse in several ways. For example, Alpert (1987) claims that the three-stage sequence in exchanges identified by Sinclair and Coulthard (1975) and Mehan (1979) is peculiar to classrooms (but cf. Tsui, 1994), that turn allocation patterns are different, that display questions are used more frequently in the classroom, and that the use of informal language is less frequent. Sinclair and Brazil (1982) argue that the responsibilities of participants are different, in that learners are restricted in the verbal functions they can perform and the teacher fills silences, speaks most of the time and determines the nature of the discourse. In this section, I will focus on the last of these - the power differential present in most classrooms in favour of the teacher.

The difference in power between the teacher and the learners, Kress (1989) argues, provides the driving force behind the institution of education, allowing the teacher to control and structure classroom interaction for the successful execution of learning (Biao, 1996a, 1996b). Whether the teacher's power derives from the socioculture of the classroom or needs to be earned (Cothran and Ennis, 1997), and however the teacher's power is manifested (see Froyen, 1993; Wright, 1987), it has been argued that this power is so crucial to the success of education that attempts associated with the Communicative Language Teaching movement to replace asymmetrical classroom discourse with the symmetries of conversation should be disregarded (Cullen, 1998).

The teacher can use his/her power in many ways to influence the teaching/learning process. For example, Wajnryb (1992) lists 23 areas in which the teacher can choose to take control. While most of these areas are concerned with the communicative function of language, the teacher's control over topics in the classroom shows his/her power concerning the conceptual function of language. Teachers determine and introduce the topic of subsequent interaction (Coulthard, 1977; Hatch and Brown, 1995; Hatch and Long, 1980; Malcolm, 1991; Stubbs, 1983); they ask questions which restrict the possible content of learners' responses (Kress, 1989); they decide whether learners' responses are relevant and reformulate 'inconvenient' responses to match their own agenda (Hatch, 1992; Hatch and Brown, 1995; Johnson, 1995); and they provide summaries of topics covered (Hatch and Brown, 1995). Hudak (1987) and Moita Lopes (1995), as we have seen, contain studies of lessons which illustrate the surprising lengths to which teachers will go to impose their agenda on learners. In fact, the importance of teacher power is such that it can be considered problematic when students try to take control of the classroom topics. For example, Nunan and Lamb (1996) identify the major problem in one piece of classroom interaction as "the teacher allows the agenda to be taken over by the student" (p. 73).

Teacher's control over the content of classroom interaction provides one great advantage for the researcher. As Sinclair and Coulthard (1975) put it in their highly influential study of classroom discourse:

> "we decided ... to begin ... with a more simple type of spoken discourse, one which has a more overt structure, where one participant has acknowledged responsibility for the direction of the

> discourse, for deciding who shall speak when, and for introducing and ending topics. We also wanted a situation where all participants were genuinely trying to communicate, and where potentially ambiguous utterances were likely to have one accepted meaning. We found the kind of situation we wanted in the classroom."
>
> (p. 6)

The high level of teacher power in the classroom therefore creates a simpler discourse situation than that of, say, everyday conversation. Although teacher power facilitates investigations of the classroom, the learners must not be ignored and any realistic system of investigation would have to be able to cope with misunderstandings, corrections and occasions when learners take control (Smith and Holdcraft, 1991; van Lier, 1988). Nevertheless, classrooms do provide a situation amenable to investigation, which make analyses of classroom interaction and topics productive and valuable.

1.4 Sequencing in classroom discourse

Having seen that research is needed into the conceptual function and that classrooms are an appropriate place to conduct such research, let us now consider what aspects of the conceptual function we should investigate. We could, for example, examine how topics are introduced, the summaries of topics which the teacher provides, or the topics which teachers identify as most suitable to teach. However, in this research I shall focus on the sequencing of content in classroom discourse, paying special attention to topics and topic progression.

Although a minimum account of discourse sequencing would need to include clause structure, clause subordination, genre and text-type in addition to content sequencing (Jonz, 1989), the role that content plays in the organisation of discourse macrostructure in particular is considerable. The sequencing and organisation of discourse is an important consideration in creating discourse and has a great impact on how discourses are understood, so how content is sequenced will affect how the discourse is formed and comprehended.

The importance of coherence in discourse sequencing (or the lack of it) has been stressed in many disparate situations of language use. Within the productive skills, coherently sequenced discourse is viewed as a vital responsibility of the speaker or writer (Brown and Yule, 1983b; Duranti, 1991) which is reflected in the stress on organisation given in "how to be a

public speaker" books (e.g. Fletcher, 1985). For the receptive skills, discourse sequence is instrumental in determining the ease of comprehension of a stretch of discourse (Jonz, 1989).

In the classroom, where the goal is the construction of meaning and knowledge, the structure of discourse may be even more important than in other situations. As we have seen, most work on classroom discourse structure has concentrated on the communicative function of language. However, several educational psychology models stress the need for coherent classroom discourse sequencing with regard to content.

Three of the most influential educational psychologists are Ausubel, Bruner and Gagné, and all of them emphasise the need for coherent content sequencing. Ausubel (1963) argues that learning does not consist of random acquisition of pieces of knowledge. Instead, learning involves seeing relationships between pieces of knowledge and fitting new knowledge into an organisation or cognitive structure. Ausubel therefore sees the main function of education as the organisation of knowledge, which can be achieved through the sequencing of content both at the micro-level of lesson stage and at the macro-level of syllabus.

Similarly, Bruner (e.g. 1960) argues that each subject area has a conceptual structure consisting of key concepts that define that subject. The existence of these structures implies that education must organise information about topics, divide information into various categories and show the relationships between these different categories of information.

Robert Gagné (e.g. 1985) also argues that relationships among sets of information are crucial to learning. He goes on to link the internal learning processes with "instructional events" that promote learning (pp. 246ff.), and these highlight the need for structuring the content of classroom interaction.

The importance of coherent structuring of content for learning is reflected in how teachers think about content and in how they are exhorted to teach. Teachers' thinking both about the content of lessons and about the teaching/learning process tends to be organised into highly structured schemata (see e.g. Strahan, 1989; Winitzky, 1992; Winitzky et al., 1994) suggesting that coherent content structure in classrooms is as important for teaching as for learning. In addition, numerous textbooks for teachers (e.g. Anderson, 1989; Arends, 1989; Brown and Wragg, 1993; Cole and Chan,

1987; Eggen and Kauchak, 1992; Gianelli, 1997; Mason, 1994) instruct teachers to explicitly describe relationships between ideas, clearly organise the content, present material in small steps in a coherent order with logical linking, and sequence material according to the accepted logic of the discipline. While such advice to teachers may facilitate the construction of knowledge in the classroom, it is left to the teacher's intuition to decide what the relationships are, how content can be clearly organised, what a coherent order is, and what a logical sequence is.

Research in this area is also rather unhelpful for teachers worried about their organisation and sequencing. Most of the research (e.g. Dunkel and Davis, 1994; Erickson, 1982; Pica and Long, 1986; Tyler, 1992; Tyler and Bro, 1992) has focused on the usefulness of metadiscoursal markers, but this still leaves the teacher to intuitively decide the relationship among ideas and thus what metadiscoursal marker is most appropriate. Other research (e.g. Gibbons, 1994; Mohan, 1986) has concentrated on how relationships between concepts can be depicted, but again, while these depictions are helpful in understanding identified relationships, how the relationships are to be identified in the first place is unclear and apparently left to intuition.

The picture which emerges from the literature, then, is a worrying one. Again and again, the usefulness, value and importance of coherent classroom discourse sequencing is emphasised, but each time decisions about which discourse sequences are coherent and which incoherent are left to intuition.

1.4.1 Teacher clarity

A further area of educational research, namely teacher clarity, highlights the need for coherent discourse sequencing. Research into teacher clarity has emerged as a key dimension of the broader research area of effective teaching.

Research into effective teaching lies in the domain of teaching as science (P. Woods, 1996), and assumes that certain teacher competencies which are more likely to lead to effective learning, irrespective of the specific classroom situation, can be identified. Methods of identifying effective teacher competencies include comparing a range of competencies with learning outcomes (e.g. Brophy and Good, 1986; Rosenshine and Furst, 1971; Rosenshine and Stevens, 1986), conducting surveys of teachers and learners into their opinions concerning effective teaching (e.g. Brown and McIntyre, 1993), and characterising intuitively identified "superior teachers"

(e.g. Finocchiaro, 1989). Whichever method is used, a recurring theme emerging from research into effective teaching is clarity.

The construct of teacher clarity and its antitheses, vagueness and confusion, are still somewhat unclear despite four decades of research. In a detailed review of research on teacher clarity, Cruickshank and Kennedy (1986) report that clarity is variously used to describe clarity of aims, clarity of presentation, clarity of speech and ease of understanding. Nevertheless, taking a broad intuitive interpretation of clarity, Cruickshank and Kennedy's review shows clear relationships between teacher clarity on the one hand and learner achievement or learner satisfaction on the other. Of the 16 papers reviewed by Cruickshank and Kennedy which compared the effects of clarity with those of other constructs, 15 identified clarity as having an important effect on learner achievement or satisfaction.

A further seven papers reviewed attempted to identify the specific instructional behaviours which learners perceive as being clear, i.e. the components of clarity. Of the behaviours identified, three stand out as recurring in several studies: six of the papers mention structuring of content as an important aspect of clarity; four are more specific in mentioning logical or step-by-step structuring; and four mention asking questions. Similarly, in a separate smaller review, Cruickshank (1985) also highlights step-by-step structuring of content.

Although many of the papers reviewed by Cruickshank and Kennedy have been criticised for poor research design and poor use of statistics, the overwhelming consensus in the papers suggests that teacher clarity, and specifically logical structuring of content, is an important, and possibly crucial, influence on classroom learning. However, the lack of clear meanings for the terms used and thus their openness to various interpretations results in difficulties in applying the findings of this research (Williams and Burden, 1997). The final piece of research reviewed by Cruickshank and Kennedy, investigating whether clarity can be taught to pre-service teachers, seems to confirm this. In this research, Gloeckner (1983) compared the clarity of teachers given 10 hours training in clarity with the clarity of untrained teachers and found little difference. Although no details of the training programme are given, the lack of clear meaning of constructs may have been a factor. If you cannot specifically describe what is meant by, say, logical structuring, it is difficult to train teachers in the construct. Again, we find a need for research into what logical structuring of content involves.

1.5 Eliciting

Although it seems clear that conceptual structuring is an important component of effective classroom discourse, how to investigate such structuring is less clear. While conceptual structuring is evident at all times in classrooms, any investigation of the structuring is likely to be most productive if we focus on an area of classroom discourse where conceptual structuring plays an even more prominent role than it normally does. Eliciting is one such area, and in this section I will show what eliciting is and why conceptual structuring is of such importance in eliciting.

Elicit is a problematic term in that, in the literature, it has been used with at least three distinct meanings. Firstly, *elicit* is used to refer to single teacher questions (e.g. Sinclair and Brazil, 1982; Willis, 1981). Secondly, Harmer (1983) appears to use *elicit* with the particular meaning of checking students' ability to produce a new language point prior to an explanation. Thirdly, and most frequently, *eliciting* refers to a series of linked teacher questions aiming to guide learners towards a particular piece of knowledge and thus resembles Socratic dialogue (Chaudron, 1988; van Lier, 1988; Watson Todd, 1997b). It is the third of these that I will be using in this study.

Eliciting as a series of questions can serve several purposes in the classroom, including leading from the known to the unknown (Perrott, 1982), encouraging students to speak (Cross, 1991; Nunan and Lamb, 1996), helping teachers judge what to do next, warming up a class, stimulating cognitive skills, and reactivating existing knowledge (Watson Todd, 1997b). As might be expected given such a variety of potential purposes, eliciting is a relatively common classroom phenomenon especially in classrooms that aim to be communicative.

The sequencing of questions in eliciting has been identified as a "key tactic" of teaching (Brown and Wragg, 1993: 18). Questions should be sequenced so that they produce a scaffolded structure to aid the students' learning (Roth, 1996). It is unclear, however, how such a scaffolded sequence of questions can be created.

Nevertheless, there have been several attempts to explain the sequencing of questions in eliciting. Most of these, however, are based on categorising the questions in an eliciting sequence according to functionally-based question types, such as the categories in Bloom's Taxonomy of Educational Objectives

(see e.g. Arends, 1989). For example, Kerry (1982) argues that a good sequence in eliciting is from lower-order to higher-order questions, while Brown and Edmondson (1984) identify eight different potential sequences based on question type. While useful for teachers, this approach is not helpful in identifying the content sequencing of questions that is the focus of this study.

A second less explicit method of describing the sequencing of questions in eliciting does relate to the conceptual metafunction of language. Orlich et al. (1998) argue that sequencing needs to enable a logical development of ideas. In such a sequence, each succeeding question should take "only one step of logic ... at a time" (Watson Todd, 1997b: 71).

The problem with such advice is that it is unclear how such a logical progression can be developed. While such logical sequencing may involve a sequence of functions, it is also likely to concern conceptual sequencing, and, in Watson Todd (1998), I suggest that this is one key element determining the effectiveness of eliciting. The importance of conceptual sequencing to effective eliciting makes it a potentially valuable area for research. At the same time, the fact that in eliciting, even in language classrooms, the conceptual metafunction of language plays an important role in communication means that eliciting provides a suitable focus for investigation into the conceptual organisation of discourse.

1.6 Summary

Let me briefly reiterate the main arguments made in this chapter:

1. Discourse is viewed as any coherent suprasentential stretch of language.
2. Two functions of language have been identified: the communicative and the conceptual, the latter though important being relatively less researched.
3. Previous research into the conceptual function in classroom discourse has been preliminary in nature.
4. Classrooms are amenable to investigation.
5. Logical coherent content sequencing is crucial to classroom learning.
6. Eliciting provides a particularly suitable focus for research into content sequencing in classroom discourse.

From these arguments I hope I have shown that the conceptual function of discourse warrants investigation, and that a useful focus for such investigation is the ways in which content is sequenced coherently and

logically. To conduct such an investigation, the sequencing of content in classroom eliciting should prove a productive context in which to conduct the research. This study, then, aims to investigate the nature of coherent and logical content sequencing in the context of eliciting in the language classroom.

In order to achieve this aim, we will need to consider the nature of content sequencing in more detail, and this necessarily entails us considering discourse topics, the focus of the next chapter. We will also need to find appropriate data for analysis, and the data collection and preparation procedures are discussed in chapter 3. Given the lack of previous research into topics and content sequencing in the classroom, it is wise to try out several methods of analysis, and these, together with their findings, are presented in chapters 4 and 5. Conducting several different analyses also allows us to compare the findings, and such a comparison is the focus of chapter 6. Finally, the implications of the findings of the research, both for investigations of topic and classroom discourse and for teaching, are discussed in chapter 7.

Chapter 2 Topics and Topic Development

2.1 Defining *topic*

In focusing on the conceptual metafunction of language, we shall be concerned with aspects of language such as *topic*, *coherence*, *relevance* and *aboutness*. Of these, *topic* is probably the one most frequently referred to in the literature but is also the one least frequently defined (Brown and Yule, 1983a).

Indeed, nearly all the previous studies which have had recourse to the concept of *topic* have used it as a pretheoretical notion. Definitions given, such as "subject" (Bygate, 1987: 117) or "whatever it is that is being talked about" (Brown and Yule, 1983b: 62), generally replace *topic* with something equally opaque and so are not helpful.

The single major exception to this apparent vagueness concerning *topic* is the work of van Dijk (especially 1977). In taking a very logically based approach to discourse reminiscent of Boole's (1854/1916) insistence that logical operations underpin the mind, van Dijk defines *topic* from a semantic perspective as a set of propositions. While others (e.g. Crookes and Rulon, 1988) have argued likewise, van Dijk goes on to argue that sets of propositions can be identified as delimiting certain areas or ranges of semantic space. Unfortunately, despite a wealth of seemingly objective logical formulae, van Dijk does not show how it can be ascertained whether two propositions fall into the same range of semantic space (Brown and Yule, 1983a).

The semantic approach to *topic* is only one of several possible approaches. From a pragmatic perspective, we can stipulate that interlocutors must be aware of and identify the same set of propositions as being a topic (Hatch, 1992). Focusing on metadiscourse, topics could be identified as stretches of language marked by lexical (e.g. *by the way*) or phonological (e.g. change in pitch) boundaries (McCarthy, 1991). Alternatively, we could go beyond linguistics and examine how "direct content considerations" (Sacks, 1968 quoted in Coulthard, 1977: 75) may be used to investigate the notion of *topic*.

In this study, I will be taking a semantic approach to *topic*. The pragmatic approach builds upon the semantic approach and adds the refinement of

dealing with different organisations of semantic space between two or more interlocutors and the problem of different interlocutors having conflicting wishes about the topic of communication. As we saw in Chapter 1, by investigating the classroom where the teacher can control the discourse, I hope that these problems can be minimised.

An approach to *topic* based on metadiscoursal boundaries is also subordinate to a semantic approach. While such boundaries may help us to identify topic change, they are not always present at such points in the discourse (Crow, 1983). Furthermore, focusing on the boundaries, as much previous research concerning topics has done (e.g. Hemphill, 1989; Richards and Schmidt, 1983; Stech, 1982), tells us nothing about the actual topic or how it is developed. Metadiscoursal boundaries, then, are incidental rather than defining characteristics of topics.

A content approach is also unhelpful. Some such approaches do not use the notion of topic. For example, in Mohan's (1986) analysis of content, the prime consideration is whether the content concerns specific, practical knowledge or general, theoretical knowledge rather than what the content is about. Other content approaches, such as Bruner (1960) who argues that educational content forms hierarchies, tend to rely on intuition and recourse to 'authority' and so are open to different interpretations.

The semantic approach to *topic*, therefore, seems to be the most justifiable approach. In following such an approach, I will define *topic* in a way similar to van Dijk (1977). However, there are some key differences in my approach.

Firstly, rather than seeing a topic as a set of propositions, I will view it as a set of concepts. In doing this, the topic itself is likely to be expressed as a concept rather than a proposition. This is in direct contrast to Keenan and Schieffelin's (1976) declaration that topics should not be expressed as noun phrases but as propositions, and is more akin to Brown and Yule's (1983a) notion of *topic entity* rather than their interpretation of *topic*. To illustrate how Brown and Yule distinguish between these two terms, an example will help. In analysing an obituary, they identify the topic entity as the person for whom the obituary was written, and the topic as "an appreciation of the noteworthy deeds and events in the life of X" (p. 138). While the topic entity is recoverable from the text unambiguously, the topic can only be identified through extensive application of contextual, cultural and other kinds of

extralinguistic knowledge, and, even then, is not an uncontroversial identification of topic. From an analyst's perspective, then, *topic entity* is a more attractive notion than Brown and Yule's notion of *topic*.

Furthermore, many studies of *topic* (e.g. Coulthard, 1977; Crow, 1983; Gardner, 1987; Mäkinen, 1992; Maynard, 1980; Rost, 1994) have, at least implicitly, used the term *topic* to refer to Brown and Yule's (1983a) *topic entity*. A key reason for this preference for expressing topics as noun phrases, in addition to its intuitive appeal, is summed up neatly by McCarthy (1991).

> "The intimate bond between topic development and the modification and reworking of lexical items already used makes the conversation develop coherently, seeming to move from sub-topic to sub-topic as a seamless whole"
>
> (pp. 69-70)

This strong relationship between lexis and topic is such that key lexical items can be indicative of topics in discourse, further suggesting that topics may be expressed as noun phrases. In this study, therefore, I will be treating *topic* as a set of concepts expressible as a noun phrase.

Even though I will be treating topics as clusterings of concepts rather than sets of propositions, I believe that a distinction between the terms *topic* and *topic entity* may still prove useful. Where I will be taking a hierarchical approach to topics, I will use the term *topic entity* to refer to the key concepts associated with relatively short stretches of discourse which may be subordinate to the *topic* which is associated with the longer stretch of discourse of which the shorter stretches of discourse form only a part. In this way, using Sinclair and Coulthard's (1975) terms, we may talk about the *topic entity* of an exchange and the *topic* of a transaction.

The second way in which this study differs from van Dijk's (1977) approach is the far smaller emphasis it places on logic. In identifying relationships between propositions in discourse, van Dijk relies exclusively on relationships which exhibit logical relations such as entailment. Such a position, however, is problematic. Several authors (e.g. Devlin, 1997; Duncan, 1983; Erickson, 1996; Wittgenstein, 1953) have argued that communication is not rule-governed and that any application of logic in describing language needs to include vagueness (note that this vagueness is

not the same as the fuzziness of fuzzy logic, see Devlin, 1997; Omron Corporation, 1991). Furthermore, a logical approach can only account for semantic relations and ignores the semantic associations which may be more important for interlocutors in seeing connections between concepts. For example, in a recent study of word associations Schmitt (1998) found that *dark* was associated with *light* and *night*, as one might expect, but that *dark* also elicited the associations of *fear*, *ages*, *bench* and *fresh*. While it is easily possible to show a logical relation between *dark* and *light* or *night*, the associations between *dark* and *fear*, *ages*, *bench* and *fresh*, are less easily explainable by logic. We could perhaps argue that *dark* causes *fear* though this is not always so, and that *ages* can have the characteristic of being *dark*, although this appears to be forcing the collocation *Dark Ages* into an inappropriate logical relation. The associations between *dark* on the one hand and *bench* and *fresh* on the other, however, seem to represent more personal relationships between concepts which are not describable through logic. It appears, then, that only some kinds of associations can be explained through logic, while others are less clear or even completely opaque. Ignoring the less explicit relationships would severely weaken any analysis of the conceptual metafunction of language. For this reason, in contrast to van Dijk, I will not take a purely logic-based approach but will take a looser approach accounting for associations as well as logical relations.

The definition of *topic* I will use in this study, then, differs from that of van Dijk (1977) in two-ways while still retaining the centrality of semantics in the study of *topic*. *Topic* is a clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create coherence and relevance. From this definition, we can see two more terms, *coherence* and *relevance*, which require definition.

2.1.1 Coherence

Cohesion and coherence are two terms inextricably linked with discourse. Discourse refers to "stretches of language perceived to be meaningful, unified and purposeful" (Cook, 1989: 156), and it is cohesion and coherence which gives a stretch of language meaning, unity and purpose. While this is widely agreed upon, how cohesion and coherence differ and how they are related are problematic.

Cohesion is generally taken to be a property of texts, and refers to "the mutual connection of components of surface text" (Bell, 1991: 165) or "the overt relationship between propositions" (Widdowson, 1978: 28). Coherence

is far less tangible in that it concerns the ways of connecting discourse which are not overt, and thus, it is argued, exists in how people interpret texts rather than in texts themselves (Yule, 1996b). Because of this, Hoey (1991) argues that cohesion is an objective quality of discourse whereas "coherence is subjective and judgments concerning it may vary from reader to reader" (p. 12; cf. Shakir, 1991). However, when giving a definition of coherence later, Hoey states that "an overwhelming consensus" of opinion concerning the level of coherence of naturally-occurring discourse can be achieved (p. 266). Even if people's subjective interpretations are involved in identifying and interpreting coherence, the overwhelming consensus suggests that coherence is not purely subjective. Coherence requires the utilisation of knowledge not given in the text, and two imaginary people with the same background knowledge should interpret coherence in the same way. The subjectivity of coherence, then, lies not in how it is interpreted, but in what background knowledge is available and how this is used.

To illustrate this argument, let us briefly examine attempts to write computer programs which aim to allow communication with humans. It should be possible to program a computer to identify cohesion in discourse since this is intrinsic to the text. For the computer to identify coherence, however, it would need background knowledge. For texts which are not pre-specified, the background knowledge needed is unpredictable and cannot be programmed (see Devlin, 1997). For a pre-specified text where the requisite background knowledge is identifiable and programmable, a computer could identify coherence. Attempts to write computer programs which can communicate with people therefore have either relied purely on cohesion with no attempt to identify meaning and frequent faux pas (e.g. the ELIZA psychotherapy program) or on cohesion and coherence in a tightly-controlled microworld requiring very little background knowledge (e.g. the SHRDLU program). (For a description of such programs, see Devlin, 1997.)

By distinguishing between cohesion and coherence on the basis of whether background knowledge is needed or not, we have simply shifted the problem of how to distinguish them. How are we to identify when background knowledge is required? For example, if two synonyms appear in a text suggesting lexical cohesion, do we identify that they are co-referents on the basis of their appearance in the text or on the basis of some background knowledge concerning the possibility of their being used as synonyms? To overcome this problem, I am going to suggest that it is helpful to think of

cohesion and coherence as two ends of a continuum rather than as a dichotomy.

Halliday and Hasan (1976), in their classic work, identify five types of cohesion, namely, conjunction, reference, substitution, ellipsis and lexical ties. Since there is a closed set of conjunctions with a limited set of purposes, identifying conjunctions does not usually rely on background knowledge, and is therefore clearly cohesion. Lexical ties, on the other hand, are more problematic. For instance, we need to consider what is a lexical tie. De Beaugrande and Dressler (1981) and Hoey (1991) include recurrence, parallelism, partial recurrence, paraphrase, proforms and ellipsis, while Yule (1996b) takes a broader view by also including "more general connections created by a number of terms which share a common element of meaning (e.g. 'money') *bought - saving - penny*" (p. 141). There is a great deal of variation in the amount of background knowledge needed to identify the different kinds of lexical ties. Exact repetition of a lexical item is apparent from the surface structure of the text and requires little interpretation or application of background knowledge. Because of this, repetition is generally a manifestation of cohesion. Paraphrase, or the use of a synonym instead of repetition of a lexical item, requires an element of background knowledge for a reader or listener to identify which lexical items are paraphrases of which other lexical items. The amount of background knowledge required is not considerable however, and so paraphrases of lexical items probably fall somewhere midway on the continuum from cohesion to coherence. Finally, Yule's consideration of items from the same semantic field as an example of lexical ties requires more background knowledge to identify connectedness. An example of a text exhibiting such connectedness through lexical ties from the same semantic field is the text from Enkvist (1990) given on page 27 below. Identifying the ties between the lexical items in such a text requires a lot of background knowledge and these ties are not apparent from the surface of the text. Such a text exhibits coherence rather than cohesion. Depending on the nature of the lexical ties, therefore, connectedness through lexis may be manifested at different points on a continuum from cohesion to coherence.

Coherence is itself somewhat confusing. Indeed, even distinguished authorities such as Henry Widdowson have confusingly presented arguments concerning coherence based on different principles. In *Teaching Language as Communication*, Widdowson (1978) identifies coherence as a property instilled in discourse by the illocutionary force of utterances. In other words, stretches of discourse, which on the surface lack coherence, can comprise a

series of illocutionary acts (or functions) which together provide the links needed to create discourse. In contrast, Widdowson (1979: 129) discusses two apparently unrelated sentences ("We will have guests for lunch. Calderon was a great Spanish writer.") and argues that a context can be created which would establish links between the two propositions and thus allow the two sentences to be interpreted coherently. In this case, the illocutionary acts performed by the two sentences are not of particular relevance to the establishment of coherence.

To relieve this confusion, Stubbs (1983) argues that there are multiple forms of coherence based on lexis, logical propositional development and speech acts. Coherence based on lexis, at least in Stubbs, is indistinguishable from the lexical ties discussed above which vary in the amount of background knowledge needed; propositional or ideational coherence (Redeker, 1990; Sanders et al., 1992) and coherence based on speech acts or interactional coherence (Lautamatti, 1990) generally require substantial background knowledge. However, examples of what would normally be characterised as ideational or interactional coherence which do not require background knowledge can be found. For example, prototypical adjacency pairs (e.g. question-answer, see Levinson, 1983) may be identifiable and interpretable without recourse to background knowledge. Cohesion and coherence, therefore, should not be viewed as distinct categories but as variations of ways of creating meaningful, unified and purposeful discourse. Most naturally occurring stretches of discourse exhibit connexity arising from both cohesion and coherence, or from several points on the cohesion-coherence continuum.

For our purposes in investigating topics, it is coherence that is likely to be more important, although cohesive devices and cohesive lexical ties will also play a role. Furthermore, between the two main types of coherence, interactional and ideational, only one is likely to play a key role in this study.

Interactional coherence, based on illocutionary force, is perhaps best illustrated by Widdowson's (1978: 29) famous example:

> "A: That's the telephone.
> B: I'm in the bath.
> A: O.K."

Widdowson argues that this stretch of discourse completely lacks cohesion, but that the coherence can be realised by looking at the illocutionary acts which the three utterances perform. Thus, A's first utterance is a request, B's reply provides an excuse for non-compliance, and A then accepts B's excuse.

For our purposes, however, interactional coherence is less important than ideational coherence. Interactional coherence is the more prominent type of coherence in intimate interactions, whereas in more formal settings, such as the classroom, ideational coherence dominates (Lautamatti, 1990; Redeker, 1990). In addition, ideational coherence is likely to be more relevant to our overall search for concept-based topics in the classroom, and it is to this that I shall now turn.

In the same way that coherence can be categorised into interactional and ideational coherence, there are ways of usefully categorising ideational coherence. A distinction can be made between context-based coherence and semantically-based coherence.

Discourse which is coherent on the basis of context is discourse in which the relationships between the propositions and the context provide coherence. In other words, a stretch of context-based coherent discourse may appear incoherent in isolation, but, put in context, coherence emerges. This fact has led to the emergence of a game played between authors of discourse analysis texts. The first author provides two sentences which he or she states cannot be coherently linked; the second author then tries to create a context linking them and disproving the first author (e.g. Widdowson, 1979: 129). This is reminiscent of syntacticians' attempts to provide a coherent context for Chomsky's "Colorless green ideas sleep furiously". The first author will always be on the losing end of the game, "because for any two sets of referents or concepts one can invent a superordinate set that includes them both" (Levinson, 1983: 315). The game also largely misses the point. While it may be possible to dream up contexts for two, three or even four sets of seemingly unconnected propositions, stretches of discourse often consist of larger numbers of propositions and the larger the number the harder it is to think of contexts which create coherence. At some stage a point will be reached where no context can connect the unrelated propositions and the discourse must be considered incoherent. For some discourse, then, where utterances are not illocutionarily or semantically related, it is the context

which provides coherence, and such discourses taken out of context become incoherent.

In contrast to context-based coherence, text grammarians, such as van Dijk, see coherence as "a semantic property of texts" (1977: 93). Such a viewpoint does not exclude the importance of context in determining coherence; rather, it regards context as providing input into semantic interpretations of coherence. Thus, from a semantic perspective, discourses which are coherent when taken in isolation obtain their coherence from their semantic properties, and, for those discourses which are only coherent when placed in context, the context provides the angle from which a semantic realisation of coherence can occur. The following example, taken from Enkvist (1990: 12) illustrates this point.

> "The net bulged with the lightning shot. The referee blew his whistle and signaled. Smith had been offside. The two captains both muttered something. The goalkeeper sighed for relief."

In this example there is no overt cohesion, but we recognise that *net, shot, offside* and so on all belong to the same semantic field, namely, a football game, and are indicative of the topic. By setting the context as the description of a football game, the semantic links between sentences are highlighted and the discourse as a whole can be interpreted as coherent.

The level of semantically-based coherence in a discourse, moreover, can be equated with the distance in logical space between the concepts in the discourse (van Dijk, 1977). Thus propositions, such as *The referee blew the whistle* and *Smith had been offside*, and concepts, such as *net*, *referee* and *offside*, which are close in logical space provide a basis for interpreting the discourse as having semantically-based coherence. On the other hand, some sets of propositions, such as the example below taken from van Dijk (1977: 149), are too far removed in logical space for coherence to be attributed.

> "I bought this typewriter in New York. New York is a large city in the USA. Large cities often have serious financial problems."

In this example, each sentence appears to belong to a different topic and thus, although cohesion through lexical ties is apparent, the concepts are distant and the discourse lacks coherence.

Semantically-based ideational coherence, then, is a property of language which delimits areas of semantic space. Combined with context-based ideational coherence, it shows how texts can be interpreted as hanging together even when there is no cohesion in the text. A topic provides a central point around which texts are coherent or hang together.

2.1.2 Relevance and Aboutness

Two other terms which are closely associated with *topic* and the conceptual metafunction of language are *aboutness* and *relevance*. In this section, I will look at these two terms in relation to different types of topic development. These types of topic development will be discussed in detail in section 2.3 below. For the moment, the following descriptions will suffice. Topic maintenance is where the same topic is retained through a stretch of discourse; topic drift occurs where each succeeding discourse act is connected in some way with the preceding one but where the overall topic of the discourse gradually changes; and topic shift involves a 'jump' between topics where two succeeding discourse acts have no apparent connection.

Stretches of monotopical discourse are by definition *about* one topic. Such aboutness is a semantic concept (Carlson, 1983; Hazadiah, 1993) whereby all the propositions or concepts in the discourse are related or semantically relevant to the same overall superordinate discourse topic. All the concepts in a monotopical stretch of discourse, then, are subordinate to the overall topic, or they are about the overall topic. Thus aboutness refers to a set of concepts all coming from the same delimitation of semantic space.

Relevance is slightly different in that propositions and concepts may be relevant to each other as well as being relevant to the superordinate topic. The idea of relevance stems from the influential work on conversational maxims of Grice (1975). He proposed four maxims, one of which is the maxim of relevance. Sperber and Wilson (1986) argue that this maxim is the key maxim to which the others are subordinate. Although Sperber and Wilson also argue that relevance is not a purely semantic idea, in this study of topics I will be looking mainly at semantic relevance.

Whereas aboutness concerns the relationship between propositions or concepts on the one hand and discourse topic on the other, relevance is involved in both relationships of the same kind as aboutness and relationships among propositions and concepts. Thus relevance differs from aboutness in that propositions and concepts may be relevant to each other as well as relevant to the superordinate topic. Aboutness therefore is a specific restricted kind of relevance.

While monotopical topic maintenance implies both aboutness and relevance, topic drift requires relevance but not aboutness. With topic drift, each proposition or concept is semantically connected to the previous one(s), but the overall superordinate topic changes through the discourse. This semantic connectedness between succeeding propositions or concepts is what makes the propositions and concepts relevant to each other. For the moment, we will not examine exactly how such semantic relevance is created, but I hope that it will become clear by the end of this study.

Topic shift, on the other hand, exhibits neither relevance nor aboutness. The two propositions on either side of the topic shift are not semantically connected to each other. In this situation, then there are no connections of relevance between succeeding propositions or concepts and the overall superordinate topic changes.

It should be noted that the relevance we are talking about here is a relatively local relevance. Even where there is topic shift and the two propositions on either side are not connected to each other, there may still be relevance at a very global level. For example, in the middle of one of the lessons recorded in this study (the source of extract K), the teacher finishes one unit of the book and starts a new unit. As the two units have completely different language and content objectives, we may reasonably expect no relevance to hold between the end of the first unit and the start of the second. However, at the global level of the objectives of the whole course, they are in a marginal way related. This global level of relevance will not be considered in this study. Instead, I will focus solely on the more local relevance.

Relevance and aboutness, then, can be summarised in terms of the extent to which they are exhibited in different types of topic development as in Table 2.1.

	Aboutness	Relevance
Monotopical topic maintenance	✓	✓
Topic drift	×	✓
Topic shift	×	×

Table 2.1 Aboutness and relevance as features of topic development

From the definition of *topic* on page 22, and the explanations of coherence and relevance, it can be seen that there is a certain circularity in their relationships. Topics create coherence and relevance, and yet it is coherence and relevance that allow us to identify topics. The relationships between the three elements are not relationships that allow us to start from one and move to the others in our analysis. Instead, we need to take a more organic approach whereby all three notions are examined together in analysing a stretch of discourse.

What topics, ideational coherence and relevance have in common is a kind of semantic connectedness. In this study, I am following van Dijk's (1977) metaphor of areas or regions in semantic space, and semantic connectedness is created when several concepts all fall into the same delimited area of semantic space. When such a clustering of concepts occurs in the same stretch of discourse, we have ideational coherence and we can say that the discourse is about a certain topic. Furthermore, such clustered concepts have a high likelihood of being relevant to each other and to the topic.

2.2 Identifying topics

Somewhat strangely, in the literature there is far more on how to identify where topics begin and end than how to identify what the topic is between such boundaries. Much of the literature on *topic* (e.g. Bergmann, 1990; Clark, 1996; Covelli and Murray, 1980; Gumperz et al., 1982; Hemphill, 1989; Richards and Schmidt, 1983) in fact does not try to identify any topics. In many studies where the topic is identified (e.g. Crow, 1983; Jefferson, 1984; Maynard, 1980; Rost, 1994; Sacks, 1971 cited in Coulthard, 1977; Shepherd, 1998; Stech, 1982), the identification appears to have been made on purely intuitive grounds. And where a method for identification is suggested, this method may rely on intuition; for example, Carlson (1983) argues that discourse topics can be identified by setting a question which the discourse answers and then identifying the discourse topic as the key point of the question.

The non-intuitive approaches to topic identification assume that topics provide a hierarchical approach to discourse. Such an approach means that the topmost elements in the hierarchy are most likely to be identified as the topic (Brown and Yule, 1983a). We can talk about topics at several levels of the discourse with topics at a global level being superordinate to subtopics at a more specific level (e.g. Hudson, 1980; van Dijk, 1977). Such hierarchies are often viewed as being organised around hyponymy as this is a "relation that holds between a specific or subordinate and a general or superordinate concept" (Cicourel, 1991: 40), although other types of superordinate-subordinate relations and the possibility of non-logical relationships should also be considered.

Taking a hierarchical approach has important implications for identifying topics. It means that we could approach topic identification either starting with the subordinate levels and taking a bottom-up approach or starting at the superordinate levels and taking a top-down approach (Watson Todd, 1998).

For bottom-up approaches, there are several alternatives. Firstly, we could take a theme-rheme approach where we identify themes and look at how these join together to form higher-level themes (Caron-Pargue and Caron, 1991). Secondly, we could focus on lexical chaining with the longest, most frequently recurring lexical chains indicating the topic (McCarthy, 1991). Thirdly, again from a lexical perspective, we could create a network of relations between the lexical items in the discourse and take density of linkage of bonds in the network as indicative of topic (de Beaugrande and Dressler, 1981; Hoey, 1991). Fourthly, we could amalgamate the various propositions of the discourse into superordinate propositions (Van Dijk, 1980, 1985; van Dijk and Kintsch, 1983).

Top-down approaches to topic assume that the interlocutors have expectations about the content and organisation of the discourse. While it may be difficult to justify the existence of clear expectations with respect to informal conversations, other more structured situations may allow clear expectations to be held. For example, in classroom discourse cultures of learning (Cortazzi and Jin, 1996) and the existence of topics fixed before the lesson through the teacher's disproportionate power (Brown and Yule, 1983b) may provide a relatively predictable structure to the discourse enabling participants to know what to expect. Holding expectations about

discourse is closely related to schema theory. Schemata, closely related to scripts (Schank and Abelson, 1977), frames (van Dijk, 1977; Minsky, 1985), expectations (Tannen, 1978) and points (Wilensky, 1986), are background knowledge structures which represent the relationships between components of knowledge (Carrell and Eisterhold, 1983; Anderson and Pearson, 1984; Cook, 1997). Although identifying schemata is problematic, for stretches of discourse where the concepts can be interpreted as filling the expectations of a single schema, that schema should be indicative of the topic.

There are then a variety of approaches to identifying topics and we will look at these in more detail in section 2.4. There has, however, been no research comparing the identification of topics through each approach and such a comparison is one goal of this study.

2.3 Topic development

In the previous section on topic identification, the discussion appears to imply that a single topic can be identified for a particular stretch of discourse. Although it is tempting to view topics as static hierarchies, the truth is that topics are dynamically developed through discourse. Identifying topics, while useful, is not enough in itself; we must also be able to follow how topics develop and change through the discourse.

Topic development is generally a preferred aspect of communication since otherwise there is a danger that interlocutors would simply keep on repeating themselves (Foppa, 1990). However, there needs to be a balance between topic development and topic maintenance (Bergmann, 1990). Too much topic maintenance may lead to repetition, but too much topic development would lead to incoherent discourse that would jump from topic to topic without anything worthwhile being said about any particular topic. In any reasonably long stretch of discourse, therefore, we should expect a mix of topic maintenance and topic development.

The balance between topic maintenance and topic development and the ways in which topics are developed have been the focus of much research, especially from the perspective of conversation analysis. Unfortunately, this has also led to a plethora of terms for describing different ways in which topics develop. Fortunately, these can be categorised into three main categories, two of which have various subcategories. These are described below and summarised in Figure 2.1.

The first category is topic maintenance (Crow, 1983), also termed topic continuation (Gardner, 1987) and topic sustained (Covelli and Murray, 1980). This is where the topic for a stretch of discourse remains constant.

A second category is topic drift (Sacks, cited in Coulthard, 1977; van Dijk, 1977), also called stepwise transition (Sacks, unpublished, cited in Atkinson and Heritage, 1984) and, confusingly, topic shift (Gardner, 1987; Hurtig, 1977), where topics imperceptibly blend into each other without any clear break. Topic drift can be subcategorised into three types. Topic shading (Hurtig, 1977) involves the expansion of the domain of the original topic; Push (Clark, 1996) involves moving from the topic into greater consideration of a subtopic; and topic fading (Hurtig, 1977) involves a drift to a new topic where at least one proposition or concept provides a link between the two topics (see Watson Todd, 1997b).

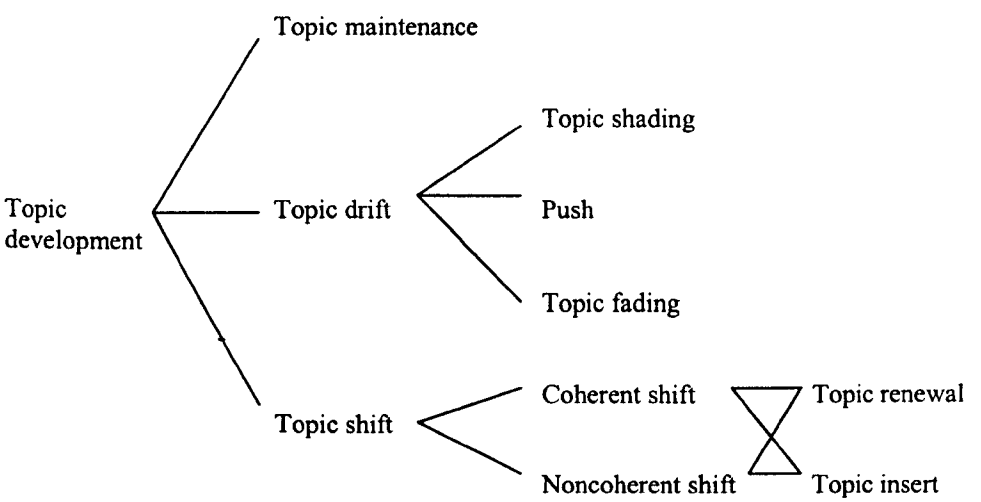


Figure 2.1 Summary of types of topic development

While topic maintenance and topic drift can be classified as continuous discourse (Keenan and Schieffelin, 1976) since there are no explicit breaks in the discourse, topic shift (Crow, 1983; Evensen, 1990; Stech, 1982) produces discontinuous discourse (Keenan and Schieffelin, 1976). Topic shift has also been termed boundaried transition (Sacks, unpublished, cited in Atkinson and Heritage, 1984), topic change (Gardner, 1987) and topically disjunctive continuation (Jefferson, 1984). It involves a jump from one topic to another with no apparent link between the two topics (except perhaps the very global relevance discussed above). Such shifts can be coherent in which

case they are clearly signalled with a metadiscoursal marker or noncoherent when there is nothing in the surface language to indicate the shift (it should be noted that Crow's (1983) use of *coherent* here is more akin to how I am using *cohesive* in this study than to my use of *coherent*). Topic shift may also involve a jump back to a previous topic of the discourse in which case it is termed topic recycling (Gardner, 1987) or topic renewal (Crow, 1983). Alternatively, it may involve a jump to a new topic which acts as a brief interlude before the previous topic is continued. In this case, it is termed topic insert (Crow, 1983), insertion sequence (Coulthard, 1977) or Digress and Return (Clark, 1996).

Nearly all of the above research has been conducted with general conversation. The applicability of the classification of topic development to other types of discourse is uncertain. This uncertainty is strengthened by Maynard (1980) who argues that there is a clear qualitative difference between narratives and topical conversational discourse.

In classrooms, eliciting, the genre studied here, is unlikely to take the form of a narrative. Rather, it is more likely to take the form of general-to-specific topical discourse (van Lier, 1988; Watson Todd, 1997a, 1997b). Furthermore, although I am not aware of any previous studies into topic progression in classroom discourse (excepting Watson Todd (1998) which supports the use of the classification of topic progression described above and is described in greater detail below), there is some circumstantial evidence that the types of topic progression identified above also apply to classroom discourse. For example, Chaudron (1983: 136) discusses teachers "simplifying their maintenance of a continuing topic, or their announcement of a new topic or subtopic". Although a few quotations like this are far from conclusive, they suggest that the types of topic progression identified in conversation might also be found in classroom discourse. Let us now turn, then, to briefly examining the approaches that are most likely to be useful in identifying topics and topic progression in classroom discourse.

2.4 Approaches to topics and topic progression in classroom discourse

Studies specifically aiming to analyse topics and topic progression in classroom discourse are few and far between. Instead of looking solely at these studies (although they will be considered), a more productive approach, and a potentially more generalisable one, is to apply analyses of

topics designed for other types of communication to the classroom. Even this approach, however, is limited since there are few approaches to topic which do not rely on the vagaries of intuition. Most of the approaches described in this section, therefore, were not originally designed to analyse either topics or classroom discourse. They are nevertheless approaches that should be applicable to nearly any kind of discourse and that may lead to insights into the nature of topics and topic progression.

2.4.1 Metadiscourse and topic shift

As we saw above, topic shift may be coherent in which case it is signalled by a metadiscoursal marker or noncoherent. Coherent topic shifts are preferred and more frequent (Crow, 1983). If we can identify metadiscoursal markers in classroom discourse, we should also be able to identify the placement of many of the topic shifts in the discourse. Although this will not help us to identify the topics between the shifts or whether the discourse between the shifts exhibits topic maintenance or topic drift, it does allow us a strong foundation for identifying breaks between topics against which other less certain methods can be evaluated.

The main function of metadiscourse is to organise discourse rather than to add propositional content (Connor, 1994). It is manifested primarily by lexical phrases which can be termed metadiscoursal markers, discourse markers (Schiffrin, 1987), language tactics (Minsky, 1985), pointers to superstructure (Evensen, 1990) and metapragmatic signals (Flowerdew, 1994). In addition to these lexical indicators of metadiscourse, prosodic features such as tone, intonation and pausing may also play an important role in metadiscourse.

Most work on metadiscoursal markers has looked at conversation and in this context a long list of lexical phrases indicating topic shift has been identified (e.g. Coulthard, 1977; Evensen, 1990; Nattinger and DeCarrico, 1992; Richards and Schmidt, 1983; Schiffrin, 1987). These phrases include "Incidentally" indicating the start of a topic insert, "Anyway" indicating the end of a topic insert, and "Speaking of that" indicating a shift to a new marginally related topic (Clark, 1996). In the classroom, however, the use of these phrases for metadiscourse is either very rare or nonexistent. Instead there seems to be a separate set of metadiscoursal markers used in the classroom.

Classroom-specific metadiscoursal markers are a key feature of the approach of Sinclair and Coulthard (1975), one of the most influential studies of classroom discourse. In Sinclair and Coulthard's model, a lesson comprises several transactions, each of which may be considered equivalent to a topic (Francis and Hunston, 1987). The transactions are typically separated by boundary exchanges. Thus any boundary exchange indicates that the teacher (who nearly always utters them) regards one stage or topic of the lesson as finished and the next as beginning. Nearly every boundary between transactions contains the following structure:

FRAME (FOCUS)

This means that boundary exchanges consist of a framing move usually, but not always, followed by a focusing move (Sinclair and Brazil, 1982).

The framing move is manifested by a closed set of metadiscoursal markers, including *OK*, *right*, *look*, *all right*, *well*, *so* and *good*. These are then followed by a measured pause. The succeeding optional focusing move may include a word referring to the content or method of teaching and a future time reference (e.g. "*Next* we *will* look at *pronouns*.") (Sinclair and Brazil, 1982). By identifying these metadiscoursal markers in classroom discourse, we should be able to identify points where there is a high likelihood of topic shift in the discourse.

One further aspect of Sinclair and Coulthard's (1975) model that may be useful in our study of topics concerns the level below transaction, namely, that of exchange. Coulthard and Brazil (1992), working within the framework of Sinclair and Coulthard's model, argue that "the exchange only carries one (potentially complex) piece of information and its polarity" (p. 74). If these pieces of information are equivalent to concepts, which provide the building blocks of topics, then identifying the key pieces of information in each exchange may allow us to follow topic progression through the discourse. Although such an analysis is far less clear-cut than the analysis of boundary exchanges to identify topic shifts, it may provide a valuable benchmark against which other methods of analysis can be compared.

The first approach to analysing topics and topic progression used in this study, therefore, derives from the work of Sinclair and Coulthard (1975). Identifying framing and focusing moves in classroom discourse should provide useful information concerning the placement of topic shifts in the

discourse. More tentatively, key concepts in the discourse may be indicated by the key pieces of information carried in each exchange.

2.4.2 Network-based approaches to topics and topic progression

Whereas Sinclair and Coulthard's (1975) model specifically aims to describe classroom discourse, other methods not designed with classroom discourse in mind may also be applicable to the classroom situation. One such approach, which may be helpful in identifying topics, involves representing the discourse as a linked network of concepts. The greater the number of links, the greater the strength of relationship between concepts. In such an approach, the higher the number of links from a particular node or concept to other nodes the higher the probability of that concept being a superordinate concept indicative of topic. Centrality of concepts to a stretch of discourse, therefore, is indicated by density of linkage in the network. There are two main models using this approach.

The first of these is a very detailed approach proposed by de Beaugrande and Dressler (1981). Their approach involves looking for links between every content word in a stretch of discourse and analysing the nature of the linkage in terms of 34 different types of relationship, such as *instrument of* and *purpose of*. While this is a very comprehensive approach to identifying topics, it has three main drawbacks. Firstly, since it takes links between all content words into account, the analysis quickly becomes unmanageable. De Beaugrande and Dressler include an analysis of a straightforward descriptive text of 31 words. The network generated from the text has 22 nodes and 21 relationships. Conducting this analysis with any longer stretch of discourse, such as those which are the focus of this study, is nearly impossible. Secondly, despite the amount of detail involved in the analysis, the actual identification of which nodes are linked and the nature of these links seem to rely solely on the vagaries of the analyst's intuition. Thirdly, some of the apparent relationships in the discourse are not identified in the analysis despite the level of detail required. For example, the statement "Empty, it [the rocket] weighed five tons" is analysed as two separate links from *the rocket*, one of which goes to *empty* and the other to *weighing five tons* (p. 105). In this way, the purpose of the original statement in indicating that it is an *empty* rather than a full rocket that weighs five tons is overlooked. These drawbacks in de Beaugrande and Dressler's approach reduce its likely value to this study. Furthermore, although possibly useful for topic identification, the networks produced following the approach provide a very static picture of discourse with no indication of sequencing of concepts, and are thus

worthless in analysing topic progression. For these reasons, de Beaugrande and Dressler's (1981) approach is not used in this study.

The second of the network-based approaches is the lexical approach of Hoey (1991). Instead of looking at all content words in a text as in de Beaugrande and Dressler's (1981) approach, this approach focuses on those lexical items which recur across sentences. Sentences which have more than a certain minimum number of recurring lexical items in common are said to be linked and a network of these linked sentences can be drawn up. Hoey's approach does not have the drawbacks of de Beaugrande and Dressler's, since it is fairly practical to use with long stretches of discourse and the procedures in the approach are relatively objective. Furthermore, density of linkage should still be indicative of topic and the approach may give some information concerning topic progression. Hoey argues that sentences with high numbers of bonds with subsequent sentences are topic-opening, while sentences with high numbers of bonds with preceding sentences are topic-closing. In addition, clusters of succeeding sentences with the same bonds in common indicate topic maintenance. Although it is difficult to follow topic drift through this approach, breaks indicating topic shift and clusters of topic maintenance may allow us to gain some insights into topic progression.

To conclude, network-based approaches where density of linkage allows us to identify topics form one set of potentially useful approaches. Of the two main approaches within this set, the one I will use in this study is that of Hoey (1991). This approach is practicable and relatively objective, increasing its reliability. It also provides indications of both the topic and two of the three types of topic progression, namely, topic shift and topic maintenance.

2.4.3 Hierarchy-based approaches to topics and topic progression

A second set of approaches to topics and topic progression assumes that discourse is organised hierarchically. While I will look in more detail at the appropriacy of taking a hierarchical approach in chapter 5, for the moment it is sufficient to realise that a hierarchical approach usually results in a model of discourse consisting of several nested levels.

2.4.3.1 Analyses at the clause level

The first two hierarchical approaches are based on analyses of components of sentences. One method of analysing sentence components results in topic-comment pairs (Connor, 1994). Whether these sentence-level topics are

equivalent to the lowest level of subtopics in discourse is unclear (see Brown and Yule, 1983a; Carlson, 1983; van Lier, 1988 for fuller discussions). Their position as the starting point of the sentence, however, suggests that an analysis of sentence topics may prove productive.

The topic-comment dichotomy, however, conflates two other dichotomies, namely, theme-rheme and given-new. The theme of a sentence is "what the sentence is about" and the rheme is "what is said about [the theme]" (Connor, 1996: 81), and both are often identified as syntactic units. Given and new, on the other hand, refer to the information contained in the sentence. Given information is that information which the speaker/writer thinks the listener/reader already knows, with new information being the self-explanatory opposite.

For both theme-rheme and given-new, it is possible to trace the progression of components of clauses through the discourse. For example, we might find that the theme of one sentence is repeated as the theme of the next sentence suggesting topic maintenance at a very local level, or we may encounter a theme which has never previously been used in the discourse suggesting potential topic shift. We may also find similar patterns of given information. Furthermore, both of the clause-level dichotomies can be considered as forming hierarchies of superordinate themes or superordinate given information respectively. These superordinates may be indicative of topics.

It is, however, unclear which of these two dichotomies is more likely to be helpful in our study of topics. Although theme often coincides with given information and rheme with new information (Halliday, 1970), the differences between the two probably make it worth conducting both analyses. In this study, therefore, I will conduct two analyses based on the clause level, one for theme and rheme, and one for given and new information.

2.4.3.2 Proposition-based hierarchies

The most fully developed approach to a proposition-based hierarchy of discourse involves the search for macrostructures (van Dijk, 1977, 1980, 1985; van Dijk and Kintsch, 1983). As with much of discourse analysis, macrostructures fall into two categories associated with the conceptual and the interpersonal metafunctions of language. Semantic macrostructures are analogous to global content schemata (Graesser et al., 1997), may be equivalent to topics of discourse (van Dijk, 1977), and focus on the

conceptual metafunction of language. Pragmatically-oriented macrostructures or macro-speech acts (van Dijk, 1977), on the other hand, investigate the communicative metafunction and are more closely associated with genre analysis and global formal schemata (Paltridge, 1995). As I am investigating the conceptual metafunction of language in this study, I will focus solely on semantic macrostructures.

Macrostructures work on the principle that discourses are hierarchically organised, and that superordinate propositions are more important than subordinate propositions in both the production and the comprehension of discourse. Identifying these superordinate propositions, van Dijk (1977, 1980) argues, is analogous to identifying topics.

In order to identify the superordinate propositions, the discourse is viewed as being an instantly conceived and perfectly formed product on which various procedures are performed to reverse the process of discourse production. An example of such a procedure or macrorule is deletion (van Dijk, 1980: 82-83), which is defined as follows:

> "Given a sequence Σ of propositions $<p_i, p_{i+1}, \ldots, p_k>$ of a text T, satisfying the normal linear coherence constraints, substitute Σ by a sequence Σ' such that each $p_{i+j} \varepsilon \Sigma$ that is not an interpretation condition (presupposition) for at least one proposition of T does not occur in Σ', whereas Σ and Σ' are further identical."

Applying a macrorule like this reduces the discourse to its most important components. Other macrorules, such as generalization and construction, also concentrate the content of the discourse into fewer propositions. By applying these macrorules recursively, the superordinate or macro propositions identified from the first application of the macrorules can be combined into macro-macropropositions until one highest-level proposition for the whole discourse is found. The output of a complete analysis of a stretch of discourse, then, is represented as a hierarchical tree of propositions.

The key problem with van Dijk's macrostructure approach, and the reason that I will not be using it in this study, is that it assumes that discourse production starts from a single superordinate proposition around which other propositions are woven to form a stretch of discourse. As such, the approach takes a completely static view of discourse that ignores the process of discourse production and the influence of anything other than superordinate

topic on the final discourse (Connor, 1990; de Beaugrande and Dressler, 1981).

Furthermore, despite devoting one whole book (van Dijk, 1980) and substantial parts of two other books (van Dijk, 1977; van Dijk and Kintsch, 1983) to the study of macrostructures, as Ventola (1987) points out, there is a lack of convincing practical demonstration of macrostructures. In van Dijk (1980), for example, most of the texts analysed are only two sentences long with few attempts to analyse any longer texts.

Although a brave attempt to objectively identify topics in discourse as superordinate propositions, I believe that the approach fails because of the unreasonable assumptions it makes about the nature of discourse. I will not, therefore, use van Dijk's macrostructure approach further in this study.

2.4.3.3 Keyword hierarchies

None of the previous approaches considered have specifically aimed at topics in classroom discourse. Some (e.g. analyses at the clause level) have very rarely been applied to classroom discourse, while another (Sinclair and Coulthard's (1975) model) does not specifically aim at describing topics. The last approach I will consider is the only approach specifically aimed at topics in classroom discourse that I am aware of, apart from that of Lemke (1989) which relies heavily on intuition; Hazadiah's (1993) study, despite the word *topic* in the title is actually an argument for an extra level in Sinclair and Coulthard's (1975) functional hierarchy. This approach is my own (Watson Todd, 1998, but see also 1997a).

In this approach, keywords are identified on the basis of frequency. The logical relations between these keywords are identified and line diagrams representative of schemata are drawn up to show these relations. The sequence in which the keywords appear in the discourse is then mapped onto the line diagrams.

Topics are identified as those concepts (taken in the article as expressed by keywords) which are most frequently linked to other concepts in the diagram, and the types of topic progression can be identified by comparing the logical relations in the line diagrams with the sequence of occurrence of the keywords in the discourse. In this way, Watson Todd (1998) is the only approach that unequivocally attempts to deal with topic progression as well as topic identification.

The approach, however, is not without problems. The main problem is whether showing connections between concepts as logical relations is valid or not (see chapter 5 for details). The definition of *topic* that I am using in this study is clustering of related or associated concepts. While my 1998 study focused solely on related concepts, in this study I intend to consider both related and associated concepts. Therefore, in addition to conducting an analysis based on logical relations, I propose to conduct a similar analysis based on associations derived from a corpus.

2.5 Conclusion

In this chapter, I have proposed that topics are clusterings of concepts which are associated or related from the perspective of the interlocutors in such a way as to create coherence and relevance. To investigate topics, we need to be able to both identify topics and follow topic progression through the discourse. To do this, I am proposing to use six methods of analysis with classroom discourse the focus for analysis. Breaks indicative of topic shift can be identified following Sinclair and Coulthard (1975). Topics may be identified as superordinate themes, superordinate given information, or densely linked nodes in a lexical network representing the discourse. Topics may also be identified as most frequently linked nodes in either relation-based hierarchies or association-based maps of discourse, and types of progression in the discourse can be followed by mapping the discourse onto these representations. These six approaches are presented in chapters 4 and 5. Before we conduct the analyses, however, we first need to examine data collection and preparation methods.

Chapter 3 Data Collection and Preparation

In the previous two chapters, I have suggested a potential approach to analysing topics in classroom discourse. More specifically, I have proposed six methods of analysing eliciting transactions: Sinclair and Coulthard's (1975) functional analysis of discourse, Hoey's (1991) lexical network analysis, analyses of theme-rheme and of given-new, and analyses of relations and of associations following Watson Todd (1998). In order to conduct such analyses, extracts of classroom discourse need to be collected. Furthermore, the extracts need to be transcribed, certain aspects such as the referents of any referring expressions need to be clarified, and the discourse needs to be divided into units. Finally, insider perspectives on the discourse need to be collected to provide triangulation by looking at multiple perspectives on the data. The purpose of this chapter is to explain how the data collection and preparation were conducted. The detailed procedures of each method of analysis are given in chapters 4 and 5 together with the results of the analyses.

3.1 Data Collection

3.1.1 The situation

For reasons of ease of access and cooperation, the data to be analysed were collected from a foundation English course at King Mongkut's University of Technology Thonburi (KMUTT), a technological university in Thailand. KMUTT is a respected government university attracting high-quality students of science and technology.

All students are required to study at least two sixty-hour courses of English with most taking three or four courses. The first course, *Fundamental English for Science and Technology*, is based around six units of *Interface* (Hutchinson and Waters, 1984), and is the source of the extracts analysed in this study.

Originally, it was planned to look at five groups of students with five different teachers. At the beginning of the semester, these five teachers were asked to choose three lessons in the course where they believed eliciting would play a major role. The most frequent of their responses were chosen as the three lessons to focus on in this study. The first was the introductory lesson for Unit 3A: Engine Types of Hutchinson and Waters (1984: 28-31). The second was the introductory lesson for Unit 4A: Robots (*ibid.*: 40-43).

The third was the introductory lesson for Unit 8B: Pumping Systems (*ibid.*: 92-95). The whole of each lesson was recorded.

3.1.2 The teachers

Although it was originally hoped to look at five teachers, only four teachers were recorded for the three lessons described above, because of timetable clashes. All four were Thai with a good level of English competence, a Master's degree in education or linguistics, and at least one year's experience in teaching the course. For the purposes of this study, the teachers are referred to as teachers A, B, C and D. Teachers A and B are male, and teachers C and D are female.

3.1.3 The students

The class size varied from 22 to 37 students. Each class comprised a group of students studying for the same degree, such as chemistry or mechanical engineering. Ages of students ranged from 17 to 22, with a mix of males and females (although teacher B's class were all male). The majority of students were lower intermediate in terms of English proficiency, and were generally hard-working and willing to cooperate. Most, however, were not prepared to take risks in speaking out in the class.

3.1.4 The recordings

The lessons used in this study were conducted in a room with two video cameras in the ceiling. This allowed both the teacher and the students to be recorded and viewed simultaneously. A microphone was placed near the teacher at the front of the class.

Using this system, there were no problems with the recording of 11 of the lessons. For one lesson (teacher D, lesson for Unit 4A), however, technical problems meant that the first two minutes of the lesson were not recorded, and thus part of the eliciting transaction is missing. The part that was recorded, however, can be analysed.

3.1.5 The extracts

Although the whole lesson was recorded, only part of the lesson was used for analysis. As we saw in Chapter 1, eliciting transactions (Sinclair and Coulthard, 1975), or transactions taking a question-and-answer format in which the teacher aims to elicit information from the students, may provide data particularly amenable to the analysis of the conceptual or ideational metafunction of language, and eliciting usually has a target predefined by the

teacher. Teachers were asked when they were going to start eliciting and what the target was. In 11 of the 12 lessons recorded, teachers said they would start eliciting at the beginning of the lesson. For the twelfth lesson (teacher C, lesson for Unit 8B), the teacher expected to start a new unit with eliciting halfway through the lesson, but also expected that this would be clearly signalled. The extracts chosen for analysis, then, were those parts of the lesson from the start of the lesson/unit to the point where the predetermined target was reached.

There were then 12 extracts collected in total, three for each of four teachers. For convenience, these extracts are labelled as shown in Table 3.1.

Extract	**Teacher**	**Unit of Hutchinson and Waters (1984)**	**Predefined target of eliciting**
A	A	3A	kinds of engines
B	B	3A	petrol engine
C	C	3A	kinds of engines
D	D	3A	kinds of engines
E	A	4A	robots
F	B	4A	robots and people at work
G	C	4A	robots and people at work
H	D	4A	robots
I	A	8B	the heart
J	B	8B	the heart
K	C	8B	the heart
L	D	8B	the heart

Table 3.1 Labelling of extracts

The extracts total 7,555 running words, with an average length of 630 words and a standard deviation of 455 words. The longest (extract A) is 1,774 words and the shortest (extract H) is 165 words. To illustrate the analyses used in this study, a single extract is used as an example for analysis throughout chapters 4 and 5. This extract is extract G which is 520 words long and is given below. This extract was chosen as being typical of most of the extracts in both length and the sorts of interaction contained in the extract. Transcriptions of other extracts are given in Appendix A, page 235.

Extract G

T: [1]**OK**. [2]I'm going to play a video [3]and we're going to see a scene from a, from a film. [4]You watch the video and tell me uh what you see, OK? [T turns on the video to show a 3-second excerpt from Star Wars Episode 1. T pauses the film.] [5]Do you know what film is it (the film)?

S: [6]The film is Star Wars Episode One.

T: [7]The film is Star Wars Episode One, right, Star Wars Episode One. [T plays the video for another ten seconds. The scene is set in a desert village. T pauses the video.] [8]Where, where do the boy take these people to? (2.0) [9]Where do this boy, this boy take these people to?

S: [10]The boy takes these people to His house.

T: [11]The boy takes these people to His place, his house. [12]Why does the boy take these people to his house?

S: [13]The boy takes these people to his house because Storm.

T: [14]Yes, there is a storm, sandstorm, right? [T plays more of the film in which three human/humanoid characters and a robot go to the boy's home.] [15]Who is the woman?

SS: [16]The woman is His (the boy's) mom.

T: [17]The woman is His (the boy's) mother, right. [T plays more of the film where one character introduces himself to the boy's mother, and then the boy takes a female character to see a robot he is building. They are followed by another robot.] (pointing at the boy in the scene) [18]What is he (the boy) talking about?

S: [19]The boy is talking about Robot.

T: [20]The boy is talking about His (the boy's) robot, right. [21]What's its (the robot's) name? [22]What's the robot's name?

SS: [23](unclear) [T plays more of the film. In the scene, the boy activates the robot he is building and this robot then walks to meet the other robot.]

T: (pointing at the robot under construction) [24]This (the robot under construction) is ...

SS: C-Three-P-O.

T: [25]The robot under construction is C-Three-P-O. [26]And how about this one (the other robot)? [27]Do you know his (the other robot's) name? (4.0) [28]It (the other robot)'s also, it's a robot also, right? [29]And what's its (the other robot's) name?

SS: [30](unclear) [T plays the next three seconds of the film where the robot's name is given.]

S: [31]The other robot's name is R-Two-D-Two.

T: [32] The other robot's name is R-Two-D-Two, right? [T plays the rest of the conversation between the two robots.] [33] OK. Why does R-Two-D-Two tell him (C-Three-P-O) that he (C-Three-P-O)'s naked, naked? [T gestures at her own body.] (2.0) [34] Why, why does R-Two-D-Two say that T-C...

SS: C-Three-P-O.

T: uh, C-Three-P-O is naked? (2.0) [35] Because he (C-Three-P-O) has no [T points to her own arms.] no skin, no skin, right? [36] Because the boy, the boy, the owner, the owner of the, the robot has no money to buy material to make his (C-Three-P-O's) skin. [37] **OK, now** [38] these are robots, C-Three-P-O and R-Two-D-Two, OK, [39] and do you think what kind of work or what kind of job can these robot do? (4.0) [40] What kind of work can robot do?

S: [41] Robot can *Tum gup kow.* {= Cook food.}

S: [42] Robot can be Servant.

T: [43] Servant, OK. Robot can be servant. [44] They (robots) can do housework, [45] they (robots) can do housework, [T mimes dusting.] [46] and what else can robots do?

S: [47] (unclear)

T: [48] Can robot cook food?

S: [49] No robots cannot cook food.

T: [50] Can robot, can robots cook food?

S: [51] I think they (robots) can (cook food) if it (the robot) has program.

T: [52] Yes, if robots have been programmed, they (robots) can cook food. [53] *Tah sahng progrum mah,* robot *samaht tum dý.* {= If you construct the program, a robot can do it (cook food).} [54] And what else can robots do? (1.0) [55] Who has seen this movie (Star Wars Episode One)? [56] Who has seen this film (Star Wars Episode One)? [57] Raise your hand. [T raises her hand as a model.] [58] *Krý mah doo laaw nung nee?* {= Who has already seen this film (Star Wars Episode One)?} [No SS raise their hands.]

F1: [59] *Doo laaw laa boo-at* head *mahk.* {= I've already seen it (the film) and I got a bad headache.}

T: [60] Robot *tum arai ka* {= What do robots do?} [61] *rawahng song tua tum arai nee nee nee?* {= between these two robots, what do they do, this this this?}

M1: [62] Robots can be *Nuk rorp.* {= Fighter.}

T: [63] Right, robots can be soldiers or warriors. [64] *Mee tua nee mah tum mý ka?* {= These robots come to do what?} [65] *Mee eek mý?* {= Is there anything else that robots can do?}

S: [66] (unclear)

T: [67] *Pen poo-ak nuk rorp mee doo-ay na ka.* {= They (robots) can also be fighters.} [68] What else can robots do?
S: [69] *Mee bunyah.* {= They (robots) have brains.}
T: [70] Ah, robot can fix machine, can fix machines, OK. [71] So robots can do so many things, so many kinds of works. [72] There's a question, [73] there's a question. (reading from a handout) [74] Will robots, will robots replace people at work?

3.2 Transcription of the extracts

Transcribing spoken discourse is neither straightforward nor truly objective. In making a transcript, there is a large amount of interpretation involved. For example, in transcribing a video as in this study, the researcher has to decide what non-verbal information to include in the transcript. More importantly, a decision has to be made about the level of detail to be shown in the transcript. There are no hard and fast rules about this, but as a general guideline, the transcript must have enough detail to serve the research purpose, but not so much as to be inaccessible or unreadable (Duranti, 1997).

For the purpose of this study, the transcript does not need to be very fine. For example, there is no need to make a phonetic transcription as pronunciation of individual words has little bearing on the study. Instead, I will use a standard orthography and try to punctuate as normally as possible. More problematically, some aspects of stress and intonation may need to be considered in certain sections of the research. For example, referring and proclaiming tone units (Brazil, 1985, 1995; Brazil et al., 1980; Sinclair and Brazil, 1982) may be helpful in distinguishing given and new information; and framing moves need to be distinguished from, say, confirmation checks based on context, pausing and intonation. Although intonation is important for these purposes, they can be shown in the transcript without obscuring the readability with levels of information about intonation. For given and new information, I could not find any points in the extracts where identification of whether information is given or new resides solely in the tone unit used. I have therefore decided not to include tone unit information in the transcript. In contrast, intonation is a key criterion in identifying framing moves. To enhance readability, however, I will not show the intonation patterns of framing moves specifically, but will highlight these moves in bold letters.

The conventions I will use in transcription are as follows:

In the left margin, conventions indicate speakers:

T: teacher
S: unidentified student, irrespective of gender
SS: students in chorus
M1: identified male student (with higher numbers indicating other identified male students)
F1: identified female student (with higher numbers indicating other identified female students)
{ to the left of speaker initials, a brace indicates simultaneous speech

In the text of the transcript, conventions concern the language used and non-verbal information important to communication.

[ ] non-verbal information e.g. [Teacher writes on the board]
( ) gloss on language e.g. (loudly), placed before the segment it applies to
(2.0) numbers in parentheses show length of pauses in seconds
" " written rather than spoken language
italics non-English speech. Thai speech is transliterated into English orthography following the conventions of Campbell and Shaveewongs (n.d.) both for clarity and because the sounds of words are important at times. Thai speech is also punctuated like English for clarity.
{= } translations of non-English speech
' ' quoted speech or for clarity e.g. wh- or 'or' questions
bold framing moves
S - P words spelt out as individual letters
superscript numbers the numbers of the T-units within each extract
subscript ellipted material
(subscript) referents for referring expressions

In addition, pseudonyms are used where students have been referred to by name.

3.3 Data preparation

Having transcribed the video recordings of the lessons, further work is needed to prepare the transcripts for analysis. Some of the approaches I shall be using in this study require the data to be divided into units before analysis. Specifically, analyses of theme-rheme and given-new progression, Hoey's (1991) lexical analysis and my own topic-based analysis (Watson Todd, 1998) all require the data to be divided into units prior to analysis. The first stage in preparing the transcripts for analysis, then, is to identify units in the transcripts.

The second stage in preparing transcripts for analysis concerns the information content of the discourse as perceived by the participants. Participants perceive more information content in the discourse than is immediately apparent from the raw data of the transcripts. In addition to the immediately apparent information, participants also perceive ellipted

material and the referents of any referring expressions. As this study focuses primarily on the conceptual or ideational metafunction of language, it is important that the full information content of the discourse as perceived by the participants is taken into account. The second stage in preparing the transcripts for analysis, then, is to identify and fill in ellipted material and to identify the referents for referring expressions.

3.3.1 Identifying units in the transcripts

To divide the transcripts into units, I had to first decide what units to use and how to identify the units. There are a variety of units which initially it appears possible to use for dividing the transcripts.

1. Turns

Turns or utterances are perhaps the easiest unit to identify from the initial transcripts, and are the unit used in conversation analysis (e.g. Jefferson, 1984; Schegloff, 1971). Identifying points where speakers change is relatively easy, and thus, from a practical perspective, turns are an appealing unit to use. However, turns vary greatly in length (from 1 to 278 words in the data collected in this study), and may be so broad as to contain two or even more transactions within one turn. There is, therefore, a lack of comparability between turns, and there is also little match between turns on the one hand and functions, content and the units usually used in analyses of the kind I will be undertaking on the other. These two problems mean that turns are an inappropriate unit for dividing up the discourse in this study.

2. Speech acts

The investigation of speech acts in linguistics dates back to Austin (1962/1976), and is an area of linguistics which has been widely investigated and which has a solid grounding. However, speech acts focus on the functional or interpersonal aspects of language use, and little attention is paid by speech act analysts to the conceptual or ideational aspects. As the relationship between these two metafunctions of language is unclear, conducting an analysis focusing primarily on conceptual aspects using discourse divided into units based on functional aspects may be inappropriate.

3. Discourse moves

The data in this study comes from classroom discourse, and therefore it might seem appropriate to use one of the ranks put forward in Sinclair and Coulthard's (1975) classic study of classroom discourse as the unit for

dividing up the transcripts. The rank most similar to the units used in the methods of analysis which I propose to apply in this study is that of move. However, as with speech acts, moves are based on functional aspects of language use and so may be inappropriate. Furthermore, while Sinclair and Coulthard's analysis of transactions based around question-answer sessions is convincing, how transactions consisting solely of teacher monologues are to be divided into moves is far less clear (see Coulthard and Montgomery, 1981). Given that some of the data used in this study consists of teacher monologues, discourse moves may be an inappropriate unit of analysis.

4. Sentences

In Hoey's (1991) lexical analysis, one of the methods to be used in this study, Hoey used sentences as the unit of analysis for his study of written discourse. While sentences may be an appropriate unit for written discourse, the applicability of sentences to spoken discourse is fraught with difficulties. Sentences are difficult to define beyond their surface features which do not appear in spoken discourse, and therefore it is impossible to identify sentences in spoken discourse and, indeed, the phrase 'sentences in spoken discourse' is oxymoronic. Sentences are therefore not an appropriate unit of analysis in this study.

5. T-units

In analysing oral discourse, several authors (e.g. Halleck, 1995; Klecan-Aker and Lopez, 1985; Larsen-Freeman, 1983) have used T-units as a parallel to the use of sentences in written discourse. A T-unit is "an independent conjoinable clause complex" (Fries, 1994: 229). Given its wide use as a unit of analysis for oral discourse and its parallels with the sentence which is used as the unit of analysis by Hoey (1991) and typically in theme-rheme and given-new progression, the T-unit would appear to be an appropriate unit of analysis in this study. However, the process of identifying T-units is not without problems, so I will now turn to how to identify T-units in the transcripts.

3.3.2 Identifying T-units in the transcripts

A T-unit can be defined as an independent clause together with all related dependent clauses (Fries, 1994). In other words, a T-unit is the maximum syntactic unit.

Although it may seem straightforward to identify independent clauses and to assign dependent clauses to independent clauses, the inherent surface

'messiness' of much oral discourse makes T-units hard to operationalise. For some stretches of discourse in the transcripts, such as example 3.1 below where boundaries between T-units are marked with double slashes, the identification of T-units is straightforward.

Example 3.1

T: // You can ask questions // and within twenty questions if you can guess correctly, if you can find out what I think of, then you win, OK? // The questions can be yes-no questions or 'or' questions. // For example, you can ask 'Can we eat it?' // I will answer just yes or no. //

(Extract E)

For much of the transcripts, however, identifying T-units is far less straightforward. To avoid total subjectivity, guidelines are needed. The following guidelines were used to identify T-units in this study. These guidelines are not meant to be applied blindly. Where, say, intonation patterns indicate otherwise (for example, a break between T-units will not occur in the middle of a tone unit), the guidelines are not followed. These, however, form a tiny minority of instances.

Guideline 1: Repetitions and paraphrases of non-self-standing information (e.g. isolated noun phrases) are included in the same T-unit as the independent clause containing the information with which it is most closely associated by intonation. For example:

Example 3.2

T: // This is its neck, right? Its neck, uh-huh, or opening. //

(Extract C)

Guideline 2: A single T-unit may cover more than one speaker turn. There are two main situations in which this can happen.

Firstly, on some occasions the teacher and the students may speak together in chorus as shown in example 3.3.

Example 3.3

T: // One, two, three.
SS: One, two, three. //

(Extract A)

That such choral speaking appears on two lines rather than one is an artifact of the transcription conventions which require that teacher and student turns are indicated separately. The two lines should therefore be counted as a single T-unit.

The second case where a single T-unit may cover more than one speaker turn is more complicated. On several occasions, the teacher uses a sentence completion instead of an interrogative to elicit an answer. The teacher expects the students to provide completion of a T-unit. This does not always happen, however. On occasion, the teacher is required to complete the T-unit him/herself as in example 3.4.

Example 3.4
T: // Yes, it will move ... forward. //

(Extract A)

Despite the pause in the middle of example 3.4, this is a single T-unit. It could be argued that example 3.4 is performing two speech acts: a question (*Yes, it will move ...*) and a response (*forward*). However, the problems of distinguishing between teacher pauses indicating a gap-fill question and pauses for other reasons make drawing divisions for T-units based on such speech acts unreliable. We will therefore count example 3.4 as a single T-unit. If we are counting a teacher self-completion as a single T-unit, it follows that we should also count a student completion of an unfinished T-unit in the same way. In example 3.5, the teacher is expecting a student completion so the first two turns should be counted as a single T-unit. The third T-unit in example 3.5, however, is probably a separate T-unit as the original T-unit has already been completed and the third turn is acting as an acknowledgement of this completion.

Example 3.5
T: //This one is
SS: bigger. //
T: Bigger. //

(Extract A)

This treatment of the third turn as a separate T-unit can be seen more clearly in example 3.6. In this example, it is clear that the teacher is not attempting to provide completion of his original T-unit and so the third turn should be treated separately.

Example 3.6

T: // The answer is ...
SS: robot. //
T: Robot? // Yes, you are right. //

(Extract E)

The reason for treating the third turn as a separate T-unit is that it is not attempting to complete the original T-unit. On a few occasions, as in example 3.7, however, both teacher and students attempt to provide completion together. Such examples are treated as a single T-unit.

Example 3.7

T: // so you see it becomes ... bigger.
SS: bigger. //

(Extract A)

Guideline 2 is only applicable to teacher initiations in the form of sentence completion. Responses to interrogatives, which form complete T-units by themselves, are treated as separate T-units.

Guideline 3: Pauses of two seconds or more are taken as indicating a boundary between T-units. It is assumed that a pause of this length indicates a break between the preceding and succeeding T-units.

Guideline 4: Framing moves are counted as separate T-units, since they are not normally conjoined with the clauses that follow them. The measured pause that is associated with a framing move (Sinclair and Brazil, 1982) makes such a move distinct from the succeeding discourse, and so it should be considered a separate T-unit. It should be noted that non-framing use of "OK", "Right" etc. is included in the same T-unit as the independent clause it is most closely associated with by intonation.

Guideline 5: The following are counted as separate T-units:

- Unclear turns.
- Counting (e.g. "One, two, three").
- Words spelt out (e.g. "B - A - L - L - O - O - N").
- Nominations.
- Information written on the board.

Where two or more independent clauses with different content are spoken simultaneously, each is counted as a separate T-unit.

Guideline 6: Exclamations are not counted as separate T-units for practical reasons. Most exclamations are of the form "Ooh" and "Aah", and it is frequently difficult to distinguish these exclamations from the general hubbub of noise present in the classroom.

Guideline 7: Verbal information from videos or audio tapes played in the classroom falls outside the discourse requiring division into T-units and is therefore not transcribed. In two of the lessons in this study (extracts G and K), the teacher plays a video tape and an audio tape respectively. In both cases, no reference is made in the classroom discourse to any specific information conveyed verbally in the video or tape, except names.

Guideline 8: T-units can be incomplete. Where appropriate (see below), ellipted material is added to T-units, and the vast majority of incomplete T-units become syntactically complete after the addition of ellipted material. Some T-units, however, remain incomplete, but this is not regarded as problematic (cf. Vavra, 2000).

Following these guidelines, T-units were identified and indicated by superscript numbers at the start of the T-unit as in example 3.8.

Example 3.8

T: [62] Can it carry things? [63] Yes it can. [64] It can help people to carry things.

(Extract E)

The nine lessons used for analysis comprise a total of 975 T-units, of which 32 are incomplete (even after supplying ellipsis). The longest extract (extract A) consists of 237 T-units, and the shortest extracts (extracts H and L) consist of 34 T-units.

3.3.3 Supplying information content

As noted above, the second stage in preparing the transcripts for analysis is to supply the information content which is perceived by the participants but which is not explicitly stated. In other words, we need to identify the referents of referring expressions and we need to identify and supply ellipted material.

3.3.3.1 Resolving reference

Firstly, all referring expressions (e.g. pronouns, pronominals), including referring expressions contained in ellipted material, are identified. For each of these, the referent is identified primarily on the basis of parallel expression. In other words, preference is given to a referent which matches to the same syntactically and semantically relevant position in another nearby T-unit (Sotillo, 1999). In some cases, gender and number information and non-verbal indications of referents are also taken into account.

Since the analyses to be used ignore reiterations of items within one T-unit, and to avoid being otiose, where the same referent is referred to by the same referring expression more than once in the same T-unit, the referent is given only once.

Referents are indicated using a subscript font in parentheses as in example 3.9.

Example 3.9

T: [34]Is it $_{\text{(the balloon)}}$ big?

(Extract B)

3.3.3.2 Identifying and supplying ellipsis

As with identifying T-units, identifying ellipted material is not straightforward. Most previous work on ellipsis (e.g. de Beaugrande and Dressler, 1981; Ordóñez and Treviño, 1999; Shapiro and Hestvik, 1995) has focused on identifying verbs in gapping formations or identifying subjects where they have been dropped. In this study, however, it is frequently the case that more than one constituent is ellipted in a T-unit. For example, take the following exchange:

Example 3.10

T: [69] What else?
S: [70] Robot.
T: [71] Robot. Robot. Mmm.

(Extract C)

In this exchange, the participants are clearly aware of more information content than is immediately apparent in the transcript. This extra information content has been ellipted in the exchange. If we assume that ellipsis is

anaphoric, we can look back at previous discourse moves to identify the ellipted material. The teacher's move in unit 69 is an initiation which implies that there has been a previous initiation trying to elicit parallel content. Looking back, we find the initiation "T: For example, how do we use engines?" with four student responses accepted by the teacher in the form "T: In a car." We can therefore supply ellipted material in unit 69 as follows where ellipted material is given in a subscript font.

T: [69] What else $_{\text{do we use engines in}}$?

The student responses and teacher follow-up would then follow a parallel pattern:

S: [70] $_{\text{We use engines in a}}$ Robot.
T: [71] $_{\text{We use engines in a}}$ Robot. Robot. Mmm.

This example is fairly straightforward, but, as with T-units, in many cases rough guidelines are needed to identify ellipted material. The guidelines are given below.

Guideline 1: Ellipted material is material which can be taken as perceived by the participants but not explicitly stated in the T-unit under consideration, and which is explicitly stated in a previous T-unit. In supplying material from a previous T-unit as ellipted material in the T-unit under consideration, only the minimum material needed is included. For example, in example 3.11, "If I release the two balloons" is not taken as ellipted.

Example 3.11
T: [149] If I release the two balloons, which one $_{\text{(balloon)}}$ will go farther?
S: [150] Bigger $_{\text{balloon will go farther.}}$

(Extract A)

Guideline 2: Where, in a previous T-unit, there is a reiteration of material taken as ellipted in the T-unit under consideration, only one iteration of the material is taken as ellipted. Where the reiteration is an exact repetition (53% of instances of reiteration in the data in this study), only one iteration is used as in example 3.12.

Example 3.12
T: [9] Where do this boy, this boy take these people to?

S: [10] This boy take these people to His house.

(Extract G)

Where the reiteration is a paraphrase at the same level of content richness (26% of instances of reiteration), the more parsimonious iteration is taken as the ellipted material as in example 3.13 where "fly" is used in preference to "move forward".

Example 3.13

T: [113] Why does it (the balloon) fly or move forward? [114] Do you have any question, [115] do you have any reason, any answer? [116] The balloon flies Because of the pressure, right?

(Extract A)

Where the two iterations differ in level of content-richness, the more content-rich is taken as the ellipted material as in example 3.14.

Example 3.14

T: [7] Do you know balloon, rubber balloon?

SS: [8] Yes I know rubber balloon.

(Extract B)

In instances such as these, the unmarked form is for the second occurrence of the reiterated material to be more content-rich than the first occurrence (19% of instances of reiteration). The marked form is for the first occurrence to be more content-rich than the second occurrence (2% of instances of reiteration).

Guideline 3: Ellipted material may carry over several turns as seen in example 3.10 above. Furthermore, as shown in example 3.13, ellipted material may refer back to material in a T-unit which is not immediately preceding. In example 3.13, units 114 and 115 do not include the ellipted material in unit 116 which is taken from unit 113. In all except one instance in the transcripts, only one or two T-units come between the T-unit which is the source of the ellipted material (unit 113 in example 3.13) and the T-unit under consideration (unit 116 in example 3.13). In the single exceptional case, 13 T-units come between the source and the T-unit under consideration (Extract B, unit 137), but common sense indicates the ellipted material since the intervening material consists of teacher treatment of an incorrect response to the same initiation as that prompting the student reply in unit 137.

Guideline 4: In supplying ellipsis, referring expressions are replaced by their referents in the ellipted material so that all lexical links between units are explicit. Example 3.13 contains an instance of this.

Guideline 5: Where necessary, 'empty' ellipted material (e.g. "there is") is used, as shown in example 3.15.

Example 3.15

T: [27] Just blow up (the balloon) and just [T holds the neck of the balloon.]
S: *kamoo-at pom.* {= tie the neck.}
T: [28] (There is) No need to *kamoo-at pom.* {= tie the neck.}

(Extract B)

Guideline 6: Where two or more different initiations directly precede a response, ellipted material is taken from the initiation which acts as the elicit rather than the initiation acting as a starter or prompt (see Sinclair and Coulthard, 1975), as shown in example 3.16.

Example 3.16

T: [T puts a transparency of R2D2 on the overhead projector.] [42] How about this one (transparency)? [43] What is his (the robot's) name?
S: [44] (The robot's name is) R-Two-D-Two.

(Extract F)

In example 3.16, unit 42 is the starter and unit 43 is the elicit, so the ellipted material in unit 44 is taken from unit 43.

Guideline 7: In a few cases, ellipted material may be added which is not given in the surrounding discourse but which is apparent from the context. In example 3.17 below, for instance, "Answer" is taken as ellipted in unit 10 even though it does not appear in the preceding discourse. Unit 10, however, is treated as a formulaic phrase which within the context of a language classroom refers to the language of the student's answer.

Example 3.17

T: [8] **OK.** [9] What are examples of, of what? Examples of anything which has pumping system. (2.0)
S: [10] (Answer) *Pahsah unggrit reu krup?* {= In English?}

(Extract J)

By resolving reference and supplying ellipted material, it is hoped that the expanded transcripts show the full information content perceived by the participants during the interaction. The expanded transcripts also highlight the potential lexical links between units thus facilitating analyses such as those of Hoey (1991) which involve lexical cohesion.

3.4 Insider perspectives on the discourse

Even though I will be conducting six different analyses and comparing the results of these analyses, this is not enough to validate any of the analyses. While conducting different methods of analysis provides theoretical triangulation in the research (Freeman, 1998), ideally we should also view the data from the perspectives of the different parties involved in the research. All of the analyses I have proposed in chapter 2 are conducted from my own perspective as researcher. We also need to gain more emic or insider perspectives on the discourse (Freeman, 1998; Watson-Gegeo, 1988), and see how the actual participants in the discourse perceive the topics and topic progression. In other words, the results from the six methods of data analysis should be compared with the perceptions of the teachers and students who generated the discourse.

To gain information concerning what and how participants are thinking and feeling during the discourse, there are three key methods available to the researcher: journals, think aloud reports, and retrospective interviews (Graham, 1997). Of these, think aloud reports cannot be used to investigate the classroom, as the extremely intrusive nature of this instrument would disrupt the production of the discourse. We are therefore left with a choice of journals or retrospective interviews.

3.4.1 Investigating teacher perspectives

In deciding which method to use to gain information about the teachers' perspectives on the discourse, the key consideration in this research was gaining the teachers' cooperation. For this reason, potentially time-consuming interviews were not possible. Teachers were therefore asked to write journals about the eliciting stage of the recorded lessons. They were asked to write a one-page reflection on their eliciting as soon as was feasible after the lesson. It was hoped that trends emerging from the journals could be related to classroom events (Wallace, 1998) and thus provide useful information concerning how the teachers perceived the discourse.

Participants' data for extract G is given as an example in Appendix B, page 265.

3.4.2 Investigating student perspectives

For the students, the need to gain cooperation was less of a consideration than for the teachers. It was therefore decided to interview students to gain a greater depth of data. The model used for the interview was to play back the video recording of the eliciting stage of the lesson, pause the video at key points (usually after 8 or 9 T-units of discourse or where the topic entity changed as judged intuitively), and ask the students for their reactions at that point (cf. Nunan, 1990; Wallace, 1998). The interviews then were retrospective but guided by the recordings of the lessons. Although it could be argued that the researcher choosing where to pause the video may have guided the students' responses, in the first interview initially students were asked to pause the video whenever they felt they wanted to say something, but no pauses were made in the first three minutes. In order to gather data on students' perspectives, it was therefore necessary for the researcher to choose where to pause the video.

For each interview, two students were interviewed together later on the same day as the lesson. A different pair of students was interviewed each time. The interviews were conducted in Thai to avoid language barriers to expression of meaning.

The interviews were semi-structured or focused (Weir and Roberts, 1994). At each point where the video was paused, the students were asked whether they were following the content of the lesson and what they thought the end goal of the eliciting was. Follow-up questions were asked as required.

3.5 Comparisons of findings

There were in total eight sources of findings in this study: the six methods of analysis and the teachers' and students' perceptions. To evaluate the usefulness of each of the methods of analysis, all of the sources were compared.

The first stage in the comparison was to compare the six methods of analysis. It should be remembered that the two methods based on Watson Todd (1998) are the only methods by which it is possible to follow topic drift. The comparison between methods therefore focused on the placement of topic shift and the identity of topic entities and topics. To judge the

validity of each of these six approaches, they were also compared with a control approach.

The second stage in the comparison was to compare the six methods of analysis with the teachers' and students' perceptions. Salient perceptions from both the teacher journals and the student interviews were identified and appended to the transcripts at the points to which they applied. These salient perceptions were then compared to the findings of each of the methods of analysis at these points.

It is hoped that these comparisons of findings will allow us to identify the key analyses which identify topics and follow topic development. Where two or more analyses produce similar findings, we may only need to conduct one of these analyses; alternatively, where an analysis produces findings which are not consistent with the findings of the other analyses, we may be able to discount that analysis, or it may lead to different insights into topics. Finally, analyses which produce findings distinct from those of the control approach or which agree most closely with the perceptions of the teachers and students are preferable.

Chapter 4 Initial Analyses of Topics

To analyse the data described in Chapter 3, I will start by using those analyses which were not specifically designed to identify topics and topic progression, but which may be of some use. These approaches are Sinclair and Coulthard's (1975) functionally-based analysis of classroom discourse, Hoey's (1991) lexical networks, and analyses of theme-rheme and given-new progression which focus on the level of clause. I will apply each of these approaches to the data collected for this study in turn.

4.1 Sinclair and Coulthard's (1975) analysis of classroom discourse

4.1.1 Background

A useful starting point in any analysis of classroom discourse is Sinclair and Coulthard (1975). Their study has been very influential in both discourse analysis and investigations of the classroom. Basically, it proposes that classroom discourse can be divided into a series of levels and describes the structure of some of these levels using an approach based in speech act theory.

The discourse levels which Sinclair and Coulthard identify are, starting with the largest, lesson, transaction, exchange, move, and act. In other words, they propose that a lesson comprises one or more transactions which in turn consist of one or more exchanges and so on. Teaching, then, is seen as having a hierarchical organisation.

Sinclair and Coulthard do not attempt to describe the structure of a lesson. For transactions, they argue that transactions normally begin and end with boundary exchanges, between which is a series of teaching exchanges for which no sequencing can be specified.

It is in describing the structure of exchanges that Sinclair and Coulthard have been most influential. The first type of exchange, the boundary exchange, consists of a framing move followed by an optional focusing move. The framing move itself comprises a marker such as 'OK', 'Right' and 'Well' usually spoken with a proclaiming intonation and high key (Coulthard, 1981, 1987; Sinclair and Brazil, 1982) and followed by a pause. The focusing move is a metastatement serving the knowledge-framing purposes of Ausubel's (1963) advance organisers.

Sinclair and Coulthard's description of the teaching exchange has probably had the largest impact of all their analysis. They argue that a typical exchange is made up of three moves: first, an initiating move (I) typically made by the teacher; second, a responding move (R) from a student; and third, a feedback move (F) by the teacher. This IRF exchange pattern is typified by the teacher asking a question, a student answering, and lastly, the teacher evaluating the answer.

This IRF pattern of exchanges is identical to the IRE pattern (initiation - response - evaluation) identified by Mehan (1979, 1985). Whereas Sinclair and Coulthard's analysis is based in speech act theory, Mehan's work uses an ethnographic approach. The similarity of the findings, despite being founded on different bases, suggests that IRF is a key characteristic of classroom discourse. Indeed, the prevalence of IRF in classroom discourse has led to suggestions that further research in this direction is unnecessary (Cazden, 1986).

In addition to identifying a key characteristic of classroom discourse, Sinclair and Coulthard's work has also been influential in other ways. Within linguistics, their approach to matching form with function and their criteria concerning what makes a good description have had a wide influence, and in pedagogy, their analysis has been related to goal structures and perceptions of teachers (e.g. Smith and Holdcraft, 1991; but cf. Woods, D. 1996).

4.1.2 Criticisms

Despite this influence, Sinclair and Coulthard's analysis has also come under criticism. Some of these criticisms have suggested changes to the details of the analysis while still working within the same framework. For example, Coulthard and Montgomery (1981) attempt to provide a similar analysis of teacher monologues although their attempt is unsatisfactory; Alpert (1987) argued that the feedback move may not exist at all in some classrooms; Sinclair (1992) broadened the nature of the feedback move and re-termed it follow-up; Coulthard and Brazil (1992) redefined the teaching exchange as I(R/I)R(F) where brackets indicate optional moves; and Hazadiah (1993) and Hoey (1993) both suggested an extra rank between the levels of exchange and transaction, which Hoey called an exchange complex. These criticisms, however, do not challenge the basic findings of Sinclair and Coulthard.

More seriously, Sinclair and Coulthard's original goal of providing an inclusive and unambiguous assignment of utterances to speech act units has had to be tempered as this was found to be an idealised state of affairs (Sinclair and Brazil, 1982). Sinclair and Coulthard's use of a hierarchical structure has also been criticised as overly idealistic and not reflective of the real world (van Lier, 1988). Most seriously perhaps, the lack of any detailed description of the context on which the analysis is based can be seen as severely weakening Sinclair and Coulthard's study (Cazden, 1986; Levinson, 1983).

4.1.3 Functional and conceptual aspects of classroom discourse

Despite these criticisms, the great influence of Sinclair and Coulthard's analysis and the wealth of evidence showing that the discourse units they describe exist in a wide range of classrooms makes it a useful starting point for analysis of classroom discourse. For the purposes of this study, Sinclair and Coulthard may prove particularly valuable as several authors have tried to equate levels in the analysis with topics. Francis and Hunston (1987), Hazadiah (1993) and van Lier (1988) all argue that transactions are topic-units (in Hazadiah's case, macro-topic-units), and that boundaries between transactions indicate topic shift. The intermediate level of Hazadiah (1993) and Hoey (1993) has also been equated with intermediate level topics (Hazadiah, 1993) with boundaries between them indicating minor topic shift (van Lier, 1988). The exchange may also have conceptual implications with Coulthard and Brazil (1992) arguing that each exchange carries only one piece of information. Similarly, in Mehan's (1979) analysis, the interactional sequences equivalent to Sinclair and Coulthard's level of exchange are organised around topics. While Sinclair and Coulthard's initial analysis was functional, the range of arguments for matches between Sinclair and Coulthard's functional units and conceptual or topic units suggests that the analysis may also have conceptual underpinnings which may be of importance to this study.

I will start the application of Sinclair and Coulthard's analysis to the data collected as described in chapter 3 by trying to identify moves.

4.1.4 Identifying moves

Starting with the initiating, responding and feedback moves, generally these can be identified readily as long as the function of the move is focused on rather than the form, speaker or length. In other words, although most questions are initiations, the teacher may ask a rhetorical question as part of

a monologue; although most initiations are performed by the teacher, students can also perform initiations with the teacher giving the response; and although many moves are one T-unit in length, some moves cover several T-units and a single T-unit may include more than one move. Thus, in the majority of instances, by focusing on functions, the moves making up teaching exchanges are identifiable.

However, in a minority of cases, such ready identification is not possible. For instance, in example 4.1, the moves are unclear.

Example 4.1

T: [8] **OK**, [9] what are examples of, of what? Examples of anything which has pumping system. (2.0)
S: [10] Answer *Pahsah unggrit reu krup?* {=In English?}
T: [11] Mmm? What are examples of anything which has The pumping system.
M1: [12] Answer in English?
T: [13] What do you mean? [14] Do you know, do you understand the pumping system?
M1: [15] Yes I understand the pumping system.
T: [16] It (pumping system) 's not *kreu-ang soop num.* {= water pump.}

(Extract J)

In example 4.1, T-unit 9 is clearly a teacher initiation. The student's answer in T-unit 10 is a question, but responses can be phrased as questions (e.g. Sinclair and Coulthard, 1975: 54). However, T-unit 10 does not appear to be an attempt to respond to the initiation in T-unit 9. T-unit 11 is a re-initiation of T-unit 9, but is again answered by a question which is not an attempt to respond. T-unit 12 is followed by two further initiations which provide no feedback or response to the student utterance. The problem in example 4.1 is how T-units 10 and 12 should be categorised. They are given in response to initiations but are themselves initiations rather than attempts to answer the teacher's initiations. If we categorise them as initiations, however, we end up with a string of six unrelated initiations, the last of which is the only one to receive a response. Categorising T-units 10 and 12 as either initiations or responses is problematic, and may reflect differing teacher and student perspectives on the discourse.

Further problems in the analysis are encountered when dealing with teacher monologues. Sinclair and Coulthard treat teacher explanations and instructions as initiations. This approach works well when an explanation is

one act long and functions as a starter in an initiation, but does not cope with long teacher explanations or instructions. As we noted above, no satisfactory alternative ways of analysing teacher monologues have been proposed. In this study, I will also leave them unanalysed but categorise them as monologues (M) to distinguish them from initiations, and take any continuous stretch of monologue moves as comprising a monologue exchange. I will categorise moves expected to elicit direct responses, either verbal or non-verbal, as initiations and moves not expected to elicit a direct response (and not functioning as framing, focusing, response or feedback moves) as monologue. Under this approach, for the data collected in this study, monologues are only performed by the teacher.

Turning now to framing and focusing moves, focusing moves never occur except directly following a framing move (Sinclair and Brazil, 1982). To identify these two kinds of moves in the discourse, therefore, the prime consideration should be given to framing moves.

Framing moves are instantiated by a closed class of words which include *OK*, *Right*, *Well*, *Look*, *Now* and *Good*, and are also indicated by intonational and paralinguistic information. To identify a framing move, these three characteristics must occur together. Thus in example 4.2, the first *OK* spoken with a high falling intonation and followed by a pause is a framing move, but the second spoken with a rising intonation is not.

Example 4.2

T: [1] **OK,** [2] for today I have a box with me. [T holds up small plastic container.] OK?

(Extract A)

With such easily identifiable characteristics, it is a relatively straightforward matter to identify framing moves in the discourse. The following move can then be examined to see if it exhibits the characteristics of a focusing move (future time reference and reference to a teaching unit).

There are, however, some cases where one of the closed class of framing words is used at the start of a T-unit, but is either spoken with a high falling intonation or is followed by a pause, but not both. An example is given in example 4.3.

Example 4.3

T: [70] Yes, it (the balloon) will move [T writes [71] "move forward" on the board] [70 cont.] forward. [T corrects his own misspelling on the board.] [72] (to self) F-O-R-W-A-R-D. (loudly) [73] It (the balloon) will move forward. [74] OK, now I will release the balloon.

(Extract A)

In this example, *OK* at the start of T-unit 74 has a high falling intonation but is not followed by any noticeable pause. Van Lier (1988) suggests that framing moves can be used as boundaries between transactions and as boundaries between units at lower levels in classroom discourse. In the first case, they mark a major topic shift and in the second a minor topic shift. Intuitively, *OK* in T-unit 74 indicates a minor topic shift. Provisionally therefore, I will identify framing moves indicating boundaries between transactions as moves exhibiting all three characteristics proposed by Sinclair and Coulthard. For markers at the start of T-units which have either a high falling intonation or are followed by a pause, I will take these as indicating a boundary between either exchange complexes or exchanges, and they will not be categorised as framing moves.

4.1.5 Identifying exchanges

Identifying monologue moves allows monologue exchanges to be identified, and similarly framing and focusing moves allow the identification of boundary exchanges. Let us now turn then to the identification of teaching exchanges.

The archetypal teaching exchange in Sinclair and Coulthard's analysis takes the form of IRF. The IRF form appears frequently in the data collected for this study as shown in example 4.4.

Example 4.4

T: [3] Unit six is about ... [I]
S: pump. [R]
T: [4] Pump. [F]

(Extract K)

In fact, about 44% of teaching exchanges in the data follow the IRF pattern, making it the most common form of exchange. Other relatively common patterns include I (where the teacher initiates but no response is given), IR (where there is a response but no feedback is given), and IRFRF and IRRF

(where there are multiple responses to an initiation). Table 4.1 shows the proportions of the various combinations of moves in exchanges.

IRF	44.2%
IR	11.6%
I	9.0%
IRFIRF	8.0%
IRIRF	6.5%
IRFRF	4.0%
IRRF	2.5%
Other	14.1%

Table 4.1 Proportions of variations of teaching exchanges

From Table 4.1, we can see that a large proportion of teaching exchanges is categorised as 'other'. This category contains 18 different permutations of initiating, responding and feedback moves in teaching exchanges, most of which occur only once in the data. Example 4.5 is an example of one such exchange.

Example 4.5

T: [17] What sound is it (the sound)? [I] [T plays a tape of a heart beating.] [I] [18] All right, (unclear) what sound is it (the sound)? [I]
SS: [19] The sound is a Heart. [R]
SS: [20] The sound is *Hoo-a jý*. {= Heart.} [R]
T: [21] The sound is a Heart. [F]
F: [22] The sound is Heart rate. [R]
T: [23] The sound is Heartbeat. [F]
SS: [24] The sound is Heartbeat. [R]
SS: [25] The sound is Beat. [R]
T: [26] Like the name of (unclear). [F]

(Extract K)

Although this variation in patterns of teaching exchanges makes identifying them more complicated, it does not conflict with Sinclair and Coulthard's analysis. Indeed, Sinclair and Coulthard's use of implicit and explicit bound initiations (i.e. re-initiations in the same exchange aiming to elicit the same information) encourages us to expect such variations.

However, some of the patterns of exchanges in the data are not covered by Sinclair and Coulthard, and a few patterns challenge the validity of their analysis.

The first of these patterns involves the interruption of exchanges. In example 4.6, the teacher makes an initiating move in T-unit 63. Before she can ask more and before the students can make a response, there is a discipline problem that needs to be treated. After this has been dealt with, the teacher gives a framing move to indicate a shift in topic which performs a similar role to *Anyway* in conversation. Finally, the teacher explicitly re-initiates the question given in T-unit 63. T-unit 67 is, therefore, a bound initiation falling into the same exchange as T-unit 63. The discourse pattern, then, is similar to the insertion sequences identified in conversation analysis (see Coulthard, 1977).

Example 4.6

T: [63] What kind of reaction? (3.0) [Two SS get up to go to the toilet.] [64] *Tum mý torng bý doo-ay gun song kon nee.* {= Why do you two have to go together?} [Several other SS make sarcastic comments.] [65] *Pood nah klee-ad jing jing.* {= You really speak nastily.} [66] **OK.** [67] Can you answer my question?

(Extract D)

If we allow exchanges to be interrupted by insertion sequences and we identify these interrupted exchanges by bound initiations before and after the insertion sequence, we find that it is not only discipline problems which can interrupt exchanges. The various questioning strategies which provide scaffolding for students' responses (see Cole and Chan, 1987; Watson Todd, 1997b) may also act as insertion sequences, as shown in example 4.7.

Example 4.7

T: [74] What kinds of engine work like this (the balloon)? [T writes [75] "Engine" on the board.] [76] Do you know what is engine?

S: [77] Engine is *Kreu-ang yon.* {= Engine.}

T: [78] Engine *keu arai kha?* {= what is it?}

S: [79] Engine is *Kreu-ang yon.* {= Engine.}

T: [80] Engine is *Kreu-ang yon.* {= Engine.} [81] What kinds of engines that work like the balloon?

(Extract D)

In example 4.7, T-units 74 and 81 are clearly bound since the latter is a near repetition of the former. T-units 75 to 80, if taken in isolation, form an exchange in themselves which provides key information to answer the questions in T-units 74 and 81. There are three ways of treating this. First, we could treat T-unit 74 as an exchange consisting solely of an initiation. Second, we could treat T-units 74 and 81 as belonging to one exchange and T-units 75 to 80 as belonging to a different exchange. Third, we could treat the whole of example 4.7 as part of one exchange with different levels of centrality.

All three possible treatments have problems. If we take the first approach, we are discounting bound initiations from our analysis. The logical conclusion of this approach is that succeeding initiating acts would have to be treated as separate initiating moves, and thus separate exchanges, since they cannot be bound.

Regarding the second and third approaches, drawing the line between exchanges becomes problematic. The continuation of the discourse in example 4.7 is given in example 4.8.

Example 4.8

S: [82] (unclear)
T: [83] *Mý chý* energy *na ka.* {= It's not energy.} [84] Engine.
S: [85] Action reaction.
T: [86] Think about when, uh, what kind of engine, what kind of uh how to say. [87] **OK**. [88] Can you compare this (action reaction), with this technique with the engine? (3.0) [89] *Payayam torp dý mý pror wah gahn tee mee pror lom mun ork laa* balloon *mun bý kahng nah, nee plee-ab tee-ab gahp* engine *chunit ný, baap ný?* {= Can you try to answer because this action where the air goes out of the balloon and the balloon goes forward, this is comparable to what kind of engine, what type of engine?}
S: [90] (unclear)
T: [91] *Bok wah arai na?* {= What did you say?}
S: [92] (unclear) *raaw raaw krup.* {= quick quick.}

(Extract D)

If we treat T-units 74 and 81 as bound initiations falling into the same exchange, we should also treat T-unit 89 as part of the same exchange which

is now also interrupted by a monologue exchange (T-unit 86) and a boundary exchange (T-unit 87).

Under the second approach, we would finish our analysis with the following exchange pattern (where braces indicate insertion sequences).

I {IRIRF} IRFIR {M} {Fr} IRIR

Such a complicated analysis appears unsatisfactory, and the inclusion of a boundary exchange within a teaching exchange contradicts Sinclair and Coulthard's analysis of the functions of boundary exchanges.

Following the third approach, we end up with exchanges nested within other exchanges. This would break the division of discourse into various hierarchical levels which underpins Sinclair and Coulthard's analysis, and thus challenges its very foundations.

Whichever approach is used to deal with example 4.7 and 4.8 is unsatisfactory and raises worrying issues about Sinclair and Coulthard's analysis. The key problem in the analysis here is the identification of the boundaries of teaching exchanges. This problem in other seemingly more straightforward stretches of discourse actually challenges the functional foundations of Sinclair and Coulthard's analysis. This is illustrated in examples 4.9 and 4.10.

Example 4.9

T: [1] What did you do this, last weekend? (1.0) [2] What did you do last weekend? After mid-term exam.

(Extract F)

Example 4.10

T: [10] *Bý tum arai RCA?* {= What did you go to RCA for?} [11] Did you watch, did you watch Star Wars Episode 1 yet?

(Extract F)

Both example 4.9 and example 4.10 consist of two consecutive initiating acts. The two acts in example 4.9 cover the same information, are bound together, and so are part of the same move. In example 4.10, on the other hand, there is no informational similarity between the two T-units and they are not bound together.

In Sinclair and Coulthard's functional approach, example 4.9 and example 4.10 would be dealt with in the same way. Consecutive initiating acts would always be included in the same initiating move since they perform the same function of eliciting a response. While this analysis is satisfactory for example 4.9, including the two T-units in example 4.10 in the same move ignores the focus and topic of the T-units and is unsatisfactory.

An alternative approach is to take the information content of the discourse into account when assigning acts to moves and moves to exchanges. Coulthard and Brazil (1992) argue that each exchange carries only one piece of information. Applying this to examples 4.9 and 4.10, we see that the two T-units in example 4.9 are concerned with the same information and so fall within the same exchange. For example 4.10, however, the two T-units have different information content and so would fall into different exchanges. The two initiating acts in example 4.10 are therefore part of different initiating moves, with a boundary between exchanges at the end of T-unit 10. Such an approach produces a more satisfactory analysis and provides grounds for deciding whether succeeding initiations are bound or whether they occur in separate exchanges.

4.1.6 Conducting an analysis based on Sinclair and Coulthard (1975)

Despite the problems in identifying exchanges, I will attempt to conduct an analysis based on Sinclair and Coulthard's (1975) model. In doing this, I will use intuitive identifications of information content in dealing with insertion sequences and immediately succeeding initiation moves.

An analysis of the first 25 T-units of extract G based on Sinclair and Coulthard is given below. In the transcript, double vertical lines indicate transaction boundaries, single vertical lines indicate exchange boundaries, and superscript italicised letters indicate types of move as follows:

Fr	framing move
Fo	focusing move
I	initiation
R	response
F	feedback
M	monologue

T: ||[1 *Fr*] **OK.** [2 *Fo*] I'm going to play a video [3 *Fo*] and we're going to see a scene from a, from a film. |[4 *M*] You watch the video and tell me uh

what you see, OK? [T turns on the video to show a 3-second excerpt from Star Wars Episode 1. T pauses the film.] |[5 I] Do you know what film is it (the film)?

S: [6 R] The film is Star Wars Episode One.

T: [7 F] The film is Star Wars Episode One, right, Star Wars Episode One. | [T plays the video for another ten seconds. The scene is set in a desert village. T pauses the video.] [8 I] Where, where do the boy take these people to? (2.0) [9 I] Where do this boy, this boy take these people to?

S: [10 R] The boy takes these people to His house.

T: [11 F] The boy takes these people to His place, his house. |[12 I] Why does the boy take these people to his house?

S: [13 R] The boy takes these people to his house because Storm.

T: [14 F] Yes, there is a storm, sandstorm, right? | [T plays more of the film in which three human/humanoid characters and a robot go to the boy's home.] [15 I] Who is the woman?

SS: [16 R] The woman is His (the boy's) mom.

T: [17 F] The woman is His (the boy's) mother, right. | [T plays more of the film where one character introduces himself to the boy's mother, and then the boy takes a female character to see a robot he is building. They are followed by another robot.] (pointing at the boy in the scene) [18 I] What is he (the boy) talking about?

S: [19 R] The boy is talking about Robot.

T: [20 F] The boy is talking about His (the boy's) robot, right. |[21 I] What's its (the robot's) name? [22 I] What's the robot's name?

SS: [23 R] (unclear) | [T plays more of the film. In the scene, the boy activates the robot he is building and this robot then walks to meet the other robot.]

T: (pointing at the robot under construction) [24 I] This (the robot under construction) is ...

SS: [R] C-Three-P-O.

T: [25 F] The robot under construction is C-Three-P-O.|

Moving back to the conceptual metafunction of language, what do the moves, exchanges and transactions in extract G tell us about topics? Identifying topics for transactions is not possible - indeed, one purpose of this study is to see how this could be done. On the other hand, identifying topic entities of moves concerns a micro-level which does not provide useful insights. It seems likely that the most productive yet feasible level of discourse to work with is that of exchanges, and doing this would involve trying to identify topic entities in exchanges. Such an approach is supported

by Coulthard and Brazil (1992), who argue that each exchange carries one piece of information.

To see how such an approach might work, let us look at the exchange in T-units 60 to 63 of extract G.

T: | [60 I] Robot *tum arai ka* {= What do robots do?} [61 I] *rawahng song tua tum arai nee nee nee?* {= between these two robots, what do they do, this this this?}
M1: [62 R] Robots can be *Nuk rorp.* {= Fighter.}
T: [63 F] Right, robots can be soldiers or warriors. |

There are three different ways in which we can identify the piece of information in this exchange.

Firstly, we could take a propositional approach and identify the piece of information as ***Robots can be fighters.*** Secondly, we could identify the concept representing new information (*fighter*) as the key piece of information. Thirdly, we could identify the most frequent concept in the exchange (*robot*) as the key concept.

Although a propositional approach works with the exchange from T-units 60 to 63, in other exchanges, such as T-units 68 to 70 shown below, it is unclear what proposition should be chosen.

T: | [68 I] What else can robots do?
S: [69 R] *Mee bunyah.* {= They (robots) have brains.}
T: [70 F] Ah, robot can fix machine, can fix machines, OK. |

In addition, taking a propositional approach would not allow us to compare the findings from an analysis based on Sinclair and Coulthard (1975) with the findings of the other analyses which identify topic entities as concepts.

Equating topic entities with new information is problematic since many exchanges (e.g. T-units 64 to 67 in extract G) contain no new information, whereas in other exchanges all of the information could be considered new (e.g. T-units 55-59 in extract G). Furthermore, focusing on new information ignores the given information which is likely to provide the continuity in discourse associated with topics.

Provisionally, I will therefore take the most frequent concept in an exchange (including ellipted material) as the piece of information carried by an exchange. In doing this, for extract G we find the information in exchanges shown in Table 4.2.

Exchange	T-units	Information
1	1-3	video
2	4	video
3	5-7	Star Wars Episode One
4	8-11	boy, people
5	12-14	storm
6	15-17	woman
7	18-20	boy
8	21-23	robot
9	24-25	C3PO
10	26-32	R2D2
11	33-36	C3PO
12	37	-
13	38-45	robot
14	46-47	robot
15	48-53	robot
16	54	robot
17	55-59	film
18	60-63	robot
19	64-67	robot
20	68-70	robot
21	71-74	question

Table 4.2 Information in exchanges in extract G

It should be stressed that the identification of the information content of exchanges in Table 4.2 is provisional. While not certain, the identification will allow us to compare the functionally-based division of discourse into exchanges with more conceptually-based analyses of topic. Conducting such a comparison may highlight parallels between functional and conceptual analyses of discourse. This comparison is conducted in chapter 6.

4.1.7 Conclusion

To conclude, Sinclair and Coulthard's analysis aims to identify the functions of stretches of discourse which can be categorised as levels in the

hierarchical structure of the whole discourse. With the data in this study, identifying acts following Sinclair and Coulthard's approach is straightforward. Identifying moves is also generally straightforward, except for dealing with consecutive initiating acts. To identify whether these fall into the same move and exchange or not, conceptual aspects need to be considered. At the level of exchange, in addition to the problems of whether consecutive initiations fall within one exchange or two, there are problems of interrupted exchanges where a functional analysis is unsatisfactory. Finally, at the level of transaction, a conceptual approach may be more productive than a functional one. The general trend, then, is that at lower levels the functional analysis of Sinclair and Coulthard is effective, but at higher levels conceptual aspects must also be considered. This study is primarily conceptually based and so may be applicable to the higher levels identified in discourse by Sinclair and Coulthard. After setting up a system of analysis, I will return to analysing the higher levels in chapter 6 to see if a conceptual approach is valuable.

4.2 Hoey's (1991) lexical analysis

Hoey's (1991) lexical analysis attempts to show how patterns of lexis which reflect text organisation can be identified through studying cohesion. Focusing on lexical cohesion, the approach identifies lexical ties which provide cohesive connections between sentences. Where these ties involve reiteration of a lexical item, the sentences are said to be linked. Two sentences which contain an above-average number of links are termed bonded sentences. Hoey argues that identifying the bonds in a text provides insight into how the text is organised.

4.2.1 Identifying lexical ties

Approaches to lexical cohesion started with the work of Halliday and Hasan (1976), and more recently have been developed by Hoey (1991). These approaches identify several ways in which lexical cohesion is created. For example, Hoey (1991) argues for the following types of lexical ties, in order of importance:

simple lexical repetition (e.g. political - political)
complex lexical repetition (e.g. historical - history)
simple mutual paraphrase (e.g. excerpt - passage)
simple partial paraphrase (e.g. attempted - try)
antonymous complex paraphrase (e.g. hot - cold)
referring expressions such as substitution, co-reference, ellipsis and deixis

In this study, in addition to these types of lexical ties, one further kind of tie was used. To account for the links between wh-questions and answers, an unspecified - specified relation was also included in the analysis. For example, in example 4.11 below there are two types of lexical ties.

Example 4.11

T: [15] Who is the woman?

SS: [16] The woman is His (the boy's) mom.

(Extract G)

In example 4.11, there is a tie between *the woman* in the question and the ellipted material in the response, and there is a second tie between the unspecified *Who* in the question and the specified *His mom.*

For Halliday and Hasan as well as Hoey, the lexical ties were applied largely to written texts. For the analysis of classroom discourse in this study, instead of using the sentence as the unit of analysis as is typical in studies of written texts, T-units are used as the unit of analysis. In the following analysis, I will largely follow Hoey's (1991) methods and will start by examining lexical ties in the first seven T-units of extract G which are shown in example 4.12.

Example 4.12

T: [1]**OK**. [2]I'm going to play a video [3]and we're going to see a scene from a, from a film. [4]You watch the video and tell me uh what you see, OK? [T turns on the video to show a 3-second excerpt from Star Wars Episode 1. T pauses the film.] [5]Do you know what film is it (the film)?

S: [6] The film is Star Wars Episode One.

T: [7] The film is Star Wars Episode One, right, Star Wars Episode One.

To conduct a lexical analysis of this example following Hoey, the first thing we need to do is to identify reiterations of lexical items (shown by the lines in example 4.12). The framing move (Sinclair and Coulthard, 1975) in T-unit 1 does not contain any lexical items which can be linked to other T-units. T-unit 2, however, contains two lexical items which are reiterated in

later units. *I* in unit 2 recurs as *me* in unit 4, and *video* is repeated in unit 4, paraphrased as *film* in units 3, 6 and 7 (the latter two as ellipted material), and substituted with a pronoun in unit 5. Thus unit 2 has one link to units 3, 5, 6 and 7 and two links to unit 4. Having dealt with unit 2, we can continue in the same manner with unit 3 and so on.

A few additional decisions in identifying links need to be made. Firstly, as far as possible, lexical items are identified as words rather than phrases. This reduces the subjectivity inherent in identifying what constitutes a lexical item. However, proper nouns such as *Star Wars Episode One* are treated as a single lexical item and phrases which are reiterated as single words or proper nouns are also taken as single lexical items (for a comparison with how concepts are identified, see chapter 5).

Secondly, following Hoey, reiterations of the same lexical item within a T-unit are ignored unless two T-units both contain multiple reiterations of the same item.

Thirdly, as mentioned above, in addition to the relations identified by Hoey as indicating reiterations, in this study I am also including an unspecified - specified relation. This is exhibited in the link between *what film* in unit 5 and *Star Wars Episode One* in unit 6.

Following these guidelines, we can draw up a table showing the number of links between the 7 T-units in example 4.12 as shown in Table 4.3.

T-unit 6	2					
T-unit 5	2	2				
T-unit 4	1	1	2			
T-unit 3	1	1	1	2		
T-unit 2	1	1	1	2	1	
T-unit 1	0	0	0	0	0	0
	T-unit 7	T-unit 6	T-unit 5	T-unit 4	T-unit 3	T-unit 2

Table 4.3 Number of links between the T-units in example 4.12

This procedure can be continued for the whole of extract G (see pp. 46-48) to produce Table 4.4.

Table 4.4 Number of links between T-units in extract G

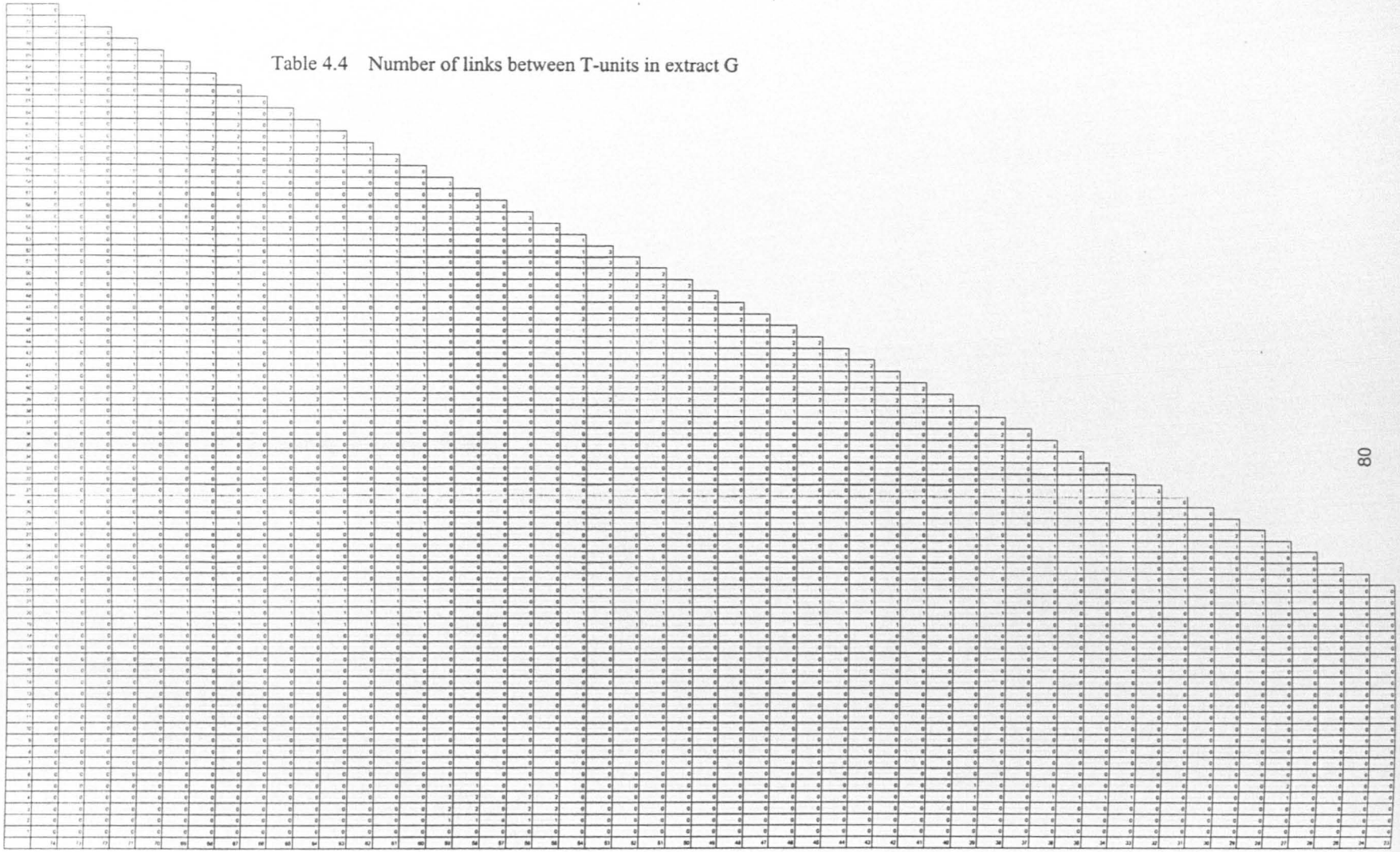


Table 4.4 cont. Number of links between T-units in extract G

The first point that is noticeable from Table 4.4 is the general paucity of links. The vast majority of links are on or near the hypotenuse of the table indicating that links generally occur between units close to each other in the example. The dearth of links in the bottom left-hand section of Table 4.4 indicates that there are very few links between units distant from each other in the example.

There are two ways in which we can look at this point in more detail. Firstly, Hoey (1991) argues that sentences (or in this case, T-units) are bonded together when they share at least three links. Under these conditions, only 1.44% of the possible combinations of T-units would be bonded. As we shall see when we draw up networks of bonds, this is probably too small a proportion to be useful for analysis and it may be better to set the requisite number of links for T-units to be considered bonded at two links.

Secondly, with most links occurring between T-units which are close to each other, we should expect a higher proportion of bonded units when we analyse shorter examples than longer examples. We can test this hypothesis by looking at the correlation between the number of possible combinations of T-units in an example and the average density of links for the example. For extract G, there are 945 links in total and there are 2,701 possible combinations of T-units. The average density of links is therefore

$$945 / 2{,}701 = 0.35$$

for the extract which is 74 T-units long. Making the same calculation for all twelve extracts in this study, we find the figures given in Table 4.5.

From Table 4.5, we find that there is a non-significant inverse relationship between the number of T-units in an extract and the average density of linkage (r = -0.17) showing that, in general, the longer the extract the lower the average density. In addition, all of the densities are considerably lower than those found by Hoey (1991) in an analysis of a 16-sentence written text (the average density in Hoey ranges from 1.02 to 1.08 depending on whether arguable cases in the analysis are included or not).

Extract	Number of T-units	Average density of linkage
L	34	0.66
H	34	0.22
J	40	0.57
K	47	0.35
G	74	0.35
C	76	0.25
I	82	0.21
E	84	0.28
D	120	0.14
F	125	0.09
B	186	0.13
A	237	0.54

Table 4.5 Comparison of number of T-units and average density of linkage

Overall then, it appears that extract G and the other extracts in this study are far more loosely connected than the extracts of written language examined by Hoey. This points towards a possible interesting contrast between spoken and written language, whereby spoken discourse which usually consists of shorter T-units than are found in written language may have fewer lexical connections between T-units (it should, however, be acknowledged that findings concerning classroom discourse may not be applicable to other types of spoken language). However, the findings concerning overall connectedness do not help us in identifying topics and topic progression. To do that, we must look in more detail at bonding between T-units.

4.2.2 Identifying bonded T-units and clusters

As we have seen, bonds exist between T-units which are linked at or above a predetermined level, which we might set at 2, 3 or 4 links. These bonds are shown in the values in Table 4.4, but the relationships between different bonded pairs of sentences are unclear. To clarify these bonds, we need to look at ways of re-representing the data in Table 4.4 focusing on bonds. Hoey himself suggests two ways of doing this.

The more recent method (Hoey, 1995) is to list the T-units as a central column with bonded T-units appearing earlier in the extract to the left of the central column and bonded T-units appearing later in the extract to the right.

Such a table for extract G with the threshold level for bonds set at 2 links is shown in Table 4.6 below.

	1	
	2	4
	3	4, 55, 56, 58, 59
2 , 3	**4**	5, 55, 56, 58, 59
4	**5**	6, 7, 27
5	**6**	7
5, 6	**7**	
	8	9, 10, 11, 12, 13
8	**9**	10, 11, 12, 13
8, 9	**10**	11, 12, 13
8, 9, 10	**11**	12, 13
8, 9, 10, 11	**12**	13
8, 9, 10, 11, 12	**13**	
	14	
	15	16, 17
15	**16**	*17*
15, *16*	**17**	
	18	*19, 20*
18	**19**	*20*
18, 19	**20**	
	21	*22*
21	**22**	
	23	
	24	25
24	**25**	
	26	
5	**27**	29, 31, 32
	28	
27	**29**	*31, 32*
	30	
27, *29*	**31**	*32*
27, *29, 31*	**32**	
	33	34, 35, 36, 38
33	**34**	35, 36, 38
33, 34	**35**	36, 38
33, 34, 35	**36**	38
	37	
33, 34, 35, 36	**38**	
	39	*40*, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 60, 61, 64, 65, 68, 71, 74
39	**40**	41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 60, 61, 64, 65, 68, 71, 74
39, 40	**41**	46, 48, 49, 50, 51, 52, 53
39, 40	**42**	43, 46
39, 40, 42	**43**	46
39, 40	**44**	45, 46
39, 40, 44	**45**	46
39, 40, 41, 42, 43, 44, 45	**46**	48, 49, 50, 51, 52, 53, 54, 60, 61, 64, 65, 68
	47	
39, 40, 41, 46	**48**	49, 50, 51, 52, 53
39, 40, 41, 46, 48	**49**	50, 51, 52, 53
39, 40, 41, 46, 48, 49	**50**	51, 52, 53

39, 40, 41, 46, 48, 49, 50	**51**	*52, 53*
39, 40, 41, 46, 48, 49, 50, *51*	**52**	*53*
39, 40, 41, 46, 48, 49, 50, *51, 52*	**53**	
39, 40, 46	**54**	60, 61, 64, 65, 68
3, 4	**55**	*56, 58, 59*
3, 4, *55*	**56**	*58, 59*
	57	
3, 4, *55, 56*	**58**	*59*
3, 4, *55, 56, 58*	**59**	
39, 40, 46, 54	**60**	61, 64, 65, 68
39, 40, 46, 54, 60	**61**	64, 65, 68
	62	63, 67
62	**63**	67
39, 40, 46, 54, 60, 61	**64**	65, 68
39, 40, 46, 54, 60, 61, 64	**65**	68
	66	
62, 63	**67**	
39, 40, 46, 54, 60, 61, 64, 65	**68**	69, 70, 71
68	**69**	
68	**70**	
68	**71**	74
	72	
	73	
39, 40, 71	**74**	

Note: Underlined bonded T-units share at least 4 links, italicised bonded T-units share 3 links, and bonded T-units in normal type-face share 2 links.

Table 4.6 Summary of the bonds in extract G

While Table 4.6 provides a more succinct representation of the bonded T-units in extract G than Table 4.4, the interrelationships between bonded T-units, especially concerning how they cluster, is still unclear. We must therefore turn to the second and earlier of Hoey's suggested methods of representing bonds.

Hoey (1991) suggested that bonds between sentences could best be represented by drawing up networks where lines between different units in the analysis indicate bonds. Before we start drawing up such networks, however, we need to consider the minimum number of links necessary for considering T-units to be bonded.

In determining the requisite level of links for bonding, Hoey (1991: 265) states:

> "Usually, the requisite number of links is three, and it is never less than three; but sometimes for texts ... in which there are a

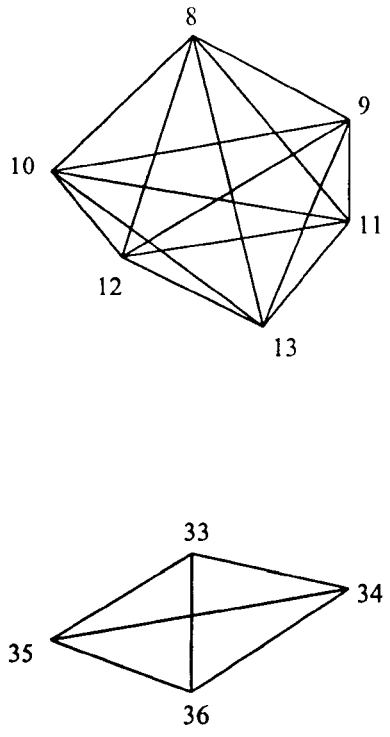
great number of repetitions ..., the threshold may be four links or more."

From this, we are left with a choice of 3 or 4 links to act as the requisite number for bonding. However, as we saw above, the extracts considered in this study have a much lower general level of connectedness through links than the texts that Hoey analysed. On this basis, we may need to ignore Hoey's guidelines and set a threshold of 2 links. Which threshold to set is not immediately apparent and may require construction of networks based on all possible thresholds to decide which is the most appropriate level.

There is, however, another way of looking at the decision behind the threshold. If, when we construct networks of bonds, we find that the networks form clusters, we may say that these clusters form topical sets as bonded sentences are semantically related (Hoey, 1991). Starting with the largest possible threshold would highlight the most strongly bonded clusters or the most semantically related clusters indicative of topical sets. If we then progressively lower the threshold, more bonds between T-units would be added to the network. As additional clusters emerge and these clusters bond with other clusters, superordinate topical sets may emerge from such cross-cluster bonds. We may then be able to identify a hierarchical structure of topics in the extract. Although such an approach presents a static view of topics, it should allow us to see where progression between topics and sub-topics occurs in the extracts and thus be of use.

Let us then start our re-representation of the data in Table 4.4 by drawing up a network of bonds where the minimum number of links necessary for two T-units to be considered bonded is 4. This network is shown in Figure 4.1.

From Figure 4.1, we can see that there are two clusters (T-units 8-10 and 33-36) of strongly bonded T-units in extract G. We can consider these two clusters to be very strongly semantically related and to be two instances of monotopical discourse.



bonds with 4 links

bonds with 5 links

Figure 4.1 Network of bonds (threshold = 4) for extract G

We can now lower the threshold of bonding to three links, and a network based on this is shown in Figure 4.2.

From Figure 4.2, we can see that several more clusters indicative of monotopical discourse have emerged in addition to the previously identified clusters. There are, however, no cross-cluster bonds between any of these nine clusters.

It is only when we reduce the threshold to two links, as in Figure 4.3, that any cross-cluster bonds emerge. In Figure 4.3, some of the previously identified clusters have emerged 'unscathed' with no additional bonds being added to the clusters (T-units 8-13; 18-20; 21-22). Others are relatively

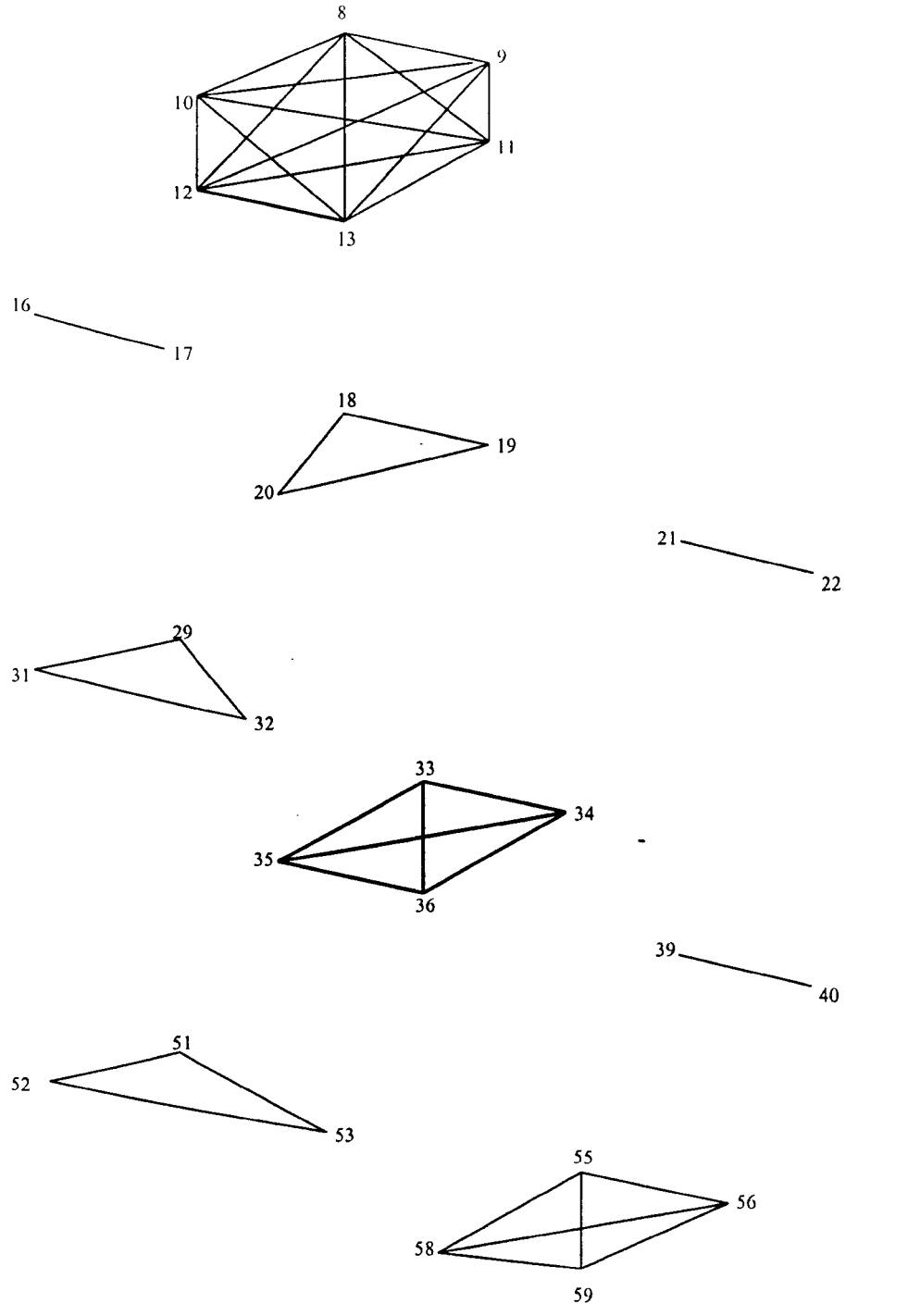


Figure 4.2 Network of bonds (threshold = 3) for extract G

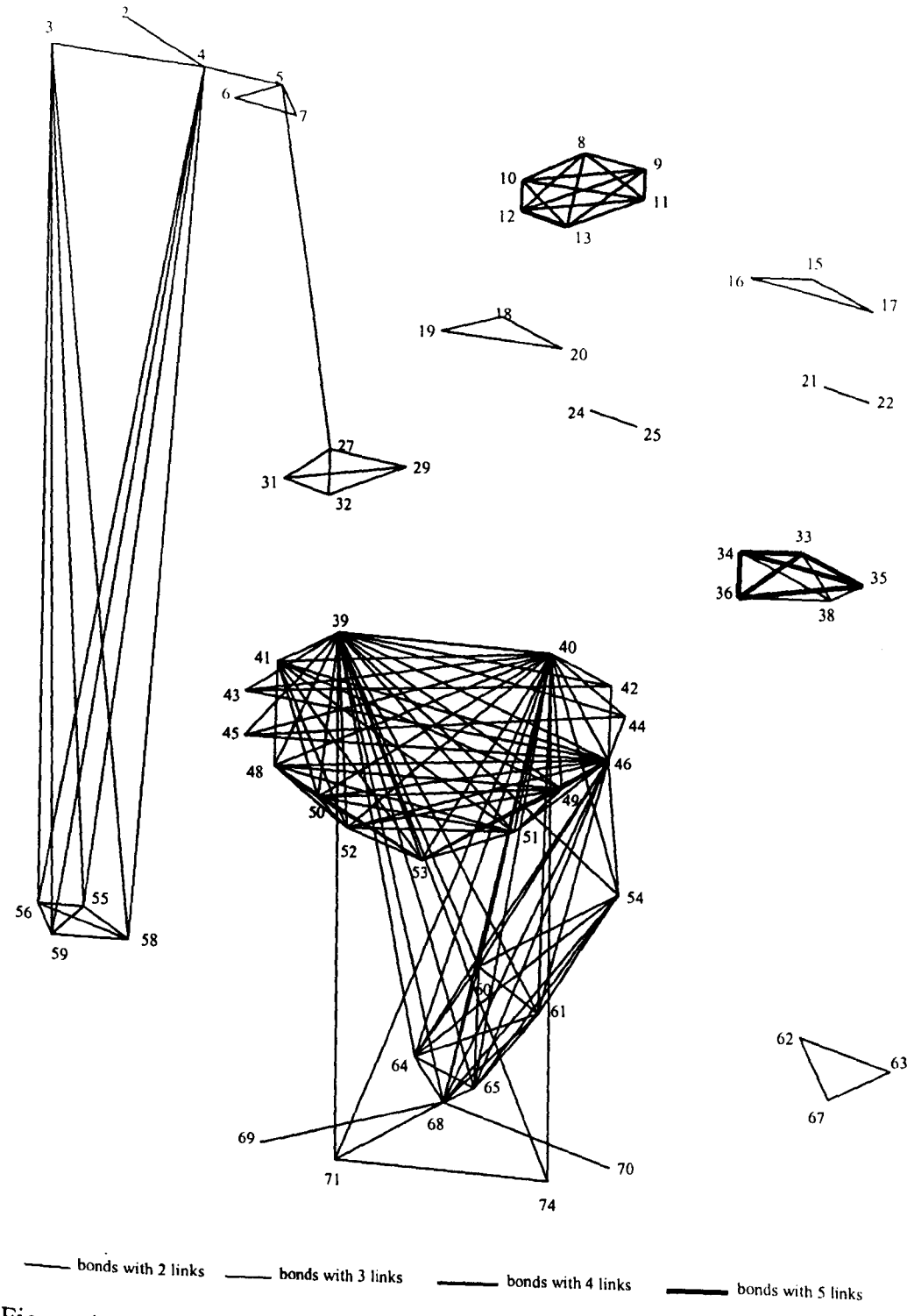


Figure 4.3 Network of bonds (threshold = 2) for extract G

unchanged with extra T-units being added to the cluster but no cross-cluster bonds being added (T-units 15-17; 33-36). The other clusters from Figure 4.2 have been joined with cross-cluster bonds in Figure 4.3 while also expanding cluster size considerably. This is most noticeable for the previously identified clusters of T-units 39-40 and 51-53 which now form part of a massive cluster extending (with breaks) from T-unit 39 to T-unit 74. Similarly, cluster 29-32 is now associated with cluster 55-59, although there are no direct bonds between these two clusters. Finally, some previously hidden clusters have emerged by lowering the threshold (T-units 24-25; 62-63 and 67).

To identify the topics of monotopical clusters, we can see which lexical items provide the links within these clusters. For example, the cluster of T-units 8-13 is linked by the proposition *the boy is taking people to his house* which contains the reiterated lexical items which create the bonds between the T-units. In the same way, we can also see how topics are nested within superordinate topics. For example, the cluster of T-units 51-53 apparent when the bonding threshold is set at 3 links is about *robots can cook food*, but is also subordinate to the topic of the very large cluster of T-units 39-61, 64-65 and 68-71 which emerges when the threshold is set at 2 links. The topic of this larger cluster is *what robots can do*. One final point about the clusters is that they largely coincide with exchanges (for example, the cluster of T-units 8-13 fits with exchanges 4-5, see p. 76). From this, then, we can draw up a table showing how clusters of lexical bonds fit with exchanges and the topics of these clusters based on frequently reiterated lexical items. This table is given as Table 4.7.

Bonded clusters (T-units)	Exchanges	Topics
3-4, 55-59	1-2, 17	*seeing a **film***
8-13	4-5	*the **boy** is taking **people** to his **house***
15-17	6	*the **woman** is the **boy's mother***
18-20	7	*the **boy** is talking about his **robot***
27-32	10	*the other robot's name is **R2D2***
33-36, 38	11	*why **R2D2** says **C3PO** is **naked***
39-61, 64-65, 68-71	13-17, 19-20	*what **robots** can do*
39-40		*work of **robots***
51-53	15	***robots cook food***
62-63, 67	18	***robots** can be **fighters***

Note: Indentation in the first column indicates sub-clusters.
Words in bold italics in the right-hand column are words which represent key concepts in the discourse (see chapter 5)

Table 4.7 Topics of bonded clusters and exchanges in extract G

From Figure 4.3 we can also identify points in the extract where there appear to be shifts or drifts in the topic. For example, the monotopical stretch of discourse from T-units 8 to 13 appears to be semantically isolated, so that we can expect some kind of break in topic immediately before T-unit 8 and after T-unit 13. Throughout the whole of extract G based on Figures 1 to 3, there appear to be breaks before the following T-units:

8, 15, 18, 21, 24, 27, 33, 39, 55, 60, 62, 64, 67, 68

These points in the extract represent jumps between semantically related clusters in the discourse. To check whether these points really represent disjunctions in topic in the discourse, there are two things we can do.

Firstly, we can look at running averages of connectedness across, say, three consecutive T-units. For example, for T-units 1-3, we find:

1-2 0 links
1-3 0 links
2-3 1 link

The total number of links between all combinations of T-units 1, 2 and 3 is 1 link giving an average connectedness of 0.33 for the three units. Following the same method for constructing averages of connectedness throughout extract G, we find the running averages of connectedness given in Table 4.8.

T-units	Average connectedness	T-units	Average connectedness	T-units	Average connectedness
1-3	0.33	25-27	0.33	49-51	2.00
2-4	1.67	26-28	1.00	50-52	2.33
3-5	1.67	27-29	1.33	51-53	3.00
4-6	1.67	28-30	0.33	52-54	1.67
5-7	2.00	29-31	1.00	53-55	0.33
6-8	0.67	30-32	1.00	54-56	1.00
7-9	1.33	31-33	1.67	55-57	1.00
8-10	4.00	32-34	2.33	56-58	1.00
9-11	4.00	33-35	5.00	57-59	1.00
10-12	4.00	34-36	5.00	58-60	1.00
11-13	4.33	35-37	1.67	59-61	0.67
12-14	2.00	36-38	0.67	60-62	1.33
13-15	0.33	37-39	0.33	61-63	1.33
14-16	0.67	38-40	1.67	62-64	1.33
15-17	2.33	39-41	2.33	63-65	1.33
16-18	1.67	40-42	1.67	64-66	0.67
17-19	1.67	41-43	1.33	65-67	0.33
18-20	3.00	42-44	1.33	66-68	0.33
19-21	1.67	43-45	1.33	67-69	1.33
20-22	1.67	44-46	2.00	68-70	1.67
21-23	1.00	45-47	0.67	69-71	1.00
22-24	0.00	46-48	0.67	70-72	0.33
23-25	0.67	47-49	0.67	71-73	0.33
24-26	0.67	48-50	2.00	72-74	0.33

Table 4.8 Running averages of connectedness for consecutive trios of T-units in extract G

We can expect that bonded consecutive trios of T-units would show high averages. Trios without ideational coherence, on the other hand, would show low averages. Based on this, points in the extract where there are topic breaks should show as large changes in the running averages in Table 4.8.

Taking large changes as being any change of at least 1.33 in value, they appear to occur before the following T-units (where positive changes occur before the first of the T-units in a trio, and negative changes occur before the last of the T-units in a trio):

2, 8, 14, 15, 18, 21, 33, 37, 38, 47, 48, 54, 55

The second method of checking is to change the unit of analysis. Although T-units are generally viewed as being the spoken equivalents of sentences (see p. 51), the comparatively low levels of connectedness apparent in the analysis based on T-units suggests that we could also consider using a larger unit of analysis such as the exchange.

4.2.3 Identifying bonded exchanges

Basing our analysis on the exchange, however, raises a problem. In Hoey's analysis, multiple reiterations of a lexical item within one unit can be linked to multiple within-unit reiterations of the same lexical item in another unit, counting them as multiple links. Applying this to extract G, however, means that, in the second half of the extract, the number of links between exchanges is dominated by multiple within-exchange reiterations of *robot*, to the extent that findings are distorted. For example, if we include multiple within-exchange reiterations in the analysis, the highest number of links between two exchanges occurs between exchanges 10 and 15. All of these links are reiterations of *robot*. From the findings at the T-unit level, there is no evidence of any bonds between T-units in these two exchanges. It therefore seems that including multiple within-exchange reiterations in the analysis has led to the positing of a very strong bond between two exchanges that is not indicated in other analyses. To avoid such distortions, multiple within-exchange reiterations are counted as single instances of a lexical item.

On this basis, we can construct a table of links between the 21 exchanges in extract G as shown in Table 4.9.

The average level of connectedness in Table 4.9 is 0.66, a figure matching the highest average density found for analyses based on T-units but still low enough to suggest setting the threshold level for bonding at 2 links. Doing this, we can construct a network of bonds as shown in Figure 4.4.

20 (68-70)	1																		
19 (64-67)	1	2																	
18 (60-63)	1	2	3																
17 (55-59)	0	0	0	0															
16 (54)	1	2	2	2	0														
15 (48-53)	1	1	1	1	1	1													
14 (46-47)	1	2	2	2	0	2	1												
13 (38-45)	2	2	2	2	0	2	4	2											
12 (37)	0	0	0	0	0	0	0	0	0										
11 (33-36)	1	1	1	1	0	1	1	1	3	0									
10 (26-32)	1	1	1	1	0	1	1	1	3	0	2								
9 (24-25)	1	1	1	1	0	1	1	1	2	0	2	1							
8 (21-23)	1	1	1	1	0	1	1	1	1	0	1	3	1						
7 (18-20)	1	1	1	1	0	1	1	1	1	0	2	1	1	1					
6 (15-17)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1				
5 (12-14)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1			
4 (8-11)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	4		
3 (5-7)	0	0	0	0	1	0	0	0	1	0	0	2	0	0	0	0	0	0	
2 (4)	0	0	0	0	3	0	0	0	1	0	0	1	0	0	0	0	0	0	2	
1 (1-3)	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3
Exchange (T-units)	21 (71-74)	20 (68-70)	19 (64-67)	18 (60-63)	17 (55-59)	16 (54)	15 (48-53)	14 (46-47)	13 (38-45)	12 (37)	11 (33-36)	10 (26-32)	9 (24-25)	8 (21-23)	7 (18-20)	6 (15-17)	5 (12-14)	4 (8-11)	3 (5-7)	2 (4)

Table 4.9 Lexical links between exchanges in extract G

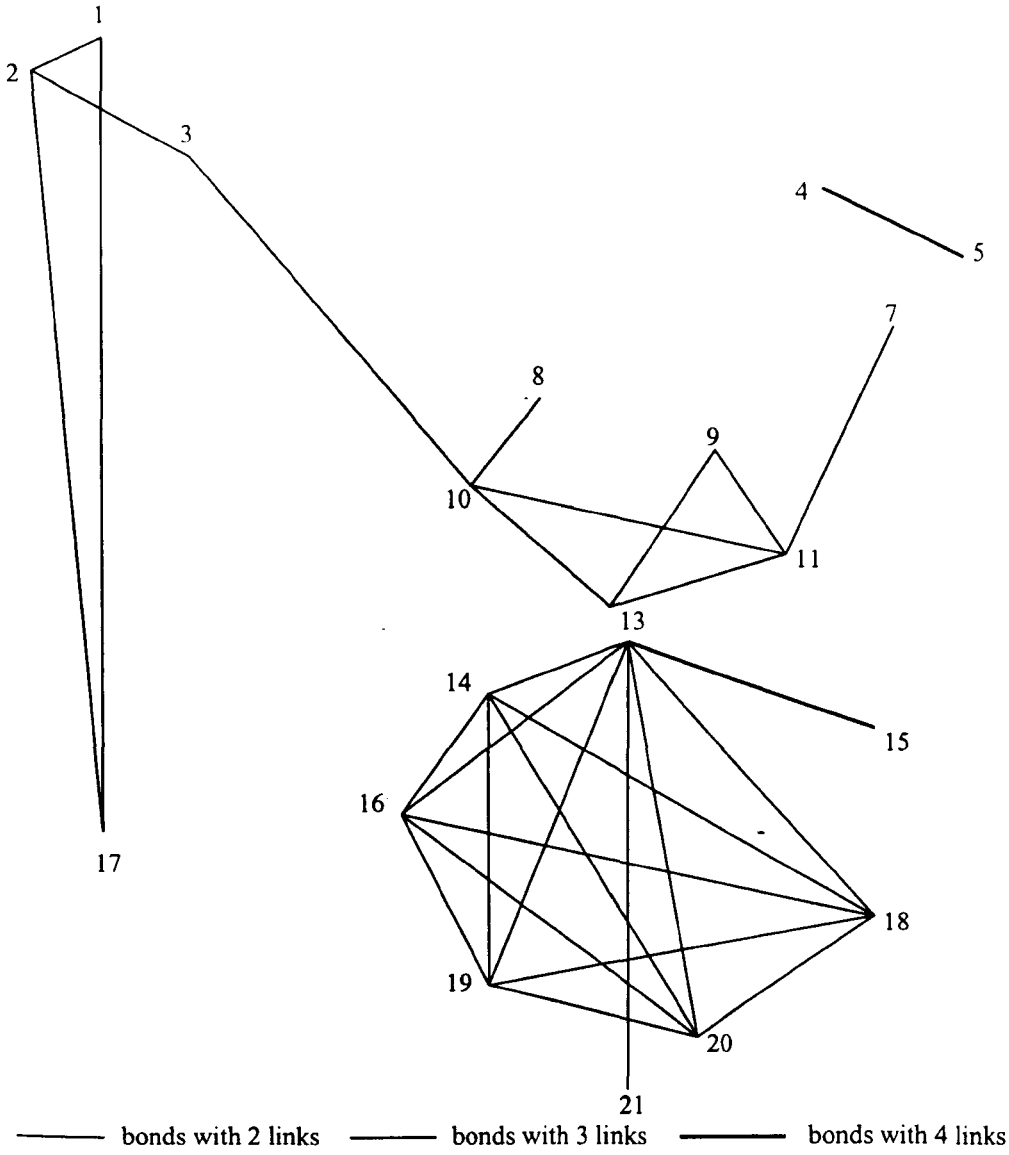


Figure 4.4 Network of bonds based on exchanges (threshold = 2) for extract G

The first thing we can notice from Figure 4.4 is that there is a lot more interconnectedness in the network than in the previous networks where the unit of analysis was the T-unit. In fact, the network looks reminiscent of the networks in Hoey (1991). Nevertheless, some of the patterns of the networks

based on analysis from T-units are carried over into the network derived from analysis of exchanges. For example, the isolated cluster of T-units 8-13 still exists as the isolated cluster of exchanges 4-5; and the very large cluster in the second half of extract G is still apparent.

Some breaks in topic progression are also apparent in Figure 4.4. There are breaks before and after exchanges 4-5, and presumably before and after the unbonded exchanges 6, 12 and 17. In terms of T-units, these breaks would occur at the following points in extract G:

8, 15, 18, 37, 38, 55, 60

However, one further point that emerges from Figure 4.4 which we have not previously been in a position to identify is the existence of a central unit. In Figure 4.4, exchange 13 is bonded to ten other exchanges which come both before and after exchange 13 in the extract. The level of bonding of exchange 13 is far higher than that of any other exchange making it the central exchange of extract G. If we wished to identify a single topic or main idea for extract G, it would probably be wise to choose a proposition or phrase containing the reiterated lexical items from exchange 13, the central exchange, such as *the kinds of work robots can do*. Although providing a potentially useful benchmark against which to compare findings from other methods of analysis, such a 'main idea' approach to topic takes a far too static perspective on discourse to be applicable in this study.

At this point, we have identified three ways in which breaks in topic can be identified. To allow comparison of these, they are summarised in Table 4.10.

Breaks identified from networks based on T-units	8, 15, 18, 21, 24, 27, 33, 39, 55, 60, 62, 64, 67, 68
Breaks identified from running averages	2, 8, 14, 15, 18, 21, 33, 37, 38, 47, 48, 54, 55
Breaks identified from network based on exchanges	8, 15, 18, 37, 38, 55, 60

Table 4.10 Summary of breaks in topic identified from three approaches

If we set the condition that, to be considered a break in topic, the break must appear under at least two of the approaches, we find that extract G contains breaks before the following T-units:

8, 15, 18, 21, 33, 37, 38, 55, 60

Conducting a lexical analysis following Hoey (1991), therefore, allows us to identify two key aspects of topic in classroom discourse:

1. Points in the discourse where there is a jump in the topic indicative of topic shift or perhaps drift.
2. Ways of expressing what the topic is, based on frequently reiterated lexical items.

Although these do not show us how topics progress through the discourse (something which can only be shown by an analysis which is more dynamic in nature than the rather static lexical analysis), they do provide the foundations on which any analysis of topic progression must be built and a valuable benchmark against which the validity of other methods of analysis can be measured.

4.3 Approaches to topic based at the clause level

A second set of approaches to identifying topics and topic progression is based on analyses of language focusing at the clause level which originate in the work of the Prague school of linguistics and M. A. K. Halliday (e.g. 1967, 1970). These approaches include theme-rheme progression and given-new progression. Despite some arguments that there is a strong correlation between theme and given on the one hand, and rheme and new on the other (e.g. Connor, 1994; Daneš, 1974; Fries, 1983, 1994), I will treat the two approaches separately. Comparison of the findings concerning topics and topic progression can then be made to see whether it is worth distinguishing between the two approaches for the purposes of analyses of topic.

4.3.1 Theme-rheme progression

Following a systemic perspective, theme-rheme analyses divide a sentence into two components. The theme is "the heading to what I am saying" (Halliday, 1970: 163), the starting point of the sentence (Brown and Yule, 1983a), or the framework for interpreting the rest of the sentence (Fries, 1994). The rheme, on the other hand, is what is said about the theme (Connor, 1996). (For a full review of distinctions in meaning in the use of the terms *theme* and *rheme*, see Peng, 1999).

Since the theme is the point of departure for a sentence, perhaps unsurprisingly in English themes are generally sentence-initial (although some other languages indicate themes in other ways, see Givón, 1983).

Beyond general agreement that themes appear early in sentences, however, theme-rheme analysts have argued for various criteria for identifying themes in English. Among the criteria used are the following:

- "the initial constituent of a clause or sentence" (Fries, 1983: 116)
- the left-most constituent of a sentence ignoring adverbials of organisation and metalinguistic comments (Brown and Yule, 1983a)
- "clause initial elements up to and including the first ideational element" (Berber Sardinha, 1997: 69, following Halliday, 1967)
- everything preceding the main verb of the main clause (McCarthy and Carter, 1994)
- the grammatical subject with all pre-subject elements serving the purpose of contextualising (Davies, 1994).

In this study, I will provisionally use Halliday's criterion for identifying themes, since, of the five criteria, it is the most influential and widely used. In addition, Halliday's criterion for theme has been effectively used in analyses of spoken language (e.g. Halliday, 1989; Hasan, 1989).

To see how Halliday's criterion for theme can be applied in practice, we can look at the first 14 T-units of extract G (given here as example 4.13). This is shown below with themes underlined and rhemes in italics. It is worth noting that, since thematic analyses focus on how speakers phrase their messages, ellipted material is not included in the analysis.

Example 4.13

T: [1] **OK**. [2] I *'m going to play a video* [3] and we *'re going to see a scene from a, from a film.* [4] You *watch the video and tell me uh what you see, OK*? [T turns on the video to show a 3-second excerpt from Star Wars Episode 1. T pauses the film.] [5] Do you *know what film is it* (the film)?

S: [6] The film is Star Wars Episode One.

T: [7] The film is Star Wars Episode One, *right, Star Wars Episode One.* [T plays the video for another ten seconds. The scene is set in a desert village. T pauses the video.] [8] Where, where *do the boy take these people to*? (2.0) [9] Where *do this boy, this boy take these people to*?

S: [10] The boy takes these people to His house.

T: [11] The boy takes these people to His place, *his house.* [12] Why does the boy take these people to his house?

S: [13] The boy takes these people to his house because Storm.

T: [14] Yes, there is a storm, sandstorm, *right*?

In analysing example 4.13, there are a few problems that need to be overcome. Having identified framing moves as separate T-units (see chapter 3), *OK* in T-unit 1 needs to be identified as a theme even though it is not clearly an ideational element, since all T-units contain themes (I am assuming that themes can be assigned to all T-units, including ones which do not contain a full clause). In T-units 2, 3, 4 and 5, the first ideational element also plays a role in the interpersonal metafunction of the discourse. Nevertheless, as the first ideational elements in each T-unit, the personal pronouns are identified as themes (but see below).

In T-unit 6, we run into a different problem. If we include the ellipsis, the first ideational element is *The film*. There are, however, two reasons for not including ellipsis in the process of identifying themes. Firstly, most writing on themes highlights the importance of the sequencing of items in sentences. In filling in ellipsis, the unmarked sequence is preferred so that T-unit 6 is *(The film is) Star Wars Episode One* rather than the marked *Star Wars Episode One (is the film)*. There is nothing, however, to indicate that the ellipsis should be added in an unmarked form except that such forms are more frequent. Identifying *The film* as the theme would entail making the suspect assumption that ellipted materials always follow unmarked forms. If we include ellipted material for the purpose of identifying theme, then, it is not clear whether the theme for T-unit 6 should be *The film* or *Star Wars Episode One*. Secondly, in addition to conducting this theme-rheme analysis, we shall also be conducting a given-new analysis where ellipsis must be taken into account. To maximise the productivity of conducting the two analyses, it may be valuable to distinguish between the two approaches by including ellipsis in one analysis but not in the other. Since it is necessary to include ellipsis in a given-new analysis, it would be more productive not to include ellipsis in a theme-rheme analysis. For these reasons, ellipsis is not considered when identifying themes, so the theme of T-unit 6 is *Star Wars Episode One*.

Continuing in the same way, identification of themes in T-units 7-13 is straightforward. In T-unit 14, however, we need to consider whether the theme is *there* or the whole clause. There are two arguments for considering the whole clause as the theme. Firstly, from the context, T-unit 14 is a rephrased acceptance of T-unit 13. Therefore, identifying something other then the confirmed *storm* as the theme in T-unit 14 is counter-intuitive. Secondly, in example 4.13 *there is* serves an existential rather than a locative function (Quirk et al., 1985). In this case it serves the purpose of identifying

"the focus of interest" (Biber et al., 1999: 952). In other words, *there is* highlights *a storm* as the focus of interest of the continuing discourse. I will therefore treat the whole of T-unit 14 as the theme.

Identifying themes and rhemes, however, is only the first step in identifying theme-rheme progression. The purpose of themes is to provide an organisation for the discourse with rhemes providing the message that pushes the communication forward (Daneš, 1974). To organise the discourse, themes may connect back to previous concepts in the discourse or they may initiate further developments in the discourse. These purposes of themes mean that, in addition to identifying the themes themselves, we also need to identify points in the surrounding discourse which are related to the themes. Different relationships with the surrounding discourse can be classified as different kinds of thematic progression, and determining the kinds of progression is the second stage in conducting an analysis of theme-rheme progression.

As many different kinds of theme-rheme progression have been identified as there are possible permutations of themes and rhemes (see e.g. Peng, 1999). However, only five types which are particularly pertinent to this study will be presented.

1. Parallel progression
   Termed thematic progression with continuous theme by Daneš (1974), parallel progression (Connor and Farmer, 1990; Schneider and Connor, 1990) involves use of the same theme for consecutive T-units.
2. Sequential progression
   Termed simple linear thematic progression by Daneš (1974), sequential progression (Connor and Farmer, 1990; Schneider and Connor, 1990) involves the rheme of one T-unit becoming the theme of the succeeding T-unit.
3. Extended parallel progression
   As with parallel progression, extended parallel progression (Connor and Farmer, 1990) involves the repetition of a previous theme. However, in extended parallel progression, the repeated theme is not the theme of the immediately preceding T-unit but is the theme of some other earlier T-unit.

4. Thematic progression with derived themes
   Thematic progression with derived themes (Daneš, 1974) occurs where the themes of consecutive T-units are not directly related to each other but are all subordinate to some unmentioned hypertheme.
5. Coherence breaks
   Where a new theme and rheme which have never been previously used in the discourse or which have not been used within the most recent 10 T-units are introduced, this is termed a coherence break (Wikborg, 1990).

In addition to these five main types of progression, we may also need an 'other' category for other varieties of theme-rheme progression between consecutive T-units, including extended sequential progression, progression where the theme of one T-unit becomes the rheme of the succeeding T-unit, and progression where the rhemes of succeeding T-units are the same.

There are various ways of representing these types of thematic progression. Perhaps the most succinct is that suggested by van Dijk (1977) who indicates T-units with angular brackets. The first item within the brackets is the theme and the second the rheme. Substituting letters for the concepts expressed in themes and rhemes, it is relatively easy to follow the kinds of progression through a text. Using van Dijk's notation, the five types of thematic progression described above can be represented as in Table 4.11.

Parallel progression	<a,b>,<a,c>,<a,d>...
Sequential progression	<a,b>,<b,c>,<c,d>...
Extended parallel progression	<a,b>,<b,c>,<c,d>,<a,e>...
Thematic progression with derived themes (where the hypertheme = a)	$<a_1,b>,<a_2,c>,<a_3,d>$...
Coherence breaks	<a,b>,<c,d>,<e,f>...

Table 4.11 Types of thematic progression

While applying parallel progression, sequential progression and coherence breaks in an analysis is fairly straightforward, extended parallel progression and thematic progression with derived themes are more problematic. For extended parallel progression, we need to set a limit to the length of the discourse which appears between the repeated themes. This limit should have psychological validity by reflecting the amount of intervening discourse that would lead to a theme being treated as a repeated theme or as a new theme. As I am unaware of any research concerning this, I have

somewhat arbitrarily set the limit for repeated themes to be considered as extended parallel progression at 10 intervening T-units.

Thematic progression with derived themes is even more problematic. As the hypertheme is unmentioned, how are we to show that consecutive themes which are not directly related are, in fact, indirectly related as co-subordinates of a hypertheme? In some ways, it could be argued that, in searching for ways of identifying topics and topic progression in this study, we are in fact looking for hyperthemes (see Brown and Yule, 1983a for a discussion of possible relationships between sentence themes, hyperthemes and discourse topics). If this is the case, it seems premature at this point to start trying to identify hyperthemes. If it is not the case that hyperthemes and topics refer to the same entity, we are still at a loss for a way to identify hyperthemes without resorting to subjective impressions. Although identifying unmentioned hyperthemes is problematic, we may still be able to use the concepts of hyperthemes and derived themes if both are explicitly mentioned in the text. For example, we may imagine a text like that in example 4.14.

Example 4.14
[1]The furniture was old. [2]The chair was battered. [3]The table was wobbly.

In example 4.14, *furniture* is the theme of sentence 1. It is also the mentioned superordinate or hypertheme of sentences 2 and 3. Since the hypertheme is mentioned, we are not faced with the problems of identifying an unmentioned hypertheme. In the analysis, therefore, I will include thematic progression with derived themes where the hypertheme is mentioned.

We are now therefore in a position to start an analysis of thematic progression, identifying themes following Halliday (1967) and classifying types of progression as parallel progression, sequential progression, extended parallel progression, thematic progression with derived themes and mentioned hyperthemes, other progression, or coherence breaks, identified in that order for analysis (i.e. if one pair of T-units exhibits parallel progression, extended parallel progression and rheme-to-rheme other progression, it will be counted as parallel progression). Let us start by applying the analysis to example 4.13 given above.

In example 4.13, the first T-unit has a theme but no rheme. The second unit does not contain the theme of the first unit, and therefore there is a coherence break between T-units 1 and 2. In fact, given the restricted number of items that can be included in framing moves such as T-unit 1, we may expect most framing moves to be preceded and succeeded by coherence breaks, especially where the contents of the two transactions separated by the framing move are different. This provides our first indication of intermethod correspondences in this study, and perhaps one of the functions of Sinclair and Coulthard's (1975) framing moves is to create coherence breaks in the discourse.

The progression between units 2 and 3 and between units 3 and 4 presents a minor problem. Should we count the progression of themes from *I* to *we* and from *we* to *you* as parallel progression, thematic progression with derived themes and mentioned hypertheme, or coherence breaks? The themes are not the same ruling out parallel progression, but both *you* and *I* could be considered components of *we*. I will count these as meronymic derived themes of the hypertheme *we*, and thus the progression from T-unit 2 to T-unit 3 and from 3 to 4 is thematic progression with derived themes and mentioned hypertheme. In T-units 4 and 5, we find identical themes exhibiting parallel progression.

The rheme in T-unit 5 becomes the theme in T-unit 6, giving sequential progression, and this new theme is repeated in T-unit 7, giving parallel progression. There are no links between T-unit 8 and the preceding discourse indicating a coherence break between T-units 7 and 8. The theme regarding location in T-unit 8 is then repeated (albeit more explicitly in units 10 and 11) in T-units 9, 10 and 11, giving three instances of parallel progression.

With T-units 12 and 13, we again face problems in the analysis. The themes identified in example 4.13 (*why* in unit 12, and *storm* in unit 13) do not occur in the preceding discourse so we should find coherence breaks between T-units 11 and 12 and between T-units 12 and 13. Looking at the ellipted material identified in the transcript, however, there are clear links between the units, and especially between T-units 12 and 13 where T-unit 13 is the answer to the question in T-unit 12. Identifying the progression between a question and its answer as a coherence break is clearly wrong. In earlier T-units, I have equated *where* and *his house* as repetitions of the same theme. To avoid a coherence break between T-units 12 and 13, we would

have to equate *why* with *storm*. Although this is a more tenuous connection between themes than that between *where* and *his house*, it seems a reasonable way to avoid the problems of a coherence break between T-units 12 and 13. The coherence break between T-units 11 and 12 cannot be avoided in the same way, and since we are not including ellipted material in the analysis, the coherence break between T-units 11 and 12 must stand. Finally, between T-units 13 and 14, there is parallel progression.

Using van Dijk's (1977) notation, the themes, rhemes and kinds of progression can be presented as in Table 4.12.

T-unit 1	<a>	
		coherence break
T-unit 2	<b_1,c>	
		thematic progression with derived themes and mentioned hypertheme
T-unit 3	<b,d>	
		thematic progression with derived themes and mentioned hypertheme
T-unit 4	<b_2,e>	
		parallel progression
T-unit 5	<b_2,f>	
		sequential progression
T-unit 6	<f>	
		parallel progression
T-unit 7	<f,f>	
		coherence break
T-unit 8	<g,h>	
		parallel progression
T-unit 9	<g,h>	
		parallel progression
T-unit 10	<g>	
		parallel progression
T-unit 11	<g>	
		coherence break
T-unit 12	<i>	
		parallel progression
T-unit 13	<i>	
		parallel progression
T-unit 14	<i>	

Table 4.12 Theme-rheme progression in extract G

Continuing in the same way for all of the data in this study, we find the frequencies for each kind of progression in each of the extracts as shown in Table 4.13.

Extract	Parallel progression	Sequential progression	Extended parallel progression	Thematic progression with derived themes and mentioned hypertheme	Other types of progression	Coherence breaks	Total
A	75	44	50	9	11	47	236
B	74	20	40	4	13	34	185
C	31	4	14	0	6	20	75
D	46	6	30	2	9	26	119
E	31	5	28	3	5	11	83
F	62	4	21	2	15	20	124
G	29	9	14	3	7	11	73
H	13	2	6	0	3	9	33
I	45	8	4	1	12	11	81
J	13	2	10	1	4	9	39
K	26	6	3	0	5	6	46
L	18	1	8	0	3	3	33
Total	463	111	228	25	93	207	1127
Percentage	41.08	9.85	20.23	2.22	8.25	18.37	

Table 4.13 Frequencies of types of theme-rheme progression

Table 4.13 is interesting for the purposes of this study in that the three most frequently occurring types of progression all have implications for the identification of topics. The most frequent two types of progression, parallel progression and extended parallel progression, can help us to identify the topic entity of a stretch of discourse (Brown and Yule, 1983a). For stretches of discourse where there is frequent repetition of a theme as evinced by high proportions of parallel progression and extended parallel progression, the repeated theme is the topic entity.

One problem with this approach is how to identify boundaries to a stretch of discourse, and this is where the third most frequent type of progression, coherence breaks, comes to the fore. Looking at example 4.13, although there is relatively frequent parallel progression, there is no single theme repeated throughout the example. To identify topic entities, we need to divide the example into shorter stretches of discourse, and coherence breaks can help us in identifying the boundaries to these.

Following this approach, example 4.13 can be divided into four shorter stretches of discourse, and the most frequently occurring theme in each of these can be identified as the topic entity, as shown in Table 4.14.

Stretch of discourse (T-units)	Topic entity
1	*OK*
2-7	*the film/Star Wars Episode One*
8-11	*where/his house*
12-14	*why/storm*

Table 4.14 Stretches of discourse and topic entities in example 4.13 identified through theme-rheme analysis

As can be seen in Table 4.14, there are some problems in identifying topic entities. For stretches 8-11 and 12-14, two topic entities have been identified, but in each case these two entities are taken as equivalent within the context. However, for T-units 2-7 there are a variety of possible themes competing for consideration as the topic entity. If we include derived themes as instances of the mentioned hypertheme, *we* is the most frequently occurring theme; if we only count repetitions of the same theme, *you* and *film/Star Wars Episode One* are the most frequent. Of these, however, *we* and *you* contribute to the interaction between the teacher and the students as well as being the first ideational elements in their respective T-units. Berry (1989, cited in Davies, 1994) has argued that such themes should be treated as a separate category of themes called interactional themes. Since these interactional themes contribute to the interaction rather than help to generate topics (McCarthy and Carter, 1994), where there are two competing potential topic entities, preference will be given to the ideational theme. The topic entity for T-units 2-7 therefore is *the film/Star Wars Episode One.*

The majority of stretches of discourse in this study do have identifiable topic entities, and these are shown in Appendix C. These provide a useful benchmark for evaluating the topics of short stretches of discourse identified through the use of other approaches. However, theme-rheme analysis does not have implications for identifying topics in discourse longer than these short stretches (despite McCarthy's (1991) claim that there is a possibility that macro-level themes exist) nor does it help us in seeing the relationships between consecutive topic entities.

4.3.2 Given-new progression

An analysis of given-new progression is conducted in much the same way as for theme-rheme progression. Initially, the sentence or T-unit is divided into two constituents, and then the patterns of relationships between the constituents are traced to identify kinds of progression. Theme-rheme and given-new analyses, however, differ in their significance and in the way the two constituents are identified.

While theme-rheme analyses are primarily linguistic in nature, most approaches to given-new analysis involve psychological as well as linguistic concerns. Although the original work concerning given-new information in English by Halliday (especially 1967) focused almost exclusively on intonation, most more recent work has taken a more psycholinguistic approach. This drift from linguistic to psycholinguistic bases to given-new analyses has resulted in some confusion as to what the terms *given information* and *new information* mean.

In the linguistic tradition, Halliday (1970) defines given information as a point of contact with what the listener knows, with new information being the point of the message (Hasan, 1989).

From a slightly more psycholinguistic perspective, Daneš (1974) and Halliday and Hasan (1976) argue that given information is information recoverable or derivable by the listener from the context or shared knowledge, and new information is non-recoverable. Similarly, Chafe (1976, 1980) says that given information is information in the consciousness of the listener whereas new information is information being introduced into the listener's consciousness.

Taking a broader and more psychological perspective, Clark and Haviland (1977) work on the premise that given information is information that the speaker believes the listener already knows and new information is information that the speaker believes the listener does not know yet. In this interpretation, whether the information is consciously activated or not is not a consideration.

A third viewpoint is that of Prince (1981) who argues that the given-new dichotomy should actually be a trichotomy of new, inferable and evoked information. Although Prince's arguments are fairly convincing, it is unclear

how her trichotomy could be used to identify types of given-new progression which are crucial to our goal of creating a dynamic model of topics.

Of the main interpretations of given-new, I will follow the second presented above (especially the interpretation of Chafe). The main reason is that research deriving from Chafe's work on given-new has potentially important implications for this study of topics. Langacker (1996) argues that the given-new distinction is related to the "current discourse space" (p. 334) or the intersection of the consciously activated knowledge of the speaker and the listener. Given information is situated within this current discourse space and provides an anchor for expanding the current discourse space to include new information. We may envisage the current discourse space as a specific area of van Dijk's (1977) semantic space which is continually expanding to include new information introduced by the speaker while also contracting at the other edge of the area to deactivate some previously given information since there are human limits on the amount of information that can be active at any given time. Such movement of the current discourse space through the universe of semantic space would be likely to produce topic drift in discourse.

Given this potential relationship between given-new information and topic progression, it would seem valuable to conduct an analysis of given-new progression. However, the variation in definitions of given and new information have also led to a wide variety of suggestions for how to identify given and new information. Although in most cases we might expect these various methods of identifying given and new information to confirm each other, we might also expect some occasions where identification from the methods conflicts. For this reason, we need to make an order of priority that determines which method will take precedence when there is a conflict. This is given below with the order reflecting both the likely validity of a method identifying given and new information and its practical applicability.

1. Ellipted material is given information (Chafe, 1980; Tomlin et al., 1997).
2. Pronominalised material is given information (Chafe, 1976, 1980; Clancy, 1980; Palmer, 1981).
3. In wh- questions, the wh- is new information (Clark and Haviland, 1977).
4. Information concerning things physically and obviously present in the context is given information (Brown and Yule, 1983a; Stockwell, 1977).

5. Given information is given a weaker stress (Chafe, 1976, 1980; Clancy, 1980; Stockwell, 1977) and a lower pitch (Chafe, 1976, 1980). It also coincides with referring tone units (Brazil, 1985, 1995; Brazil et al., 1980; de Beaugrande and Dressler, 1981). In contrast, sentence stress is placed on the last item conveying new information (Palmer, 1981).
6. Noun phrases with definite articles are given information (Haviland and Clark, 1974).
7. Material which has been previously mentioned in the discourse is given information (Brown and Yule, 1983a; Stockwell, 1977).
8. Given information appears before new information (Jonz, 1989).
9. New information is marked in other ways than those given above (Cumming and Ono, 1997).

It should be noted that two approaches in the literature are not considered in this study. Firstly, although Winter (1982) claims that subordinate clauses are frequently given information, following Givón (1995) I will identify given and new information at levels below the clause, which means that most clauses are likely to contain both given and new information. Secondly, Tomlin et al. (1997) somewhat idiosyncratically only consider noun phrases as given or new and ignore other parts of speech, but I will consider all parts of speech in this analysis.

We are now in a position to identify given and new information in the data. A sample of the identification of given and new information is given in example 4.15 with given information underlined and new information in italics; for purposes of comparison, the data analysed are the same as in example 4.13 above.

Example 4.15

T: [1]***OK***. [2]I *'m going to play a video* [3]and we *'re going to see a scene from a, from a film.* [4]You watch the video *and tell me uh what you see, OK*? [T turns on the video to show a 3-second excerpt from Star Wars Episode 1. T pauses the film.] [5]*Do* you *know what film is* it (the film)?

S: [6]The film is *Star Wars Episode One.*

T: [7]The film is Star Wars Episode One, *right,* Star Wars Episode One. [T plays the video for another ten seconds. The scene is set in a desert village. T pauses the video.] [8] *Where, where do* the boy take these people to? (2.0) [9] Where do this boy, this boy take these people to?

S: [10]The boy takes these people to *His house.*

T: [11] The boy takes these people to His place, his house. [12] *Why* does the boy take these people to his house?

S: [13] The boy takes these people to his house because *Storm*.

T: [14] *Yes*, there is a storm, *sandstorm*, *right*?

Identifying given and new information is less problematic in many ways than identifying themes and rhemes. The framing move spoken with heavy stress in T-unit 1 is new information. The pronouns in T-units 2-5 are given and derivable from the context, while *watch the video* has been previously mentioned in the discourse and contains a definite noun phrase. *The film* in T-units 6 and 7 is ellipted material, and so on.

One notable exception to normal patterns of given-new information in example 4.15 concerns T-unit 14. Existential *there* usually introduces new information (Biber et al., 1999) and *storm* being preceded by an indefinite article supports this. However, as we saw above, T-unit 14 is a rephrased acceptance of the student's utterance in T-unit 13. The *storm* in T-unit 14 is not new, but is material previously mentioned in the preceding T-unit. Instead, it is the acceptance of this material (as indicated by *Yes*) and the expansion into full-sentence form that are new.

Comparing examples 4.13 and 4.15, we can see that theme and given information coincide for 8 of the 14 T-units (after categorising the subject of the polar question as a theme). Given- that several other studies have identified a coincidence between theme and given information on the one hand, and between rheme and new information on the other (Brown and Yule, 1983a; Connor, 1994; Daneš, 1974; Fries, 1994), this seems a surprisingly low amount of coincidence. For all of the data in the study theme coincides with given information for 63.78% of T-units, again a relatively low figure. Looking through the data, most of the 397 instances of non-coincidence involve framing moves, wh- questions, short-answer responses with ellipsis, and T-units containing no given information. On the other hand, in polar questions, repetitions and reviews, and teacher follow-up moves, theme and given information generally coincide. It should be noted that there are potential problems with counting repetitions of, say, questions as given information since the repetition may be required because the learners did not hear the question the first time and thus the information in the question is not activated in their consciousness. However, the repetition may also be used as a prompt for an answer (Watson Todd, 1997b) in which case we can assume that the information is already

activated. Since it is difficult to know which of these cases is true in each instance, I have decided to count all repetitions as being given information unless there is clear phonological evidence to think otherwise. Further research into classroom discourse, especially where it follows an IRF pattern (Sinclair and Coulthard, 1975) is needed to see if the normally strong coincidence between theme and given information is not prevalent in the classroom.

For the purposes of this study, however, we need to focus on given-new progression rather than comparing themes and given information. Although there has been some previous work on given-new progression (e.g. Firbas, 1987; Goldberg, 1983; Rutherford, 1987), the range and depth of the work is noticeably less than that on theme-rheme progression and no comprehensive taxonomies of kinds of given-new progression have been suggested. Since the given-new distinction divides sentences or T-units into two components in ways parallel to theme and rheme, it is worth investigating whether the same kinds of progression can be applied.

Before we can do this, however, there is a problem to be overcome. In theme-rheme analyses, the two components of a T-unit are distinct and separate. In a given-new analysis, on the other hand, each component can be discontinuous. For example, in T-unit 5 *you* and *it* are given but separated by the new *know what film*. Discontinuities, although perhaps regardable as marked forms, are far from rare in English. Quirk et al. (1985), for example, give examples of discontinuous noun phrases, adjectival phrases and prepositional phrases. If we can draw a parallel between these syntactic discontinuities and given-new discontinuities, then perhaps the latter should not be considered an obstacle to further analysis. Furthermore, although given information generally precedes new information in a T-unit, this criterion for identifying given information should not be given a high priority. Whether the information is in the consciousness of the listener is the key consideration in identifying given information rather than its placement in a T-unit. I will therefore treat given and new information as two components of a T-unit irrespective of their placement in the unit.

We are therefore in a position to consider the implications of drawing a parallel between the various kinds of theme-rheme progression and potential kinds of given-new progression. Let us start with the simplest: parallel progression. The first potential example of parallel progression in example 4.15 is the use of *you* and *video/film* as given information in T-units 4 and 5.

In T-unit 4, these two concepts are linked by the direction to watch, while in T-unit 5 they are linked by a request about knowledge. In other words, while two given concepts in T-units 4 and 5 remain constant, the way in which they are connected changes. We could represent this change in connection as in Figure 4.5.

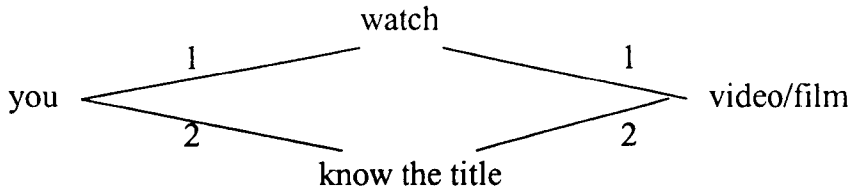


Figure 4.5 Different connections between given *you* and *video/film*

The given information in both T-units provides a point of contact with the preceding discourse, but the way in which the two concepts are connected changes from a given connection to a new connection and is the point of the message in each of the T-units. (It is also worth noting that we would not have identified this if we had not accepted the existence of discontinuous given information.) We might argue that, from a small-scale perspective of topics, the given information indicates the topic while the new information provides information about some aspect of this topic. As the given information does not change, we have topic maintenance with new-information connections providing information about different aspects of the maintained topic. Although the argument here is based on only two T-units, it seems reasonable to propose that it may hold for longer stretches of discourse as well. It would therefore seem worthwhile to look for given-to-given progression between T-units.

Applying the other kinds of thematic progression in the same way looks similarly promising. For example, for sequential progression, the ellipted given information in T-unit 13 leads us to the new *Storm*. The teacher then takes this new information up in T-unit 14 and adds more new information to it. The focus of the discourse has then drifted (again albeit on a very small scale) from *the boy taking these people to his house* to *a storm*. Sequential progression of given information apparently seems related to topic drift.

Indeed, all of the kinds of thematic progression can be adapted to analyse given-new progression. To distinguish between the two taxonomies of

progression, I will rename the types of thematic progression for when they are used in given-new analyses as in Table 4.15.

Thematic progression	Given-new progression
parallel progression	parallel given progression
sequential progression	sequential given progression
extended parallel progression	extended parallel given progression
thematic progression with derived themes and mentioned hypertheme	progression following superordinate given information
other types of progression	other types of given progression
coherence breaks	given-new coherence breaks

Table 4.15 Types of given-new progression

Let us now examine the types of given-new progression in example 4.15 in more detail. T-unit 1 contains only new information which is not included in T-unit 2, indicating a given-new coherence break between the first two T-units. The progression between T-units 2, 3 and 4 works in the same way as for theme-rheme analysis (see p. 103) with the progression being progression following superordinate given information. The given *you* in T-unit 4 is repeated in T-unit 5 giving parallel given progression, and similarly *the film* in T-unit 5 is repeated in T-units 6 and 7. Although T-unit 8 contains given information from the context of the video, none of the information in T-unit 8 links back to information in previous T-units indicating a given-new coherence break between T-units 7 and 8. The given information from the video is then repeated in T-units 9, 10, 11, 12 and 13 (as ellipsis in the last four of these and with added given information in the last two) for parallel given progression. Finally, the new information in T-unit 13 is repeated as given information in T-unit 14 indicating sequential given progression.

Again using van Dijk's (1977) notation, the given-new progression in example 4.15 can be presented as follows, where the letters before the comma denote given information and those after the comma new information.

T-unit 1	<a>	
		coherence break
T-unit 2	<b_1,c>	
		progression following superordinate given information
T-unit 3	<b,d>	
		progression following superordinate given information
T-unit 4	<b_2c,e>	
		parallel given progression
T-unit 5	<b_2c,e>	
		parallel given progression
T-unit 6	<c,f>	
		parallel given progression
T-unit 7	<cf>	
		coherence break
T-unit 8	<g,h>	
		parallel given progression
T-unit 9	<gh>	
		parallel given progression
T-unit 10	<hg>	
		parallel given progression
T-unit 11	<hg>	
		parallel given progression
T-unit 12	<hg,i>	
		parallel given progression
T-unit 13	<hg,i>	
		sequential given progression
T-unit 14	<i>	

Note: The information before the comma in angle brackets is given information with new information after the comma. Where there is no comma, the T-unit contains given information only.

Table 4.16 Given-new progression in extract G

Even though we have seen that theme and given information do not coincide very frequently, comparing the types of given-new progression shown above with the types of theme-rheme progression on page 104, we can see that the types of progression identified are the same except for between T-units 11 and 12, where the coherence break identified in the theme-rheme analysis is replaced by parallel given progression dependent on ellipted material.

Continuing to identify types of given-new progression in the same way for all the data, we can show the frequency of the different kinds of progression in Table 4.17.

Extract	Parallel given progression	Sequential given progression	Extended parallel given progression	Progression following superordinate given information	Other types of given progression	Given-new coherence breaks	Total
A	120	35	36	0	3	42	236
B	94	26	27	5	2	31	185
C	31	8	12	1	4	19	75
D	36	18	29	3	3	30	119
E	34	3	31	5	2	8	83
F	73	19	12	2	5	13	124
G	48	7	8	2	0	8	73
H	17	2	4	0	3	7	33
I	48	15	9	0	2	7	81
J	18	5	10	0	1	5	39
K	31	6	4	0	0	5	46
L	25	3	2	0	1	2	33
Total	575	147	184	18	26	177	11127
Percentage	51.02	13.04	16.33	1.60	2.31	15.71	

Table 4.17 Frequency of types of given-new progression

The total percentages in Table 4.17 are remarkably similar to the total percentages for types of theme-rheme progression shown in Table 4.13 (p. 105). Indeed, the correlation between the two sets of figures is significant ($r = 0.97$, $p < 0.001$). Again, parallel given progression, extended parallel given progression and given-new coherence breaks are the three most frequent types of progression with the preferred sequential given progression which moves discourse forward (Newman, 1985; Rutherford, 1987) less frequent. As we saw with theme-rheme progression, these three most frequent types of progression are important for identifying topic entities and dividing the discourse into shorter stretches.

Following the same approach used for theme-rheme analysis, example 4.13 can be divided into three shorter stretches of discourse with the most frequently occurring given information which is primarily ideational rather than interactional identified as the topic entity, as in Table 4.18.

Stretch of discourse (T-units)	Topic entity
1	-
2-7	*the film*
8-14	*the boy takes these people*

Table 4.18 Stretches of discourse and topic entities in example 4.13 identified through given-new analysis

In example 4.13, the most frequently occurring given information is more easily identified than was the most frequently occurring theme. Continuing in the same way for all of the data in this study, we find the breaks in the discourse and the topic entities shown in Appendix C. These breaks and topic entities may be more useful for our purposes than those identified through theme-rheme analysis for two reasons. Firstly, they are more closely linked theoretically to topics, especially through the idea of current discourse space, than those identified through theme-rheme analysis. Secondly, their identification has been less problematic than was the case in theme-rheme analysis. Nevertheless, the given-new analysis is not enough on its own for us to be able to be confident in our ability to identify topics and topic progression. As with theme-rheme analysis, the overall approach presents a fairly static perspective on discourse. In addition, there are a number of stretches of discourse where the topic entity is not identifiable through this approach, and the link between frequently occurring given information and topic entities needs confirmation through other methods of analysis.

Chapter 5 Topic-based Analyses of Classroom Discourse

The methods of analysis in chapter 4, while hopefully productive, were not specifically designed to analyse topics. The methods used in this chapter, on the other hand, do aim to identify topics and topic progression. In chapter 2, we defined *topic* as a clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create coherence and relevance. The key aspects of this definition which will be investigated in this chapter are how concepts are clustered and the nature of the associations and relations which produce clustering.

Following van Dijk (1977), we may imagine semantic knowledge as existing in semantic space. The closer two concepts are in this semantic space the more connected they are and the lower the chance of a move between the two concepts being perceived as a topic shift. A promising approach to identifying topics and topic progression through discourse would therefore involve investigating the distance in semantic space between the various concepts in a stretch of discourse. Where concepts are very close and recur frequently, we may identify topic maintenance; where there is a movement across semantic space in short jumps, we may identify topic drift; and where there is a single large jump between two concepts separated by a large distance in semantic space, we may identify topic shift.

The method of analysis that I will be using in this chapter attempts to create models of semantic space for the set of key concepts in a stretch of discourse, and then follows the progression of concepts from the discourse within these models.

5.1 The background of topic-based analysis

The analysis I will be using in this chapter is based on two of my own articles (Watson Todd, 1997a, 1998). In this section, I will present a brief summary of these articles.

The first of the articles (Watson Todd, 1997a) involved analysing a vignette (see van Lier, 1988) of a fairly long stretch of classroom discourse centred around eliciting. Initially, the key points of the vignette were rewritten as propositions (here taken to mean 'statement'). In writing these propositions, attempts were made to phrase them as similarly as possible (e.g. by using the

same subject and syntactic structure in creating the propositions) to make them amenable to logical analysis. Entailment relations between the various propositions were then identified enabling the propositions to be placed into a hierarchy. The sequence of occurrence of the propositions in the stretch of discourse was then mapped onto this hierarchy. Finally, the patterns of text organisation (Hoey, 1983) or larger patterns in text (McCarthy, 1991), such as general - specific - more specific, were identified.

While this paper represents an initial attempt at identifying patterns of topic progression in classroom discourse which achieved some of its goals, it is also fraught with problems. Key among these are how to identify the key points of the discourse and how to rewrite these as comparable propositions. The second paper (Watson Todd, 1998) attempts to overcome these problems.

In the second article, a verbatim transcript of classroom discourse covering both eliciting and explanation was used instead of a vignette. From this transcript, key concepts were identified based on frequency (cf. Scott, 1997). In this way, a more rigorous method of identifying key points in the discourse was used, and since these key points are concepts rather than propositions, there are few problems of comparability between different key points. After the key concepts had been identified, entailment relations between the concepts were identified, a hierarchy of these relations was constructed, and the sequence of occurrence of the concepts in the discourse was mapped onto the hierarchy, as in Watson Todd (1997a). Topics were identified as being the more superordinate and more frequently occurring concepts, and topic progression through semantic space was followed by assigning distances to the different connections in the hierarchy.

The second approach to a topic-based analysis in Watson Todd (1998) is probably more satisfactory than the first. The implications of the study, for instance, suggest that the ratio of movements between concepts to T-units, the distance in semantic space as measured by the moves between concepts mapped onto the hierarchy, and the use of metadiscoursal markers are all apparently helpful indicators of topics and topic progression and of the comprehensibility of the discourse.

As with the first article, however, the second is not without problems. For example, what kinds of surface features can be considered key concepts? How can a topic shift be identified unambiguously? And are there any other

possible logical relations that should be considered in addition to hyponymic entailment? I hope that this chapter will provide some answers to these questions.

One further potential problem with both analyses which is fundamental is their reliance on logical relations. I will examine the arguments for and against the use of logical relations in detail later, but for the moment it is enough to point out that, if the definition of *topic* concerns both relations and associations, then so should the analysis. We therefore need to look at how we can base an analysis on associations as distinct from the one based on relations.

In this chapter, I propose to conduct analyses based largely on the topic-based analysis of Watson Todd (1998). Initially, I will follow this approach, but will attempt to make it more rigorous and to consider a range of logical relations. As a second approach, I will attempt to do the same analysis again but this time will use associations instead of relations.

5.2 Conducting a topic-based analysis using logical relations

In conducting a topic-based analysis using logical relations, there are six main stages: identifying key concepts, identifying relations between concepts, putting the concepts into a hierarchy based on these relations, mapping the discourse onto the hierarchy, measuring the distance in the hierarchy of each move between concepts, and finally identifying the topics and topic progression. In this section, I will look at each of these in turn, using extract G again (see pp. 46-48) as the sample data.

5.2.1 Identifying key concepts

Before we can identify key concepts, we must first decide what a concept is. Although it is tempting to equate concepts with words, it should be remembered that a concept is a psychological construct whereas a word is a linguistic phenomenon. Because of this difference, a straightforward equivalence between concepts and words may not be appropriate.

Fairly uncontroversial evidence for there being no automatic one-to-one match between concepts and words concerns proper nouns and translations. In extract G, the phrase *Star Wars Episode One* occurs relatively frequently (and is the only aspect of *Star Wars* referred to in this extract). This proper noun is a noun phrase referring to the title of a film. While some film titles may consist of a single word, the number of words in a film title is arbitrary.

The film is a single concept irrespective of the number of words in the title identifying this concept. *Star Wars Episode One* should therefore be treated as a single concept rather than four separate concepts, one for each word.

The second kind of evidence for no automatic match between words and concepts concerns translations. In the transcripts in this study, Thai is used occasionally. In translating the Thai into English, the number of words may change while the number of concepts remains constant. A clear example of this can be found in extract B, where the Thai word *bahngfý* is translated as *traditional Thai rocket*. Even though I am using the English words in the analysis for clarity, the fact that the translation of *bahngfý* consists of three words rather than one is merely an indicator of the lack of a direct translation for the concept in English rather than evidence for three separate psychological units. *Traditional Thai rocket* is therefore treated as a single concept.

While these two cases are uncontroversial, there is some evidence that less unambiguously unitary multi-word units should also be treated as single concepts. Much recent work in lexical linguistics has focused on the use of multi-word units (e.g. Fernando, 1996; Lewis, 1993; Nattinger and DeCarrico, 1992; Schmitt, 2000). It is generally agreed that there is a range of fixedness in the use of multi-word units with idioms being the most fixed. To take the classic example, *kick the bucket* allows for no variation in the lexis of the phrase. Where no variation is allowed, the idiom can be taken as representing a single concept. Again, this is relatively uncontroversial, but the problem is how much variation should be allowed before a multi-word unit stops being considered a single concept and starts being treated as several separate concepts.

In this study, to avoid the need to rely on the analyst's personal interpretations, I will only consider two cases of multi-word units representing single concepts (other than the proper noun and translation cases discussed above). The first is where the multi-word units show no variance in the data. In other words, if two content words always occur in exactly the same pattern with respect to each other in the same multi-word unit in any given extract (or in a reasonably long stretch of discourse within one extract), they will be considered a single concept. For instance, in extract G, both *cook* and *food* occur 7 times. On each of these occasions, they occur in the phrase *cook food* and there is no variation in this phrase. *Cook food* is therefore considered a single concept. Although in this case the

object is separable from the verb (i.e. if *food* is omitted, there is no change in meaning), in other extracts there are non-separable multi-word units which need to be treated as single concepts (e.g. *lunar module* in extract H).

The second case concerns occasions where a single word is explicitly equated with a multi-word unit by one of the participants in the discourse. For example, in extract B T-unit 60, the teacher explicitly equates *open end* with *neck*. In this example, we have a single concept being expressed as either a single-word noun or as a compound noun. Similarly, in extract A *release* is equated with *let go*. In such situations, if the single word represents a single concept, then the equivalent multi-word unit must also represent the same single concept.

As well as considering whether two or more words can represent a single concept, we also need to consider whether a single word can be used to represent more than one concept. The existence of numerous polysemous words suggests that this is so, but we also need to consider the discourse-specific referents of words. For example, in extract G, the word *people* occurs 7 times. In the first 6 occurrences, *people* is used in the pattern *the boy takes these people to*, and refers to a specific group of people in the video. For the last occurrence, *robots replace people at work*, *people* refers to the general entity of people around the world. These two uses of *people*, then, are referring to different concepts and should be treated separately.

Following these guidelines, the general pattern that we find in the extracts is that in the majority of cases, a single word represents a single concept, but on occasions, two or more synonyms may refer to the same concept, a multi-word unit may represent a single concept, and a single word may refer to two different concepts.

Even with these provisos, the vast majority of words in the extracts do not represent concepts. There are three categories of words which do not serve the conceptual metafunction of language and which will be ignored in our analysis.

Firstly, function or grammar words (Lewis, 1993; Read, 2000), such as *the*, *of* and *if*, do not convey any concepts. Rather, they allow the concepts to be expressed in ways which are grammatically and syntactically acceptable. Except where function words form part of a multi-word unit expressing a concept, these words are not taken into consideration in the analysis.

Secondly, there are words which serve interactional rather than conceptual purposes. In the analysis of theme-rheme progression above (see p. 106), we saw that some themes did not help to generate topics but rather contributed to the interaction. These interactional themes are typically expressed by pronouns, such as *I* and *you* or as proper nouns for names of people. Since such words are not usually used for the conceptual metafunction (and are exclusively interactional in the data in this study), they are not included in the analysis.

The third category of words which are excluded from the analysis is more complex. To explain this category, we have to turn to schema theory. Schemata are "previously acquired knowledge structures" (Carrell and Eisterhold, 1988: 76) which represent the relationships between components of knowledge (Anderson and Pearson, 1984). These knowledge structures can be of two types (Carrell and Eisterhold, 1988; Kitao, 1990). Formal or textual schemata provide rhetorical organisation, whereas content schemata are knowledge structures organising relationships between concepts. While content schemata are clearly important in this study, the value of textual schemata to our analysis is less clear. The search for textual schematic structures is the goal of genre analysis (Bhatia, 1993; Nwogu, 1991; Swales, 1990) and text approaches (Biber, 1988; Crombie, 1985a, 1985b; Paltridge, 1996; Winter, 1994). In such textual schematic approaches, a wide range of words is used to describe the structure of texts, including *problem*, *reason*, *cause*, *example*, *focus* and *introduction*. These words serve Halliday's (1970) textual metafunction of language rather than the ideational metafunction which is the focus of this study. In other words, such terms do not apply to the concepts which form the components of knowledge in content schemata.

Another similar category of words is that of words which describe the relationship between concepts rather than the concepts themselves. For example, in extract G we can identify the concepts *robot* and *R2D2*. It is relatively easy to show that *R2D2* is a hyponym of *robot* (see 5.2.2 below). We also know that *R2D2* is the *name* of a *robot*. While *robot* and *R2D2* are incontrovertibly key concepts in extract G, *name* is more problematic. Although it is a content word occurring 6 times in the extract, each time it appears to be describing the nature of a hyponymic relationship rather than being a separate concept node in itself. Other words in the extracts which serve the function of describing relationships rather than expressing concepts include *work* (in the context of *What kind of work can robot do?*) and *use*.

As with the words related to textual schemata, such relation-focused words serve the textual metafunction of language. Both of these categories of words will be excluded from the analysis.

Having narrowed down the types of words that can be used to express concepts, we can now look at extract G to see what words are potentially includable in our analysis. These, together with their frequency of occurrence, are given in Table 5.1.

robot	24	who	4	storm	2	mom	1
robots	22	why	4	video	2	money	1
boy	15	naked	3	bad	1	mother	1
what	15	servant	3	brains	1	movie	1
C3PO	9	skin	3	buy	1	place	1
cook	7	where	3	construct	1	play	1
food	7	woman	3	fighter	1	programmed	1
people	7	work	3	fighters	1	raise	1
episode	6	construction	2	hand	1	replace	1
film	6	fix	2	headache	1	sandstorm	1
star	6	housework	2	how	1	scene	1
wars	6	owner	2	machine	1	soldiers	1
R2D2	5	program	2	machines	1	warriors	1
house	4	see	2	material	1	watch	1
seen	4						

Table 5.1 Frequencies of potential concept words in extract G

The list in Table 5.1 is derived from a word frequency count, and includes ellipted material and referents of referring expressions (see chapter 3), since these are in the consciousness of the participants if not in the surface language. The first stage in identifying key concepts from this list of potential concepts is to identify those words which are part of multi-word units referring to single concepts (see above), such as *Star Wars Episode One* and *cook food*.

The second stage is to deal with the wh-words. In some cases, such as the *who* in *Who is the woman?* in T-unit 15, the wh- word is used in a closed or convergent question (see Moore, 1989; Watson Todd, 1997b). In such questions with only one possible answer, the wh- word functions as a cataphoric referring expression with the answer being the referent. In such cases, the wh- word and the answer can be considered to be referring to the same concept. For open-ended or divergent questions where there is a range

of possible answers, however, it is not possible to justify a match between the wh- question and the answer. For example, in extract G, the question *What kind of work can robot do?* elicits five different answers and it is impossible to say which the wh- word is referring to. In these cases, the wh- word and the answer cannot be considered as referring to the same concept.

In addition to matching wh- words to content words, we also need to identify reiterations of the same concept expressed using different content words. The main types of reiteration that we will need to consider are simple repetition (e.g. *film - film*), complex repetition (e.g. *program - programmed*), simple paraphrase (e.g. *film - movie*), complex paraphrase (e.g. *servant - housework*), hyponymic repetition (e.g. *woman - the boy's mother*), and co-reference (e.g. *the film - Star Wars Episode One*) (see Hoey, 1991 for details and algorithms for identifying the different kinds of reiteration). In identifying reiterations, there are two key points to bear in mind. Firstly, we are looking for reiterations of concepts rather than words. Secondly, the reiterations are context-specific. In other words, although a dictionary of synonyms or a thesaurus may help us in identifying paraphrases, these are primarily identified based on the specific discourse context rather than on generic meanings.

Following Hoey's (1991) algorithms for identifying reiterations of words and the guidelines given above for identifying concepts, we can group together some lexical items which express the same concept, and so make the word frequency count given in Table 5.1 into a concept frequency count which is given in Table 5.2. In counting frequencies of concepts, reiterations of the same concept within a single T-unit are counted as single occurrences.

robot/robots 28
boy/owner 12
C3PO/the robot/the robot under construction/what (the boy is talking about)/(one of) these two robots 12
R2D2/the other robot/(one of) these two robots 11
Star Wars Episode One/what film/this film/this movie 7
cook food 7
house/place/where 6
these people 6
naked/no skin 4
servant/housework 4
seen 4

storm/sandstorm/why	3
the woman/mother/mom/who	3
video/scene from a film	3
fighter/fighters/soldiers/warriors	3
program/construct a program/be programmed	3
see/watch	2
a bad headache	1
brains	1
buy material	1
fix machines	1
raise a hand	1
money	1
skin	1
people	1
play	1
replace	1
at work	1

Table 5.2 Frequencies of concepts in extract G

From Table 5.2, we can see that several of the concepts in extract G are referred to in several different ways. For clarity, in the succeeding discussion I will use only one wording for each concept. The wording chosen will be the most frequent (e.g. *boy* rather than *owner*), and, if this is the same, the one containing the most information (e.g. *Star Wars Episode One* rather than *this film*).

The list of concepts in Table 5.2 is a list of all concepts in extract G. Some of these concepts are unlikely to be important to us in our search for topics since they occur only once or appear in an aside to the main flow of the discourse (e.g. *a bad headache* in T-unit 59). We need, then, to be able to distinguish between these less relevant concepts and the key concepts that are important to our study of topics. There are two main principles on which the identification of key concepts may be based.

Firstly, key concepts are frequent in the discourse. There are two ways of considering such frequency. The first and more simple method is to measure the frequency in the discourse with no reference to any benchmark outside the discourse. Essentially, this is what Hoey (1991) does when he sets thresholds for including lexical items in his analysis (see chapter 4). The

second, more complex method is to compare the frequency of occurrence of the concept in the discourse with the frequency of occurrence of the same concept in other discourses. For example, Scott (1997) uses a chi-square analysis to compare the observed frequency of a word in a given text with the expected frequency of the same word in a reference corpus. Significant differences between the observed and expected frequencies indicate a keyword. While the second method is a more valid method of identifying keywords, the difficulties in applying it (e.g. setting up a suitable reference corpus, calculating expected frequencies) make it impractical for this study. I will therefore use a simple discourse-specific frequency count in this study.

The second method of identifying key concepts is based on their salience. In spoken language, intonation, pausing and discourse markers can all be used to indicate keyness (Scott, 2000). One problem with using such criteria to identify keyness is that the key items identified are not specifically conceptual. For example, using the three criteria given above, framing moves would be identified as key. They probably should be considered key items in discourse, but their keyness is textual rather than conceptual. Instead of using these linguistic criteria to identify salient keyness, in this study I will use more classroom-specific criteria that are likely to be used exclusively for conceptual keyness. The first criterion is that of concepts written on the board by the teacher. The most common use of board writing is to highlight points (Watson Todd, 1997b) or make them salient. The second criterion is concepts which the teacher and/or students spell out letter by letter. Again, such classroom behaviour makes the spelt words more salient.

Key concepts therefore are identified in two ways. Firstly, any concepts which occur more frequently than a given threshold are considered key. Secondly, any concepts made salient by being written on the board or spelt out are considered key.

For extract G, the threshold chosen for a concept to be considered key is 3. We need to identify a threshold which generates enough key concepts to allow progression between concepts in the discourse to be identified, but which does not generate so many key concepts as to create many instances of concept-to-concept progression which does not reflect topic progression. For the extracts in this study, the threshold chosen ranges from 2 (extracts H, J and L) through 3 (extracts C, G, I and K) to 4 (extracts A, B, D, E and F). This variation in threshold is proportional to the variation in the number of

words in each extract (r = 0.75). With a threshold of 3 occurrences and no concepts being made salient through being written on the board or spelt out, the key concepts in extract G are:

robot
boy
C3PO
R2D2
Star Wars Episode One
cook food
house
these people
naked
servant
seen
storm
woman
video
fighter
program

One further refinement of thresholds is needed to ensure coverage of concepts in insertion sequences as well as in the main thread of the discourse. We can identify such points by looking for stretches of discourse which contain no concepts when the threshold is set high, but in which new concepts emerge when the threshold is reduced. For example, in extract A, T-units 103-107 do not contain any of the key concepts identified in the extract when the threshold is set at a minimum of 4 occurrences. These T-units concern discipline and are inserted into the main thread of the discourse. Reducing the threshold to 2 occurrences for T-units 103-107 of extract A (and I will use a minimum of 5 consecutive T-units for such stretches of discourse), we find that *technician* should also be included as a key concept for the extract. This refinement of the guidelines for identifying key concepts does not affect extract G.

5.2.2 Identifying relations between key concepts

Having identified the key concepts, we now need to find a way of identifying their distance in semantic space from each other. In this section, following Watson Todd (1998) I will take distance in semantic space to refer

to how closely related the concepts are. In section 5.3, I will take distance in semantic space to mean how closely associated the concepts are.

To be able to identify relations between concepts, there are two main things to consider: the types of relations which may hold between any two concepts, and how each of the relations between two concepts may be fitted together to show relative distance in semantic space between all concepts. Since the choice of types of relations that can be used is partially dependent on how they will be fitted together, I will start with the second of these considerations.

In considering how a set of relations may fit together, we need to examine the different ways in which information can be organised (see Watson Todd, 2002a). A variety of ways of representing information organisation in applied linguistics has been suggested, including grids or matrices, flow charts, algorithms, hierarchies, and networks (Burgess, 1994; Graney, 1992; Mohan, 1986). Of these, flow charts and algorithms can be used to describe processes and narratives (Hearst, 1994) and grids are useful for comparisons (Burgess, 1994). To represent patterns of organisation of conceptual knowledge, we are left with a choice of hierarchies (and their mathematically equivalent alternatives such as Venn-Euler diagrams, see Lipschutz, 1964; Watson Todd, 1997a), networks, and combinations of these (see e.g. Collins and Quillian, 1969; Strahan, 1989).

For our purposes in this section, using hierarchies is preferable as the basis of hierarchical construction is relations between the concepts at the nodes in the hierarchies. In constructing a network, on the other hand, relations between the concepts at the nodes do not need to be considered.

A further reason for choosing hierarchies is their extensive and productive use in linguistics and discourse analysis, starting with Roget's lexicographic use of a hierarchy to organise his thesaurus. Within discourse analysis, hierarchical models of organisation have been usefully applied to schemata (e.g. Long, 1989; Mann and Thompson, 1988; Slavin, 1994), scripts (e.g. Abbott et al., 1985; Whitney, 1998); anaphora (e.g. Langacker, 1996); rhetorical management of discourse (e.g. Tomlin et al., 1997); overall structure of texts (e.g. Goetz and Armbruster, 1980; Grabe, 1984); and, significantly for our purposes, classroom discourse (e.g. Sinclair and Coulthard, 1975; Smith and Holdcraft, 1991) and organisation of classroom tasks (e.g. Cole and Chan, 1987; Erickson, 1982; Mohan, 1986). This

extensive history of valuable use of hierarchies suggests that taking a hierarchical approach to information organisation may also be useful in this study. Furthermore, the arguments for taking a hierarchical approach to topics given in chapter 2 suggest that constructing hierarchies to fit relations between key concepts together is an appropriate approach.

If we are taking a hierarchical approach, we next need to consider the types of relations that can form a hierarchy. Although we may need to consider many different kinds of relations as none can be said to be necessarily more important than any others (Lönngren, 1989), the relation which is most usually associated with hierarchies, because by its nature it produces ranks of specificity, is hyponymy (Cruse, 1986). Hyponymy is the "relation that holds between a specific or subordinate and a general or superordinate concept" (Cicourel, 1991: 40) and involves the inclusion of the meaning of the hyponym within the meaning of the superordinate (Bell, 1991; Hoey, 1991). It is definable logically through unilateral implication (Lyons, 1977; Malmkjaer, 1991). In other words, if X entails Y (if X is true, Y is true) but Y does not entail X (if Y is true, X is not necessarily true), we can say that X is the hyponym and Y the superordinate. Hyponymy can also be defined through componential analysis, whereby a hyponymic relationship exists if the componential formula for the hyponym contains all the features present in the formula of the superordinate plus at least one other feature (Leech, 1981). Classic examples of hyponymy include those comprising the Linnaean classification of living things, such as *lion* is a hyponym of *feline*, and *oak* is a hyponym of *tree*.

There are, however, several problems with hyponymy, some of which can be easily solved while others present more of a problem. Among the more easily solved problems is Cruse's (1986) observation that in some sentences the entailment is in the 'wrong' direction. As an example of such sentences, he gives *It's not red* which entails *It's not scarlet* and, he argues, for these sentences *red* entails *scarlet*. This conclusion is not true, as the entailment is actually from *not red* to *not scarlet*. This is termed the contrapositive proposition of the original statement and is, in fact, directly equivalent to a direction of implication from *scarlet* to *red* (Lipschutz, 1964).

More serious problems are raised by the rigorous logical basis of hyponymy. In line with most modern semantics (see for example the papers in the collection edited by Lappin (1996)), hyponymy is defined in strict logical terms. There are, however, persuasive arguments against basing relations

between concepts on strict logic. Firstly, there are several arguments against thinking being based on logic, including the problems of the enormous amount of information needed for a semantics of logic (Johnson-Laird, 1986; van Dijk, 1977, 1980; Wittgenstein, 1953), the fact that much thinking is inductive rather than deductive (Allwood, 1986), and the errors that people regularly make in logical reasoning (Best, 1999; Reisberg, 1997; Wason and Shapiro, 1971). Secondly, there are arguments against language being based on logic, including the much wider range of meanings in natural language than in logic (Fodor, 1970), the crucial role played by context and the knowledge of the world in language use (Devlin, 1997), and the inability of logic to cope with both the deep and the surface structure of language (Fodor, 1970; Lyons, 1977).

To overcome these problems with logic, we need to take a less rigorous approach to relations between concepts and thus I will use *hyponymy* in a way akin to McCarthy's (1988) use of *inclusion*. This looser interpretation of hyponymy includes relations such as quasi-hyponymy, para-hyponymy and pseudo-hyponymy as well as strictly logical hyponymy in our analysis. Quasi-hyponymy involves hyponymic relations between concepts which are not of the same type, such as *knife* is a quasi-hyponym of *cutlery* (Cruse, 1986). Para-hyponymy is relations based on expected rather than canonical characteristics of concepts, so that *dog* is a para-hyponym of *pet* (*ibid.*). Pseudo-hyponymy is a context-specific interpretation of hyponymy. For example, the phrase *watches, cameras and other presents* implies that *watch* is a pseudo-hyponym of *present* (Carter and McCarthy, 1988). More amusingly, Samuel Johnson in his dictionary angered many Scots by including *oats* as a pseudo-hyponym within the superordinate class *food for animals* rather than *food for men* (Bell, 1991). Including these less rigorous interpretations of hyponymy in our analysis is more likely to capture how the discourse participants view the relations between concepts.

One further problem with hyponymy involves deciding on the superordinate for any concept. For example, *dog* is a hyponym of *animal*, but is also a hyponym of the intermediate superordinates *canine*, *mammal* and *vertebrate*. In our analysis, however, given that we are dealing with a limited set of concepts for each extract, we can only look for relations within this set of concepts, and thus this problem should be easily overcome.

Although hyponymy is the relation most closely associated with hierarchies, it is not the only relation that can be used to connect two concepts. There are several other types of relation that we will also need to consider.

Some of these other relations can be defined in hyponymic terms. Synonymy and antonymy, for instance, can both be considered as involving a hyponym-superordinate relationship, albeit with the superordinate being at different distances above the hyponym. For example, the synonyms *frigid* and *freezing* are co-hyponyms of the close superordinate *cold*, while the antonyms *freezing* and *boiling* are co-quasi-hyponyms of the more distant superordinate *temperature* as seen as a scale. Although most instances of synonymy in this study will be dealt with at the stage of identifying concepts, being able to define other relations in terms of hyponymy may aid in the construction of hierarchies.

Other non-hyponymic hierarchical relations will also need to be considered in this study. These include meronymy or part-whole relations (e.g. *arm* is a meronym of *body*). Meronymy can be defined as follows:

> "X is a meronym of Y if and only if sentences of the form *A Y has Xs/an X* and *An X is a part of a Y* are normal when the noun phrases *an X, a Y* are interpreted generically"
>
> (Cruse, 1986: 160)

We will also include both permanent and explicitly stated temporary entity-characteristic relations (e.g. *large* is a characteristic of *elephant*) and possessor-possessed relations (e.g. *the Falkland Islands* is possessed by *the United Kingdom*) in the analysis. As with hyponymy, these relations will be applied somewhat loosely, so that we may need to consider, say, pseudo-meronymy. In using this less strict approach, in some cases we may find that two different relations are possible between the same two concepts. For example, we have seen that *knife* is a quasi-hyponym of *cutlery*, but the relationship could also be considered quasi-meronymy. This is not considered a problem, since, in this study, it is the existence of a relationship between two concepts, rather than the nature of the relationship, that is considered important.

Having seen the types of relation that can be used in this study, let us look at how the key concepts identified from extract G can be related. The majority of the pairs of concepts are not directly related, either through general

semantic definitions or within the specific context of the discourse (for example, there is no relation holding between *people* and *program*, or between *storm* and *R2D2*). There are, however, 21 relations holding between the key concepts in extract G.

There are three clear examples of hyponymy: between *video* and *Star Wars Episode One*, between *robot* and *R2D2*, and between *robot* and *C3PO*. There are also several reasonably clear-cut examples of meronymy, if we consider the various things seen in the film (*people*, *storm*, *boy*, *woman*, *house*, *C3PO*, and *R2D2*) to be parts of the whole film (*Star Wars Episode One*). Similarly, the four examples of possessor-possessed relation in the discourse are relatively straightforward: from general knowledge, a *robot* has a *program*, and from the discourse, the *boy* has a *house*, a robot called *C3PO*, and a *woman* for his mother (it should be noted that the direction of possession for the last of these is solely due to the specific references to *his mother* in the extract). One characteristic (*naked*) of the entity *C3PO* is also relatively easy to identify.

More problematic are the relations between *robot*, on the one hand, and three uses of robots, *fighter*, *servant* and *cook food*, on the other. Uses could be considered characteristics of an entity, and thus we would have three entity-characteristic relations with *robot* the superordinate in each. However, we could also easily imagine a description of a fantasy computer game which states 'There are three kinds of fighter: a robot, a human, and a tiger'. In this situation, *robot* would be a context-specific subordinate of *fighter*. The relationship between *robot* and *fighter* is perhaps best illustrated through a Venn diagram as in Figure 5.1.

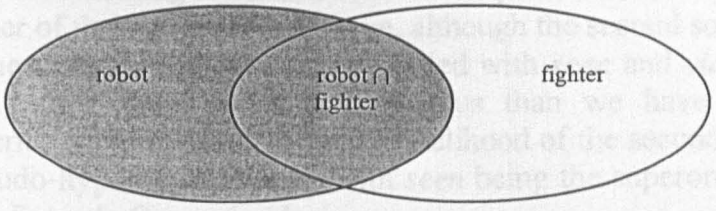


Figure 5.1 The relationship between *robot* and *fighter*

In Figure 5.1, neither *robot* nor *fighter* can be considered the superordinate. However, the discourse in extract G is only concerned with the shaded part of Figure 5.1, and within this part *robot* ∩ *fighter* (*robot* intersect *fighter*) is

a subordinate of *robot*. As we are concerned with the relations between concepts as used in a specific stretch of discourse, in this case extract G, we will identify *robot* as the superordinate and *fighter*, *servant* and *cook food* as subordinates.

Similarly discourse-specific is the relationship between *program* and *cook food*. We have already seen that both of these are subordinate to *robot*. Within extract G (T-units 51-53), we also see that there is a discourse-specific relationship between *program* and *cook food*. This relationship involves some form of implication as shown in the following two sentences where the first is true but the second is not necessarily true within the context.

If a robot can cook food, it must have been programmed.

If a robot is programmed, it must be able to cook food.

From this pseudo-hyponymic relation, we can see that *program* is the superordinate of *cook food*.

The final relation that we need to identify is the most problematic. Of the 16 key concepts identified in extract G, we have found relations for 15 of them. The as-yet-unrelated key concept is *seen*. In T-units 55-59 of extract G where *seen* occurs, it co-occurs with *Star Wars Episode One* and may also be related to *video*, suggesting a relationship between these key concepts. However, *seen* is not related by meronymy, entity-characteristic or possessor-possessed to either *Star Wars Episode One* or *video*. Turning to hyponymy, trying to create if-sentences to relate these concepts produces an unclear situation:

If something is seen, it is Star Wars Episode One.

If something is done in Star Wars Episode One, it is seen.

Neither of these is necessarily true, although the second seems more likely to be true than the first (the same applied with *seen* and *video*). Although not wholly true, being even less rigorous than we have been up to now concerning hyponymy, the higher likelihood of the second sentence suggests a pseudo-hyponymic relation with *seen* being the superordinate of both *Star Wars Episode One* and *video*.

Within extract G, then, for the 16 key concepts identified, there are 21 hierarchical relations between concepts, some of which are clearer and easier to identify than others. While some of these relations hold generally, being based on world knowledge (e.g. the three clear examples of hyponymy),

most of the relations identified are discourse-specific (e.g. the possessions of *boy* and the uses of *robot*).

5.2.3 Putting the concepts into a hierarchy

Since all the relations identified are hierarchical, it is relatively straightforward to put them into a single hierarchy. In doing this, for purposes of clarity, intersecting lines should be avoided as far as possible. A hierarchy for the key concepts in extract G, using the relations identified above, is given in Figure 5.2.

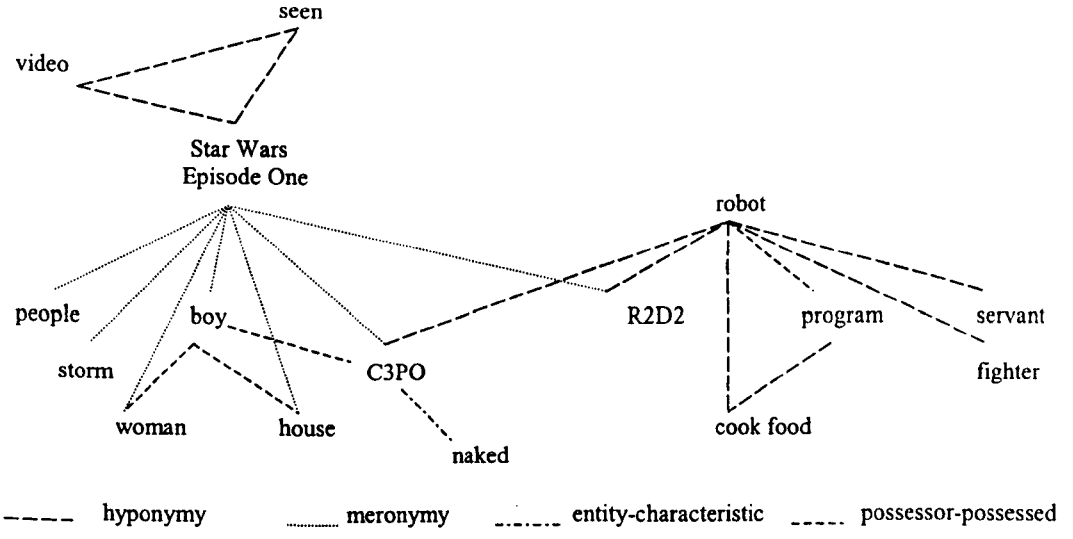


Figure 5.2 A hierarchy for the key concepts in extract G

Figure 5.2 is not a 'pure' hierarchy based on a single logical relation. Because of the use of several different logical relations together, a single concept can be subordinate to two or three superordinate concepts. For example, *C3PO* is a meronym of *Star Wars Episode One*, a hyponym of *robot*, and a subordinate possessed by *boy*. This mix of relations and multiple subordination in the hierarchy is not considered problematic.

5.2.4 Mapping the discourse onto the hierarchy

Having created a hierarchy based on the relations between the various key concepts, we are now in a position to map the occurrences of these concepts in the discourse onto the hierarchy.

At first sight, mapping the discourse onto the hierarchy would appear to be straightforward. We could simply identify the occurrence of each of the key

concepts in the discourse and put this sequence of occurrences into the hierarchy. While this is a potentially valid approach, it can lead to an excessive number of repetitive moves between concepts around the hierarchy. To illustrate this, let us have a look at the short stretch of discourse in example 5.1 taken from extract G.

Example 5.1

T: [8] **Where, where** do the **boy** take these **people** to? (2.0) [9] **Where** do this **boy**, this **boy** take these **people** to?

S: [10] The **boy** takes these **people** to His (the **boy's**) **house**.

T: [11] The **boy** takes these **people** to His (the **boy's**) **place**, his (the **boy's**) **house**. [12] **Why** does the **boy** take these **people** to his (the **boy's**) **house**?

S: [13] The **boy** takes these **people** to his (the **boy's**) **house** because **Storm**.

T: [14] Yes, there is a **storm, sandstorm**, right?

In example 5.1, all occurrences of key concepts have been given in bold type. If we count every possible occurrence of a key concept as a potential move, we find that the 7 T-units in example 5.1 contain 5 key concepts and 29 potential moves between concepts (25 if we do not count moves to the same concept), with 6 of these moves being from *boy* to *people* (one for each of T-units 8 to 13) and another 5 being from *people* to *boy*. This large proportion of moves to T-units is due to the recurrence of a single shared proposition. Including all of these moves between concepts in our analysis reduces the significance of each move and is also problematic for more important reasons.

Before we continue, let us pause and look again at how we identified concepts. The purpose of our analysis here is to try to create a simulation of semantic space as perceived by the participants. For this reason, we included both ellipted material and referents as concepts, since they are in the consciousness of the discourse participants. We will also therefore have to include ellipted material and referents as occurrences of concepts when we map the discourse onto the hierarchy.

However, we do not need to include every occurrence of a concept in this process. We can assume that constantly recurring concepts are activated in semantic space already and can concentrate on the introduction of new concepts into the discourse. The key factor that is important for our analysis is how each newly introduced concept links in semantic space to the immediately preceding concept. To retain the significance of moves between

concepts in the hierarchy, therefore, I will not include every move that occurs in the discourse. Instead, I will concentrate on those moves which introduce new concepts or new links between concepts. To do this, some guidelines are needed.

Guideline 1: The first occurrence of a concept in the discourse is always counted as a move.

Guideline 2: For each first occurrence of a concept, the move to this concept is from the concept immediately preceding it in the discourse.

Guideline 3: Repetitions of the same concepts in two (or more) immediately succeeding T-units are not counted as moves.

Guideline 4: Repetitions of the same two (or more) concepts in non-immediately succeeding T-units are counted as moves except where the intervening T-unit is an unclear T-unit.

Guideline 5: All other occurrences of concepts are counted as moves.

Let us see how these guidelines can be put into practice. The application of guidelines 1 to 3 can be seen from example 5.1 above. T-unit 8 contains the first occurrence of *house* (referred to by the cataphoric *where*), *boy* and *people*. Following guideline 1, we therefore have three moves, the first from the preceding *Star Wars Episode One* to *house*, the second from *house* to *boy*, and the third from *boy* to *people*.

Following guideline 3, the repetition of *house*, *boy* and *people* in T-units 9, 10 and 11 are not counted as moves. The next new concept is *storm* (referred to by the cataphoric *why*). The concept immediately preceding *storm* is *house*. Following guideline 2, we need to include a move from *house* to *storm*. Working backwards, we therefore need a move from *people* to *house* before we can move on to *storm*.

Guideline 4 can be illustrated from T-units 53 to 58 of extract G given as example 5.2 below.

Example 5.2

T: [53] *Tah sahng progrum mah,* robot *samaht tum dý.* {= If you construct the program, a robot can do it $_{\text{(cook food)}}$.} [54] And what else $_{\text{can robots do}}$?

(1.0) [55] Who has seen this movie (Star Wars Episode One)? [56] Who has seen this film (Star Wars Episode One)? [57] Raise your hand. [T raises her hand as a model.] [58] *Krý mah doo laaw nung nee?* {= Who has already seen this film (Star Wars Episode One)?}

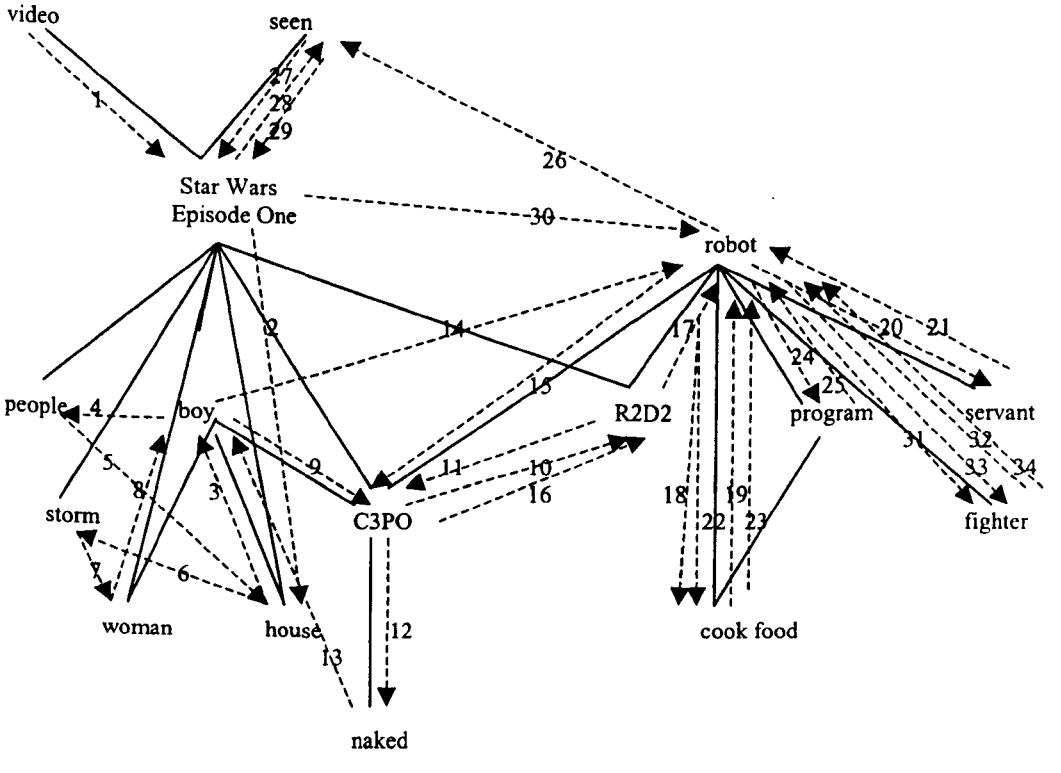
In this example, *seen* and *Star Wars Episode One* in T-unit 55 are new information since the preceding discourse has concerned robots. We therefore have moves from *robot* to *seen*, and from *seen* to *Star Wars Episode One*. There is no move in T-unit 56 following guideline 3. T-unit 57 contains no key concepts but is an intervening T-unit between T-units 56 and 58. Following guideline 4, we therefore need to include a move from *Star Wars Episode One* (the preceding concept) to *seen*, and from *seen* to *Star Wars Episode One* in T-unit 58.

Applying these guidelines to the whole of extract G, we find that there are 34 moves in the extract as shown in Table 5.3.

Move	T-unit in which the move occurs	Starting concept	End concept
0	2	-	video
1	5	video	Star Wars Episode One
2	8	Star Wars Episode One	house
3	8	house	boy
4	8	boy -	people
5	11	people	house
6	12	house	storm
7	15	storm	woman
8	16	woman	boy
9	18	boy	C3PO
10	26	C3PO	R2D2
11	33	R2D2	C3PO
12	33	C3PO	naked
13	36	naked	boy
14	38	boy	robot
15	38	robot	C3PO
16	38	C3PO	R2D2
17	39	R2D2	robot
18	41	robot	cook food
19	42	cook food	robot
20	42	robot	servant
21	46	servant	robot
22	48	robot	cook food

23	51	cook food	robot
24	51	robot	program
25	54	program	robot
26	55	robot	seen
27	55	seen	Star Wars Episode One
28	58	Star Wars Episode One	seen
29	58	seen	Star Wars Episode One
30	60	Star Wars Episode One	robot
31	62	robot	fighter
32	64	fighter	robot
33	67	robot	fighter
34	68	fighter	robot

Table 5.3 Moves between concepts in extract G



relations between concepts

moves between concepts (numbers indicating the number of the move in the discourse)

Figure 5.3 Moves between concepts mapped onto the hierarchy for extract G

While Table 5.3 is useful in showing which concepts succeed other concepts in the discourse, to make it more applicable to our search for topics, we can map these moves onto the hierarchy. This is shown in Figure 5.3.

Although it is difficult to follow the moves between concepts in the hierarchy, Figure 5.3 does highlight one key aspect important in our search for topics. From Figure 5.3, we notice that *robot* is the concept most frequently linked by moves between concepts. In addition, *boy* is also frequently linked, and *seen* and *Star Wars Episode One* form an insertion sequence between moves 26 and 30. Mapping the moves between concepts onto the hierarchy may therefore give us some ideas concerning identifying topics, topic entities and topic progression. I will look at this point in more detail in section 5.2.6.

5.2.5 Measuring the distance of each move

The purpose of constructing a hierarchy is to make a model of semantic space for the key concepts in the discourse. Distances within the hierarchy should reflect distances in semantic space. Having identified each of the moves between key concepts in the discourse, we are now in a position to judge the distance in semantic space of each move by looking at the number of hierarchical relations each move involves. In this way, we may be able to identify topic shift where moves cover a large distance in the hierarchy, topic drift where there is a series of succeeding small moves, and topic maintenance where the same concept(s) recurs over a stretch of discourse. The key problem here is how to translate the relations in the hierarchy into distances of moves between concepts.

The first point that we can make about distances of moves is something that was implicitly assumed when we were identifying moves between concepts. Where succeeding T-units contain the same concept, there was no move identified. In other words, we are assuming a distance of 0 between a concept and itself.

In Watson Todd (1998), I made some tentative suggestions for assigning distances to moves between concepts. A value of 1 was assigned to moves between a key concept and its immediate superordinate or subordinate concept. Thus, in extract G, the first move from *video* to *Star Wars Episode One* is assigned a distance of 1.

In this study, I will include two other situations where the distance between moves is assigned a value of 1. Firstly, moves between co-subordinates of the same immediate superordinate are considered as being separated by a distance of 1. Thus move 4 in extract G from *boy* to *people*, which are co-meronyms of *Star Wars Episode One*, has a distance of 1. Secondly, any move between two concepts occurring in the same T-unit in the discourse is assigned a value of 1, since the two concepts should be closely related in the speaker's mind in order to occur in the same T-unit. In extract G, move 13 from *naked* to *boy* is given a distance of 1 even though *boy* is not the immediate superordinate of *naked*. This second case holds except where it is explicitly stated that the two concepts are unrelated, such as in the invented teacher turn *Let's now turn from robots to flowers*.

The distance of other moves between concepts in the hierarchy can be found by counting the number of hierarchical links of the shortest path in the hierarchy between the two concepts. In extract G, move 14 from *boy* to *robot* would be given a value of 2 (moving from *boy* to *C3PO* to *robot*), and move 26 from *robot* to *seen* would be given a value of 3 (moving from *robot* to *R2D2* to *Star Wars Episode One* to *seen*).

A final consideration for identifying distances between moves, even though there are no examples in extract G, is for moves between concepts for which no path linking them can be found in the hierarchy. In Watson Todd (1998), I stated that such moves have "an indeterminate but large value" (p. 315). In this study, I will assume that such moves have a value of 6 (since we may find reasons for assigning such moves a specific value), which is a higher value than the distance between any two concepts linked by a pathway in a hierarchy in this study.

Following these methods for identifying distances of moves, we can assign values to the moves in extract G, and these are given in Table 5.4.

Move	T-unit in which the move occurs	Starting concept	End concept	Distance of move
0	2	-	video	-
1	5	video	Star Wars Episode One	1
2	8	Star Wars Episode One	house	1
3	8	house	boy	1

4	8	boy	people	1
5	11	people	house	1
6	12	house	storm	1
7	15	storm	woman	1
8	16	woman	boy	1
9	18	boy	C3PO	1
10	26	C3PO	R2D2	1
11	33	R2D2	C3PO	1
12	33	C3PO	naked	1
13	36	naked	boy	1
14	38	boy	robot	2
15	38	robot	C3PO	1
16	38	C3PO	R2D2	1
17	39	R2D2	robot	1
18	41	robot	cook food	1
19	42	cook food	robot	1
20	42	robot	servant	1
21	46	servant	robot	1
22	48	robot	cook food	1
23	51	cook food	robot	1
24	51	robot	program	1
25	54	program	robot	1
26	55	robot	seen	3
27	55	seen	Star Wars Episode One	1
28	58	Star Wars Episode One	seen	1
29	58	seen	Star Wars Episode One	1
30	60	Star Wars Episode One	robot	2
31	62	robot	fighter	1
32	64	fighter	robot	1
33	67	robot	fighter	1
34	68	fighter	robot	1

Table 5.4 Distances of moves between concepts in extract G

5.2.6 Identifying topics and topic progression from the analysis

The purpose of conducting the analysis is to gain insights into topics and topic progression in the classroom. For extract G, several tentative suggestions can be made from the analysis. In looking at these, I will start by showing how the analysis sheds light on topic progression.

In chapter 2, we saw that there are three main types of topic progression: topic maintenance, topic drift, and topic shift. I will look at how each of these can be identified from the analysis.

Topic maintenance is where a coherent stretch of discourse is about the same topic. We might, however, distinguish topic maintenance occurring at several different levels. How long does a stretch of discourse need to be for it to be considered to stand alone as monotopical discourse which is not subsumed into some other type of topic progression at a higher level? One T-unit is too short, but should the minimum length of an unsubsumed stretch of monotopical discourse be 2 T-units, 5 T-units, 10 T-units or some other figure? A stretch of monotopical discourse 2 T-units long is less striking than a stretch 20 T-units long, yet both exhibit topic maintenance. A stretch 2 T-units long may indicate topic insertion or may be subsumed into a longer stretch of topic drift, whereas a stretch of monotopical discourse 20 T-units long probably stands as topic maintenance at higher levels of analysis as well. In this way, we can see how a hierarchical approach to topics affects our identification of topic progression.

In extract G, the 3 T-unit stretch from T-unit 2 to T-unit 4 concerns *video*, and the 16 T-unit stretch between T-units 39 and 54 is about *robot*. It seems likely that, given its relatively short length with no surrounding discontinuities, the former can be subsumed into another type of topic progression at a higher level, whereas the latter is likely to still be identified as topic maintenance at higher levels.

It should also be noted that the topics of these two stretches were identified in different ways. In T-units 2 to 4, *video* is the only key concept to occur and must therefore be the topic of these T-units. In T-units 39 to 54, on the other hand, *cook food*, *servant* and *program* co-occur with *robot*. *Robot* is the first of these 4 key concepts which occurs in the stretch of discourse and occurs in every T-unit in the stretch (except for the unclear T-unit 47), and it is therefore identified as the topic of these T-units.

Let us now turn to whether these instances of topic maintenance are subsumed into a different type of topic progression at higher levels of analysis. The initial maintenance of *video* changes to *Star Wars Episode One*, and then to various components of the latter. From our analysis, all of the moves up to move 14 at T-unit 38 only cover a distance of 1, and we are therefore not concerned with topic shift where greater distances are covered.

The first part of this stretch of discourse involves a steady progression from the more general *video* at the top of the hierarchy to more specific concepts. It is therefore an example of Push topic drift (see p. 33) where there is more concentration on sub-topics as the discourse progresses.

From T-units 8 to 38, the discourse stays at the same level of specificity, but covers a wide range of concepts, all of which are components of *Star Wars Episode One*. This is an example of topic fading, one kind of topic drift, where new topics at the same level of specificity are steadily introduced as the discourse progresses.

At T-unit 38, we have the first move with a distance greater than 1. Although the distance of 2 is not very large, it is an indication of topic shift, albeit not a massive jump to a completely unrelated topic. This shift is coherent shift, since it is marked by a framing move in T-unit 37.

The ensuing discourse is about a single topic, *robot*, up to T-unit 55, where there is another shift, this time with a distance of 3 indicating a more abrupt shift. This time the shift is not marked and is thus noncoherent. After 5 more T-units about *seen* and *Star Wars Episode One*, there is another noncoherent shift back to the topic of *robot*. The stretch of discourse from T-unit 55 to T-unit 59 is therefore an insertion sequence (in the terms of conversation analysis) with the first shift at T-unit 55 being topic insert and the second at T-unit 60 being topic renewal. It should be noted that, for topic renewal to be identified, the pre- and post-insertion stretches of discourse need to concern the same topic (i.e. there should be a move with a value of 0, or possibly 1, between the pre- and post-insertion stretches of discourse if the insertion were not there). From T-unit 60 to the end at T-unit 74, the topic of *robot* is maintained. These different kinds of progression at different levels are shown in Figure 5.4 below.

In the discussion so far, I have been using the term *topic* very loosely. In chapter 2, however, I made a distinction between *topic entity* and *topic*, with *topic entity* referring to a key concept which is central to a short stretch of discourse and which may be subsumed into a larger *topic*. On this basis, the initial *video* in extract G would be regarded as a topic entity and *robot* in the second half of the extract is a topic.

While the second half of extract G is relatively unproblematic regarding the identification of the topic, the first half is less clear. From the start of the

extract up to T-unit 38, there are a potential nine topic entities: *video*, *Star Wars Episode One*, *house*, *boy*, *people*, *storm*, *woman*, *C3PO* and *R2D2*. To decide which of these should be identified as topics, we can look back at the hierarchy in Figure 5.3. Following de Beaugrande and Dressler's (1981) guideline that density of linkage is indicative of topics, we can see that *boy* is the most densely linked concept with *house*, *C3PO* and *R2D2* next most densely linked. While *house* is closely associated with *boy* in the discourse, *C3PO* and *R2D2* are separate (until T-unit 36) with the topic drifting from *boy* in T-units 8-17, through *C3PO* in T-units 18-25, to *R2D2* in T-units 26-32, before linking the three concepts together in T-units 33-36.

We still need to consider how to analyse T-units 1-7 which provide the context for the discourse before the more specific discussion concerning *boy*. In this section, *video* is the first concept mentioned and appears in every T-unit, so this will be identified as the topic. These relationships between topic entities and topics at different levels of analysis are summarised in Figure 5.4.

From Figure 5.4, I hope that it can be seen that a topic-based analysis using relations is a promising method of analysis for identifying topics and topic progression. Both topic entities and topics can be identified, and topic progression can be followed. The application of the analysis is of course not restricted to extract G; similarly useful findings can be derived for all of the extracts. Before we start thinking that our search for an approach to topics and topic progression is finished, however, we must remember that the use of relations as a basis for analysis is only one possible approach, and we also need to compare the results with findings from other methods of analysis. In the next section, I will look at conducting a similar approach based on associations, and in the next chapter, I will compare the findings from all of the approaches.

T-unit	Topic entity	Topic progression	Topic	Summary
2–4	video	drift (Push)	video	
5–7	Star Wars Episode One	drift (Push)		
8–11	boy + house + people	drift (fading)		
12–14	storm		boy	topic drift from *video* to *boy*, *C3PO* and *R2D2* (all within the context of *Star Wars Episode One*)
15–17	boy + woman			
18–25	C3PO		C3PO	
26–32	R2D2		R2D2	
33–36	boy + C3PO + R2D2			
		shift (coherent)		
37–54	robot	maintenance	robot	topic maintenance of *robot*
		shift (insert)		
55–59	seen + Star Wars Episode One		seen + Star Wars Episode One	topic insertion
		shift (renewal)		
60–74	robot	maintenance	robot	topic maintenance of *robot*

Figure 5.4 Summary of topic entities, topics and topic progression in extract G as identified from a hierarchical approach

5.3 Conducting a topic-based analysis using associations

Topics, as defined in this study, are associated or related clusterings of concepts. In the previous section, we looked at how clusterings of concepts may arise through relations. The purposes of this section are to examine how associations may lead to clusterings of concepts or topics, and to trace topic progression through the use of associations.

Much of the work that is required to fulfil these purposes has already been accomplished in the previous section. To identify topics and follow topic progression using associations of concepts, we need to identify the concepts that may be associated and look at the movements between the concepts in the discourse. These two stages of analysis have been examined in sections 5.2.1 and 5.2.4 (especially Table 5.3) respectively. I will therefore use these findings from the previous section in the current analysis based on associations.

There are, however, certain stages of analysis which have not been completed and which will be examined in detail below. These are identifying and measuring associations between concepts, creating maps of semantic space based on associations, and from these maps identifying topics and topic progression.

5.3.1 Identifying and measuring associations between concepts

Associations are primarily cognitive rather than linguistic (although there has been research into linguistic aspects of associations e.g. Carter, 1987; Schmitt, 1998). In a network or connectionist approach to cognition, associations may be viewed as being responsible for much cognitive structure (Winitzky et al., 1994).

Traditionally associations have been elicited through stimulus-response tests, where the first word given in response to a particular stimulus is taken as being an associate of that stimulus. Such an approach was one of the first research methods used in psychology, and it reached its heyday in the nineteen sixties and seventies, when research was conducted into the relationships between association, on the one hand, and age (e.g. Palermo, 1965), reading speed (e.g. Jay, 1968), reading retention (e.g. Shima, 1970), and sentence construction (e.g. Prentice, 1967), on the other.

Such stimulus-response associations, however, have been criticised. In accepting single associations as sufficient evidence of cognitive structure,

common and uncommon associations of a word are not differentiated, leading to acceptable associations between, for example, *dark* and *bench* (Schmitt, 1998). To overcome this problem, Schmitt advocates using the first three responses to a stimulus as associates and assigning measures of likelihood of a response truly being an associate by comparing it to a large number of other responses. To assign such measures of association, Schmitt elicited three responses to 17 stimuli for 100 subjects. To conduct such large-scale stimulus-response research to elicit associations in the context of this study is impractical. We must therefore look for other ways of identifying associations.

If we view discourse as a network of connected nodes, each representing a concept, as in the approach advocated by de Beaugrande and Dressler (1981), there is some evidence that these networks reflect and are reflected by similar networks between concepts in the mind (Givón, 1995). In other words, associations between concepts in discourse are reflective of the associations that are responsible for cognitive structure.

Following Gleason and Ratner (1998), I will assume that the similarities between associations in discourse and in cognition occur because associations are built through exposure. Thus, for example, a child may hear the words *bottle* and *milk* occurring together in discourse frequently, and from this frequent exposure build an association between the two words. This is, in fact, a connectionist explanation of how associations develop (see Fodor and Pylyshyn, 1988).

To be able to identify associations, then, we need to see the language that a person has experienced. While it is impossible to ever collect all the language experience of a person, a very rough guide to this experience can be found in a corpus (Hunston, 2002). Although not equivalent to any particular person's language experience, a corpus sampling the language as a whole can allow the identification of the association patterns that most people hold (Aston and Burnard, 1998).

It should be noted, however, that there are arguments against linking cognitive structure and corpus data. For example, Widdowson (1996) points out that corpus data provides no insights into prototypes, so although *pea* is not the most frequently occurring vegetable in most corpora, it is the most cited prototype of native English speakers.

Nevertheless, the difficulties of gaining any insights into a person's experience of language other than through a corpus mean that I will take a corpus approach to identifying associations. If we are doing this, we need to decide what corpus to use. If frequency of co-occurrences leads to associations, then the corpus we should be using is one which provides a wide coverage of a range of discourse types and genres which may reflect how users of English are exposed to the language (with the lack of any corpus of English used by Thais, I am assuming that the Thais' exposure to English will be similar to that of other users of English). For this reason, I will use the British National Corpus (2000) as the corpus in this study, since the 100 million words it contains come from a variety of discourse types and genres.

A corpus linguistic approach to identifying associations, then, can be used instead of stimulus-response tests, and yields very similar associations (Hearst, 1997). Using such an approach, for example, Scott (1997) argued that pairs of key words in the same text are associated.

Having decided to take a corpus approach and chosen the corpus, the next stage to consider is how to identify the co-occurrences in the corpus which lead to associations. Most previous research into co-occurrences in corpora has focused on identifying collocations. Identifying collocations usually involves examining the discourse up to four words either side of the node word for which collocates are being sought (Sinclair, 1991). Associates, however, may be located at a greater distance from the node word than collocates, and therefore I will look at co-occurrences up to ten words either side of the node word.

	boy	cook food	house	servant	naked	people	robot	C3PO	R2D2	Star Wars	video	fighter	storm	seen	woman
program	-2.0	*	-5.3	*	-0.3	-6.1	1.8	*	*	*	0.6	*	-0.5	-3.2	*
woman	4.4	*	5.7	3.7	21.4	-12.0	-0.7	*	*	*	-1.9	-0.4	-1.6	4.8	
seen	4.6	*	-1.3	-0.3	9.3	1.7	-0.5	*	*	0.3	3.9	0.7	-2.0		
storm	-0.5	14.3	0.2	*	0.3	-1.3	*	*	*	*	-0.4	1.0			
fighter	-0.1	*	-2.6	*	*	-3.0	*	*	*	*	2.7				
video	-1.0	*	-0.2	*	0.5	-3.9	2.1	*	*	*					
Star Wars	*	*	*	*	*	*	54.4	319.3	159.6						
R2D2	*	*	*	*	*	*	79.1	*							
C3PO	*	*	*	*	*	*	158.2								
robot	0.1	*	-2.3	0.9	1.4	-1.3									
people	-10.5	*	-2.0	-2.4	-2.5										
naked	6.1	*	-1.4	4.5											
servant	0.1	*	5.0												
house	0.4	6.7													
cook food	*														

Note:
* indicates pairs of concepts which do not co-occur in the British National Corpus
Star Wars was used as a search term in the corpus rather than *Star Wars Episode One*, due to the recentness of the latter term.

Table 5.5 Z-scores for associations between concepts in extract G

To identify whether a particular co-occurrence is important or not (whether it is common or uncommon in Schmitt's (1998) terms), we need to compare the observed frequency of co-occurrences of two items in the corpus against the expected frequency of them co-occurring given the overall frequency of each of the items in the corpus. In other words, we do not simply want to make a frequency count of co-occurrences, but need to examine the number of co-occurrences while accounting for the likelihood of the two items occurring together by chance (Biber et al., 1998). To do this, we can use the z-score (see Shei, 1999), which is calculated as follows:

$$z = \frac{O - E}{\sigma}$$

where O = the observed frequency of co-occurrence;
E = the expected frequency of co-occurrence;
σ = the standard deviation of the potential associate in the corpus.

Using the British National Corpus (2000), z-scores were calculated for each of the pairs of concepts identified in extract G (see p. 127). For concepts expressed through polysemous words, the query term used was restricted to the part of speech of the word expressing the concept (e.g. *well* as a noun in extract J). Table 5.5 shows the z-scores for extract G.

Table 5.5 shows the likelihood that any two concepts are associated. Two concepts which never co-occur in the corpus should not be considered associated, and negative scores indicating a lower observed than expected frequency are unlikely to show associations. But should we take all positive z-scores to be indicative of associations?

Referring back to research on collocations as a benchmark, Barnbrook (1996) recommends that z-scores greater than 3.0 represent collocations. Associations can be considered weaker than collocations and so we can set a lower minimum z-score for identifying associations. The number of associations which would be identified between concepts in extract G for a variety of threshold levels of z-scores is given in Table 5.6.

Threshold z-score	Number of associations
3.0	18
2.0	20
1.0	24
0.0	34

Table 5.6 Number of associations in extract G for different threshold levels of z-scores

If we assume that the number of associations between concepts is comparable to the number of relations between the same concepts, and remembering that we identified 21 relations between concepts for extract G, then it seems likely that we should set a threshold of 1.0 or 2.0 to identify associations. For all extracts (except extract H), the number of associations identified with a threshold z-score of both 1.0 and 2.0 is higher or equal to the number of relations. To retain comparability between associations and relations while also setting the threshold lower than that needed to identify collocations, I will set the minimum z-score to identify an association at 2.0. Doing this we find that the following pairs of concepts are associated.

C3PO - Star Wars
R2D2 - Star Wars
C3PO - robot
R2D2 - robot
Star Wars - robot
woman - naked
storm - cook food
seen - naked
house - cook food
naked - boy
woman - house
servant - house
woman - seen
seen - boy
naked - servant
woman - boy
seen - video
woman - servant
fighter - video
video - robot

Intuitively some of these associations seem anomalous. There are several potential reasons for this. Firstly, we may have been unlucky enough to have identified one of the "instances of untypical, irregular, accidental, and possibly erroneous phenomena" in the corpus (Aston and Burnard, 1998: 36-37). The strong association between *storm* and *cook food* comes from a single co-occurrence which is highlighted due to the overall low frequency of *cook food* in the British National Corpus. Secondly, specific language forms which do not necessarily lead to associations between their parts may be involved. For example, 3 of the 4 instances of co-occurrence between *naked* and *servant* come from the title *The Naked Civil Servant*. Thirdly, the nature of the corpus chosen may affect the results in two ways. We have already seen that the search term *Star Wars* needs to be used instead of *Star Wars Episode One* because of the recent nature of the latter title. In the later film, a boy played a major role, whereas in the earlier films boys were almost entirely absent. We might then expect some recent co-occurrences of *boy* and *Star Wars Episode One* which even a corpus as recent as the version of the British National Corpus used in this study (2000) would not contain. More seriously, the selection of texts included in the British National Corpus, with 90% coming from written sources and most of these being informative texts which are unlikely to contain many instances of the concepts from extract G, may have led to some over- and under-reporting of associations (e.g. the non-co-occurrence of *cook food* with either *woman* or *servant* which might be more prevalent in nineteenth century literature). Despite these anomalies, many of the associations identified seem intuitively reasonable (e.g. *Star Wars - robot*, *servant - house*). It seems therefore worth continuing with the analysis to see what emerges.

5.3.2 Creating maps of semantic space based on associations

While the list of pairs of concepts considered associated based on the z-scores is potentially useful for analysing movements between concepts in the discourse, using these raw z-score data is not the only possible approach. In the analysis based on relations, we fitted all the identified relations together into a single hierarchy. We can use a similar approach with associations, although the pattern joining them will be a network rather than a hierarchy. Before I attempt to construct a network from the paired associates, I will give a brief overview of the nature and uses of networks.

The use of networks to represent the organisation of concepts in cognitive psychology grew out of a dissatisfaction with hierarchies. In 1969, Collins and Quillian posited a hierarchical tree with extra concepts attached to nodes to represent conceptual organisation. They further

posited that response times taken to link two concepts would be proportional to the distance between the two concepts in the tree. Their initial results suggested that this was true. However, Conrad (1972) took an alternative approach. Instead of predetermining the relationships between concepts as a hierarchy, he asked subjects to describe certain concepts (analogous to eliciting associations) and drew up networks for the concepts based on these descriptions. He then showed that these networks were better predictors of response times than Collins and Quillian's hierarchies. Since Conrad's research, networks have been influential as ways of representing conceptual organisation in cognitive psychology.

There have, however, been several progressively more complicated variations on the theme of networks to represent the organisation of concepts. Initial models proposed a spreading activation between nodes in a network (e.g. Collins and Loftus, 1975). Refinements to this spreading activation model incorporated strengths of connections (usually indicated by weightings) between nodes in the model (e.g. Anderson, 1980). Further refinements allowed the weightings of connections to be negative (i.e. inhibitory) as well as positive, and allowed activation to move backwards and forwards along connections rather than only spreading outwards. Yet further refinements led to less intuitive connectionist models where concepts are represented by combinations of connections between nodes rather than by the nodes themselves (McClelland and Rumelhart, 1985).

Research into networks as representations of conceptual organisation has shown that they predict results well and so are likely to be better representations of how people organise knowledge than hierarchies. For example, there has been a lot of research using networks which has produced results matching people's reading performance and acquisition (e.g. Ans *et al.*, 1998; McClelland and Rumelhart, 1981; McEneaney, 1994; Seidenberg, 1992). Other aspects of linguistics which have been validly modelled by networks include child acquisition of German articles (McWhinney *et al.*, 1989), learning of regular and irregular forms of English verbs (Ellis and Schmidt, 1998), and ability to identify word boundaries (Christiansen *et al.*, 1998). More controversially, research contrasting network models and rule-based models has generally shown that networks explain human thinking better (e.g. Hunt, 1989; Ney and Pearson, 1990).

Given that the assumptions underpinning a corpus approach to identifying associations reflect a connectionist approach and that many networks are

built from associations, the findings above supporting the use of networks to describe conceptual organisation suggest that an analysis based on associations could be valuable.

It is not, however, a straightforward matter to draw up a network. While it is possible to take each of the 16 concepts in extract G as a node each radiating 15 connections to other nodes, the resulting network of 240 connections would be very unclear and probably unhelpful. Alternatively, for the 16 nodes representing concepts in extract G we could use only those 24 connections which represent associations. While somewhat clearer, such an approach adds no extra information to that contained in Table 5.5 and still ignores the vast majority of potential connections between concepts, especially those which are inhibitory.

One approach which does take all connections into account while also providing a relatively clear picture is the use of ultrametric (additive and extended) trees. Such trees represent the similarities between items in a set in terms of the distance between concepts (Corter, 1996). In an ultrametric tree, the distance (d') between three objects *a*, *b* and *c* is shown by

$$d'(a,b) \quad \text{MAX}\,[d'(a,c), d'(b,c)]$$

An example may make things clearer. If we have three objects *a*, *b* and *c* whose dissimilarities are measured as given in Table 5.7, we can draw up a tree showing these dissimilarities as in Figure 5.5.

	a	*b*
c	4	8
b	8	

Table 5.7 An example of dissimilarities between three objects

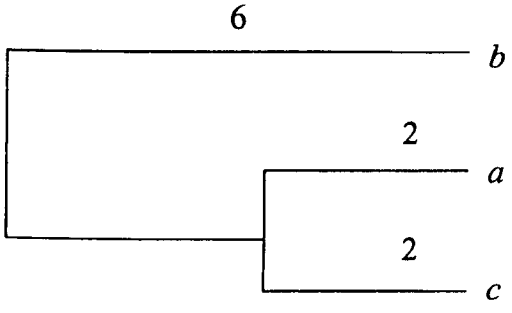


Figure 5.5 A tree drawn up from Table 5.7

Following the horizontal paths in Figure 5.5, we find the distances of dissimilarity given in Table 5.7. Figure 5.5 shows that *a* and *c* are more closely connected to each other than either is to *b*, suggesting a small clustering of objects. In this way, a tree can provide a clear picture taking all connections between pairs of concepts into account.

The tree model used in this study is an extended tree model to cope with the large number of similarity-connected concepts, and the program *EXTREE* (Corter, 1986) was used to fit this model to the data.

In fitting the model to the data, the z-scores in Table 5.5 were used. Positive z-scores show similarity between concepts, and negative z-scores show dissimilarity. To allow pairs of concepts which do not co-occur in the corpus to be included in the model (i.e. pairs indicated by * in Table 5.5, such as *Star Wars* and *boy*), these were assigned a value of -100.0, a value which is markedly lower than any score for pairs of concepts which do co-occur and which is equivalent in magnitude to the most closely associated pairs of concepts. Inputting these z-score data for extract G into an extended tree model produces the tree given in Figure 5.6.

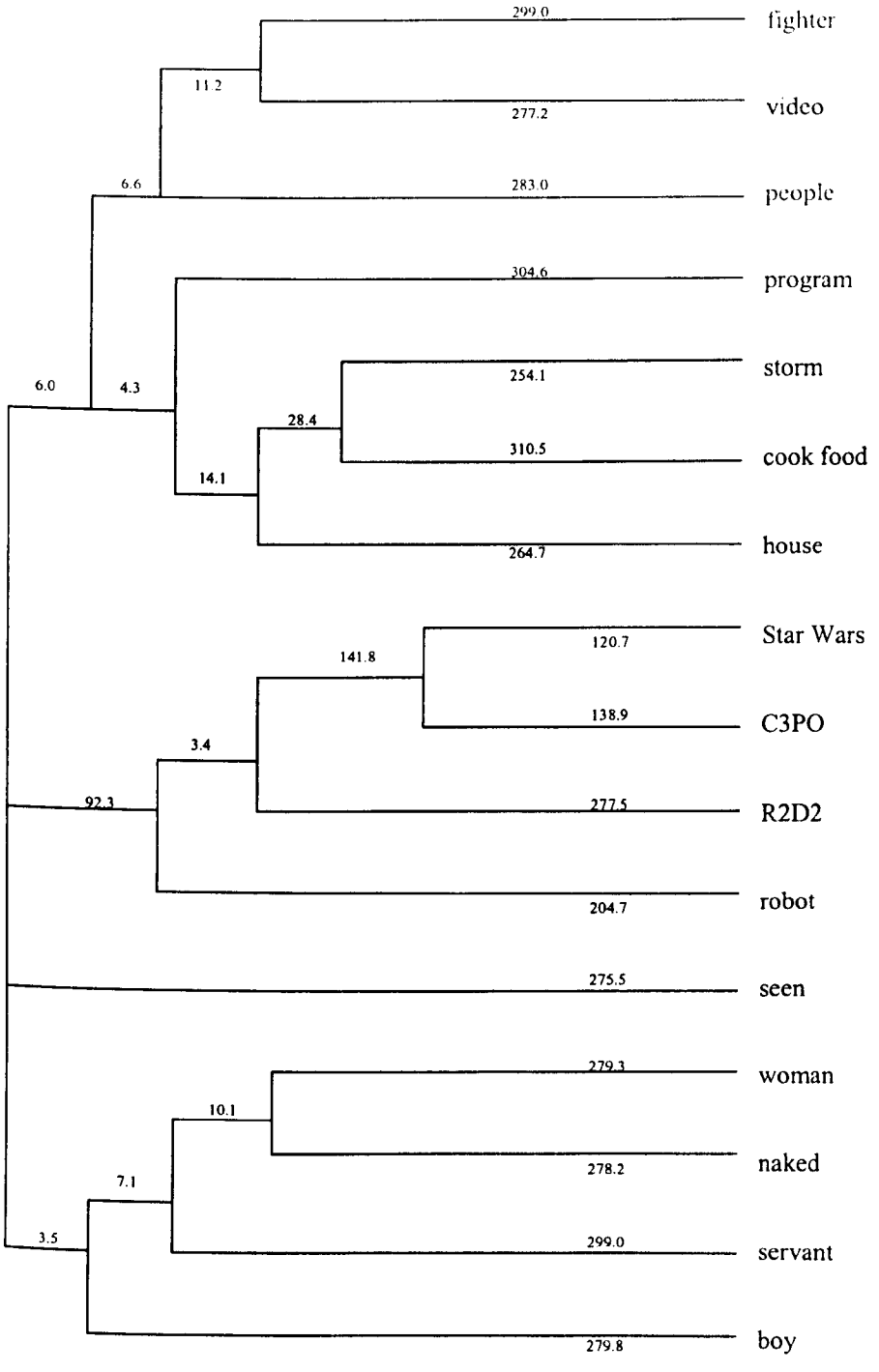
From Figure 5.6, we can see that *Star Wars* and *C3PO* are the most connected pair of concepts (which we already knew from the z-scores) and that *Star Wars*, *C3PO*, *R2D2* and *robot* are the most closely associated cluster of concepts. On the other hand, *seen* is the concept which is least associated with any cluster, something not apparent from Table 5.5. The clusters shown in Figure 5.6 are summarised below.

((*fighter* + *video*) + *people*)
(((*storm* + *cook food*) + *house*) + *program*)
(((*Star Wars* + *C3PO*) + *R2D2*) + *robot*)
(*seen*)
(((*woman* + *naked*) + *servant*) + *boy*)

It should be noted that pairs of concepts which never co-occur in the corpus can be located within the same cluster (e.g. *C3PO* and *R2D2*). This is because, in drawing up the extended tree, the relationships between all concepts are considered, and thus the co-occurrence of both *C3PO* and *R2D2* with *Star Wars* leads to an implicit relationship between the first two concepts.

Given the definition of *topic* used in this study as involving clusterings of concepts, using a tree analysis may be an appropriate approach in that it is

the only method of analysis used in this study that explicitly generates clusters of concepts which may be indicative of topics.



Note: Numbers refer to distances between concepts

Figure 5.6 Extended tree diagram of associations for extract G

The tree analysis also fits with our aim of producing models of semantic space. Although the tree in Figure 5.6 looks superficially like a hierarchy on its side, it is actually far more akin to a network. While most network approaches use weightings to indicate strengths of connections between concepts, it is also possible to use distances (e.g. Hofstadter, 1996) with longer distances indicating less strong connections. We can therefore view Figure 5.6 as a model of semantic space for the 16 concepts in extract G based on corpus data. The values in the figure show the distance between nodes in the tree with shorter distances representing stronger connections between nodes.

5.3.3 General and discourse-specific associations

We have identified associations using two methods: z-scores between concepts in a corpus, and clusterings in a tree analysis based on these z-scores. Both of these methods aim to identify general associations between concepts. We can apply these methods of identifying associations to the moves between the concepts in extract G given in Table 5.3. Doing this gives us Table 5.8.

Move	T-unit	Starting concept	End concept	z-score	Same clustering in tree analysis
0	2	-	video		
1	5	video	Star Wars Episode One	*	N
2	8	Star Wars Episode One	house	*	N
3	8	house	boy	0.4	N
4	8	boy	people	-10.5	N
5	11	people	house	-2.0	N
6	12	house	storm	0.2	**Y**
7	15	storm	woman	-1.6	N
8	16	woman	boy	**4.4**	**Y**
9	18	boy	C3PO	*	N
10	26	C3PO	R2D2	*	**Y**
11	33	R2D2	C3PO	*	**Y**
12	33	C3PO	naked	*	N
13	36	naked	boy	**6.1**	**Y**
14	38	boy	robot	0.1	N
15	38	robot	C3PO	**158.2**	**Y**
16	38	C3PO	R2D2	*	**Y**
17	39	R2D2	robot	**79.1**	**Y**
18	41	robot	cook food	*	N
19	42	cook food	robot	*	N
20	42	robot	servant	0.9	N
21	46	servant	robot	0.9	N

22	48	robot	cook food	*	N
23	51	cook food	robot	*	N
24	51	robot	program	1.8	N
25	54	program	robot	1.8	N
26	55	robot	seen	-0.5	N
27	55	seen	Star Wars Episode One	0.3	N
28	58	Star Wars Episode One	seen	0.3	N
29	58	seen	Star Wars Episode One	0.3	N
30	60	Star Wars Episode One	robot	**54.4**	**Y**
31	62	robot	fighter	*	N
32	64	fighter	robot	*	N
33	67	robot	fighter	*	N
34	68	fighter	robot	*	N

Bold indicates moves considered as associations.
* indicates pairs of concepts which do not co-occur in the British National Corpus.

Table 5.8 General associations between moves in extract G

From Table 5.8, we can see that a very low proportion of the moves between concepts are moves between associated concepts. Only 14.7% of moves between concepts are associated by z-score from the corpus, and only 26.5% are between concepts associated by clusterings in the tree analysis. Although these figures are lower than the average for all of the extracts (overall, 38.2% of moves are between concepts associated by z-score; 43.6% are between concepts associated by clusterings; and 57.5% are between concepts associated by either z-score or clustering), they suggest that we have not seen the full picture. Such low proportions of moves between concepts being moves between associated concepts (even the 57.5% of moves between associated concepts for all extracts) imply a very fragmented discourse where a large proportion of moves are between unassociated concepts.

The reason why there is such a low level of connectedness exhibited in the analysis is that we have only been considering general associations. We have yet to consider discourse-specific associations.

While general associations are important in providing a schematic framework for communication, they do not lead to the generation of new understandings that occurs when two generally unassociated concepts are linked together. To generate such a semantic link, we need to link the two

concepts syntactically in the same T-unit (Givón, 1995). It should be pointed out that some concepts have a preferred syntactic structure, and thus the generation of new temporary associations involves syntactic and semantic links working hand in hand. Where such temporary associations do not match the typical general associations we have identified through z-scores and clusterings in the tree analysis, they are discourse-specific associations. Including these discourse-specific associations (where concepts are associated by co-occurring in a T-unit) in our analysis, for extract G we find the associations given in Table 5.9.

Move	T-unit	Starting concept	End concept	z-score	Same clustering in tree analysis	Co-occurrence in a T-unit
0	2	-	video			
1	5	video	Star Wars Episode One	*	N	**Y**
2	8	Star Wars Episode One	house	*	N	N
3	8	house	boy	0.4	N	**Y**
4	8	boy	people	-10.5	N	**Y**
5	11	people	house	-2.0	N	**Y**
6	12	house	storm	0.2	**Y**	**Y**
7	15	storm	woman	-1.6	N	N
8	16	woman	boy	**4.4**	**Y**	**Y**
9	18	boy	C3PO	*	N	**Y**
10	26	C3PO	R2D2	*	**Y**	**Y**
11	33	R2D2	C3PO	*	**Y**	**Y**
12	33	C3PO	naked	*	N	**Y**
13	36	naked	boy	**6.1**	**Y**	**Y**
14	38	boy	robot	0.1	N	N
15	38	robot	C3PO	**158.2**	**Y**	**Y**
16	38	C3PO	R2D2	*	**Y**	**Y**
17	39	R2D2	robot	**79.1**	**Y**	**Y**
18	41	robot	cook food	*	N	**Y**
19	42	cook food	robot	*	N	**Y**
20	42	robot	servant	0.9	N	**Y**
21	46	servant	robot	0.9	N	**Y**
22	48	robot	cook food	*	N	**Y**
23	51	cook food	robot	*	N	**Y**
24	51	robot	program	1.8	N	**Y**
25	54	program	robot	1.8	N	**Y**
26	55	robot	seen	-0.5	N	N
27	55	seen	Star Wars Episode One	0.3	N	**Y**
28	58	Star Wars Episode One	seen	0.3	N	**Y**

29	58	seen	Star Wars Episode One	0.3	N	**Y**
30	60	Star Wars Episode One	robot	**54.4**	**Y**	N
31	62	robot	fighter	*	N	**Y**
32	64	fighter	robot	*	N	**Y**
33	67	robot	fighter	*	N	**Y**
34	68	fighter	robot	*	N	**Y**

Bold indicates moves considered as associations.

Table 5.9 General and discourse-specific associations between moves in extract G

We can take an approach aiming to maximise the number of moves identified as being connected by associations by saying that two concepts are associated if:

1. they have a z-score of at least 2.0 OR
2. they fall into the same clustering in the tree analysis OR
3. they appear together in a single T-unit in the extract.

While the first two of these aim to identify general associations between concepts, the third is more likely to identify discourse-specific associations. Including discourse-specific associations in our analysis, we find far fewer examples of non-associated moves between concepts. In fact, there are only four non-associated moves (move 2 at T-unit 8, move 7 at T-unit 15, move 14 at T-unit 38, and move 26 at T-unit 55). Of the 30 moves in extract G which are identified as associated in Table 5.9, only one (move 30) is associated solely through the general associations. This suggests that the discourse-specific associations generated by semantically linking two concepts in the same T-unit far outweigh the general associations based on the corpus z-score and extended tree clusters. In this instance, however, extract G is not representative of all of the extracts. Overall, 20.49% of the associated moves between concepts are identified solely through general associations. Although a minority of associated moves, this figure suggests that we cannot rely solely on discourse-specific associations to identify associated moves. I will therefore use the findings in Table 5.9 concerning both general and discourse-specific associations in trying to identify topics and follow topic progression.

5.3.4 Identifying topics and topic progression based on associations

In Table 5.9, we can see which moves in extract G are connected by associations and which are not. Following the same principles as in

section 5.2, we may assume that there are topic shifts at those four moves between concepts which are non-associated. Of these (moves 2, 7, 14 and 26), we may also say that move 2 shows the biggest shift since there is no evidence in the corpus that *Star Wars* and *house* co-occur, while move 14 shows the smallest shift as the z-score is positive.

While the validity of these shifts will be investigated in more detail in the next chapter, intuitively it seems that moves 14 and 26 may represent shifts. Move 2 is perhaps more akin to a plane shift (Sinclair and Brazil, 1982), whereby the discourse progresses from talk about a film to the contents of the film, than a topic shift. Move 7, on the other hand, probably does not represent a shift. Intuitively, the move from *storm* to *woman* is of the same type as move 6 and move 9, namely a move between two items of content in the film. If we do not identify moves 6 and 9 as shifts, it seems unsatisfactory to identify move 7 as a shift. Notwithstanding this problem, in this study I will continue looking at the implications of the analysis for topics and topic progression.

For stretches of discourse without breaks, in section 5.2 we identified the concepts most frequently occurring in the discourse as indicative of topic. We can use the same approach in the analysis based on associations. Taking the stretch of discourse from move 14 to move 26 (T-units 38 to 54) as an example, the concepts occurring in the moves are *robot - C3PO - R2D2 - robot - cook food - robot - servant - robot - cook food - robot - program - robot*, with *robot* occurring in every T-unit except the unclear T-unit 47. It therefore seems likely that the topic of this stretch of discourse concerns *robot*.

Turning to the other concepts that occur in this stretch of discourse, we may view their relationships with *robot* by looking at the clusters of concepts produced in the extended tree analysis and given below.

Cluster 1 *fighter, video, people*
Cluster 2 *storm, cook food, house, program*
Cluster 3 *Star Wars, C3PO, R2D2, robot*
Cluster 4 *seen*
Cluster 5 *woman, naked, servant, boy*

Using these clusters, we may draw the following implications concerning the stretch of discourse from T-unit 38 to 54. The first T-unit of the stretch (T-unit 38) appears to be consolidating the position of cluster 3 as central to the succeeding discourse since it contains three of the four concepts in that cluster. T-units 41 and 48-53 concern concepts in cluster 3 (*robot*) and cluster 2 (*cook food*), suggesting an attempt to highlight a potentially new association between these two clusters. Similarly, T-units

42-45 may highlight a potentially new association between cluster 3 (*robot*) and cluster 5 (*servant*).

Using the same approach for the whole of extract G, we may draw the implications given in Table 5.10 below.

T-units	**Concepts indicative of topic**	**Other concepts present**	**Topic progression**
1-7	*Star Wars*	*video*	Linking of clusters 1 and 3
			Shift
8-14	*boy*	*house, people, storm*	Linking of clusters 1, 2 and 5 (within the context of *Star Wars*)
			Shift?
15-36	*boy, R2D2, C3PO*	*woman, naked*	Drift from cluster 5 to cluster 3
37			Shift
38-54	*robot*	*C3PO, R2D2, cook food, servant, program*	Linking of cluster 3 with clusters 2 and 5
			Shift
55-74	*robot*	*seen, Star Wars, fighter*	Initially linking clusters 4 and 3, then linking clusters 3 and 1

Table 5.10 Summary of topics and topic progression in extract G based on associations

Basing an analysis on associations is promising in that clusters of concepts can be identified from an extended tree. Although the method seems promising, in practice problems with the corpus used mean that some anomalous associations were generated and other potential associations were missed. Whether the differences between the analysis and intuitions are valid problems needs to be investigated by triangulating the analysis with other analyses. This triangulation of analyses is the purpose of the next chapter.

Chapter 6 Comparisons of Analyses and Perspectives

In the previous two chapters, I have conducted six separate analyses of the classroom discourse data collected for this study (one based on Sinclair and Coulthard (1975), a theme-rheme progression analysis, a given-new analysis, an analysis based on Hoey (1991), and two analyses following Watson Todd (1998), one focusing on relations and one focusing on associations). All of these analyses, I hope I have shown, have led to insights into topics and possibly topic progression. In addition to the analyses, as discussed in chapter 3, I have also collected data concerning the teachers' and students' perspectives on the discourse.

A quick glance at the findings of the analyses suggests that, while there are similarities between the six approaches, there are also some differences. The usual justification in research for conducting several analyses is theoretical triangulation, which involves using several different perspectives on the same data (Allwright and Bailey, 1991; Brown, 2001; Freeman, 1998). Similarly, the justification for collecting teacher and student perspectives in addition to the discourse data is data triangulation, which involves collecting data from several different sources (Brown, 2001; Freeman, 1998).

The purpose of triangulation is "to maximize the possibility of obtaining credible findings by cross-validating them" (Brown, 2001: 228). The concept stems from the field of surveying where more precise identifications of location can be made by increasing the number of reference points. In carrying this concept into applied linguistics research methodology, two points need to be made, one relatively unimportant and one crucial to the use of triangulation.

The first less important point is that, while three referenced points are sufficient for identifying a location in three-dimensional space, in conducting research there is potentially no limit to the number of different perspectives we could take (Stubbs, 1983). Conducting a comparison of six analyses may therefore be useful.

The second vital difference between the use of triangulation in surveying and in research is that, whereas in surveying the goal is to reach a more precise definition of something which can be objectively measured, in applied linguistics research this is generally not a realistic goal. In surveying, a location identified is either correct or incorrect, but in linguistic research two conflicting perspectives can both be valid at the

same time. In using triangulation in applied linguistics research, therefore, we should not always be attempting to "support a finding by showing that independent measures of it agree with it" (Miles and Huberman, 1984: 235, quoted in Brown, 2001: 228).

This point is especially important in this study. As we saw in chapter 2, although there may be a general consensus concerning propositional coherence and topics in discourse, differing opinions are not necessarily invalid since there is no objectively correct interpretation of coherence or topics. Because of this, attempting to compare the different analyses and perspectives to identify the most valid interpretation of topic may not be a valid approach. In comparing the different analyses and perspectives, points of agreement and disagreement will be highlighted. While the analysis which agrees the most with other analyses is likely to be the most reliable indicator of topics and topic progression, it is not necessarily the only valid indicator. It is equally interesting to examine an analysis which stands in disagreement to the others as doing this may lead to insights into how differing interpretations of topic can occur in discourse. For such conflicting methods of analysis, we also need to be able to judge whether conflicting findings present a different view of topics or whether they concern some other aspect of discourse.

On this basis, I intend to conduct five different comparisons in this chapter. The first is to compare where in the discourse discontinuities likely to be indicative of topic breaks are identified by the six methods of analysis. Secondly, I will compare the topic entities identified by the six methods. Thirdly, I will compare these findings with a control analysis based on random insertion of breaks after every n^{th} T-unit. Fourthly, the patterns of topic progression, especially from the two analyses based on Watson Todd (1998), will be compared. Lastly, the findings from the six analyses will be compared with the teachers' and students' perspectives on the discourse.

6.1 Comparison of discontinuities in the analyses

To compare the analyses, we need some basis on which to conduct the comparison. One aspect which all six analyses have in common is that they all identify discontinuities in the discourse, providing a basis for comparison. In the six analyses, discontinuities are identified as in Table 6.1.

Sinclair and Coulthard (1975)	breaks between transactions breaks between exchanges
Hoey (1991)	coincidence of at least two of: breaks identified from networks based on T-units, breaks identified from running averages, and breaks identified from networks based on exchanges
Theme-rheme progression	coherence breaks
Given-new progression	given-new coherence breaks
Topic-based analysis (relations)	moves between concepts with a relational distance greater than 1
Topic-based analysis (associations)	moves between concepts with a z-score of less than 2.0 and placement in different clusters in the tree analysis and non-co-occurrence in a single T-unit

Table 6.1 Methods of identifying discontinuities in the six approaches

The comparative positions of discontinuities based on the six methods of analysis for all extracts are given in Appendix C. Following the ways in which discontinuities can be identified in each method of analysis, for all extracts studied we find the frequencies of discontinuities shown in Table 6.2.

Method of analysis	Total number of discontinuities	Number of discontinuities/100 T-units
Sinclair and Coulthard (1975): exchanges	279	24.50
Theme-rheme progression	207	18.17
Given-new progression	175	15.36
Topic-based analysis (relations)	83	7.29
Topic-based analysis (associations)	79	6.94
Hoey (1991)	72	6.32
Sinclair and Coulthard (1975): transactions	26	2.28

Table 6.2 Frequencies of discontinuities identified in the six analyses

One point we need to decide about before continuing the comparison is whether to use breaks between transactions or breaks between exchanges as the discontinuities in the analysis based on Sinclair and Coulthard (1975). The number of discontinuities based on transactions is much lower than in any other method, whereas the number of exchange breaks is a little higher than the number of coherence breaks identified in theme-rheme progression. Since the number of exchange breaks is more compatible with the findings of other analyses than the number of transaction breaks, I will use exchange breaks as the basis for comparing discontinuities in the analysis based on Sinclair and Coulthard (1975)

with discontinuities found in the other analyses. Nevertheless, I will still cross-check some of the findings from comparisons with the breaks between transactions.

A second issue emerging from Table 6.2 is that the methods of analysis fall into two groups based on the frequencies of discontinuities identified. The first group (the analysis based on exchange breaks, theme-rheme progression and given-new progression) identify over twice the number of discontinuities identified by analyses in the second group (the two topic-based analyses and the analysis based on Hoey (1991)). This is a point we will come back to later.

To compare the placement of discontinuities identified, there are two main approaches we can use. The first follows the traditional use of triangulation where we assume that agreement of findings concerning placement of discontinuities between the different analyses indicates stronger evidence for a 'real' discontinuity. The second approach takes the view that all discontinuities identified are probably valid, irrespective of whether they match the findings of other analyses or not. In taking this second approach, we can attempt to see how the different placements of discontinuities are related. I will use both approaches but place a greater emphasis on the second.

In taking the first approach, we find that there are only 13 places where the findings concerning discontinuities from all six analyses coincide. Interestingly, only one of these coincides with a break between transactions.

Similarly, there are only 14 places where the findings from five of the six analyses coincide. For these, the number of times each of the different analyses does not identify a discontinuity identified in the other five approaches is given in Table 6.3.

Method of analysis	Frequency of not identifying a discontinuity identified by the other five analyses
Hoey (1991)	7
Topic-based analysis (relations)	5
Given-new progression	1
Theme-rheme progression	1
Exchange breaks	0
Topic-based analysis (associations)	0

Table 6.3 Frequencies of each analysis not identifying a discontinuity

To some extent, the figures in Table 6.3 are as might be expected. The overall higher frequency of discontinuities identified by using the first three approaches suggests that they are less likely to be the 'missing' analysis in identifying discontinuities. However, the lack of any points at which association-based topic-based analysis is the 'missing' analysis suggests that this approach may have a higher validity.

Turning to the 'opposite' way of analysing the data, we can look at those points where only one of the analyses identifies a discontinuity. The overall frequencies for this are given in Table 6.4.

Method of analysis	Frequency of unique discontinuities in each analysis	Unique discontinuities as a percentage of total number of discontinuities identified
Exchange breaks	108	38.71%
Topic-based analysis (associations)	15	18.99%
Theme-rheme progression	25	12.08%
Topic-based analysis (relations)	10	12.05%
Hoey (1991)	8	11.11%
Given-new progression	12	6.86%

Table 6.4 Frequencies of discontinuities unique to one method of analysis

From Table 6.4, we can see that by far the highest number of unique discontinuities is identified using the analysis based on exchange breaks (Sinclair and Coulthard, 1975). In part, this is likely to be due to the overall higher frequency of discontinuities identified in this approach. Nevertheless, the high frequency of uniquely identified discontinuities suggests that we cannot rely on an exchange break analysis to identify shifts in topic. It should be pointed out, however, that only one of these uniquely identified discontinuities involves a break between transactions. Although this seems to suggest that transaction breaks are related to topic discontinuities (confirming Francis and Hunston's (1987) claim that transactions are topic units), when we take into account the low frequency of transaction breaks at points where all six analyses agree, we cannot in fact tell very much about the relationship between transactions and topics.

In contrast, given-new progression is the approach which is proportionately least likely to identify a discontinuity at points not coinciding with the findings of other analyses. This suggests that given-new progression may be more reliable than the other methods in identifying discontinuities.

Turning now to association-based topic-based analysis, the approach identified as potentially the most valid when we looked at agreement of placement of discontinuities between approaches, there is a relatively high number of uniquely identified discontinuities, suggesting low validity for this approach. This contrast between the findings in Table 6.3 and Table 6.4 suggests that no single approach should be relied on to identify topic discontinuities. In other words, although those points in the discourse where all six or even five of the analyses agree are most likely to be the locations of topic shift, to identify these points we would have to conduct all six analyses and there seems to be no more parsimonious way to 'guarantee' identifying the location of topic shifts.

	Sinclair and Coulthard (exchanges)	Theme-rheme progression	Given-new progression	Hoey (1991)	Topic-based analysis (relations)	Topic-based analysis (associations)
Sinclair and Coulthard (exchanges)	100.00	42.29	34.77	20.43	20.43	18.30
Theme-rheme progression	57.00	100.00	73.43	14.49	24.15	19.32
Given-new progression	55.43	86.86	100.00	14.86	24.00	22.86
Hoey (1991)	79.17	41.67	36.11	100.00	30.56	37.50
Topic-based analysis (relations)	68.67	60.24	50.60	26.51	100.00	39.76
Topic-based analysis (associations)	64.56	50.63	50.63	34.18	41.77	100.00

Table 6.5 Comparison of placement of discontinuities in each analysis

Instead of attempting to find the most valid method of identifying discontinuities in classroom discourse, let us turn to how each of the different approaches is related. To do this, we need to look at the extent to which each pair of analyses matches in their identification of discontinuities. The percentages of discontinuities in one approach which match with those identified in another approach are given in Table 6.5. In this table, figures represent the number of discontinuities coinciding in

the two analyses as a percentage of the total number of discontinuities in the analysis on the left. For example, the bottom-left space shows the number of discontinuities coinciding in topic-based analysis (associations) and the analysis based on Sinclair and Coulthard (1975) divided by the total number of discontinuities identified in topic-based analysis (associations) multiplied by 100.

While some points in Table 6.5 are interesting (such as the high percentages of matches between theme-rheme progression and given-new progression), the patterns of relationships are not clear, largely due to the disparate total frequencies of discontinuities identified in the different methods of analysis. To gain a clearer picture of how each pair of approaches is related, we can conduct an analysis similar to that we conducted in chapter 5 to compare associations. From the figures in Table 6.5, we can calculate z-scores to see how closely related each pair of analyses is. From these z-scores, we can construct an extended tree to gain an overall picture of the relationships between the different methods of analysis. The z-scores are given in Table 6.6.

	Topic-based analysis (associations)	Topic-based analysis (relations)	Hoey (1991)	Given-new progression	Theme-rheme progression
Sinclair and Coulthard (exchanges)	7.65	8.67	9.86	9.25	10.76
Theme-rheme progression	7.18	9.62	4.97	23.59	
Given-new progression	8.52	8.78	2.11		
Hoey (1991)	10.44	7.82			
Topic-based analysis (relations)	12.02				

Table 6.6 Z-scores for comparing placement of discontinuities

From Table 6.6, we can see that all of the z-scores (with the possible exception of the relationship between given-new progression and the analysis based on Hoey (1991)) are strikingly high, suggesting strongly held relationships between the different approaches. As in Table 6.5, it is also noticeable that the strongest relationship appears to hold between theme-rheme progression and given-new progression.

An extended tree diagram using the z-scores in Table 6.6 is given in Figure 6.1.

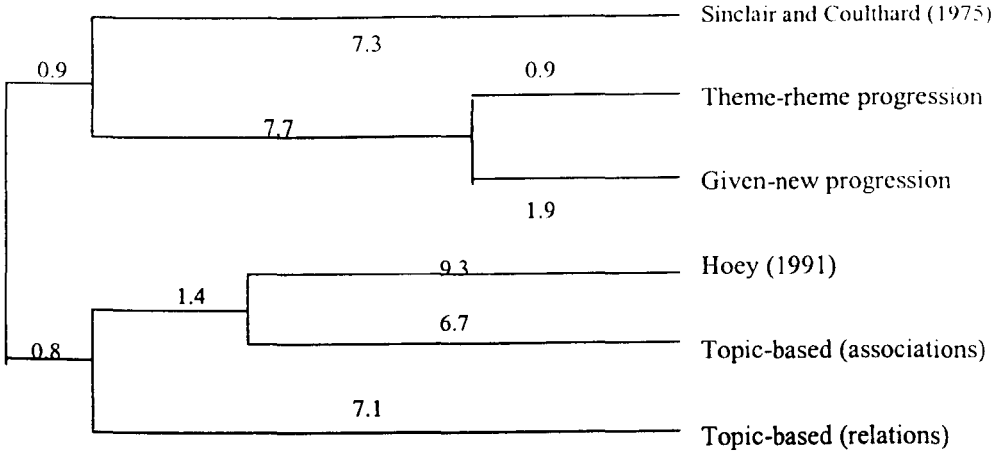


Figure 6.1 Extended tree diagram of placement of discontinuities in the different analyses

The tree analysis confirms that given-new progression and theme-rheme progression are the two most closely related approaches, despite the attempt to contrast these two methods by including ellipted material in one but not in the other. Also related to this pair of approaches, though far less strongly, is Sinclair and Coulthard's analysis of exchange breaks.

The other three approaches also form a group with the analysis based on Hoey (1991) more closely related to association-based topic-based analysis than either is to relation-based topic-based analysis.

Using the same conventions as in chapter 5, we can therefore summarise Figure 6.1 as follows.

> ((Theme-rheme progression + Given-new progression) + Exchange breaks)
> ((Hoey, 1991 + Association-based topic-based analysis) + Relation-based topic-based analysis)

An interesting aspect of Figure 6.1 is the overall pattern of relations. Reading from the top to the bottom of the figure, we can see a progression from more functionally-oriented analyses to more conceptually-oriented analyses. The most functionally-oriented approach is the exchange breaks of Sinclair and Coulthard (1975), an approach which does not take content into account (except in deciding whether succeeding initiations fall into the same exchange). At the other extreme, relation-based topic-based analysis is the most 'semantic' of the

approaches, especially in its emphasis on logical relations between concepts. This pattern of functional to conceptual bases for analysis supports the existence of the theoretical constructs of the communicative and conceptual metafunctions of language posited by Widdowson (1984).

It is less clear, however, how Figure 6.1 can help us in our search for topics and topic progression. In constructing Figure 6.1, no preference for any given analysis has been made. Instead of using Figure 6.1 to decide which of the approaches we should favour in identifying topic discontinuities in discourse, we should focus on how Figure 6.1 provides insights into the methods of analysis rather than the topic discontinuities themselves. For the moment, I will leave the identification of topic discontinuities, but will return to this aspect of topic later.

6.2 Comparison of topic entities

A second finding which all methods of analysis have in common is identifying topic entities, although we have to reinterpret the findings from some analyses in order to identify topic entities as single concepts to facilitate comparison. The ways in which topic entities expressed as single concepts can be identified in each of the approaches are as in Table 6.7.

Sinclair and Coulthard (1975)	most frequent concept (including ellipted material) in an exchange
Hoey (1991)	most linked lexical item in a monotopical cluster
Theme-rheme progression	most frequent theme in a stretch of discourse between two coherence breaks
Given-new progression	most frequent given information in a stretch of discourse between two given-new coherence breaks
Topic-based analysis (relations)	most frequent key concept in a stretch of monotopical discourse
Topic-based analysis (associations)	most frequent key concept in a stretch of monotopical discourse

Table 6.7 Methods of identifying topic entities in the six approaches

The key principle for identifying topic entities, then, is to identify the most frequent element within stretches of discourse between discontinuities. The usual approach to dealing with qualitative data, however, is to look for both frequent and salient issues (see e.g. McDonough and McDonough, 1997). To take account of saliency, in circumstances where a relatively frequent, but not the most frequent, element is highlighted in the discourse by being written on the board or

spelt out, this element will displace the most frequent element as topic entity.

Using these guidelines of frequency and saliency, topic entities can be identified for all T-units for four of the approaches (the analysis based on Sinclair and Coulthard (1975), theme-rheme progression, and the two topic-based analyses), if we assume that unclear utterances have the same topic entity as the surrounding discourse. For given-new progression, there are stretches of discourse containing no given information covering a total of 44 T-units; and there are stretches of discourse which contain no lexical items linking to any other item covering 6 T-units when we consider the analysis based on Hoey (1991). Given that there are 1139 T-units in total in the data, these figures are comparatively low and do not affect the comparison of approaches unduly.

In taking the approach focusing primarily on frequency of elements to identify topic entities, there are a few problems. Firstly, although the methods of analysis in this study were chosen for their potentiality in identifying topic entities, it is not clear whether the elements identified in each approach should be truly considered topic entities. Identifying topic entities, however, is one potential purpose of comparing the different approaches, so for the present, I will assume that the elements identified in each method of analysis represent topic entities and will evaluate this position after conducting the comparison.

A second and more serious problem is that identifying a single topic entity in each stretch of discourse between discontinuities assumes that each of these stretches of discourse exhibits topic maintenance and takes no account of the potential occurrence of topic drift. Unfortunately, the majority of the approaches provide few insights into topic progression, and in those approaches which do provide such insights (especially the two topic-based approaches), where topic drift is identified it is often unclear exactly where one topic entity ends and another one starts. To facilitate the comparison of the topic entities identified in each approach, identifying a single topic entity for each stretch of discourse is helpful. In this section, then, I will assume single topic entities for stretches of discourse, and in a later section, I will examine the kinds of topic progression, including topic drift, more closely.

Bearing these points in mind, we are in a position to conduct a comparison of the topic entities identified in each method of analysis (and these are shown in Appendix C, p. 269). The approach I will take in

conducting the comparison will be similar to that taken in comparing discontinuities in section 6.1.

The first point to make is that topic entities can be identified for all T-units for four of the approaches (the analysis based on Sinclair and Coulthard (1975), theme-rheme progression, and the two topic-based approaches), if we assume that unclear utterances have the same topic entity as the surrounding discourse. For given-new progression, there are stretches of discourse between discontinuities containing no given information covering a total of 44 T-units, and there is a similar 6 T-units containing no lexical items linking to any other item when we consider the analysis based on Hoey (1991). Given that there are 1139 T-units in total in the data, these figures are comparatively low and do not affect the comparison of approaches unduly.

To start the comparison, we can look for points in the discourse at which all six methods of analysis agree. For all extracts, there are 180 T-units where the topic entity identified is the same in all six approaches. This is 15.80% of the total number of T-units in the data. For these 180 T-units, the topic entity identified is likely to be one for which there would be a general consensus. This, however, represents only a small part of the data, so we must also examine points at which the different approaches disagree.

One way in which we can identify such points is to look for points where one of the methods of analysis identifies a topic entity which is not identified in any of the other approaches at that point. The numbers of such uniquely identified topic entities varies between approaches as shown in Table 6.8.

Method of analysis	**Number of T-units in which the method of analysis uniquely identifies a topic entity**
Theme-rheme progression	433
Sinclair and Coulthard (1975)	272
Hoey (1991)	150
Given-new progression	137
Topic-based analysis (relations)	121
Topic-based analysis (associations)	54

Table 6.8 Number of uniquely identified topic entities in each approach

The figures in Table 6.8 are suggestive concerning the reliability of identification of topic entities in each approach. The high number of T-

units in which theme-rheme progression identifies a topic entity different from that identified in any of the other approaches leads us to question the reliability of this approach. Looking through the data, it does in fact seem as if theme-rheme progression is operating on different principles in identifying topic entities from the other approaches. Despite attempts to reduce their impact, interactional themes (such as *I* and *you*) are frequently identified as topic entities, and even when the theme identified as topic entity is conceptual in nature, it often seems anomalous. Based on the classroom data in this study, then, there does not seem to be a very close relationship between T-unit theme and discourse topic.

Similarly, the analysis based on Sinclair and Coulthard (1975) identifies unique topic entities fairly frequently. In this case, however, the topic entities identified seem reasonable from an intuitive perspective. The reason for the relatively frequent identification of unique topic entities seems to concern the nature of the segments of discourse which form the basis for their identification rather than the principles on which identification is made. As we saw in section 6.1, the high frequency of exchange breaks in the data means that the analysis based on Sinclair and Coulthard divides the data into a larger number of generally shorter stretches of discourse than any other approach. The topic entities in this analysis, then, are the topic entities of stretches of discourse shorter than in other analyses. This, in turn, precludes the identification of topic drift using this approach since the topic entities identified are discrete. The relatively high number of unique topic entities identified for the shorter stretches of discourse suggests that identification of topic entities is dependent on the length of the stretch of discourse, and appears to confirm a hierarchical approach to topics whereby viewing discourse at different levels leads to different identifications of topic and topic entity.

One further aspect emerging from the data concerns both the points at which all six approaches agree on a topic entity and the points at which one approach uniquely identifies a topic entity. Looking through the data, it seems that the former are proportionately more frequent and the latter proportionately less frequent in the shorter extracts. We can check this by looking at the number and proportions of both types of point in each extract, and this is done in Table 6.9.

Extract	Length in T-units	Number of T-units in which all six approaches identify the same topic entity	% of T-units in which all six approaches identify the same topic entity	Number of T-units for which a topic entity unique to one approach is identified	% of T-units for which a topic entity unique to one approach is identified
A	237	0	0.00	322	135.86
B	186	22	11.83	188	101.08
C	76	31	40.79	65	85.53
D	120	23	19.17	105	87.50
E	84	2	2.38	88	104.76
F	125	3	2.40	182	145.60
G	74	32	43.24	49	66.22
H	34	4	11.76	32	94.12
I	82	14	17.07	94	114.63
J	40	2	5.00	28	70.00
K	47	30	63.83	8	17.02
L	34	17	50.00	6	17.65

Note: Figures concerning unique topic entities can be greater than the number of T-units because each identification of a unique topic entity in each approach is counted. So for example, if for a particular T-unit both theme-rheme progression and given-new progression identify distinct topic entities not identified by any of the other approaches, this T-unit would be counted twice.

Table 6.9 Topic entities identified in each of the extracts

The figures in Table 6.9 seem to confirm the intuition that shorter extracts contain proportionately more points where the approaches agree and fewer points where unique topic entities are identified. We can confirm this by calculating the correlation coefficients between the number of T-units in each extract, on the one hand, and the percentage of T-units at which all six approaches identify the same topic entity and the percentage of T-units for which a topic entity unique to one approach is identified, on the other. Comparing length of extract and agreement between all six approaches, we find $r = -0.49$ ($p = 0.10$), suggesting that the longer the extract the lower the proportion of T-units where all six approaches agree on the topic entity. Comparing length of extract and identification of unique topic entities, we find $r = 0.65$ ($p < 0.05$), showing that longer extracts contain proportionately more points at which topic entities unique to one approach are identified. If we assume that the amount of agreement between approaches concerning topic entities represents the extent of consensus likely to occur concerning identification of topic

entities, then shorter stretches of eliciting are likely to lead to less potential for participants to identify conflicting topics concerning topic entities. It should be emphasised that the assumption behind this conclusion is questionable, and I shall look at participants' perceptions in more detail in section 6.5.

Turning now to the second purpose of triangulation discussed above, we can compare the topic entities identified in each possible pair of approaches. The numbers of T-units for which each pair of approaches identifies the same topic entity are given in Table 6.10.

	Topic-based (associations)	Topic-based (relations)	Hoey (1991)	Given-new progression	Theme-rheme progression	Sinclair and Coulthard
Sinclair and Coulthard	578	541	524	518	416	1139
Theme-rheme progression	397	427	360	455	1139	
Given-new progression	688	644	619	1095		
Hoey (1991)	792	700	1133			
Topic-based (relations)	807	1139				
Topic-based (associations)	1139					

Table 6.10 Frequencies of commonly identified topic entities for each pair of approaches

From Table 6.10, it appears that theme-rheme progression is the approach which has the least agreement with other approaches concerning identification of topic entities. The two topic-based approaches and the analysis based on Hoey (1991), on the other hand, have the highest levels of agreement with each other.

As in section 6.1, we can represent these relationships between approaches diagrammatically by using an extended tree diagram. Unlike section 6.1, however, there is no need to convert the raw data into z-scores first, since the total potential number of points of similarity is roughly the same for all six methods of analysis. An extended tree diagram using the raw data in Table 6.8 and showing the relationships

between the six approaches concerning identification of topic entities is given in Figure 6.2.

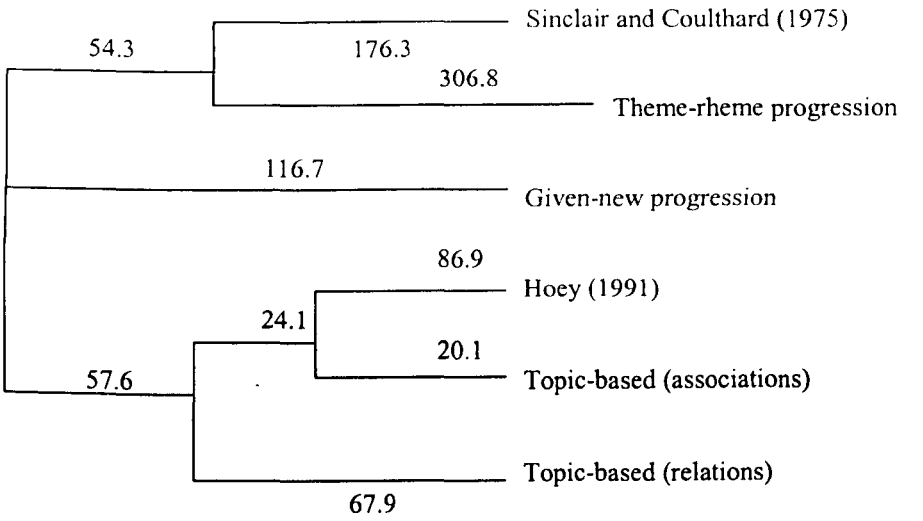


Figure 6.2 Extended tree diagram for identification of topic entities

The first thing we may notice in Figure 6.2 is that, reading from top to bottom, we find the same sequence of approaches as we did for the tree diagram concerning discontinuities in Figure 6.1. In Figure 6.1, there was little relative difference in the length of lines in the diagram. In Figure 6.2, however, there is a noticeable difference between how closely related the clustering of approaches at the bottom of the diagram are and the distances to the analysis based on Sinclair and Coulthard (1975) and, especially, theme-rheme progression. In other words, although it presents no case for preferring any particular method of analysis, Figure 6.2 confirms the earlier finding that topic entities identified through theme-rheme progression and through Sinclair and Coulthard are markedly different from the topic entities identified from the other approaches. It is also worth noting the isolated position of given-new progression in the extended tree diagram which may reflect the distinct nature of identifying given information when compared to the other approaches.

In this section we have seen that the principle underlying the identification of topic entities and the lengths of the stretches of discourse for which topic entities are identified both have an influence on what topic entities are identified. Because of this, those approaches which are most similar concerning average length of stretches of discourse between discontinuities and the principle they use for identifying topic entities are the approaches for which there is most agreement.

6.3 Comparison with a control method of analysis

It could be argued that all of the findings presented so far in this study are based on preconceptions and assumptions about how people produce and process discourse which may not reflect reality. In other words, although I have tried to show how the present research is based on findings of previous research, an extreme iconoclast could claim that none of this previous research is certain and therefore the present research is also suspect. It seems unlikely that anyone would take such a position and nearly all discourse studies take no account of this issue. It is a point, however, that I would like to address in this study.

From an extreme logical positivist perspective, perhaps the only finding indicating that the present research is valid in its approaches is the relatively high z-scores given in Table 6.6. Nevertheless, even on this point, an argument could be made that the lack of any control method of analysis weakens the findings. In this section, then, I would like to suggest how we might attempt such a control analysis and compare the findings from the six approaches with a control approach. Doing this would allow us to test the validity of the six methods of analysis used in this study.

The control analysis should be comparable in nature to the six 'experimental' approaches against which it is being compared. The easiest way to do this is to control for the number of discontinuities identified in the control approach. For the six methods of analysis in this study, there is a total of 895 discontinuities, or an average of 149.17 discontinuities per approach. Given a total discourse length of 1139 T-units, this means that on average there is one discontinuity identified in each analysis every 7.64 T-units. To create a comparable control approach, we could aim to identify discontinuities at points whose distance is 7.64 T-units apart. To simplify matters, for the control approach I will identify discontinuities after each 8th and 7th T-unit alternatively (in other words, after T-units 8, 15, 23, 30, 38 ...).

Doing this, we find that the control approach identifies a total of 145 discontinuities in the data. Of these, 88 or 60.69% are uniquely identified by the control approach, a proportion of uniquely identified discontinuities far higher even than the more numerous exchange breaks. We can also compare the discontinuities identified in the control approach with each of the other six methods of analysis, and this is done in Table 6.11, showing the numbers of coinciding discontinuities, the proportions of these as percentages of all discontinuities, and z-scores.

	Number of coinciding discontinuities	Coinciding discontinuities as % of total in control analysis	Coinciding discontinuities as % of total in applicable analysis	z-score
Sinclair and Coulthard (1975)	41	28.28	14.70	1.13
Theme-rheme progression	23	15.86	11.11	-0.69
Given-new progression	18	12.41	10.29	-0.96
Hoey (1991)	8	5.52	9.64	-0.41
Topic-based (relations)	10	6.90	12.66	-0.19
Topic-based (associations)	13	8.97	18.06	1.05

Table 6.11 Comparison of discontinuities in the control analysis and the six approaches

From Table 6.11, we can see that the percentages of coinciding discontinuities are lower than the ones found when we compared each pair of approaches. More importantly, the z-scores for the control analysis are also lower and are all near zero, as we would expect when there is no relationship between two approaches. This confirms that there is almost no chance that the amount of coincidence between the six approaches in identifying discontinuities is due to random factors.

So far, the comparison with the control analysis has yielded findings such as we might expect. Given that the control analysis is not based on any principles held in common with the other approaches, we should expect the comparison of placement of discontinuities to produce z-scores around zero. In addition to comparing the analyses on the basis of discontinuities, we can also attempt to identify topic entities in the control analysis, and this should yield more interesting findings. When we compared the identification of topic entities in the six approaches in section 6.2, we found that two of the approaches yielded results different from the other four approaches. I suggested two different reasons for this, one for each of the two differing approaches. The differences may be due to the length of stretches of discourse and placement of discontinuities, or they may be due to the principles behind the identification of topic entities. Comparing topic entities identified in the control approach with those identified in the six approaches may allow us to see which of these two posited reasons has more effect.

Before we can do this, however, we need to consider how to identify topic entities in the control approach. The most straightforward method is probably to take the most frequently occurring lexical item as the topic entity with adaptations to account for the saliency of board writing and spelt words. In addition, where two lexical items are of similar frequency, I will take the item which is directly referred to in speech, as opposed to being ellipted, as the topic entity. It should be noted that, unlike the identification of discontinuities in the control approach, the identification of topic entities using this method is biased. Basing the identification of topic entities on frequency of lexical items is similar to the identification of topic entities in the analysis based on Hoey (1991) and the two topic-based approaches. Nevertheless, any method of identifying topic entities will be somewhat biased and this is probably the simplest and crudest method. If a method of identifying topic entities as crude as this were to produce results similar to any of the six approaches, the validity of the similar approach would be cast into doubt.

Using this method of identifying topic entities, we can identify topic entities for each of the short stretches of discourse between discontinuities that were identified in the control analysis. We find that there are 362 T-units (or 31.78% of the extracts) where the control approach uniquely identifies a topic entity, a figure higher than for any of the other approaches with the exception of theme-rheme progression (see Table 6.8).

We can also compare the topic entities identified in the control approach with each of the other methods of analysis. Doing this, we find the numbers of T-units for which the control approach identifies the same topic entity as other approaches given in Table 6.12.

	Number of T-units where the topic entity identified coincides with the control approach
Sinclair and Coulthard (1975)	516
Topic-based analysis (associations)	505
Topic-based analysis (relations)	492
Hoey (1991)	475
Given-new progression	407
Theme-rheme progression	326

Table 6.12 Frequencies of commonly identified topic entities for the control approach and each of the six approaches

From Table 6.12, we can see that the control approach appears to have most in common with the analysis based on Sinclair and Coulthard (1975) and with the two topic-based analyses, and least in common with theme-rheme progression. To check this, we can combine the data in Table 6.12 with the data in Table 6.10 to find the frequencies of commonly identified topic entities for each possible pair of seven approaches, and from this combined data we can construct another extended tree as shown in Figure 6.3.

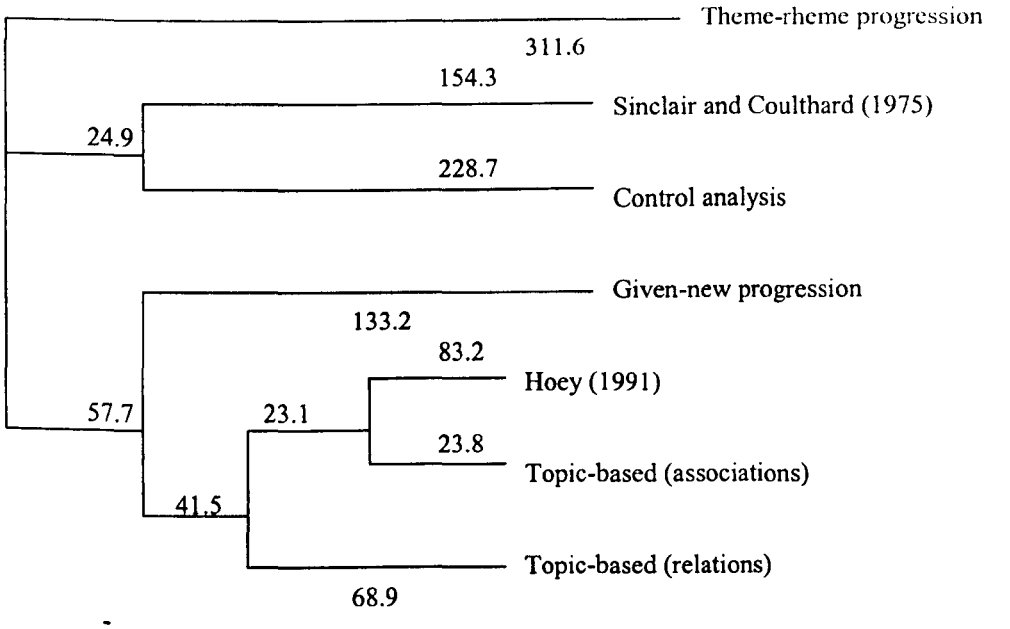


Figure 6.3 Extended tree diagram for commonly identified topic entities in the six approaches and the control approach

As with the other extended tree diagrams, the pattern in Figure 6.3 from top to bottom is from functionally-oriented to conceptually-oriented approaches. It can also be seen that the addition of the control approach to the diagram has not affected the overall pattern of clusterings between methods of analysis. A further similarity between Figures 6.2 and 6.3 is that theme-rheme progression is still separated from the other approaches, including the control approach, confirming that the different basis for identifying topic entities in theme-rheme progression leads to the identification of distinct topic entities. Whether the topic entities identified through theme-rheme progression are truly concerned with topics is unclear, although the large gap between theme-rheme progression and the other approaches suggests that theme-rheme progression may focus on some aspect of discourse not directly related to topics.

In Figure 6.3, with the exception of theme-rheme progression, the control approach is the approach most distant from the others. Although in the tree diagram, the control approach appears related to the analysis based on Sinclair and Coulthard (1975), following the distances in the tree the latter is actually more closely related to given-new progression and the two topic-based analyses. The lack of any close relationship between the control approach and the other approaches, despite the similar basis for identifying topic entities, suggests that the placement of discontinuities is also important in identifying topic entities and strengthens the argument that there is a meaningful basis for the identification of discontinuities and topic entities, especially for the clustering of approaches which consists of given-new progression, the analysis based on Hoey (1991) and the two topic-based analyses.

6.4 Comparison of topic progression

So far in making comparisons, I have concentrated on what it is easiest to compare rather than what it may be important to compare. Discontinuities and topic entities can be identified relatively objectively in all six approaches and allow for quantitative comparisons to be made. In this section, however, I will attempt to compare the identification of topics/topic entities at different levels and the types of topic progression in the discourse. Such a comparison is necessarily more qualitative and, to keep the comparison to a reasonable length, I will focus mostly on extract G. I will look at each of the approaches in turn.

Sinclair and Coulthard (1975): It was hoped that an analysis based on Sinclair and Coulthard (1975) would allow us to identify topic shifts and topics at different levels. If we assume that transactions are topic-units, then framing moves indicating transaction boundaries might also have indicated topic boundaries or topic shifts. Unfortunately, the findings concerning topic discontinuities in section 6.1 do not support such a position. Another potential insight into topics based on Sinclair and Coulthard's approach is that topics might be identifiable at different levels of discourse. In section 6.2, in identifying topic entities for T-units, we were in effect identifying the topics of moves (which match most closely with T-units). We might also identify topics at the levels of exchange, exchange complex and transaction providing a hierarchical description of topics. It is, however, unclear how topics should be identified at these levels, so the analysis based on Sinclair and Coulthard is not as productive as it promises.

Approaches based at the clause level: Theme-rheme progression and given-new progression provide fewer potential insights into topic progression. Stretches of discourse with consistent parallel progression (or parallel given progression) should exhibit topic maintenance, and sequential progression (or sequential given progression) may indicate topic drift. Given the lack of relationship between themes and topic entities shown in sections 6.2 and 6.3, however, given-new progression is more likely to provide valid insights than theme-rheme progression.

Looking for parallel given progression in extract G, T-units 8 to 13, 15 to 17, 18 to 20, 21 to 25, 26 to 32, 33 to 36, 39 to 54, and 60 to 71 exhibit topic maintenance. There is, however, little evidence of sequential given progression indicating topic drift.

Hoey (1991): This approach has little to say concerning topic progression. Nevertheless, we can say that clusters of successive T-units linked by a large number of the same bonds indicate topic maintenance. Furthermore, T-units with a high number of links to later T-units are likely to be topic-opening, and T-units predominantly linking to earlier T-units may be topic-closing. One further point is that, using different thresholds to create networks may indicate topics at different levels of specificity.

Using these guidelines with extract G, we find that T-units 8 to 13 have a large number of the same bonds indicating topic maintenance, with T-unit 8 being a topic-opening T-unit and T-unit 13 being topic-closing. There is a similar pattern in T-units 33 to 36. Finally, T-unit 39 is topic-opening with 21 bonds to later T-units. From the networks, the clusters at T-units 8 to 13 and 33 to 38 appear to stand alone. The networks also confirm T-unit 39 as the topic-opening T-unit of a loosely connected stretch of discourse extending to the end of the extract with the notable exception of the inserted T-units 55 to 59, which appear to exhibit topic renewal from T-units 3 and 4.

Topic-based analyses: The two topic-based analyses differ from the other approaches in that they specifically aim to identify topic progression. For extract G, the topic progression identified by the topic-based analysis using relations is given in Figure 5.4 (p. 145) and that identified by the analysis using associates in Table 5.10 (p. 162). At first glance, these two figures have little in common. Although they share discontinuities after T-units 36 and 54, there are also three topic shifts which are not shared. However, the minor breaks indicative of topic entity are similar, although the stretches of discourse for each in Table 5.10 are longer. If we consider that the linking of different clusters of concepts apparent in much of

Table 5.10 can represent either topic maintenance or topic drift, then the parallels between the two figures become clearer. The problems of distinguishing between topic maintenance and topic drift in the association-based analysis are recompensed by the details it gives concerning the nature of topic maintenance. In the relation-based analysis, T-units 37 to 54 and 60 to 74 simply show topic maintenance of *robot*. The association-based analysis adds more information by showing what aspects of the *robot* topic are being discussed while it is being maintained.

A similar pattern holds for most of the other extracts. The identification of topic shifts frequently coincides for the two topic-based analyses; and the topic drift and maintenance identified in the relation-based approach parallels the linking of clusters of concepts in the association-based approach. The notable exception to this pattern is extract B, where from T-unit 19 to T-unit 143, the shifts identified in the two approaches do not coincide and there is a subsequent poor match between types of topic progression identified.

The overall findings concerning topic progression, then, are that, for much of the discourse, the topics and types of topic progression identified by the analyses match. For example, in extract G T-units 8 to 13 appear to concern the topic *boy*, and T-units 38 to 54 and 60 to 74 exhibit topic maintenance of *robot*. There are, however, several points at which the findings of the various approaches differ (for example, T-units 1 to 7 in extract G). It may be that the extent of agreement between the different approaches parallels that between the various participants in the discourse. So, for example, at points where the analyses agree we might expect a consensus concerning topics and topic progression among the participants. To check this supposition, we need to compare the findings of the analyses with the teachers' and students' perceptions of the discourse.

6.5 Comparison of analyses and participants' perceptions

In the previous four sections, I have been conducting theoretical triangulation in comparing the findings of the six methods of analysis. In this section, I will conduct data triangulation by comparing the findings of the six approaches with the participants' perspectives on the discourse. I will again focus primarily on extract G, and the data from the teacher journals and student interviews for this extract are given in Appendix B (p. 265). I will, however, also examine what we can learn from all of the extracts.

Before looking at the data concerning participants' perspectives, it is worth noting that these findings are generally disappointing. To gain insights into teachers' perspectives, teachers were asked to write a journal concerning the eliciting stage and focusing on content after each lesson (see section 3.4, pp. 60-61). Unfortunately, probably because of unclear instructions, the journals for the first lessons (extracts A to D) covered the whole lesson rather than focusing on eliciting. Even in the later journals where there was a greater focus on the eliciting stage of the lessons, the teachers discussed the classroom atmosphere, discipline problems and a range of other pedagogic issues in addition to structuring content. Furthermore, the teachers' journals concern the whole of the eliciting stage and only rarely consider specific points in the discourse.

For the students' perspectives, since the students were interviewed while watching a video of the eliciting stage of the lessons, their comments directly concerned eliciting. There were, however, two key problems in obtaining useful data. Firstly, each interview involved a different pair of students, meaning that each interview was the first time that the students had met the researcher. This caused some problems of nervousness, and students were not sure what was expected of them in the interviews. Secondly, since the interviews were retrospective, students were asked to recall what they were thinking at various points in the lesson. Time constraints precluded more than minimal training in retrospection (see Ericsson and Simon, 1993; Matsumoto, 1993). Because of this, at several points, student comments in the interviews showed that they were not thinking only about the point of the lesson which had just been shown in the video but were also considering information from later points in the lesson as well.

Despite these problems, the data on the participants' perspectives do produce some interesting insights and are worth pursuing. Let us, then, look at what the teacher journals and student interviews concerning extract G can tell us.

For extract G, although the teacher journal does not consider any specific points in the discourse, there are a couple of interesting insights into the teacher's perceptions of the nature and purpose of eliciting. Firstly, there is some evidence that this teacher focuses primarily on content while planning how to conduct an eliciting session. She explicitly considers the end-goal of eliciting, which in this case is the topic of the lesson, *robots*, and implicitly considers how to activate students' background knowledge concerning this topic:

"The topic of the lesson was 'Robots'. So why didn't I find something about robots which was very interesting and 'in the news'. I chose a film entitled 'Star Wars Episode 1' since I knew that there were absolutely robots in it."

Secondly, she appears to consider the main purpose of eliciting to be increasing interest and motivation in the students:

"What should I do to make the lesson more interesting when eliciting responses from the students for the first 5-10 minutes of the session?"

"It seemed very interesting and enjoyable for the students to watch a film rather than look at the pictures in handouts. This technique was effective and helped increase the students' motivation."

In the interview, students showed awareness of several issues relevant to this study. Firstly, they were aware of a range of potential purposes in teaching. Initially, they considered the playing of the video in extract G to be a listening exercise:

"I thought it was going to be a listening exercise where we would watch the video and the teacher would check how much we had understood."

[The teacher was asking questions] "To check whether we could understand details, to check our understanding."

Later, however, they changed to identifying the teacher's key purpose of introducing the content of the lesson:

"at first it was about our listening comprehension, but now the teacher asked, she was checking our understanding. She asked whether we could understand from listening but I didn't think the teacher would emphasise listening comprehension now."

"now the lesson was different and focuses on the differences between robots and people and how they are different, because the questions and answers were in the video such as the boy didn't have enough money to buy, but this is background for us, so that we think and we also listen."

Secondly, the students were aware of the potential purposes of eliciting. They viewed one purpose of eliciting as activating existing knowledge:

> "When I think about how she introduced the topic, it was appropriate, because it provided appropriate background knowledge."

> "I've already seen this film and there is, well, the teacher asked us about this film and I've already seen it and there's robots in most of the film."

They also perceived eliciting as aiming to motivate the students:

> "She made us interested to learn in this lesson."

Thirdly, the students showed awareness of the topics of the discourse. This awareness was exhibited at the level of lesson:

> "The teacher said straight out that this was about robots, so I knew the lesson was going to be about robots."

The students were also aware of topics at more specific levels:

> "At this point, the teacher said that, well, I thought that robots from this point on would be about how to use robots and the teacher would talk about the future, something like that, that the teacher would talk about methods, would recommend ways that robots can be used, something like that."

A fourth aspect that the students were aware of is the basis on which they identified topics. They viewed frequency of occurrence of concepts in the discourse (e.g. "The teacher keeps on asking") and saliency (e.g. "The teacher emphasised robots") as indicative of topics.

A last issue is the students' awareness of the extent to which the topic is under the control of the teacher. All of their comments in the interview focused on what the teacher said and asked, and ignored the content of student responses in the classroom.

To summarise the findings from extract G, the teacher clearly considers the end-goal and purposes of eliciting, and the students show awareness of purposes in teaching, purposes of eliciting, topics, bases for identifying topics, and the extent of the teacher's control. These findings are

generally confirmed by the data on participants' perspectives from the other extracts. For example, from the teachers' journals, other teachers focus on the content of eliciting and how to lead students to the end-goal (e.g. for extract L, "I try to lead them to the word I've prepared"). Other teachers also view motivation and interest as the purpose of eliciting (e.g. for extract B, "the teacher had to try to motivate them").

Similarly, the student interviews for the other extracts support the awareness shown by the students concerning extract G. Students show awareness of potential purposes in teaching; for example, for extract F the students initially thought that the teacher was just chatting with them and only later realised that the purpose was to elicit the topic of *robots*. There are several quotations showing awareness of the purposes of eliciting, such as "making the lesson fun" (extract D), "leading to a depth of understanding" (extract J), "stimulating thinking" (extract K), and "increasing interest" (extract L). In addition to their awareness of pedagogical issues, students also exhibit awareness concerning topics, specifically of what the topics of the discourse are (a point I will return to below), and of bases for identifying topics (e.g. "concentrating on *water pump* for some time" in extract I, "the teacher has prepared a transparency" in extract F, and "she wrote it on the board" in extract L).

There are two further points emerging from the participant data from other extracts which merit detailed examination. Firstly, there are several points where problems with connectivity in the discourse occur. While some of these relate more to language use than to relationships between concepts (e.g. the teacher's description of the heart as being two pumps rather than one in extract I), there is one point at which both the teacher and the students identify problems with the links between concepts.

In extract D (as in extract A to C), the teacher starts the lesson by using a balloon with the goal of moving on to engines. Having asked students to release the balloons, the teacher tries to elicit why the balloon flies. In T-unit 63, she asks "What kind of reaction?" - a question which, according to the interview, students did not know how to answer. She then asks a second question concerning why the balloon goes forward (T-units 68 and 69) which she quickly answers herself (T-unit 71), and then proceeds on to the topic of *engines* (T-unit 73 onwards). At this point, in the interview students stated, "I was confused. The teacher has changed from balloons to engines, but I don't know how engines are relevant." The teacher herself is also aware of the confusion and attempts to rectify the problem using Thai in T-unit 89 (see example 4.8, p. 71). In her journal she also writes, "My students don't understand what I want them to do. I

just can't bridge the activity (blowing the balloon) with the topic." In the discourse after T-unit 89, the students realise that the topic of *balloon* has been dropped and that the discourse has moved on to the new topic of *engines*, but they are still unclear how the two topics are related. This confusion continues after the lesson, as in the interview the students asked the researcher to explain the connection between the two topics. This example illustrates how unclear linking of concepts may correlate with confusion in the teaching/learning process.

The second key point emerging from the participant data in the other extracts concerns the topics that the students identify. For example, in extracts I, J and K, the students identify the change in topic from *pump* to *heart*; and in extract A, the students identify the topic at various stages as *balloon*, *pressure*, *action reaction* and *engine*. In identifying topics, the students prefer to use single words or short phrases where possible (*cf.* one student's identification of *the uses of robots* in extract G). This suggests that the students view topics as concepts rather than as propositions. Furthermore, the students readily accept that topics change as the discourse progresses, implying a dynamic view of topics.

Although I had hoped to compare the topics or topic entities identified by the participants with those identified in the six approaches, this was not possible. Because of the problems with data collection, participants identified topics for only a relatively small proportion (roughly 20%) of the total discourse, making comparisons unreliable. Where they did identify topics, these matched the topic entities identified in at least one of the six approaches about half of the time, with students' topics most frequently matching those identified in the analysis based on Hoey (1991) and the two topic-based approaches. In the cases where the topics identified by the students and by the approaches did not match, the students were generally identifying topics at a level in a hierarchy of topics higher than that of topic entity. In other words, the students tend to report topics at the level of exchange complex or transaction rather than at the level of exchange, although this may be an artifact of the interview technique. Although the students' identifications of topics are interesting, the problems of the sparseness of their identification of topics, of matching the topics identified with T-units, and of the level in a hierarchy for which topics are identified mean that we cannot use these findings to evaluate the validity of the six methods of analysis used in this study.

Despite these problems with using the data concerning the participants' perspectives, the insights into the discourse given by the participants have important implications for this study. The participants' perspectives

confirm some of the assumptions underlying the study: that links between concepts are important in discourse, that topics can be represented by concepts, that topics are dynamic, and that frequency and saliency of concepts in the discourse are indicative of topics. Finally, there is some evidence that we should focus on fairly long stretches of discourse in identifying topics. This suggests that those methods of analysis which divide the discourse into longer stretches or which allow hierarchies of topics to be generated (the two topic-based analyses and the analysis based on Hoey (1991)) may be the most productive approaches for identifying topics.

6.6 Conclusion

In this chapter, I have been following two different approaches in comparing the six methods of analysis conducted in chapters 4 and 5. On the one hand, I have tried to treat all of the analyses as potentially valid; and, on the other, I have tried to identify those analyses which provide the most valid descriptions of topics. There are three key findings which favour the latter approach. Firstly, the consistent pattern of functional to conceptual clusterings in all extended tree diagrams suggests a difference between these two metafunctional approaches to language analysis. Secondly, some of the approaches provide useful descriptions of topic progression. Thirdly, the participants' perspectives on topic, especially those of the students, favour some approaches. These three findings suggest that the analysis based on Sinclair and Coulthard (1975) and theme-rheme progression are less valid in identifying topics and following topic progression than given-new progression, which, in turn, is less valid than the analysis based on Hoey (1991) and the two topic-based analyses. By comparing the analyses, then, we have been able to identify those analyses which are most likely to be productive in conducting research into topics and topic progression.

Chapter 7 Identifying Topics and Implications

In the previous chapter, we saw that three of the methods of analysis, namely, the analysis based on Hoey (1991) and the two topic-based approaches, appeared to produce more valid findings than the other three analyses. In this chapter, I will focus on the three more productive approaches and attempt to identify relatively definitive topic entities and topics and to follow topic progression based primarily on these methods of analysis. I will then move on to consider the implications of such identifications for linguistics researchers and for language teachers.

7.1 Identifying topics and topic progression

Focusing on the three most promising approaches, in this section I intend to use theoretical triangulation in a traditional way to identify the topics in the extracts of classroom discourse used in this study. In chapter 6, I conducted triangulation of the methods of analysis primarily to identify which approaches produced the most valid findings. In this chapter, I will compare the findings of these more valid approaches to identify points of consensus indicative of topics and to see what insights are generated at those points where the approaches disagree.

7.1.1 Topic entities identified in the three approaches

Before we attempt to identify any topic, it should be noted that there is generally a high level of agreement between the three methods of analysis. Based on the summary of the data shown in Appendix C (p. 269), Table 7.1 shows the extent to which the analysis based on Hoey (1991) and the two topic-based approaches identify the same topic entities.

From Table 7.1, we can see that overall the three methods of analysis agree in their identification of topic entities for over half of the data in this study. Furthermore, most of the disagreement between the approaches is due to artifacts of the methods of analysis. There are three main causes of differing identifications of topic entities.

Extract	T-units where all 3 approaches agree		T-units where 2 approaches agree		T-units where the 3 approaches disagree	
	Number	%	Number	%	Number	%
A	79	33.33	141	59.49	17	7.17
B	94	50.54	83	44.62	9	4.84
C	58	76.32	15	19.74	3	3.95
D	108	90.00	7	5.83	5	4.17
E	54	64.29	30	35.71	0	0.00
F	45	36.00	62	49.60	18	14.40
G	51	68.92	17	22.97	6	8.11
H	25	73.53	9	26.47	0	0.00
I	24	29.27	52	63.41	6	7.32
J	27	67.50	13	32.50	0	0.00
K	33	70.21	14	29.79	0	0.00
L	17	50.00	14	41.18	3	8.82
TOTAL	615	53.99	457	40.12	67	5.88

Table 7.1 Extent of agreement in identification of topics between the analysis based on Hoey (1991) and the two topic-based approaches

Firstly, where one method identifies more discontinuities in the discourse, the stretches of discourse in which topic entities are identified in that approach will be shorter. These differences in lengths of stretches of discourse can result in the topic entities identified in two approaches representing different levels in a topic hierarchy. For example, in extract A for T-units 1 to 20 (see Appendix C, p. 269), both the method based on Hoey (1991) and the relation-based topic-based approach identify a single topic entity. The association-based topic-based approach, on the other hand, identifies two discontinuities after T-units 3 and 4. This third approach then identifies distinct topic entities for the short stretches of discourse bounded by these discontinuities. The other two approaches, however, subsume these short stretches of discourse into a much longer stretch for which a topic entity higher up the topic hierarchy is identified.

Secondly, the different methods may identify discontinuities at slightly different points while still identifying the same topic entities around these discontinuities. For example, in extract D (see p. 276) the relation-based topic-based approach identifies *volunteer* as the topic entity of the first 10 T-units at which point there is a discontinuity. *Balloon* is identified as the topic entity for the succeeding discourse. The other two approaches identify the same topic entities but place the discontinuity after T-unit 12, resulting in different identifications of topic entities for T-units 11 and 12.

Thirdly, the basis for identifying topic entities in the analysis based on Hoey (1991) and the two topic-based approaches are different. Hoey (1991) stresses identification on the linguistic basis of words (see p. 79), whereas the two topic-based approaches use the more psychological basis of concept (see pp. 119-123). This results in the identification of different topic entities. For example, in extract A for T-units 145 to 178, the analysis based on Hoey identifies the topic entity as *balloon*, whereas the two topic-based approaches identify it as *big balloon*.

These three causes for variations in the topic entities identified in the three approaches may parallel differences in how people identify topics. When asked to identify the topic(s) in a stretch of discourse, a person may interpret *topic* as referring to different levels in the topic hierarchy, may place the start of a new topic at different points in the discourse, or may express any identified topic as a single word, as a phrase or even as a proposition.

More problematically for our purposes of trying to identify topics, even where two people identify a topic at the same level in the hierarchy, for the same stretch of discourse, and expressed in the same way, the topics identified may differ. These differences may be due to differences in the schematic structures of the two people or differences in what the two people perceive as salient in the discourse.

These problems with topic identification are similar to the problems in evaluating coherence that were discussed in chapter 2. We saw that coherence is inherently subjective leading to variations between people in judgments concerning coherence, but also that an overwhelming consensus of opinion can be achieved. Coherence does not reside in discourse, but is created by interaction with the discourse. As long as people's schemata and other bases for interpreting discourse are similar, their opinions concerning coherence are likely to be similar.

This argument also applies to topics. Topics are identified through interaction with the discourse and can differ between people. However, high levels of agreement concerning topics can also be achieved, as evinced by the high reliability ratings for topic identification questions on reading tests (see e.g. Peirce, 1992). Discourse which is relatively incoherent ideationally may lead to variation in topic identification, but coherent discourse is likely to lead to a general consensus concerning the topics.

While we might hope for agreement between people concerning the topics of a stretch of discourse, we have to accept that there never can be an unambiguously conclusive identification of topics for any discourse. From the perspective of this study, while we may be more confident of our identification of topics at points where all six approaches and the participants agree than at points where there is a wide range of topics identified, even at the former points we cannot say that we have definitively identified the topics of the discourse. The most that it is possible to hope for in this study, therefore, is that the amount of agreement between methods of analysis reflects the amount of agreement that there would be between two people faced with the same discourse. On the basis of the data in this study, we cannot, however, tell if this is the case or not.

7.1.2 Topic entities and higher-level topics

My argument so far has largely been based on the topic entities shown in Appendix C (p. 269). In this section, I intend to look at how topic entities combine to form topics by examining how lower levels in a topic hierarchy combine to form higher levels. While four of the methods of analysis used in this study (the approaches based on Sinclair and Coulthard (1975) and Hoey (1991), and theme-rheme and given-new progression) allow only bottom-up identifications of topics, the two topic-based approaches take both bottom-up and top-down considerations into account. Although the identification of key concepts and topic entities in these approaches is bottom-up, the hierarchies in the relation-based approach and the networks in the association-based approach can be considered representative of generalised schemata and thus allow top-down considerations a role. In the following discussion, the approach I will take will be primarily bottom-up, but I will also consider the schematic hierarchies and networks of the topic-based approaches.

Topic entities, or the outstanding concepts in stretches of discourse bounded by discontinuities, are probably one of the smallest units in a topic hierarchy. Below topic entities, we might consider sentence topics or themes, but, as we saw in chapter 6, identification of topics through theme-rheme analysis has little in common with the other methods of analysis. Instead, as we shall see when we look at sub-topical progression, it is probably more useful to consider concepts as being at a lower level than topic entities, although it is unclear whether concepts should be considered part of a topic hierarchy.

If topic entities are at a low level in a topic hierarchy, we need to examine how they combine to form larger units. For convenience, in the following

discussion, I will assume that there are three levels in a topic hierarchy, namely topic entities, mid-level topics and macro-topics. This is a simplification for the purposes of clarifying the discussion, and other real discourse may generate more intermediate levels or may be viewed as a continuum of levels rather than a series of discrete levels.

The first point to make is that, while a certain concept may be the topic entity of a short stretch of discourse, it may also act as a mid-level topic and the macro-topic at the same time. Indeed, most mid-level topics and macro-topics also serve as topic entities for some stretch of discourse. There are three main ways in which topic entities may become topics at higher levels.

Firstly, where a stretch of discourse is isolated from the rest of the discourse by two discontinuities indicating topic shift and the only concept(s) within the stretch serve as topic entity(ies), the topic entity of the stretch of discourse would also serve as the mid-level topic and macro-topic of the same stretch of discourse if we wished to identify these for reasons of parallelism with other stretches of discourse. For example, the two topic entities, *seen* and *Star Wars Episode One*, both also act as mid-level topics for the isolated stretch of topic insertion in extract G from T-unit 55 to 59 (see Figure 5.4, p. 145).

Secondly, for relatively long stretches of discourse containing no discontinuities and exhibiting topic maintenance, the topic entity is likely to also be a topic at a higher level. In cases such as these, there may be other recurring concepts in the discourse which are not identified as topic entities. For example, in extract G, T-units 37 to 54, there is a single topic entity for the stretch of 18 T-units, namely, *robot*. Other concepts, such as *cook food* and *servant*, recur in the stretch of discourse. This stretch exhibits topic maintenance of *robot* at the mid-level and probably macro-level, but at a sub-topical level exhibits topic drift from *cook food* to *servant* and back to *cook food* within the topic entity level framework of *robot* (or linking clusters 2 and 5 with the predominant cluster 3 in the association-based approach).

Thirdly, several topic entities may be subsumed by a mid-level or macro-topic in the same way that *cook food* and *servant* are subsumed by *robot*. For example, in extract E the teacher asks students to ask questions to guess an object he is thinking of. Some of these questions (in T-units 38 to 53) are about whether the object has *wheels*, whether it is used *every day*, and whether it is used in a *car*. These three concepts are identified as topic entities in the analysis based on Sinclair and Coulthard (1975), but

the three approaches we are focusing on all identify *object* as the topic entity for this stretch of discourse. In this way, T-units 38 to 53 in extract E follow the same pattern as T-units 37 to 54 in extract G discussed above with topic maintenance of *object* and sub-topical drift. Later in extract E, however, another question concerns whether the object can carry things, and *carry* is identifies as a topic entity for T-units 61 to 65 by the two topic-based approaches. For the same T-units, the analysis based on Hoey (1991) identifies *object* as the topic entity. The succeeding discourse then identifies *object* as the topic entity in all three approaches. *Carry* has the same ideational import as *wheels*, *every day* and *car*, but has a level of prominence in the discourse that raises it to the level of topic entity. It is, however, subsumed by *object* as a mid-level topic.

There may be occasions when we identify a higher level topic that never acts as a topic entity. For example, in extract G for T-units 8 to 36, a variety of topic entities (*boy*, *woman*, *robot*, *C3PO*, *R2D2*) are identified. In the relation-based topic-based approach, all of these concepts are meronyms of *Star Wars Episode One* (see Figure 5.2, p. 134). Although some of these topic entities, such as *boy* and *R2D2*, may act as mid-level topics, none provide the coherence for the whole stretch of discourse necessary to be considered macro-topics. Instead, at a macro-level, they may be subordinate to topic maintenance of *Star Wars Episode One* (which does not appear as a topic entity) with mid-level drift between meronyms.

In the topic drift of T-units 8 to 36 of extract G, we can see how different concepts may be identified as the topic entity, mid-level topic and macro-topic of a stretch of discourse. For T-units 37 to 54, on the other hand, the topic entity, *robot*, is the mid-level topic and macro-topic as well (see Figure 7.1, p. 202 which illustrates this). In this latter case, the stretch of discourse exhibits topic maintenance, the type of progression for which topic identification is the most straightforward.

Given that it is possible for topic entities to be mid-level topics and for mid-level topics to be macro-topics, is it possible to identify one of the macro-topics in extract G as the overall topic of the whole extract? At first glance, it seems potentially possible. In extract L, for example, the relation-based topic-based approach identifies only one topic entity for the whole extract, and this would presumably serve as mid-level topic, macro-topic and overall topic for the whole extract. Using the analysis based on Hoey (1991), we saw that a single exchange (T-units 38 to 45) was central to the whole of extract G and could be considered indicative of topic (see p. 96). Furthermore, in conducting a relation-based topic-

based analysis, we found that *robot* was the most frequently linked concept while also being superordinate to many other concepts, suggesting that *robot* could be the overall topic.

Before we state that *robot* is the overall topic of extract G, we need to consider whether we should identify a single overall topic for the extract. Extract G is simply one part of a lesson. Although it is interactionally coherent in that the whole extract serves the discourse function of eliciting, there is no reason why the whole of extract G should also be ideationally unitary and fall under a single overall topic. Indeed, given that eliciting usually involves a move from a familiar starting point to the focus of the lesson, we may actually expect single-topic eliciting to be the exception to the norm. The two macro-topics we have identified above (topic drift within the context of *Star Wars Episode One*, and topic maintenance of *robot*) act as the familiar starting point and the focus of the lesson. Although *robot* is more central to the lesson as a whole, we cannot say that extract G has a single overall topic. Instead, it would be better to categorise extract G as involving a move from *Star Wars Episode One* to *robot*.

Similar moves from an initial topic to the lesson focus are apparent in extracts C and D (from *balloon* to *engine*), F (from *Star Wars Episode One* to *robot*), and I and K (from *pump* to *heart*). For the longer extracts A and B, there are three macro-topics: for A, *balloon*, *big balloon* and *action reaction* (the focus of the lesson, *engine*, only becomes explicitly focused on in the last T-unit of the extract; and for B, *balloon*, *action reaction* and *engine*. For the other three extracts, H, J and L, all of which are short, there is only one macro-topic (*heart*, *pumping system* and *heart* respectively).

The pattern throughout the extracts, then, seems to be one where the shortest extracts may have only one macro-topic, the medium-length extracts have two macro-topics, and the longest extracts have three. This pattern suggests a correlation between length of extract and number of macro-topics ($r = 0.88$; $p < 0.001$), which in turn suggests a preferred length of 40 to 80 T-units for a macro-topic at this level in the topic hierarchy.

7.1.3 Topic hierarchies

In the previous section I have attempted to identify topics at various levels in a topic hierarchy. We have seen that short stretches of discourse have topic entities which combine to form mid-level topics and, at the highest level in this study, macro-topics. I am not, however, suggesting

that it is not possible to have levels of topics higher than the macro-topics discussed above. Although the macro-topics appear to be the highest topic level in the extracts investigated in this study, if we expand the range of our data to include whole lessons or units comprising four or five lessons, we may find topics at higher levels which subsume the macro-topics identified in this study. What topics we identify, then, are dependent on the length of the discourse that we are investigating.

In this section, I would like to suggest ways in which these different levels in a topic hierarchy may be manifested in some of the methods of analysis used in this study. Although this study does not present clear findings concerning these points, they may warrant further research.

The first method that warrants further investigation concerning its relationship to topics is that of Sinclair and Coulthard (1975). In chapter 4, we saw that some authors have tried to equate levels in the discourse hierarchy of Sinclair and Coulthard with topics. It has been suggested that transactions are macro-topic units (Hazadiah, 1993), and that exchanges in carrying only one piece of information (Coulthard and Brazil, 1992) may help us in identifying either concepts or topic entities. The intermediate level between exchange and transaction (variously called exchange complex (Hoey, 1993), topic framework (Hazadiah, 1993) and sequence (Sinclair and Brazil, 1982)) has been equated with mid-level topics (Hazadiah, 1993). Furthermore, van Lier (1988) has suggested that the boundary exchanges of Sinclair and Coulthard between transactions are indicative of major topic shifts (or shifts between macro-topics in the terms of this study), and that there may be other less prominent boundaries (manifested by, for example, 'OK' without a following measured pause) indicative of minor topic shift between mid-level topics.

In this study, the evidence for such claims is far from convincing. In chapter 6 we saw that evidence for placement of topic shifts at transaction boundaries was ambiguous. The lack of any convincing match between the topic entities identified using the analysis based on Sinclair and Coulthard and those identified in the other approaches suggests that, if exchanges equate with a level in a topic hierarchy, it is with a level lower than that of topic entity as used in this study. The intermediate level of exchange complex was not investigated in this study since methods of identifying this level are unclear and vary between authors.

Given the lack of any convincing evidence showing a relationship between Sinclair and Coulthard's levels of discourse and the levels in a topic hierarchy, why is this worthy of further investigation?

The first reason is that several aspects of applying Sinclair and Coulthard's method of analysis in this study are unsatisfactory, most notably the need to put all teacher monologues into a single exchange irrespective of length. Recent work on teachers' instructions (Chaiyasuk, 2002; Watson Todd et al., mimeo) has suggested ways of identifying exchanges in teacher monologues. Further developments, including clearer guidelines for identifying exchange complexes, may help to provide more clear-cut evidence for whether levels of discourse and topic hierarchies are related.

A second and probably more serious set of uncertainties concerning Sinclair and Coulthard's approach is that they do not explicitly define the levels in their hierarchy and that, in some cases, it is less than clear how levels should be identified. In the series of studies that followed the publication of Sinclair and Coulthard, a functional investigation of lectures (Coulthard and Montgomery, 1981) proposed a different set of levels, namely, transaction, sequence and member. The interesting point about this study is that, while transactions were identified from the surface features of boundary exchanges as in Sinclair and Coulthard, sequences were identified based on the introduction of concepts, and members were equivalent to syntactic T-units. If this basis for identifying the different levels is applicable to Sinclair and Coulthard, then their seemingly purely functional approach has a conceptual basis warranting further investigation of potential parallels between the two hierarchies: discourse and topic.

Finally, although this study has suggested that the link between the discourse levels of Sinclair and Coulthard and the levels in a topic hierarchy is less clear than might be expected, the fact that several authors have posited such a link means that further less ambiguous evidence for the lack of any such link is needed before we can state that the two hierarchies are not related. At present, we can only say that any such relationship is uncertain.

A second method of analysis that may provide interesting insights into topics at different levels in a topic hierarchy is that of Hoey (1991). In chapter 4, we saw that there may be a central exchange which bonds to many other exchanges coming both before and after it (see p. 96). Another point that is worth considering is whether a network of bonds with a high threshold such as Figure 4.1 (p. 87) highlights topics at higher levels in a topic hierarchy than a network of bonds with a low threshold (e.g. Figure 4.3, p. 89).

A moment's consideration, however, shows that there is unlikely to be any useful relationship between these points and the levels in a topic hierarchy. The networks in Hoey's analysis are based on the number of lexical items appearing in different T-units or exchanges, and pay no attention to what the lexical items are. In other words, those T-units or exchanges with a high number of bonds are those which contain the most lexical items which appear in other T-units or exchanges. Although these lexical items may include the macro-topics, there is no way in the approach to distinguish the lexical item representing the macro-topic from the other repeated lexical items in a T-unit or exchange.

An approach which is more insightful concerning the higher levels in a topic hierarchy is the relation-based topic-based approach. We have already seen (section 5.2.4, pp. 134-139) that in extract G the concept which is involved in the most moves between concepts in the discourse is likely to be important. This method of identifying important concepts applies to nearly all of the macro-topics in the extracts used in this study. In every case, the concept most frequently linked by moves between concepts is identified as a macro-topic, and for those extracts with more than one macro-topic, generally the most frequently linked concepts are the macro-topics. The exceptions to this pattern are either general concepts providing contextualisation at the start of an extract (such as *Star Wars Episode One* in extract G) or a final concept which is the focus of a lesson. In the latter case, such as *robot* in extract E, the focus of the lesson is identified but not explored in the eliciting part of the lesson.

Perhaps more surprisingly, macro-topics may also be identifiable from the hierarchies of concepts (such as Figure 5.2, p. 134) even before the moves between concepts have been drawn onto the hierarchy. Generally, the macro-topics in the discourse appear relatively high in the hierarchy as either the most or the second most superordinate concepts. Also, the macro-topics are usually those concepts with the highest numbers of subordinate concepts in the hierarchy. Again, the main exceptions to this pattern are those concepts which are the identified but not explored foci of lessons.

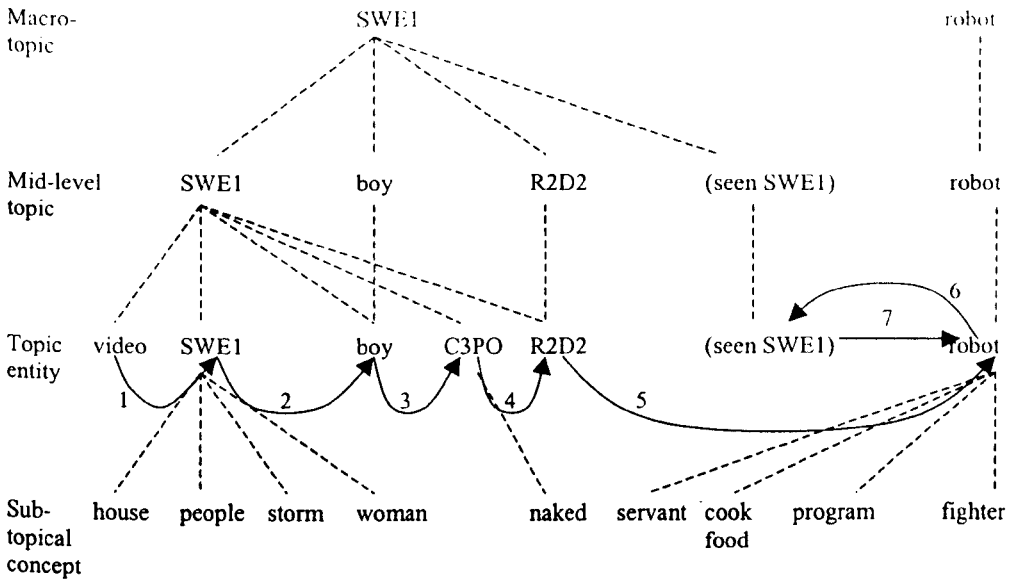
If these patterns hold for other stretches of discourse, they allow top-down identification of the higher levels in a topic hierarchy. Instead of looking at how topic entities combine to form topics, we can examine the generalised schema represented in the hierarchy of concepts to identify macro-topics.

A final approach that needs to be discussed concerning its relationship to levels in a topic hierarchy is the association-based topic-based approach. For extract G, we find that the two macro-topics, *Star Wars Episode One* and *robot*, occur in the same cluster (see p. 155). This case, however, appears to be the exception to the general pattern. In the vast majority of extracts where there is more than one macro-topic, the two or three macro-topics occur in different association clusters. The general pattern in the extracts, then, is a move from one cluster to another, perhaps indicative of a drift across semantic space.

This indication of a drift in semantic space is actually crucial to our understanding of topics. Throughout this section, I have been discussing topic hierarchies, and, as Hudson (1980) points out, it is tempting to view these as static. They are, however, dynamic in nature.

The two topic-based approaches provide ways of conceptualising dynamic topic hierarchies. Firstly, using the association-based topic-based approach as a metaphor, we can view the dynamic nature of topics as a drift between different clusters of concepts, each of which represents an area of semantic space. Within each cluster, there may be a more central concept representing a higher level in a topic hierarchy with concepts more peripheral to the cluster being lower down the topic hierarchy.

Secondly, we could use the relation-based topic-based approach, and specifically the hierarchy of concepts with the moves between concepts drawn onto it such as Figure 5.3 (p. 138), as an analogy. The main difference between such a hierarchy and a hierarchy of topics is that the latter may include the same concepts at several different levels. Basing the topic hierarchy on the topic entities identified in the relation-based topic-based approach, we can draw up a hierarchy of topics for extract G as shown in Figure 7.1.



Note: SWE1 = Star Wars Episode One
Dotted lines show topic relationships
Numbered arrows show moves between topic entities in the discourse

Figure 7.1 A dynamic topic hierarchy for extract G

In Figure 7.1, we can see that the first half of extract G (the left half of Figure 7.1) has a more complicated structure than the second half. This is indicative of the more complicated nature of topic drift than of topic maintenance. The first four moves between topic entities are all between concepts co-subordinate to *Star Wars Episode One* indicating topic drift. Move 5, however, jumps to a topic entity subordinate to a different macro-topic representing topic shift. Such shifts are also apparent in moves 6 and 7 around the topic insertion of *seen* and *Star Wars Episode One*. The general pattern of topic maintenance in the second half of the extract is shown by the existence of a single concept at the different levels of the topic hierarchy, with sub-topical drift shown by the range of sub-topical concepts under *robot* as a topic entity. A hierarchical representation of the topics in extract G shows how the progression through the topics in the discourse involves moves between topics/topic entities within one hierarchy or moves between one hierarchy of topics (that under the macro-topic of *Star Wars Episode One*) and another (that under the macro-topic of *robot*). In this way, the hierarchy presents a dynamic view of topics.

7.1.4 Guidelines for identifying topics and topic progression

While I hope that a figure such as Figure 7.1 provides a clear and valid summary of the topics and topic progression in extract G, the procedures used to reach this summary are scattered throughout chapters 3 to 7. In this section, therefore, I intend to provide a list of the procedures I have used to identify topics and topic progression in this study. These are as follows:

- from a transcription, identify T-units (see section 3.3.2);
- identify referents of referring expressions (see section 3.3.3.1);
- identify and supply ellipsis (see section 3.3.3.2);
- identify key concepts (or lexical items, see section 4.2) based on frequency and saliency (see section 5.2.1);
- identify moves between concepts (see section 5.2.4);
- identify moves representing discontinuities in the discourse, using one or more of the following analyses (see section 6.1): the analysis based on Hoey (1991) (see section 4.2), the relation-based topic-based approach (see section 5.2.6), the association-based topic-based approach (see section 5.3.4), and possibly given-new progression (see section 4.3.2);
- identify topic entities for the stretches of discourse between discontinuities (see sections 5.2.6, 5.3.4 and 6.2);
- if a stretch of discourse is isolated by two discontinuities representing major topic shift, the topic entity also acts as a higher-level topic (see section 7.1.2);
- for a long stretch of discourse with a single topic entity, the topic entity also acts as a higher-level topic with topic maintenance and possible sub-topical drift (see section 7.1.2);
- for discourse with several minor discontinuities, a superordinate concept is likely to be a higher-level topic with topic drift (see section 7.1.2);
- a concept most frequently linked by moves between concepts in a concept hierarchy is likely to be a higher-level topic (see section 7.1.3);
- a superordinate concept with many subordinate concepts in a concept hierarchy is likely to be a higher-level topic (see section 7.1.3);
- a second macro-topic in an extract is likely to be located in a different association cluster from the first macro-topic (see section 7.1.3).

Following these guidelines should help in identifying topics in discourse. It should be remembered, however, that any such identification can never be definitive.

7.2 Methodological implications

In this study, I have purposely used a variety of theoretical frameworks in my attempt to identify topics and topic progression. Because of this, there are a variety of methodological implications arising from this study. In this section, I will move the focus away from topics and look at the methodological implications which are applicable to other areas of applied linguistics.

7.2.1 Implications specific to a given approach

In this study, I used six main methods of analysis to identify topics. In attempting to apply these methods, implications concerning each of them have arisen and I will look at the implications of each approach in turn. I will also examine the implications of the methods used to compare the approaches.

The first method of analysis is that of Sinclair and Coulthard (1975). This approach is founded in speech act theory and the descriptions it produces are functional. However, the need to take conceptual considerations into account to determine whether succeeding initiating moves fall into the same exchange suggests that the implementation of the method is not purely based on functions. The extent to which conceptual concerns are involved in the approach may warrant further research and could provide insights into the relationship between functional and conceptual aspects of language. Although it may seem that further research into a method that is over a quarter of a century old may not be worthwhile, the recent award for an article based on this approach (Nassaji and Wells, 2000) suggests that such research may be productive.

The second approach used in this study was Hoey's (1991) lexical analysis. While I followed Hoey (1991) closely in applying this method, one aspect which I did not follow was the setting of a particular threshold to identify bonded T-units. By lowering the threshold for bonds (see Figures 4.1, 4.2 and 4.3, pp. 87-89), we can clearly identify the clusters of T-units which are most closely bonded and distinguish these from more loosely bonded clusters. We can also see an emerging pattern of interconnectedness as the threshold is reduced. I believe that reducing the threshold in this way produces a richer picture of the lexical networks in discourse than setting a predetermined threshold, and may be a more fruitful approach in other studies involving lexical networks.

The findings in this study concerning the two approaches based at the clause level, theme-rheme progression and given-new progression, stand

in contrast to much of the literature. Firstly, the literature claims a strong correlation between theme and given information on the one hand, and between rheme and new information on the other (e.g. Brown and Yule, 1983a; Connor, 1994; Daneš, 1974; Fries, 1983, 1994). In this study, however, less than two-thirds of themes coincided with given information. Although some of the difference is due to the inclusion of ellipted material in the given-new analysis, the differences in the topic entities identified in the two approaches suggest that they have different bases. The high coincidence in the identification of discontinuities, on the other hand, does suggest similarities. The findings therefore suggest that claims for a strong correlation between the two approaches may be overstated, but that some relationship may exist. The exact nature of this relationship may warrant further investigation.

A second aspect of theme-rheme analysis which this study sheds light on concerns whether sentence-level topics (or themes in this study) are equivalent to the lowest level in a topic hierarchy. Previous work in this area (e.g. Brown and Yule, 1983a; Carlson, 1983; van Lier, 1988) has not reached any firm conclusions concerning this. On the evidence of this study, however, it appears that sentence topics or themes are not related to discourse topics. This conclusion is especially apparent in the lack of a relationship between the topic entities identified through theme-rheme progression and those identified by the other approaches.

The final two methods used in this study are the two topic-based approaches. Both of these approaches attempt to generate a representation of semantic space or a content schema for the concepts in the discourse (in the form of a hierarchy for the relation-based approach and a network for the association-based approach). Although based on the concepts in the discourse, these representations use methods from outside the discourse to relate the concepts. For the association-based approach, the representation of the discourse is general and would be the same for any stretch of discourse containing the same concepts, irrespective of how these concepts were sequenced. The sequencing of the concepts becomes important when we map the moves between the concepts onto the network. In doing this, we found that it was necessary to account for context-specific associations generated by syntactic proximity between concepts as well as the general associations in the network. Although no context-specific relations were added to the hierarchy of the relation-based approach in mapping the moves onto the hierarchy, the hierarchy itself was generated using both general and context-specific information. For example, while *C3PO* will always be identified as a *robot* indicating a general relationship between the two concepts, *C3PO* is not usually

naked indicating a context-specific relationship. In both approaches, therefore, both general and context-specific relations and associations were taken into account. The proportions of and relationship between general and context-specific concerns in discourse could have wide-reaching implications, especially for schema theory. It would therefore appear to be an area clearly needing further investigation.

Focusing on the association-based approach, in chapter 5 (pp. 152) we saw that there were several anomalous associations generated by the corpus used in this study. These problems with the corpus severely weakened the analysis. Nevertheless, in chapter 6 the association-based approach appeared to be a relatively valid method of identifying topics, suggesting that further research into this method, probably involving more careful selection of the corpus, would be worthwhile.

In chapter 6, the six methods of analysis were compared, and I believe that two of the approaches to comparing the analyses could be productive in other comparisons of approaches. Firstly, the use of extended trees allows several approaches to be compared simultaneously focusing on either their similarities (as in this study) or their differences. Secondly, the use of a control approach can strengthen our certainty concerning findings arising from approaches based on largely untested theoretical frameworks and can provide insights into the nature of approaches which produce distinct findings.

Because of these methodological implications, I hope that this study may have a potential impact beyond the relatively narrow area of topics that is its specific focus.

7.2.2 Implications for natural language processing

A major area of research associated with applied linguistics is natural language processing, or the manipulation and analysis of naturally occurring language using computers. Much of the work in this area concerns the discourse level of linguistics (as defined in this study, see p. 5). For example, among the concerns of natural language processing are anaphora resolution, automatic text summarisation, discourse segmentation, and information extraction.

In examining discourse, natural language processing research often uses approaches familiar in applied linguistics. For instance, Hearst (1994) and Kozima (1993) both use lexical cohesion as the basis for computer segmentation of text, and Strzalkowski et al. (1998) attempt to produce automatic summaries of news based on genre templates.

Automatic text summarisation is perhaps the application for which this study has most relevance. A typical approach to computer-based summarisation involves three stages: topic identification, topic interpretation and summary generation (Hovy and Lin, 1997). The first of these three stages is clearly an area where there may be implications arising from this study.

Computer identification of topics is fraught with problems, primarily because of the need for a model of world knowledge to account for semantics (Devlin, 1997; Polanyi, 1993). Such world knowledge is unprogrammable, and thus natural language processing researchers usually rely on cohesion, rather than interactional coherence, together with resources such as thesauruses, as the basis of their identification of topics. Since the present study focuses on ideational coherence, it may be worth considering whether any of the methods of analysis or findings have implications for natural language processing. In doing this, I will assume that the data are either written discourse or spoken discourse which has already been divided into T-units.

To start with, we can consider which of the methods of analysis is potentially programmable. Three of the approaches can quickly be ruled out: the analysis based on Sinclair and Coulthard (1975) because of the need for researcher interpretation to identify exchanges; given-new progression because of the need to identify some given information from the context or from shared expectations of the interlocutors (information deriving its given status from the preceding discourse could theoretically be identified by computer); and the relation-based topic-based analysis because of the reliance on world knowledge in identifying logical relations between concepts.

Themes and rhemes in theme-rheme progression, on the other hand, could be identified by computer. Because of the syntactic nature of themes, a reliable parsing program would allow their identification. However, the lack of relationship between themes and topics shown in chapter 6 suggests that this approach will not be very productive as a basis for topic identification.

The analysis based on Hoey (1991) could also be conducted by computer. Indeed, a similar approach was used by Hearst (1994) and Kozima (1993) in their attempts to segment discourse. With a thesaurus to identify paraphrases of lexical items and algorithms for referent and ellipsis resolution, a program could be constructed to identify links, and counting

these would generate the networks of bonds and running averages of links. From these, discontinuities and topic entities could be identified automatically, but it is unclear how a computer could be programmed to combine the identified topic entities into topics. This approach may therefore be more appropriate to computer-based discourse segmentation than to topic identification.

Even more promising is the association-based topic-based analysis, since most of the methodology of this approach requires computers. With a suitable corpus, clusters of concepts could be generated, and these would allow discontinuities, topic entities and diagrams of topic progression such as that given in Table 5.10 to be identified. Because of the extra dimension of topic progression, the association-based analysis may prove to be the most productive of the methods for computer-based topic identification and may warrant further research.

One further aspect of this study has potentially important applications for natural language processing. In this study I have viewed topics as being expressible by concepts rather than propositions, and the evidence from the students' identification of topics supports this view (see p. 189). Concepts are usually expressed as a single word, which are easily identifiable by computer (and the guidelines given in chapter 5, pp. 120-123, for identifying exceptions to this are generally programmable). Viewing topics as concepts is therefore more facilitative of computer-based topic identification than viewing topics as propositions. This study then provides some evidence for taking an easier and potentially more productive approach to topics in natural language processing.

7.3 Pedagogical implications

A starting point to this study was that content or conceptual concerns are vital to classroom learning, especially how the content is structured. In chapter 1, (p. 14), we saw that teachers need to describe relationships between ideas, clearly organise content, present material in small coherent logical steps, and sequence material according to the accepted logic of the discipline. We also saw that logical step-by-step sequencing producing a scaffolded structure in eliciting is a key tactic in teaching (p. 16). However, it seems that how to do all these things is left to a teacher's intuition. We can now ask whether this study has provided any foundations on which to guide or evaluate these aspects of teaching.

There are some problems in this study which make it unclear whether the study has anything to add to teachers' intuitions concerning these aspects of teaching.

Firstly, the disappointing data from the participants (see section 6.5, p. 185) means that we are unable to evaluate the effectiveness of any of the eliciting in this study. Because of this, we are unable to identify whether any of the patterns emerging from the findings are likely to be more effective in promoting learning.

Secondly, the complex nature of the procedures used in the study makes it unlikely that any teachers can apply them to improve the logical sequencing of material in their teaching.

Thirdly, most of the methods used in this study can only be used to analyse eliciting post teaching. The methods need transcriptions of eliciting sections of lessons to be made before any evaluation of the logical sequencing of the eliciting can be made. Because of this, they cannot be used to improve teaching by, for example, helping teachers to plan more logical steps of teaching. The only two methods which could potentially be used in planning teaching are the two topic-based approaches. If in the planning stage teachers can predict the concepts that will be covered in their teaching, producing a general representation of the relationships between these concepts may help a teacher to plan a logical organisation of the concepts for the lesson. Of the two topic-based approaches, however, the association-based approach requires time and resources in preparing the general representation beyond those available to nearly all teachers. The only method that may be of some practical use to teachers, then, is the relation-based approach.

In implementing the relation-based approach in this study, I have been quite rigorous in my attempts to identify relations between concepts. It seems likely, however, that a reasonably reliable representation of the relationships between concepts could be produced even with a much looser interpretation of how to identify relations. Evidence for this comes from the initial impetus behind the first of my two articles which led to this study (Watson Todd, 1997a), where the idea originated in trying to draw up hierarchies of relations between concepts while observing episodes of peer eliciting (see Watson Todd, 2002b). A loose version of the relation-based approach may therefore be applicable in real-time observation of teaching enabling teachers to gain feedback on the logical sequencing of material in their teaching. It also seems likely that, with some initial training, teachers could apply a loose version of the relation-based approach in their planning to improve logical sequencing. Further research into the feasibility and effectiveness of such training and

planning is needed, however, before we can clearly see whether this study has direct pedagogical applications.

7.4 Conclusion

In chapter 2, I proposed the following definition of *topic*: a clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create coherence and relevance. We are now in a position to evaluate the validity and usefulness of this definition. To do this, I will look at each of the terms in the definition in turn.

Clustering: This study assumes that topics are delimitations of semantic space within which is a group or cluster of concepts. This assumption is most clearly illustrated by the association-based topic-based analysis which produced clusters of concepts. In comparing the discourse against these clusters, we found that context-specific associations also need to be considered in using the approach. These context-specific associations suggest that clusters of concepts can be produced through discourse in addition to the more generalisable clusters of associations identified through the corpus-based analysis. In other words, associations and relations in semantic space may be constantly forming and shifting under the influence of the ongoing discourse. This constant forming and shifting does not militate against the use of clusters, but does imply that these clusters are dynamic in nature, hence the use of the more dynamic term *clustering* rather than the more static *cluster*.

Concepts: As we saw in chapter 2 (p. 20), some authors (e.g. Keenan and Schieffelin, 1976) explicitly argue that topic should be expressed as propositions rather than concepts. However, I believe that the potential productiveness of this study together with the participants' use of concepts to express topics justifies an approach to topics based on concepts.

Associated or related: In chapter 5, I implemented two versions of the topic-based analysis, one based on relations and the other on associations. The similarities in the findings from these two analyses (see chapter 6) suggest that they may share some common basis. However, the differences between them, especially in how they conceptualise topic progression, are also interesting. The relationship between associations and relations is one which warrants further investigation, and it may also be worthwhile attempting an analysis which combines the two, although at present it is unclear how this could be done.

From the perspectives of the interlocutors: Earlier in this chapter, I stressed the fact that there can never be any definitive identification of topics. Although a general consensus may be reached, we may also expect some informants to hold idiosyncratic ideas concerning topics. This does not devalue a study like this which attempts to identify topics, but it does mean that such a study can never be conclusive. More research attempting to elicit participants' perspectives is needed (especially as this is probably the weakest aspect of the present study), and it may be particularly interesting to try to identify the bases underlying idiosyncratic identifications of topics.

To create: Creation implies a dynamic process, and the shifts and drifts of topics through semantic space necessarily entail a dynamic view of topics. The production of discourse, especially its context-specific aspects, does not confirm an existing model of semantic space, but actively generates new models.

Coherence: In this study, I have taken coherence as meaning the covert ways of connecting discourse which cannot be detected directly from the surface of the discourse, in contrast to the overt connections of cohesion. I have suggested that coherence and cohesion should not be viewed as two distinct entities, but as ends of a continuum of connectedness. There is some circumstantial evidence for this viewpoint. The analysis based on Hoey (1991) is founded on lexical cohesion. The two topic-based analyses, on the other hand, in searching for connections based on principles outside the discourse, focus on coherence. The close match between the findings of these three approaches can be explained in two ways. Either there is a surprisingly high coincidence between coherence and cohesion in the data of this study; or coherence and cohesion are less distinct than has generally been argued previously. Given that highly coherent discourse does not need to rely exclusively on cohesion for connectedness, the first explanation seems unlikely. It may therefore be that coherence and cohesion are the two ends of a connectedness continuum. If this is the case, and given the apparent validity of an approach based on lexical cohesion, it may be appropriate to replace *coherence* in the definition of *topic* with the superordinate term *connectedness*.

Relevance: In chapter 2 (pp. 29), we saw that stretches of monotopical discourse exhibit both aboutness and relevance, and drifting discourse exhibits relevance. I also argued that relevance could be applied at several different levels, with seemingly unrelated concepts potentially exhibiting relevance at a very global level. Earlier in this chapter, I examined the

nature of topic progression at different levels in a topic hierarchy, whereby, for instance, a monotopical stretch of discourse at one level may involve topic drift at a lower level. Aboutness and relevance would therefore seem to be dependent on the level in a topic hierarchy at which the discourse is viewed. Because of this, when we talk about topics creating relevance, we are implying that we are viewing discourse at a certain level in a topic hierarchy.

From the discussion of the terms in the definition of *topic* used in this study, the definition which was derived from theory appears to be supported by the findings of this study. From one perspective, this is to be expected as the study was largely founded on the same principles as the definition. From another perspective, however, it could be argued that, by using the approach based on Sinclair and Coulthard (1975), the two approaches based at the clause level, and the analysis based on Hoey (1991), none of which are based on the same principles, this study has evaluated the definition. The unsuitability of the first three of these approaches for investigating topics lends support for the definition. The potential effectiveness of the last of these approaches suggests that we need to replace *coherence* in the definition with *connectedness*. We are therefore now in a position to more strongly assert that *topic* can be defined as a clustering of concepts which are associated or related from the perspective of the interlocutors in such a way as to create connectedness and relevance.

As a final point, I would like to consider whether this study has met the goals I set in Chapter 1 (pp. 3, 17-18). This study stems from an article I wrote (Watson Todd, 1998) outlining a relation-based topic-based analysis. The main goal of the article and of this study is to shed light on topics and topic progression in classroom discourse. I hope that, although no definite conclusions have been, or indeed can be, reached, this study has led to a clearer understanding of this goal.

Since this study is essentially a much expanded follow-up to an initial application of relation-based topic-based analysis, the fact that this analysis emerges as one of the most promising in this study is heartening confirmation for my earlier work. I also intended this study to provide an improvement of the methodology of the original article. I aimed to make improvements in four main areas: increasing the amount and variety of data to be analysed, comparing the approach against other methods of analysis, making the methodology more rigorous, and attempting to account for associations as well as relations. I hope that by examining twelve extracts instead of a single extract as in the original article, by

comparing the relation-based topic-based analysis with five other approaches (and with a control approach) in chapter 6, by setting explicit guidelines especially in chapters 3 and 5, and by designing and implementing an association-based topic-based analysis, I have overcome the main weaknesses of the approach described in the original article.

In these ways, I believe that this study has met its goals and hope that it will lead to further research that sheds more light on both topics and classroom discourse.

References

Abbott, V., Black, J. B. and Smith, E. E. (1985) The representation of scripts in memory. In Komatsu, L. K. (ed.) (1994) *Experimenting with the Mind: Readings in Cognitive Psychology*. Pacific Grove, CA: Brooks/Cole. pp. 195-208.

Allwood, J. (1986) Logic and spoken interaction. In Myers, T., Brown, K. and McGonigle, B. (eds.) (1986) *Reasoning and Discourse Processes*. London: Academic Press. pp. 67-91.

Allwright, D. (1988) *Observation in the Language Classroom*. London: Longman.

Allwright, D. (1989) How Important are Lessons, Anyway? Lancaster-Leeds Language Learning in Large Classes Research Project Report no. 12.

Allwright, D. and Bailey, K. M. (1991) *Focus on the Language Classroom: An Introduction to Classroom Research for Language Teachers*. Cambridge: Cambridge University Press.

Allwright, R. L. (1980) Turns, topics and tasks: Patterns of participation in language learning and teaching. In Larsen-Freeman, D. (ed.) (1980) *Discourse Analysis in Second Language Research*. Rowley, MA: Newbury House. pp. 165-187.

Alpert, B. R. (1987) Active, silent and controlled discussions: Explaining variations in classroom conversation. *Teaching & Teacher Education* vol. 3 no. 1 pp. 29-40.

Anderson, J. (1980) *Cognitive Psychology and its Implications*. San Francisco: Freeman.

Anderson, L. W. (1989) *The Effective Teacher: Study Guide and Readings*. New York: McGraw-Hill.

Anderson, R. C. and Pearson, P. D. (1988) A schema-theoretic view of basic processes in reading comprehension. In Carrell, P., Devine, J. and Eskey, D. (eds.) (1988) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 37-55.

Ans, B., Carbonnel, S. and Valdois, S. (1998) A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review* vol. 105 pp. 678-723.

Arends, R. I. (1989) *Learning to Teach*, international edition. Singapore: McGraw-Hill.

Aston, G. and Burnard, L. (1998) *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Atkinson, J. M. and Heritage, J. (eds.) (1984) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Austin, J. L. (1962/1976) *How to Do Things with Words*. Oxford: Oxford University Press. Republished by Oxford University Press, 1976, edited by Urmson, J. O. and Sbisa, M.

Ausubel, D. P. (1963) *The Psychology of Meaningful Verbal Learning*. New York: Grune and Stratton.

Barnbrook, G. (1996) *Language and Computers: A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.

Bell, R. T. (1991) *Translation and Translating: Theory and Practice*. London: Longman.

Berber Sardinha, A. P. (1997) *Automatic Identification of Segments in Written Texts*. Unpublished PhD thesis, University of Liverpool.

Berger, C. R. and Bradac, J. J. (1982) *Language and Social Knowledge: Uncertainty in Interpersonal Relations*. London: Edward Arnold.

Bergmann, J. R. (1990) On the local sensitivity of conversation. In Markova, I. and Foppa, K. (eds.) (1990) *The Dynamics of Dialogue*. London: Harvester Wheatsheaf. pp. 201-226.

Berry, M. (1989) Thematic options and success in writing. In Butler, C., Cauldwell, R. and Channell, J. (eds.) (1989) *Language and Literature - Theory and Practice: A Tribute to Walter Grauberg*. Nottingham: University of Nottingham.

Best, J. B. (1999) *Cognitive Psychology*, fifth edition. Belmont, CA: Wadsworth.

Bhatia, V. K. (1993) *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Biao, Z. (1996a) Misconceptions: Clarifying the concept of control. *English Teaching Forum* vol. 34 no. 3-4 pp. 4-11, 29.

Biao, Z. (1996b) On 'control' in the second language teaching classroom. *The Teacher Trainer* vol. 10 no. 2 pp. 8-10.

Biber, D. (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Pearson.

Boole, G. (1854/1916) *Collected Logical Works Volume II: Laws of Thought*. London: The Open Court Publishing Company.

Brazil, D. (1985) Phonology: Intonation in discourse. In van Dijk, T. A. (ed.) (1980) *Handbook of Discourse Analysis, Volume 2*. New York: Academic Press.

Brazil, D. (1995) *A Grammar of Speech*. Oxford: Oxford University Press.

Brazil, D. C., Coulthard, M. R. and Johns, C. (1980) *Discourse Intonation and Language Teaching*. London: Longman.

British National Corpus (2000) *British National Corpus: World Edition* (computer software). Oxford: Oxford University Press.

Brophy, J. and Good, T. (1986) Teacher behavior and student achievement. In Wittrock, M. (ed.) (1986) *Handbook of Research on Teaching*, third edition. New York: Macmillan. pp. 328-375.

Brown, G. and Wragg, E. C. (1993) *Questioning*. London: Routledge.

Brown, G. and Yule, G. (1983a) *Discourse Analysis*. Cambridge: Cambridge University Press.

Brown, G. and Yule, G. (1983b) *Teaching the Spoken Language: An Approach Based on the Analysis of Conversational English*. Cambridge: Cambridge University Press.

Brown, G. A. and Edmonson, R. (1984) Asking questions. In Wragg, E. C. (ed.) (1984) *Classroom Teaching Skills*. London: Croom Helm. pp. 97-120.

Brown, J. D. (2001) *Using Surveys in Language Programs*. Cambridge: Cambridge University Press.

Brown, S. and McIntyre, D. (1993) *Making Sense of Teaching*. Buckingham: Open University Press.

Brumfit, C. (1994) Understanding, language, and educational processes. In Brown, G., Malmkjaer, K., Pollitt, A. and Williams, J. (eds.) (1994) *Language and Understanding*. Oxford: Oxford University Press. pp. 22-33.

Brumfit, C. (1995) Teacher professionalism and research. In Cook, G. and Seidlhofer, B. (eds.) (1995) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*. Oxford: Oxford University Press. pp. 27-41.

Bruner, J. (1960) *The Process of Education*. Cambridge, MA: Harvard University Press.

Burgess, J. (1994) Ideational frameworks in integrated language learning. *System* vol. 22 no. 3 pp. 309-318.

Bygate, M. (1987) *Speaking*. Oxford: Oxford University Press.

Cambell, S. and Shaweevongs, C. (n.d.) *The Fundamentals of the Thai Language*. Bangkok: Marketing Media Associates.

Carlson, L. (1983) *Dialogue Games: An Approach to Discourse Analysis*. Dordrecht, Holland: D. Reidel.

Caron-Pargue, J. and Caron, J. (1991) Psychopragmatics vs. sociopragmatics: The function of pragmatic markers in thinking-aloud protocols. In Verschueren, J. (ed.) (1991) *Pragmatics at Issue: Selected Papers of the International Pragmatics Conference*, Antwerp, August 17-22, 1987, Volume 1. Amsterdam: John Benjamins. pp.29-36.

Carrell, P. L. and Eisterhold, J. C. (1988) Schema theory and ESL reading pedagogy. In Carrell, P., Devine, J. and Eskey, D. (eds.) (1988) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 73-92.

Carter, R. (1987) *Vocabulary: Applied Linguistics Perspectives*. London: Allen & Unwin.

Carter, R. and McCarthy, M. (1988) *Vocabulary and Language Teaching*. London: Longman.

Cazden, C. B. (1986) Language in the classroom. *Annual Review of Applied Linguistics* vol. 7 pp. 18-31.

Chafe, W. L. (1976) Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. N. (ed.) (1976) *Subject and Topic*. New York: Academic Press. pp. 25-55.

Chafe, W. L. (1980) The deployment of consciousness in the reproduction of a narrative. In Chafe (1980) *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex. pp. 9-50.

Chaiyasuk, I. (2002) *Investigating the Functional Structures of Teacher Monologues*. Unpublished MA thesis. Bangkok: King Mongkut's University of Technology Thonburi.

Chaudron, C. (1983) Foreigner talk in the classroom - an aid to learning? In Seliger, H. W. and Long, M. H. (eds.) (1983) *Classroom Oriented research in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 127-143.

Chaudron, C. (1988) *Second Language Classrooms: Research on Teaching and Learning*. Cambridge: Cambridge University Press.

Christiansen, M. H., Allen, J. and Seidenberg, M. S. (1998) Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes* vol. 13 pp. 221-268.

Cicourel, A. V. (1991) Semantics, pragmatics, and situated meaning. In Verschueren, J. (ed.) (1991) *Pragmatics at Issue: Selected Papers of the International Pragmatics Conference*, Antwerp, August 17-22, 1987, Volume 1. Amsterdam: John Benjamins. pp. 37-66.

Clancy, P. M. (1980) Referential choice in English and Japanese narrative discourse. In Chafe (1980) *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex. pp. 127-202.

Clark, H. H. (1996) *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. and Haviland, S. E. (1977) Comprehension and the given-new contract. In Freedle, R. O. (ed.) (1977) *Discourse Production and Comprehension: Discourse Processes: Advances in Research and Theory Volume 1*. Norwood, NJ: Ablex. pp. 1-40.

Cole, P. G. and Chan, L. K. S. (1987) *Teaching Principles and Practice*. Sydney: Prentice Hall.

Collins, A. M. and Loftus, E. F. (1975) A spreading activation theory of semantic processing. *Psychological Review* vol. 82 pp. 407-428.

Collins, A. M. and Quillian, M. R. (1969) Retrieval time from semantic memory. In Komatsu, L. K. (ed.) (1994) *Experimenting with the Mind: Readings in Cognitive Psychology*. Pacific Grove, CA: Brooks/Cole. pp. 185-192.

Connor, U. (1990) Discourse analysis and writing/reading instruction. *Annual Review of Applied Linguistics* vol. 11 pp. 164-180.

Connor, U. (1994) Text analysis. *TESOL Quarterly* vol. 28 no. 4 pp. 682-684.

Connor, U. (1996) *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing*. Cambridge: Cambridge University Press.

Connor, U. and Farmer, M. (1990) The teaching of topical structure analysis as a revision strategy for ESL writers. In Kroll, B. (ed.) (1990) *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press. pp. 126-139.

Conrad, C. (1972) Cognitive economy in semantic memory. *Journal of Experimental Psychology* vol. 92 pp. 149-154.

Cook, G. (1989) *Discourse*. Oxford: Oxford University Press.

Cook, G. (1997) Schemas. *ELT Journal* vol. 51 no. 1 p. 86.

Cortazzi, M. and Jin, L. (1996) Cultures of learning: language classrooms in China. In Coleman, H. (ed.) (1996) *Society and the Language Classroom*. Cambridge: Cambridge University Press. pp. 169-206.

Corter, J. E. (1986) *EXTREE: A Program for Fitting the Extended Tree Model*. Computer software program available at: http://www.columbia.edu/~jec34/extree.exe.

Corter, J. E. (1996) *Tree Models of Similarity and Association*. Thousand Oaks, CA: Sage.

Cothran, D. J. and Ennis, C. D. (1997) Students' and teachers' perceptions of conflict and power. *Teaching & Teacher Education* vol. 13 no. 5 pp. 541-553.

Coulthard, M. (1977) *An Introduction to Discourse Analysis*. London: Longman.

Coulthard, M. (1981) Developing the description. In Coulthard, M. and Montgomery, M. (eds.) (1981) *Studies in Discourse Analysis*. London: Routledge & Kegan Paul. pp. 13-31.

Coulthard, M. (1987) Intonation in discourse. In Coulthard, M. (ed.) (1987) *Discussing Discourse. Discourse Analysis Monograph 14*, University of Birmingham. pp. 44-62.

Coulthard, M. and Brazil, D. (1992) Exchange structure. In Coulthard, M. (ed.) (1992) *Advances in Spoken Discourse Analysis*. London: Routledge. pp. 50-78.

Coulthard, M. and Montgomery, M. (1981) Originating a description. In Coulthard, M. and Montgomery, M. (eds.) (1981) *Studies in Discourse Analysis*. London: Routledge & Kegan Paul. pp. 1-12.

Covelli, L. H. and Murray, S. O. (1980) Accomplishing topic change. *Anthropological Linguistics* vol. 22 no. 9 pp. 382-389.

Crombie, W. (1985a) *Discourse and Language Learning: A Relational Approach to Syllabus Design*. Oxford: Oxford University Press.

Crombie, W. (1985b) *Process and Relation in Discourse and Language Learning*. Oxford: Oxford University Press.

Crookes, G. and Rulon, K. A. (1988) Topic and feedback in native speaker/nonnative-speaker conversation. *TESOL Quarterly* vol. 22 no. 4 pp. 675-681.

Cross, D. (1991) *A Practical Handbook of Language Teaching*. London: Cassell.

Crow, B. K. (1983) Topic shifts in couples' conversations. In Craig, R. T. and Tracy, K. (eds.) (1983) *Conversational Coherence: Form, Structure and Strategy*. Beverley Hills, CA: Cage Publications. pp. 136-156.

Cruickshank, D. R. (1985) Applying research on teacher clarity. In Anderson, L. W. (1989) *The Effective Teacher: Study Guide and Readings*. New York: McGraw-Hill. pp. 285-291.

Cruickshank, D. R. and Kennedy, J. J. (1986) Teacher clarity. *Teaching & Teacher Education* vol. 2 no. 1 pp. 43-67.

Cruse, D. A. (1986) *Lexical Semantics*. Cambridge: Cambridge University Press.

Cullen, R. (1998) Teacher talk and the classroom context. *ELT Journal* vol. 52 no. 3 pp. 179-187.

Cumming, S. and Ono, T. (1997) Discourse and grammar. In van Dijk, T. A. (ed.) (1997) *Discourse as Structure and Process: Discourse Studies: A Multidisciplinary Introduction Volume 1*. London: Sage. pp. 112-137.

Daneš, F. (1974) Functional sentence perspective and the organization of the text. In Daneš, F. (ed.) (1974) *Papers on Functional Sentence Perspective*. Prague/The Hague: Academia/Mouton. pp. 106-128.

Davies, F. (1994) From writer roles to elements of text: interactive, organisational and topical. In Barbara, L. and Scott, M. (eds.) (1994) *Reflections on Language Learning*. Clevedon, Avon: Multilingual Matters. pp. 170-183.

De Beaugrande, R. and Dressler, W. (1981) *Introduction to Text Linguistics*. London: Longman.

Devlin, K. (1997) *Goodbye, Descartes: The End of Logic and the Search for a new Cosmology of the Mind*. New York: John Wiley.

Duncan, S. (1983) Rules and strategy in conversation. In Reis, H. T. (ed.) (1983) *Naturalistic Approaches to Studying Social Interaction: New Directions for Methodology of Social and Behavioral Science Number 15*. San Francisco: Jossey-Bass.

Dunkel, P. A. and Davis, J. N. (1994) The effects of rhetorical signaling cues on the recall of English lecture information by speakers of English as a native or second language. In Flowerdew, J. (ed.) (1994) *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press. pp. 55-74.

Duranti, A. (1991) Four properties of speech-in-interaction and the notion of translocutionary act. In Verschueren, J. (ed.) (1991) *Pragmatics at Issue: Selected Papers of the International Pragmatics Conference*, Antwerp, August 17-22, 1987, Volume 1. Amsterdam: John Benjamins. pp. 133-150.

Duranti, A. (1997) *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Eggen, P. D. and Kauchak, D. (1992) *Educational Psychology: Classroom Connections*. New York: Macmillan.

Ellis, N. C. and Schmidt, R. (1998) Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes* vol. 13 pp. 307-336.

Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Enkvist, N. E. (1990) Seven problems in the study of coherence and interpretability. In Connor, U. and Johns, A. M. (eds.) (1990) *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 9-27.

Erickson, F. (1982) Classroom discourse as improvisation: relationships between academic task structure and social participation structure in lessons. In Anderson, L. W. (ed.) (1989) *The Effective Teacher: Study Guide and Readings*. New York: McGraw-Hill. pp. 165-174.

Erickson, F. (1996) Ethnographic microanalysis. In McKay, S. L. and Hornberger, N. H. (1996) *Sociolinguistics and Language Teaching*. Cambridge: Cambridge University Press. pp. 283-306.

Ericsson, K. A. and Simon, H. A. (1993) *Protocol Analysis: Verbal Reports as Data*, revised edition. Cambridge, MA: MIT Press.

Evensen, L. S. (1990) Pointers to superstructure in student writing. In Connor, U. and Johns, A. M. (eds.) (1990) *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 169-182.

Fernando, C. (1996) *Idioms and Idiomaticity*. Oxford: Oxford University Press.

Finocchiaro, M. (1989) *English as a Second/Foreign Language: From Theory to Practice*, fourth edition. Englewood Cliffs, NJ: Prentice Hall.

Firbas, J. (1987) On two starting points of communication. In Steele, R. and Threadgold, T. (eds.) (1987) *Language Topics: Essays in Honour of Michael Halliday Volume 1*. Amsterdam: John Benjamins. pp. 23-46.

Fletcher, L. (1985) *How to Design & Deliver a Speech*, third edition. New York: Harper & Row.

Flowerdew, J. (1994) Research of relevance to second language lecture comprehension - an overview. In Flowerdew, J. (ed.) (1994) *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press. pp. 7-29.

Fodor, J. A. and Pylyshyn, Z. W. (1988) Connectionist and cognitive architecture: a critical analysis. *Cognition* vol. 29 no. 1 pp. 3-71.

Fodor, J. D. (1970) Formal linguistics and formal logic. In Lyons, J. (ed.) (1970) *New Horizons in Linguistics*. Harmondsworth: Penguin. pp. 198-214.

Foppa, K. (1990) Topic progression and intention. In Markova, I. and Foppa, K. (eds.) (1990) *The Dynamics of Dialogue*. London: Harvester Wheatsheaf. pp. 178-200.

Francis, G. and Hunston, S. (1987) Analysing everyday conversation. In Coulthard, M. (ed.) (1987) *Discussing Discourse. Discourse Analysis Monograph 14*, University of Birmingham. pp. 107-148.

Freeman, D. (1998) *Doing Teacher Research: From Inquiry to Understanding*. Pacific Grove, CA: Heinle & Heinle.

Fries, P. H. (1983) On the status of theme in English: arguments from discourse. In Petöfi, J. S. and Sözer, E. (eds.) (1983) *Micro and Macro Connexity of Texts*. Hamburg: Helmut Buske Verlag. pp. 116-152.

Fries, P. H. (1994) On theme, rheme and discourse goals. In Coulthard, M. (ed.) (1994) *Advances in Written Text Analysis*. London: Routledge. pp. 229-249.

Froyen, L. A. (1993) *Classroom Management: The Reflective Teacher-Leader*, second edition. New York: Macmillan.

Gagné, R. M. (1985) *The Conditions of Learning and Theory of Instruction*, fourth edition. New York: Holt, Rinehart and Winston.

Gardner, R. (1987) The identification and role of topic in spoken interaction. *Semiotica* vol. 65 no. 1/2 pp. 129-141.

Gianelli, M. C. (1997) Thematic units: Creating an environment for learning. In Snow, M. A. and Brinton, D. M. (eds.) (1997) *The Content-Based Classroom: Perspectives on Integrating Language and Content*. New York: Longman. pp. 142-148.

Gibbons, J. (1994) Cycle and activity structures in second language instruction. *System* vol. 22 no. 3 pp. 319-339.

Givón, T. (ed.) (1983) *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.

Givón, T. (1995) Coherence in text vs. coherence in mind. In Gernsbacher, M. A. and Givón, T. (eds.) (1995) *Coherence in Spontaneous Text. Typological Studies in Language 31*. Amsterdam: John Benjamins. pp. 59-116.

Gleason, J. B. and Ratner, N. B. (1998) Language acquisition. In Gleason, J. B. and Ratner, N. B. (eds.) (1998) *Psycholinguistics*, second edition. Fort Worth, TX: Harcourt Brace. pp. 347-408.

Gloeckner, G. (1983) *An Investigation into the Effectiveness of a Preservice Teacher Clarity Training Unit in two Different Experimental Settings*. Unpublished doctoral dissertation, The Ohio State University.

Goetz, E. T. and Armbruster, B. B. (1980) Psychological correlates of text structure. In Spiro, R. J., Bruce, B. C. and Brewer, W. F. (eds.) (1980) *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education*. Hillsdale, NJ: Lawrence Erlbaum. pp. 201-220.

Goldberg, J. A. (1983) A move toward describing conversational coherence. In Craig, R. T. and Tracy, K. (1983) (eds.) *Conversational Coherence: Form, Structure and Strategy*. Beverly Hills, CA: Sage. pp. 25-46.

Grabe, W. (1984) Written discourse analysis. *Annual Review of Applied Linguistics* vol. 5 pp. 101-123.

Graham, S. (1997) *Effective Language Learning*. Clevedon, Avon: Multilingual Matters.

Graney, J. M. (1992) A framework for using text graphing. *System* vol. 20 no. 2 pp. 161-167.

Grice, H. P. (1975) Logic and conversation. In Fogelin, R. J. (1978) *Understanding Arguments: An Introduction to Informal Logic*. New York: Harcourt Brace Jovanovich. pp. 329-343.

Gumperz, J. J., Aulakh, G. and Kaltman, H. (1982) Thematic structure and progression in discourse. In Gumperz, J. J. (ed.) (1982)

Language and Social Identity. Cambridge: Cambridge University Press. pp. 22-56.

Halleck, G. B. (1995) Assessing oral proficiency: a comparison of holistic and objective methods. *The Modern Language Journal* vol. 79 no. 2 pp. 223-234.

Halliday, M. A. K. (1967) Notes on transitivity and theme in English. *Journal of Linguistics* vol. 3 no. 1 pp. 37-81, no. 2 pp. 199-244.

Halliday, M. A. K. (1970) Language structure and language function. In Lyons, J. (ed.) (1970) *New Horizons in Linguistics*. Harmondsworth: Penguin. pp. 140-165.

Halliday, M. A. K. (1973) *Explorations in the Functions of Language*. London: Edward Arnold.

Halliday, M. A. K. (1989) *Spoken and Written Language*, second edition. Oxford: Oxford University Press.

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London: Longman.

Halliday, M. A. K. and Hasan, R. (1989) *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*, 2nd edition. Oxford: Oxford University Press.

Harmer, J. (1983) *The Practice of English Language Teaching*. London: Longman.

Hasan, R. (1989) *Linguistics, Language and Verbal Art*, second edition. Oxford: Oxford University Press.

Hatch, E. (1992) *Discourse and Language Education*. Cambridge: Cambridge University Press.

Hatch, E. and Brown, C. (1995) *Vocabulary, Semantics and Language Education*. Cambridge: Cambridge University Press.

Hatch, E. and Long, M. H. (1980) Discourse analysis, what's that? In Larsen-Freeman, D. (ed.) (1980) *Discourse Analysis in Second Language Research*. Rowley, MA: Newbury House.

Haviland, S. E. and Clark, H. H. (1974) What's new? Acquiring new information as a process in comprehension. In Komatsu, L. K. (ed.) (1994) *Experimenting with the Mind: Readings in Cognitive Psychology*. Pacific Grove, CA: Brooks/Cole. pp. 363-369.

Hazadiah, M. D. (1993) Topic as a dynamic element in spoken discourse. In Baker, M., Francis, G. and Tagnini-Bonelli, E. (eds.) (1993) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins. pp. 55-74.

Hearst, M. A. (1994) Multi-paragraph segmentation of expository text. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*. Available at: http://www.csi.uottawa.ca/tanka/ArtDB/bibliography.html.

Hearst, M. (1994) Multi-paragraph segmentation of expository text. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*. Available at: http://www.csi.uottawa.ca/tanka/ArtDB/bibliography.html.

Hearst, M. (1997) *Introduction to Content Analysis*. Available at: http://www.sims.berkeley.edu/academics/courses/is202/f97/Lecture14/Lecture14.html.

Hemphill, L. (1989) Topic development, syntax and social class. *Discourse Processes* vol. 12 no. 3 pp. 267-286.

Hoey, M. (1983) *On the Surface of Discourse*. London: George Allen & Unwin.

Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. (1993) The case for the exchange complex. In Hoey, M. (ed.) (1993) *Data, Description, Discourse: Papers on the English Language in Honour of John McH Sinclair on his Sixtieth Birthday*. London: HarperCollins. pp. 115-138.

Hoey, M. (1995) The lexical nature of intertextuality: a preliminary study. In Wårvik, B., Tanskanen, S.-K. and Hiltunen, R. (eds.) (1995) *Organization in Discourse: Proceedings from the Turku Conference, Anglicana Turkensia No. 14*. Turku: University of Turku. pp. 73-94.

Hofstadter, D. R. (1996) Analogy-making, fluid concepts, and brain mechanisms. In Clark, A. and Millican, P. J. R. (eds.) (1976) *Connectionism, Concepts and Folk Psychology: The Legacy of Alan Turing Vol. 2*. Oxford: Clarendon Press. pp. 195-248.

Holec, H. (1985) You did say 'oral interactive discourse'? In Riley, P. (ed.) *Discourse and Learning*. London: Longman. pp. 21-34.

Hovy, E. and Lin, C. Y. (1997) Automated text summarization in summarist. In *ACL/EACL Workshop on Intelligent Scalable Text Summarization*. Ottawa: ACL. pp. 18-24.

Hudak, G. (1987) Student knowledge and the formation of academic discourse: A case study. In Smyth, J. (ed.) (1987) *Educating Teachers: Changing the Nature of Pedagogic Knowledge*. London: The Falmer Press. pp. 55-72.

Hudson, R. A. (1980) *Sociolinguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunt, E. (1989) Connectionist and rule-based representations of expert knowledge. *Behavior Research Methods, Instruments and Computers* vol. 21 no. 1 pp. 88-95.

Hurtig, R. (1977) Toward a functional theory of discourse. In Freedle, R. O. (ed.) (1977) *Discourse Production and Comprehension*. Norwood, NJ: Ablex. pp. 89-106.

Hutchinson, T. and Waters, A. (1984) *Interface: English for Technical Communication*. London: Longman.

Jantz, R. (1989) Teaching concepts for higher-level thinking. In Arends, R. I. (1989) *Learning to Teach*, international edition. Singapore: McGraw-Hill. pp. 318-360.

Jay, S. S. (1968) The effect of word associations on the recognition of flashed words. Paper presented at the American Educational Research Association Conference, Chicago, 6th - 10th February 1968.

Jefferson, G. (1984) On stepwise transition from talk about a trouble to inappropriately next-positioned matters. In Atkinson, J. M. and Heritage, J. (eds.) (1984) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press. pp. 191-222.

Johnson, K. E. (1995) *Understanding Communication in Second Language Classrooms*. Cambridge: Cambridge University Press.

Johnson, R. K. (1990) Developing teachers' language resources. In Richards, J. C. and Nunan, D. (eds.) (1990) *Second Language Teacher Education*. Cambridge: Cambridge University Press. pp. 269-281.

Johnson-Laird, P. N. (1986) Reasoning without logic. In Myers, T., Brown, K. and McGonigle, B. (eds.) (1986) *Reasoning and Discourse Processes*. London: Academic Press. pp. 13-49.

Jonz, J. (1989) Textual sequence and second-language comprehension. *Language Learning* vol. 39 no. 2 pp. 207-249.

Keenan, E. O. and Schieffelin, B. B. (1976) Topic as a discourse notion: A study of topic in the conversation of children and adults. In Li, C. N. (ed.) (1976) *Subject and Topic*. New York: Academic Press. pp. 335-384.

Kennedy, J. (1996) Classroom explanatory discourse: A look at how teachers explain things to their students. *Language Awareness* vol. 5 no. 1 pp. 26-39.

Kerry, T. (1982) *Effective Questioning*. London: Macmillan

Kitao, S. K. (1990) Textual schemata and English language learning. *Cross Currents* vol. 17 no. 2 pp. 147-154.

Klecan-Aker, J. S. and Lopez, B. (1985) A comparison of T-units and cohesive ties used by first and third grade children. *Language and Speech* vol. 28 no. 3 pp. 307-315.

Kozima, H. (1993) Text segmentation based on similarity between words. In *Proceeding of the 31st Annual Meeting of the Association for Computational Linguistics*. Ottawa: ACL. pp. 286-288.

Kress, G. (1989) *Linguistic Processes in Sociocultural Practice*, second edition. Oxford: Oxford University Press.

Langacker, R. W. (1996) Conceptual grouping and pronominal anaphora. In Fox, B. (ed.) (1996) *Studies in Anaphora*. Amsterdam: John Benjamins. pp. 333-378.

Lappin, S. (ed.) (1996) *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Larsen-Freeman, D. (1983) Assessing global second language proficiency. In Seliger, H. and Long, M. (eds.) (1983) *Classroom Oriented Research in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 287-304.

Lautamatti, L. (1978) Observations on the development of the topic in simplified discourse. In Connor, U. and Kaplan, R. B. (eds.) (1987) *Writing across Languages: Analysis of L2 Text*. Rowley, MA: Addison-Wesley. pp. 87-113.

Lautamatti, L. (1990) Coherence in spoken and written discourse. In Connor, U. and Johns, A. M. (eds.) (1990) *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 29-39.

Leech, G. (1981) *Semantics: The Study of Meaning*, second edition. London: Penguin.

Lemke, J. L. (1989) *Using Language in the Classroom*, second edition. Oxford: Oxford University Press.

Levinson, S. C. (1983) *Pragmatics*. Cambridge: Cambridge University Press.

Lewis, M. (1993) *The Lexical Approach: The State of ELT and a Way Forward*. Hove: Language Teaching Publications.

Lipschutz, S. (1964) *Schaum's Outline of Theory and Problems of Set Theory and Related Topics*. Singapore: McGraw-Hill.

Long, D. R. (1989) Second language listening comprehension: A schema-theoretic perspective. *The Modern Language Journal* vol. 73 no. 1 pp. 32-40.

Long, M. H. (1983) Inside the "black box": Methodological issues in classroom research on language learning. In Seliger, H. W. and Long, M. H. (eds.) (1983) *Classroom Oriented Research in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 3-35.

Lönngren, L. (1989) How can the meaning of a text be represented? In Conte, M.-E., Petöfi, J. S. and Sözer, E. (eds.) (1989) *Text and Discourse Connectedness: Proceedings of the Conference on Connexity and Coherence, Urbino, July 16-21, 1984*. Amsterdam: John Benjamins. pp. 93-102.

Lyons, J. (1977) *Semantics Volume 1, Volume 2*. Cambridge: Cambridge University Press.

Mäkinen, K. (1992) Topical depth and writing quality in student EFL compositions. *Scandinavian Journal of Educational Research* vol. 36 no. 3 pp. 237-247.

Malamah-Thomas, A. (1987) *Classroom Interaction*. Oxford: Oxford University Press.

Malcolm, I. G. (1991) "All right then, if you don't want to do that ...": Strategy and counter-strategy in classroom discourse management. *Guidelines* vol. 13 no. 2 pp. 1-17.

Malmkjaer, K. (1991) Semantics. In Malmkjaer, K. (ed.) (1991) *The Linguistics Encyclopedia*. London: Routledge. pp.389-398.

Mann, W. C. and Thompson, S. A. (1988) Rhetorical Structure Theory: toward a functional theory of text organization. *Text* vol. 8 no. 3 pp. 243-281.

Marland, P and Osborne, B. (1990) Classroom theory, thinking and action. *Teaching & Teacher Education* vol. 6 no. 1 pp. 93-109.

Mason, A. (1994) By dint of: Student and lecturer perceptions of lecture comprehension strategies in first-term graduate study. In Flowerdew, J. (ed.) (1994) *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press. pp. 199-218.

Matsumoto, K. (1993) Verbal-report data and introspective methods in second language research: state of the art. *RELC Journal* vol. 24 no. 1 pp. 32-60.

Maynard, D. W. (1980) Placement of topic changes in conversation. *Semiotica* vol. 30 no. 3/4 pp. 263-290.

McCarthy, M. (1988) Some vocabulary patterns in conversation. In Carter, R. and McCarthy, M. (eds.) *Vocabulary and Language Teaching*. London: Longman. pp. 181-200.

McCarthy, M. (1991) *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.

McCarthy, M. and Carter, R. (1994) *Language as Discourse: Perspectives for Language Teaching*. London: Longman.

McClelland, J. L. and Rumelhart, D. E. (1981) An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychological Review* vol. 88 pp. 375-407.

McClelland, J. L. and Rumelhart, D. E. (1985) Distributed memory and the representation of general and specific information. In Komatsu, L. K. (ed.) (1994) *Experimenting with the Mind: Readings in Cognitive Psychology*. Pacific Grove, CA: Brooks/Cole. pp. 278-303.

McDonough, J. and McDonough, S. (1997) *Research Methods for English Language Teachers*. London: Arnold.

McDonough, S. H. (1981) *Psychology in Foreign Language Teaching*. London: George Allen & Unwin.

McEneaney, J. E. (1994) Does Parallel Distributed Processing provide a plausible framework for modeling early reading acquisition? Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, 4th - 8th April 1994.

McWhinney, B., Leinbach, J., Taraban, R. and McDonald, J. (1989) Language learning: cues or rules? *Journal of Memory and Language* vol. 28 pp. 255-277.

Mehan, H. (1979) *Learning Lessons: Social Organization in the Classroom*. Cambridge, MA: Harvard University Press.

Mehan, H. (1985) The structure of classroom discourse. In van Dijk, T. A. (ed.) (1985) *Handbook of Discourse Analysis Volume 3: Discourse and Dialogue*. London: Academic Press. pp. 120-131.

Minsky, M. (1985) *The Society of Mind*. New York: Simon & Schuster.

Mohan, B. A. (1986) *Language and Content*. Reading, MA: Addison-Wesley.

Moita Lopes, L. P. da (1995) What is this class about? Topic formulation in an L1 reading comprehension classroom. In Cook, G. and Seidlhofer, B. (eds.) (1995) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*. Oxford: Oxford University Press. pp. 349-362.

Moore, K. D. (1989) *Classroom Teaching Skills: A Primer*. New York: Random House.

Nassaji, H. and Wells, G. (2000) What's the use of 'triadic dialogue'? An investigation of teacher-student interaction. *Applied Linguistics* vol. 21 no. 3 pp. 376-406.

Nattinger, J. R. and DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Ney, J. W. and Pearson, B. A. (1990) Connectionism as a model of language learning: parallels in foreign language teaching. *The Modern Language Journal* vol. 74 no. 4 pp. 474-482.

Nunan, D. (1990) The language teacher as decision maker: a case study. In Brindley, G. (ed.) (1990) *The Second Language Curriculum in Action*. Sydney: National Centre for English Language Teaching and Research. pp. 9-35.

Nunan, D. and Lamb, C. (1996) *The Self-Directed Teacher: Managing the Learning Process*. Cambridge: Cambridge University Press.

Nwogu, K. N. (1991) Structure of science popularizations: A genre-analysis approach to the schema of popularized medical texts. *English for Specific Purposes* vol.10 no. 2 pp. 111-123.

Omron Corporation (1991) *Clearly Fuzzy*. Kyoto, Japan: Omron Corporation.

Ordóñez, F. and Treviño, E. (1999) Left dislocated subjects and the pro-drop parameter: a case study of Spanish. *Lingua* vol. 107 nos. 1-2 pp. 39-68.

Orlich, D. C., Harder, R. J., Callahan, R. C. and Gibson, H. W. (1998) *Teaching Strategies: A Guide to Better Instruction*, fifth edition. Boston: Houghton Mifflin.

Palermo, D. S. (1965) *Word Associations as Related to Children's Verbal Habits*. ERIC Report no. UM-CHGD-8.

Palmer, F. R. (1981) *Semantics*, second edition. Cambridge: Cambridge University Press.

Paltridge, B. (1996) Genre, text type, and the language learning classroom. *ELT Journal* vol. 50 no. 3 pp. 237-243.

Peirce, B. N. (1992) Demystifying the TOEFL® reading test. *TESOL Quarterly* vol. 26 no. 4 pp. 665-691.

Peng, W. (1999) *Lexical Patterning, Key Words, and the Theme-Rheme System*. Unpublished PhD thesis, University of Liverpool.

Perrott, E. (1982) *Effective Teaching: A Practical Guide to Improving Your Teaching*. London: Longman.

Pica, T. and Long, M. H. (1986) The linguistic and conversational performance of experienced and inexperienced teachers. In Day, R. R. (ed.) (1986) *Talking to Learn: Conversation in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 85-98.

Polanyi, L. (1993) Linguistic dimension of text summarization. In *Working Notes of the Dagstuhl Seminar on Summarizing Text for Intelligent Communication*. Available at: http://www.csi.uottawa.ca/tanka/ArtDB/bibliography.html.

Prentice, J. L. (1967) *Intraverbal Associations in Sentence Behavior*. ERIC Report no. BR-6-1784.

Prince, E. F. (1981) Toward a taxonomy of given-new information. In Cole, P. (ed.) (1981) *Radical Pragmatics*. New York: Academic press. pp. 223-255.

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

Read, J. (2000) *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Redeker, G. (1990) Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* vol. 14 no. 3 pp. 367-381.

Reisberg, D. (1997) *Cognition: Exploring the Science of Mind*. New York: W. W. Norton.

Richards, J. C. and Lockhart, C. (1994) *Reflective Teaching in Second Language Classrooms*. Cambridge: Cambridge University Press.

Richards, J. C. and Schmidt, R. W. (1983) Conversational analysis. In Richards, J. C. and Schmidt, R. W. (eds.) (1983) *Language and Communication*. London: Longman. pp. 117-154.

Rosenshine, B. and Furst, N. (1971) Research on teacher performance criteria. In Smith, B. O. (ed.) *Research in Teacher Education*. Englewood Cliffs, NJ: Prentice Hall. pp. 37-72.

Rosenshine, B. and Stevens, R. (1986) Teaching functions. In Wittrock, M. (ed.) (1986) *Handbook of Research on Teaching*, third edition. New York: Macmillan. pp. 376-391.

Rost, M. (1994) *Introducing Listening*. London: Penguin.

Roth, W.-M. (1996) Teacher questioning in an open-inquiry learning environment: interactions of context, content and student responses. *Journal of Research in Science Teaching* vol. 33 no. 7 pp. 709-736.

Rutherford, W. E. (1987) *Second Language Grammar: Learning and Teaching*. London: Longman.

Sacks, H. (1967-1971) Mimeo lecture notes.

Sanders, T. J. M., Spooren, W. P. M. and Noordman, L. G. M. (1992) Toward a taxonomy of coherence relations. *Discourse Processes* vol. 15 no. 1 pp. 1-35.

Schank, R. C. and Abelson, R. P. (1977) *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schegloff, E. A. (1971) Notes on a conversational practice: Formulating place. In Giglioli, P. P. (ed.) (1972) *Language and Social Context: Selected Readings*. London: Penguin. pp. 95-135.

Schiffrin, D. (1987) *Discourse Markers*. Cambridge: Cambridge University Press.

Schmitt, N. (1998) Quantifying word association responses: what is native-like? *System* vol. 26 no. 3 pp. 389-401.

Schmitt, N. (2000) *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schneider, M. and Connor, U. (1990) Analysing topical structure in ESL essays: Not all topics are equal. *Studies in Second Language Acquisition* vol. 12 no. 4 pp. 411-427.

Scott, M. (1997) PC analysis of key words - and key key words. *System* vol. 25 no. 2 pp. 233-245.

Scott, M. (2000) Focusing on the text and its key words. In Burnard, L. and McEnery, T. (eds.) (2000) *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Amsterdam: Peter Lang. pp. 103-121.

Seidenberg, M. S. (1992) Connectionism without tears. In Davis, S. (ed.) (1992) *Connectionism: Theory and Practice*. New York: Oxford University Press. pp. 84-122.

Shakir, A. (1991) Coherence in EFL student-written texts: Two perspectives. *Foreign Language Annals* vol. 24 no. 5 pp. 399-411.

Shapiro, L. and Hestvik, A. (1995) On-line comprehension of VP-ellipsis: syntactic reconstruction and semantic influence. *Journal of Psycholinguistics Research* vol. 24 no. 6 pp. 517-532.

Shei, C. (1999) *Extracting Collocations*. Available at: http://www.dai.ed.ac.uk/homes/shei/collocation/node4.html.

Shepherd, J. D. (1998) *Storytelling in Conversational Discourse: A Collaborative Model*. Unpublished PhD thesis. University of Birmingham.

Shima, F. (1970) *Research on Word Association in Connected Discourse*. ERIC Report no. TR24.

Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1992) Priorities in discourse analysis. In Coulthard, M. (ed.) (1992) *Advances in Spoken Discourse Analysis*. London: Routledge. pp. 79-88.

Sinclair, J. McH. and Brazil, D. (1982) *Teacher Talk*. Oxford: Oxford University Press.

Sinclair, J. McH. and Coulthard, R. M. (1975) *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Slavin, R. E. (1994) *Educational Psychology: Theory and Practice*, fourth edition. Boston, MA: Allyn and Bacon.

Smith, P. and Holdcraft, D. (1991) Towards a computational theory of speech acts. In Verschueren, J. (ed.) (1991) *Pragmatics at Issue: Selected Papers of the International Pragmatics Conference*, Antwerp, August 17-22, 1987, Volume 1. Amsterdam: John Benjamins. pp. 211-228.

Sotillo, C. (1999) Reading texts: processing discourse. Available at: http:// www.hcrc.ed.ac.uk/~cathy/discourse.htm.

Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stech, E. L. (1982) The analysis of conversational topical sequence structures. *Semiotica* vol. 39 no. 1/2 pp. 75-91.

Stockwell, R. P. (1977) *Foundations of Syntactic Theory*. Englewood Cliffs, NJ: Prentice Hall.

Strahan, D. B. (1989) How experienced and novice teachers frame their views of instruction: An analysis of semantic ordered trees. *Teaching & Teacher Education* vol. 5 no. 1 pp. 53-67.

Strzalkowski, T., Wand, J. and Wise, B. (1998) A robust practical text summarization. In *AAAI 98 Spring Symposium on Intelligent Text Summarization*. Available at: http://www.csi.uottawa.ca/tanka/ArtDB/bibliography.html.

Stubbs, M. (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Blackwell.

Stubbs, M. (1986) *Educational Linguistics*. Oxford: Blackwell.

Swales, J. M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Tannen, D. (1978) The effect of expectations on conversation. *Discourse Processes* vol. 1 no. 2 pp. 203-209.

Thompson, S. (1994) Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes* vol. 13 no. 2 pp. 171-186.

Tomlin, R. S., Forrest, L., Pu, M. M. and Kim, M. H. (1997) Discourse semantics. In van Dijk, T. A. (ed.) (1997) *Discourse as Structure and Process: Discourse Studies: A Multidisciplinary Introduction Volume 1*. London: Sage. pp. 64-111.

Tsui, A. B. M. (1994) *English Conversation*. Oxford: Oxford University Press.

Tyler, A. (1992) Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly* vol. 26 no. 4 pp. 713-729.

Tyler, A. and Bro, J. (1992) Discourse structure in non-native English discourse: The effect of ordering and interpretive cues on perceptions of comprehensibility. *Studies in Second Language Acquisition* vol. 14 no. 1 pp. 71-86.

Van Dijk, T. A. (1977) *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.

Van Dijk, T. A. (1980) *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction and Cognition*. Hillsdale, NJ: Lawrence Erlbaum.

Van Dijk, T. A. (1985) Semantic discourse analysis. In van Dijk, T. A. (ed.) (1985) *Handbook of Discourse Analysis Volume 2: Dimensions of Discourse*. London: Academic Press. pp. 103-136.

Van Dijk, T. A. and Kintsch, W. (1983) *Strategies of Discourse Comprehension*. New York: Academic Press.

Van Lier, L. (1988) *The Classroom and the Language Learner*. London: Longman.

Van Lier, L. (1989) Classroom research in second language acquisition. *Annual Review of Applied Linguistics* Vol. 10 pp. 173-186.

Vavra, E. (2000) *Definitions of the "T-unit"*. Available at: http://curie.pct.edu/courses/evavra/ED498/Essay009_Def_TUnit.htm.

Ventola, E. (1987) *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*. London: Frances Pinter.

Wajnryb, R. (1992) *Classroom Observation Tasks: A Resource Book for Language Teachers and Trainers*. Cambridge: Cambridge University Press.

Wallace, M. J. (1998) *Action Research for Language Teachers*. Cambridge: Cambridge University Press.

Wason, P. C. and Shapiro, D. (1971) Natural and contrived experience in a reasoning problem. In Komatsu, L. K. (ed.) (1994) *Experimenting with the Mind: Readings in Cognitive Psychology*. Pacific Grove, CA: Brooks/Cole. pp. 398-403.

Watson-Gegeo, K. A. (1988) Ethnography in ESL: Defining the essentials. *TESOL Quarterly* vol. 22 no. 4 pp. 575-592.

Watson Todd, R. (1997a) Textual patterns in teachers' eliciting. *RELC Journal* vol. 28 no. 1 pp. 1-14.

Watson Todd, R. (1997b) *Classroom Teaching Strategies*. London: Prentice Hall.

Watson Todd, R. (1998) Topic-based analysis of classroom discourse. *System* vol. 26 no. 3 pp. 303-318.

Watson Todd, R. (2002a) Hierarchies and networks in applied linguistics. *rEFLections* vol. 4 pp. 49-69.

Watson Todd, R. (2002b) Finding a topic for research. *ThaiTESOL Focus* vol. 15 no. 2 pp. 20-24.

Watson Todd, R., Chaiyasuk, I. and Tantisawetrat, N. (mimeo) *A functional analysis of teachers' instructions*. Bangkok: King Mongkut's University of Technology Thonburi.

Weir, C. and Roberts, J. (1994) *Evaluation in ELT*. Oxford: Blackwell.

Whitney, P. (1998) *The Psychology of Language*. Boston: Houghton Mifflin.

Widdowson, H. G. (1978) *Teaching Language as Communication*. Oxford: Oxford University Press.

Widdowson, H. G. (1979) *Explorations in Applied Linguistics*. Oxford: Oxford University Press.

Widdowson, H. G. (1984) *Explorations in Applied Linguistics 2*. Oxford: Oxford University Press.

Widdowson, H. G. (1996) *Linguistics*. Oxford: Oxford University Press.

Wikborg, E. (1990) Types of coherence breaks in Swedish student writing: Misleading paragraph division. In Connor, U. and Johns, A.

M. (eds.) (1990) *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 131-148.

Wilensky, R. (1986) Points: A theory of the structure of stories in memory. In Grosz, B. J., Jones, K. S. and Webber, B. L. (eds.) (1986) *Readings in Natural Language Processing*. Los Altos, CA: Morgan Kaufmann. pp. 459-473.

Williams, M. and Burden, R. L. (1997) *Psychology for Language Teachers: A Social Constructivist Approach*. Cambridge: Cambridge University Press.

Willis, J. (1981) *Teaching English through English: A Course in Classroom Language and Techniques*. London: Longman.

Winitzky, N. (1992) Structure and process in thinking about classroom management: An exploratory study of prospective teachers. *Teaching & Teacher Education* vol. 8 no. 1 pp. 1-14.

Winitzky, N., Kauchak, D. and Kelly, M. (1994) Measuring teachers' structural knowledge. *Teaching & Teacher Education* vol.10 no. 2 pp. 125-139.

Winter, E. O. (1982) *Towards a Contextual Grammar of English: The Clause and its Place in the Definition of Sentence*. London: George Allen & Unwin.

Winter, E. (1994) Clause relations as information structure: two basic text structures in English. In Coulthard, M. (ed.) (1994) *Advances in Written Text Analysis*. London: Routledge. pp. 46-68.

Wittgenstein, L. (1953) *Philosophical Investigations*, translated by Anscombe, G. E. M. New York: Macmillan.

Wong-Fillmore, L. (1985) When does teacher talk work as input? In Gass, S. M. and Madden, C. G. (eds.) (1985) *Input in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 17-50.

Woods, D. (1996) *Teacher Cognition in Language Teaching: Beliefs, Decision-Making and Classroom Practice*. Cambridge: Cambridge University Press.

Woods, P. (1996) *Researching the Art of Teaching: Ethnography for Educational Use*. London: Routledge.

Wright, T. (1987) *Roles of Teachers & Learners*. Oxford: Oxford University Press.

Yule, G. (1996b) *The Study of Language*, second edition. Cambridge: Cambridge University Press.

Appendix A Transcripts of extracts

Extract A

T: [1] **OK**, [2] for today I have a box with me. [T holds up small plastic container.] OK? [3] $_{\text{I have}}$ A box. [T laughs.] [4] Can you guess? [5] There are ten pieces of toy, ten pieces of children's toys here. [6] Can you guess? [7] What are they $_{\text{(the toys)}}$? [8] Can anyone guess? Yes? Children's toys.

F1: [9] $_{\text{The toys are}}$ Jigsaw.

T: [10] Pardon.

F1: [11] $_{\text{The toys are}}$ Jigsaw.

T: [12] $_{\text{The toys are}}$ Jigsaw? [13] No, $_{\text{the toys are}}$ not a jigsaw. [14] Ah, for this toy you can blow up air inside it $_{\text{(the toy)}}$. (2.0)

S: [15] (unclear)

F: [16] $_{\text{The toys are}}$ Balloon. [T doesn't hear]

T2: [17] And $_{\text{you can}}$ make it $_{\text{(the toy)}}$ bigger.

F2: [18] $_{\text{The toys are}}$ Balloon.

T: [19] $_{\text{The toys are}}$ Balloon? [20] Yes, $_{\text{the toys are}}$ the balloon. [21] So I have uh [T laughs, opens box and pulls out balloons.] many balloons here. OK? [22] And I would like one of you to come out and blow up a balloon, blow up a balloon. Yes? [23] $_{\text{I would like}}$ A volunteer? [T laughs. SS smile and look uneasy.] Yes. [T writes [24] "a balloon(s)" on the board.] [25] Yes? Balloons, a balloon. [26] I'd like you to blow up $_{\text{a balloon}}$. [T writes [27] "blow up" on the board.] [28] What does it mean by blow up the balloon?

S: [29] $_{\text{Blow up means}}$ *Pow*. {= Blow up.}

T: [30] $_{\text{Blow up means}}$ *Pow*, {= Blow up,} yes, uh-huh. [31] You blow air inside the balloon and make it $_{\text{(the balloon)}}$ [T moves hands apart.]

SS: bigger.

T: [32] $_{\text{You make the balloon}}$ Bigger, yes. [33] Can anyone please? [T holds up balloon.] [34] I need a volunteer. [No SS move.] [35] *Ow, suner mah tha yahng nun*. {= OK, if it's like that, volunteer a friend.} [One S moves.] [36] OK, who'd like to come up and blow up a balloon? Yes. [All SS look at the volunteer.] [37] You are so strong. [38] Yes, he is quite strong. [M1 comes out to the front of the class and takes the balloon from T.] [39] (T and M talk together unclearly) [M1 starts to blow up the balloon, halfway through he looks quizzically at T. The other SS laugh. M1 stops blowing up the balloon.] [40] OK, enough, thank you, thank you. [M1 passes the half-blown up balloon to T.] [41] **OK**, [42] now what's your name please?

M1: [43] $_{\text{My name is}}$ Kittipong.

T: [44] Huh? $_{\text{Your name is}}$ Nettipong?

M1: [45] My name is Kittipong.
T: [46] Your name is Kittipong, yes. [47] Now, Kittipong blew up a balloon
T: already, [48] so you see, it (the balloon) becomes bigger
SS: bigger
T: [49] If I release the balloon, I let go the balloon, what will happen? [50] Can you guess? [SS make wave motions with their hands. T copies the motion.] [51] What does it (wave motions with hands) mean here?
S: [52] (unclear)
T: [53] Huh? *Tum nee kroo mý roo keu arai.* {= I don't know what doing this (wave motions with hands) is.} [T makes wave motions with his hands.] [54] What does it (wave motions with hands) mean? OK. [55] I ask Kittipong to blow up a balloon? Yes. [T holds the balloon high.] [56] Now it (the balloon)'s bigger now. [57] You see there is some (2.0) [T points at the balloon.] [58] He blow up what inside the balloon?
SS: [59] He blows up Air.
T: [60] He blows up Air, yes. [61] There is some air inside the balloon. [62] When, can you guess, when I let go the balloon, I release the balloon, what will happen? [Some SS make wave motions again.]
F: [63] The balloon will Move forward.
T: [64] Uh-huh, you say it (the balloon) will move forward, [65] it (the balloon) will, it (the balloon) will (1.5) [T makes bird wing movements with his hands.]
SS: fly.
T: [66] The balloon will Fly, yes. [67] Will the balloon fly Forward or backward?
SS: [68] The balloon will fly Forward.
T: [69] Ah, yes. You say it (the balloon) will move forward, right? [70] Yes, it (the balloon) will move [T writes [71] "move forward" on the board] [70 cont.] forward. [T corrects his own misspelling on the board.] [72] (to self) F-O-R-W-A-R-D. (loudly) [73] It (the balloon) will move forward. [74] OK, now I will release the balloon. [75] Now uh the neck of the balloon is ... closed, right? [76] The neck is closed, uh-huh? [77] And what happens when the neck is closed. [78] OK, I draw a picture for you. [T draws a picture of a balloon on the board.] [79] This is [T writes [80] "neck" with an arrow pointing to the neck of the balloon in the picture.] [79 cont.] the neck of the balloon, right? [T writes [80 cont.] "is closed" next to "neck".] [81] The neck is ... closed. OK. [82] Uh, I tell you that when Kittipong blow up air inside the balloon, there is air in the balloon or what we call [T writes "p _ _ _ _ _ _ _" on the board.] inside the balloon.
M2: [83] There is Pump inside the balloon.
T: [84] There is Pump inside the balloon? [85] No, there is not pump inside the balloon.
S: [86] There is Pressure inside the balloon.

T: [87] There is Pressure, yes, pressure inside the balloon. [88] How do you spell pressure?

SS: [89] P-R-E-S-S-U-R-E.
T: P-R-E-S-S-U-R-E. [90] Pressure. [T fills in [91] "pressure" in the blanks on the board.] [92] So, when the neck is closed, how is the pressure? [T adds radial arrows inside the balloon to the picture.] [93] The pressure in all directions [T writes [91 cont.] "in all directions" on the board after "pressure".] [93 cont.] are ... the same or different?

S: [94] The pressure in all directions are Different.
SS: [95] The pressure in all directions are Same.

T: [96] Uh, when the neck is closed here, [T adds more arrows inside the balloon on the board.] the pressure in all directions are
SS: same.
T: [97] The pressure in all directions are The same, OK? Right? [T adds [91 cont.] "is the same" after "pressure in all directions" on the board.] [98] Pressure is the same in all directions. [99] The pressure is the same in all directions. [100] Now, OK, you told me that, if I release the balloon, if I let go the balloon, it will
S: [101] (unclear)
T: [100 cont.] move forward. [102] Could you speak loudly? [103] Everyone, could you speak loudly when I ask you, OK? [104] *Gloo-a arai?* {= What are you afraid of?} [105] *Kow yoo kahng nýy* {= He (the technician) is inside} [106] *mý yoo kahng nork.* {= he (the technician) isn't outside.} [T points to the technician sitting in a small room at the back of the class.] [107] OK, you think that it (the balloon) will move forward, right? [108] Now everyone let's see. [T prepares to release balloon.] [109] OK, you can count.

M3: [110] One
SS: One two three
T: One two three [T releases the balloon which flies away.] [111] OK, so it (the balloon) [T points at "move forward" on the board.] moves forward. [[71 cont.] T adds "It" (the balloon) before "move" and "s" after "move" on the board. T writes [112] "Why does it (the balloon) fly/move forward?" on the board.] [113] Why does it (the balloon) fly or move forward? (1.5) [114] Do you have any question, [115] do you have any reason, any answer? [SS talk together.] [116] The balloon flies Because of the pressure, right? Because of the pressure, yes, because of the pressure of the air inside. [117] Kittipong, I will ask you to blow the balloon up again. [M1 stands up.] [118] Now the neck is (unclear) the neck is open, right, [119] the neck is open, [120] and what happen here? [121] What happen? [T draws a picture of a balloon with an open neck on the board. He adds arrows showing air leaving the balloon.]

SS: [122] Air goes out balloon
T: Air yes, air goes out of the balloon, right? Uh-huh. [123] Escape. [T writes [124] "escape" on board.] [125] Right, *dee tee soot* {= the best is} escape, escape, escapes [[124 cont.] T adds "s" after "escape" on the board.] [125 cont.] or goes out [T writes [126] "goes out" on the board.] [125 cont.] and goes out of the balloon through … what? Through [T draws an arrow pointing to the open neck of the balloon on the board.]
F: neck.
T: The neck, uh-huh, [T writes [127] "neck" on the board.] [125 cont.] or we call through the … opening. [T writes [128] "opening" on the board.] [125 cont.] Yes, OK? [129] When I release the balloon, I let go the balloon, the air inside, right? The air inside, what happens to the air inside?
F3: [130] The air inside Goes out.
SS: Escapes goes out
T: escapes or goes out of the balloon through the … neck through the uh through the neck or through the
SS: opening
T: opening. [131] And what happened to the balloon?
SS: [132] (unclear)
T: [133] Uh-huh, it (the balloon) moves [T draws arrows next to the balloon with an open neck on the board.]
SS: forward
T: forward. Yes. [T writes [134] "moves forward" on the board.] [135] It (the balloon) moves forward. [136] Now Kittipong, [M1] I ask you to blow up this one (balloon) again, the big one (balloon), yes, [137] and uh I'd like one of you to blow up the small one (balloon). [138] Oh Mongkut [M4] please come out. [139] You blow up the small one (balloon). [M1 and M4 come to the front of the class and start blowing up the balloons.] [140] OK. Yes (unclear) [M4 has problems blowing the balloon up] [141] Yes, can you help me? [142] Yes, you blow up this balloon. [M5 comes to the front of the class. T gives M5 another small balloon. M4 manages to start blowing up his balloon.] [143] Oh, OK, thank you, thank you. [M5 sits down again] OK. [T takes the small balloon from M4. The balloon bursts. T and SS laugh. T gives M4 another small balloon to blow up. T takes the big balloon from M1 who sits down. M4 gives T the small balloon.] [144] OK, thank you. [145] You see, the two balloons are … Are they (the two balloons) the same size?
SS: [146] No small
T: No, this one (balloon) is small smaller. [147] This one (the balloon)
SS: bigger

T: is bigger. [148] If I release the two balloons, which one (balloon) will go farther?

SS: [149] Bigger balloon will go farther
T: [150] The bigger one (balloon) will go farther? [151] OK, let's try. [152] Which one (balloon) will go faster?

SS: [153] Small balloon will go faster
SS: [154] Big balloon will go faster
T: [155] You say the smaller one (balloon) will go faster? [156] The bigger one (balloon) will go
SS: farther
T: farther and will go slower. [157] OK, let's try. [158] OK, everyone, let's see.

SS: [159] One two three
T: One two three [T releases the two balloons. The big one flies swiftly and erratically. The small one's neck does not open and it floats gently to the floor. T and SS laugh. T picks up the unsuccessful small balloon.] [160] What happened? [161] What happened? (to M1) [162] Can you blow up that one (balloon) again? [M1 picks up balloon from the floor and starts to blow it up.] (to all SS) [163] So we cannot prove, right, [164] we cannot prove. [165] Uh **OK**, [166] now can you blow this one (balloon)? [T passes another big balloon to M5.] [167] Can you blow up a small one (balloon)? [T passes another small balloon to M6.] [168] Thank you. [T takes the big balloon from M1] (to M5) [169] OK you can keep it with you. [T takes the small balloon from M6.] (to M6) [170] OK. Thank you. (to all SS) [171] You say if I release the two balloons you s- you think the bigger one (balloon) will go farther [172] but you think the small one (balloon) will go faster. OK. [173] Let's uh let's count everyone.

SS: [174] One two three.
T: One two three. [T releases the two balloons.] [175] What happened? [176] Which one (balloon) go farther?

SS: [177] Bigger.
SS: The bigger one (balloon) goes farther.
T: [178] The bigger one (balloon) goes farther. OK, the bigger one. [179] Why does the bigger balloon go farther? [SS talk together.] [180] So you have two questions to answer. (reading from the board) [181] Why does the balloon fly or move forward? [182] You get the answer already, right? [T points to drawings of balloons on the board.] [183] This (the drawings on the board) is the answer to the first question. [184] When you, **OK**, [185] when the neck is closed, you close the neck, the pressure is the same in all directions. [186] The air could not escape, right, [187] the air could not

escape. [188] The pressure is the same, [189] but when you open the neck, uh-huh, what happens to the air inside, to the pressure inside?

F: [190] The air inside Escape.

T: [191] Yes, the air escapes or goes out of the balloon [192] and it (the air) moves the balloon forward. [193] What we call (unclear) action [T writes [194] "action" on the board.] [193 cont.] and

SS: reaction.

T: reaction, yes. [T writes [194 cont.] "reaction" on the board.] [195] Action and reaction. Action and reaction. [196] When the air goes out of the balloon, it (the air) moves the balloon forward. [197] And the big balloon and the small one (balloon), [T writes [198] "The big balloon/The small one (balloon)" on the board.] [197 cont.] which one (balloon) moves farther? [199] Which one (balloon) moves farther?

F1: [200] The big one (balloon) moves farther.

T: [201] The big one (balloon) moves farther. [202] Why does the big one (balloon) move farther than the small one (balloon)? [SS talk together.] [203] It is because of what? Because of ...

M3: air inside.

T: [204] Air, you think air inside, right?

SS: [205] There is More air inside.

T: [206] Uh-huh. Yes, there is more air, much more air inside the big balloon than

SS: the small one (balloon)

T: the small one (balloon). [207] Or we can say that there is more [T points to "pressure" on the board.]

SS: pressure small one (balloon)

T: pressure in the bigger balloon than in the smaller balloon. [208] Uh, can you think of this principle (unclear) the principle. [209] The principle of what? [T points to board.] Action and reaction can be used in a vehicle, right? Vehicle, [T writes [210] "vehicle" on the board.] [209 cont.] right? [211] We can use this principle, the principle of action reaction in some kinds of vehicle. [212] What is that vehicle? (3.0)

S: [213] (unclear)

T: [214] Pardon?

M7: [215] Pilot is that vehicle.

T: [216] Pilot. Pilot is a person. (6.0) [217] We use this principle, action reaction in a vehicle, in a machine, in an engine. [218] What engine uses this uh principle? [SS talk together. T points to M8.] Yes?

M8: [219] Aeroplane uses this principle.

T: [220] Aeroplane uses this principle. In an aeroplane, yes, in an aeroplane. [T writes [221] "aeroplane" on the board.] [220 cont.] In an aeroplane, yes. [222]

What kind of engine is used in an aeroplane? [223] What kind of engine is used in an aeroplane? (1.5) Huh?

S: [224] Jet engine is used in an aeroplane.

T: [225] Jet engine is used in an aeroplane. [226] So this [T points to the picture of the balloon with an open neck on the board.] action and reaction [T underlines "action reaction" on the board.] is used in a ... Yes? In a ...

S: jet engine.

T: jet engine. [T writes [227] "jet engine" on the board.] [228] Is it (action reaction) used only in a jet engine? ... [229] No, it (action reaction) can be used with a ... that can fly to the moon, fly to the Mars, fly to space.

S: [230] Action reaction can be used with a Rocket.

T: [231] Action reaction can be used with a Rock what?

S: [232] Action reaction can be used with a Rocket engine.

T: [233] Action reaction can be used with a Rocket, yes. [234] Action reaction can be used with a Jet engine or ...

S: rocket.

T: rocket. [T writes [235] "rocket" on the board.] [236] Action reaction can be used with a Jet engine or rocket engine. [237] So today we are going to study about engines, many kinds of engines.

Extract B

T: [1] **So**, [2] shall we start? [3] Are you ready? Yes, OK. [4] Uh, have you ever played the balloon? [5] When you were a child, have you ever blown up a balloon?

S: [6] Of course [I have blown up a balloon.]

T: [7] Do you know balloon, rubber balloon?

SS: [8] Yes [I know balloon.]

T: [9] Yes [you know balloon.] [10] Have you ever blown up a balloon?

SS: [11] Yes [I have blown up a balloon.]

T: [12] OK, when you blow up a balloon, what happens?

M1: [13] [The balloon] Explode. [SS laugh.]

T: [14] It [(the balloon)] will explode? [15] But before [the balloon] explode, it [(the balloon)] will be (1.0)

SS: big.

T: [16] [The balloon will be] Bigger, right? Bigger and bigger, right?

M2: [17] Yes [the balloon will be bigger.]

T: [18] Yes [the balloon will be bigger.] [19] I want one volunteer to blow up this rubber balloon. [T holds up a balloon.]

S: [20] (unclear) *num lý*. {= saliva}

T: [21] [I want] One volunteer. [22] A new one. It [(the balloon)] is a new one. [23] [I want] One volunteer please.

SS: [24] Rubi, Rubi [is a volunteer.]

M2: [25] Rubi *ork bý ler-y*. {= go to the front immediately.}

T: [26] Who is Rubi? [SS point to M3.] OK. [SS clap.] [27] Just blow up and just [T holds the neck of the balloon.]

S: *kamoo-at pom*. {= tie the neck.}

T: [28] [There is] No need to *kamoo-at pom*. {= tie the neck.} [29] Just to hold this, the open end, just hold the open end [30] and see what happens to the balloon. [M3 stands up and walks to the front of the class. T gives M3 the balloon. M3 blows it up.] [31] OK. Stop, stop. [T indicates that M3 should stop blowing air into the balloon.] [32] OK. Thank you. [T takes the balloon from M3.] [33] OK, let's see this balloon. [34] Is it [(the balloon)] big?

SS: [35] Yes [the balloon is big.]

M2: [36] [The balloon is] Normal.

T: [37] [The balloon is] Small?

M2: [38] [The balloon is] Normal.

T: [39] [The balloon is] Normal, OK. [40] Let's see what happens. [41] Do you think what happens inside this balloon?

SS: [42] Air.

T: [43] The air. The air inside the balloon. What happens to the air inside the balloon? (2.0) [44] The air inside, the air inside the balloon, what, what, what happens to the air?

S: [45] (unclear)

T: [46] Huh?

S: [47] (unclear) pressure.

T: [48] The air pressure, right. The pressure inside the balloon is what? (2.0) [49] The pressure inside this balloon, what happens to the pressure inside this balloon? [50] The pressure inside this balloon is Higher? Higher than outside? No? (1.0)

M2: [51] The pressure inside this balloon is Higher. Higher than outside.

T: [52] Mmm? Mmm? **OK**. [53] This is the balloon, right? [T draws a balloon on the board.] [54] Then the air inside this balloon, [T draws arrows radiating from the centre of the balloon.] what happens? (3.0) [55] It goes to all directions, right?, to all directions inside the balloon [56] and the pressure is equal, right? [57] The pressure inside the balloon is equal in all directions, right?, in all, in all directions. [58] And then, if I release this open end, what will happen do you think? [59] What will happen do you think? (1.0) [60] If I release this open end, the neck, what happens?

S: [61] (unclear)

T: [62] Pardon?

S: [63] (unclear)

T: [64] The air will go out or escape through this open end, right? Right? [65] And what happen to the balloon?

M2: [66] The balloon will

T: [67] It (the balloon) will

M4: [66 cont.] fly.

M5: [67 cont.] small.

T: [68] It (the balloon) will be smaller and smaller, right, [69] and what happen?

M6: [70] The balloon will [M6 makes wavy motions with his hands.]

M2: move.

T: [71] The balloon Will move. [72] The balloon will move In which direction? In which way?

M2: [73] I don't know which direction the balloon will move.

T: [74] You don't know which direction the balloon will move?

M2: [75] Everyone don't know which direction the balloon will move.

T: [76] So if I release here, the balloon will move which way, this way, this way or this way? [T makes motions with his hands vertically, to the left and to the right of the balloon he is holding.]

SS: [77] No.

SS: [78] I don't know.

[SS make upward motions with their hands.]

T: [79] It should be opposite to (1.0) to, to this open end, right? [80] The direction, the direction will, will be opposite to this hole. OK? [81] The air will escape through this open end, OK? [82] Let's see what happens. [T releases the balloon which flies up to hit the ceiling and then falls to the ground.] Right? Right?

M6: [83] *Mý tun krup.* {= It was too quick for me to see.}

T: [84] Huh? OK, you come, you come. [85] If you want to do it again, you come out. [M6 comes to the front of the class. T gives M6 the balloon. M6 blows up the balloon to a reasonable size.] [86] Stop. [M6 hands the balloon to T.] [87] Or you want to release it (the balloon) by yourself? [T offers the balloon back to M6 who takes it.] [88] OK. Let's see the direction. [M6 releases the balloon which flies around randomly.] [89] OK. Thank you. [M6 sits down. T holds up the balloon.] [90] This produces what? [91] This produces what?

S: [92] Rubber.

T: [93] No. [94] But it's like what do you think about this kind of working, this kind of work?

M2: [95] I think Reaction.

T: [96] Huh?

M2: [97] I think Reaction.

T: [98] You think Reaction, very good. [M2 holds his arms in the air in victory and cheers himself. T writes [99] "Reaction" on the board.] [100] It's about action reaction sequence, right? [101] Why? [102] OK, this is the reason why the balloon move forward, right? because of the

S: reaction

T: action reaction sequence. [103] Think about the engine, the engine.

S: [104] (unclear)

T: [105] Huh? What kind of engine works like this (with action reaction sequence)?

S: [106] Jet engine works with action reaction sequence.

T: [107] Jet engine works with action reaction sequence. OK, jet engine, good.

S: [108] Jet engine. [T writes [109] " Engine Types " on the board. T writes [110] "Jet" on the board joined by a line to "Engine Types".]

T: [111] OK, think about the jet engine. [112] What else works with action reaction sequence? What else?

S: [113] (unclear)

T: [114] Missile works with action reaction sequence?

S: [115] Rocket works with action reaction sequence.

T: [116] Rocket works with action reaction sequence, rocket.

M2: [117] Rocket works with action reaction sequence.

T: [118] Rocket works with action reaction sequence. [T writes [119] "Rocket" on the board joined by a line to "Engine Types".]

S: [120] Aeroplane works with action reaction sequence.

T: [121] Pardon?

S: [122] Aeroplane works with action reaction sequence.
T: [123] Aeroplane works with action reaction sequence. [124] Aeroplane is a kind of engine?
M2: [125] Concorde.
T: [126] Is an aeroplane a kind of engine?
S: [127] (unclear)
T: [128] Huh?
S: [129] (unclear)
T: [130] No. An aeroplane is not a kind of engine, [131] but rocket can be a kind of engine. Rocket engine. [132] But we cannot say aeroplane engine, no. [133] But aeroplane, aeroplane engine use what kind of engine? [134] Aeroplane uses Jet engine or rocket engine, right? [135] Anything else about the engine? [136] Do you know?
S: [137] *Bahngfý.* {= Traditional Thai rocket works with action reaction sequence.}
T: [138] *Bahngfý, bahngfý,* OK, *bahngfý.* {= Traditional Thai rocket, traditional Thai rocket, OK, traditional Thai rocket works with action reaction sequence.} [139] How does it work? *Bahngfý.* {= Traditional Thai rocket.} [T writes [140] "Bangfai" on the board joined by a line to "Engine Types".]
M2: (while T is writing) [141] *Bahngfý.* {= Traditional Thai rocket.} F - A - I.
T: [142] *Bahngfý* {= Traditional Thai rocket } can work like this (with action reaction sequence). OK. Good. [143] What else works with action reaction sequence? (3.0) [144] How about the, a car? [145] What kind of engine, what kind of engine is used in a car?
M2: [146] Quarter mile car.
T: [147] Huh?
M2: [148] Quarter mile.
T: [149] What do you mean?
M2: [150] *Rot tee kaang doo-ay kwahm raaw soong mahk krup tee mun torng brayk doo-ay gahn chý rom choo cheep.* {= The car which races very fast which has to use a parachute to break.}
T: Ah.
S: [151] (unclear)
T: [152] OK. Anything else? (1.5) [153] I want normal car. [154] In normal car, what type of engine is used?
S: [155] (unclear)
T: [156] Huh? What type of engine is used? (4.0) In normal car.
M2: [157] *Rot kayai* air, *bloy pood laa go wing.* {= The car expands with air, farts and drives away.}
T: [158] Engine. About the engine.
S: [159] Diesel engine is used.

T: [160] Diesel engine is used. OK, diesel engine. [161] Diesel engine is used in, in, in, in what kind of vehicle? [T writes [162] "Diesel" on the board joined by a line to "Engine Types".] (3.0) [163] Diesel engine is used In a car?

S: [164] *Rot buntook.* {= Diesel engine is used in a Lorry.}

T: [165] Diesel engine is used in a Lorry, right? [166] *Rot buntook* {= Lorry} is lorry. [167] Anything else about the engine type?

S: [168] Train.

T: [169] Plane.

S: [170] Train.

T: [171] Train, train. [172] What types of engine can be used in a train? [173] Diesel engine can be used in a train?

S: [174] *Ý num.* {= Steam engine can be used in a train.}

T: [175] *Ý num wah arai?* {= Steam is what?}

S: [176] *Ý num* {=steam} is Steam.

T: [177] Steam. Steam engine can be used in a train. [T writes [178] "Steam" on the board joined by a line to "Engine Types".] [179] Anything else you know about the engine, the type of engine?

S: [180] Gasoline.

T: [181] Huh?

S: [182] Gasoline.

T: [183] Gasoline, or we can say petrol, right?, petrol. [T writes [184] "Petrol" on the board joined by a line to "Engine Types".] [183 cont.] Petrol engine. [185] Now, I would like to, I would like you to see this diagram about the petrol engine, petrol engine.

Extract C

T: [1] **OK.** [2] Shall we start our lesson now? Good. Mmm. [3] What is it (the balloon)? [T reaches into a big bag which obviously contains something quite big. T pulls out a balloon.]

SS: Wow.

T: [4] What's in my hand? [5] What is it (the balloon) called in English? [T holds up the balloon.]

M: [6] *Look pong.* {= Balloon is in your hand.}

T: [7] *Look pong.* {= Balloon is in my hand.} [8] What is balloon called In English. In English.

SS: [9] Balloon is called Balloon in English.

T: [10] Balloon is called Balloon in English, right. [11] Can you spell the word balloon?

SS: [12] B-A-L (Many SS spell balloon at different speeds so the letters become unclear)

T: B...B-A-L-L-O-O-N. [T writes [13] "Balloon" on the board.]

SS: B-A-L-L-O-O-N.

T: [14] OK, balloon. [15] **OK,** [16] can the balloon fly? [17] Do you think it (the balloon) can fly?

M: [18] No the balloon cannot fly.

F: [19] Sometimes the balloon can fly.

T: [20] Sometimes the balloon can fly, OK. [21] So let's see. [22] I need a volunteer. [23] Who wants to be a volunteer? ... [24] Yes, come on. [SS look at each other.] [25] Come on, volunteer, please. [SS talk together. F1 stands up and comes to the front.] [26] OK, thank you. OK. [T points to the neck of the balloon.] [27] This is its (the balloon's) neck, right? Its (the balloon's) neck, uh-huh, or opening. [28] Now our volunteer will release the neck. [T passes the balloon to F1. F1 releases the balloon, but the neck stays closed and the balloon floats gently to the ground.] [29] Woah. It (the balloon) cannot fly. [F1 picks up the balloon. T lets it float in the air again. F1 takes the balloon, opens the neck and releases the balloon so that it flies quickly in random directions. T picks up the empty balloon. F1 walks back to her seat.] [30] And (3.0) why the balloon fly like this? [31] Why the balloon fly like this? Uh-huh, [32] **OK,** [33] work in groups of three or four.

F: [34] Work in groups of Four.

T: [35] Work in groups of Three or four?

SS: [36] Work in groups of Four.

T: [37] OK, work in groups of four. [38] And try to find the reason why this balloon fly like this, OK? [SS talk together.] Mmm, right. OK. [T puts a picture of a balloon flying by expelling air on the overhead projector.] [39] **OK,** [40] why does the balloon fly like this? Question

number one. [41] And question number two, what kinds of engines work in the same way? [42] *Yahng raak hý tahm keu tum mý tum mý ka* {= The first (question) to be asked is why, why} fly, *tum mý mun teung bin?* {= why can it (the balloon) fly?} [43] Find reasons. [44] And what kinds of engines work in the same way? [45] Engine. What does it mean? Engine.

F2: [46] (unclear)

T: [47] Engine. Engine is the machine, uh-huh, that uses heat or other kinds of energy to produce power. Uh-huh. [48] For example, steam engine, petrol engine. [49] Duangdeuan (F2), right? [48 cont.] And jet engine. [50] Ekchalor (M1) OK. [M1 stands up and gives a piece of paper to T.] (to self) [51] Mmm, where is it? (to all SS) [52] OK. Engine in Thai what does it mean? (2.0) Engine. (2.0) [53] *Kreu-ang juk kreu-ang yon kreu-ang yon tee row chý* {= Engine, engine, engine that we use} [54] *yahng chen row chý kreu-ang yon yahng ný bahng?* {= for example, how do we use engines?}

SS: [55] (unclear)

T: [56] We use engines In a car.

M: [57] We use engines in a Train.

T: [58] We use engines In a train, OK.

M: [59] (unclear)

T: [60] Electric what? Electric. Electric *chý reu plow?* {= yes or no?} [61] Electric *mý chý.* {= no.} [62] *Mee arai bahng ka?* {= What is there?}

S: [63] (unclear)

T: [64] We use engines in a *Rot týy nah,* {= machine for pulling rice ploughs,} OK. [SS talk together.] [65] *Arai eek ka?* {= What else do we use engines in?}

F: [66] We use engines in a Pump.

T: [67] We use engines in a Pump, mmm. [68] What else do we use engines in?

M: [69] We use engines in a Robot.

T: [70] We use engines in a Robot. Robot. Mmm. [T nods her head.] [71] Yes, robot, we use petrol to make it (robot) move. OK. [72] So what does, why does the balloon fly like this? Question number one [73] and number two, what kinds of engines work in the same way? [74] Work in groups of four, so now. [75] Sit in groups of four. Right now. [SS rearrange chairs and start working in groups.]

Extract D

T: [1] **OK**. [2] Now I want three volunteers. Three volunteers.
S: [3] You want Three volunteers.
SS: [4] (unclear)
S: [5] *Hý tum arai krup?* {= What are you going to give us to do?}
S: [6] *Sahm, sahm.* {= You want Three, three volunteers.}
T: [7] *Sahm* volunteer *keu arai?* {= Three volunteer is what?}
S: [8] *Ahsahsamuk.* {= Three volunteer is Volunteer.}
T: [9] *Ahsahsamuk.* {= Three volunteer is Volunteer.}
S: [10] *Tum arai krup, tum arai?* {= To do what, to do what?} [SS talk together. Three SS stand up and come to the front. Other SS clap. T gives each volunteer a balloon.]
T: [11] Listen now please. [12] I want you to (very loudly) [13] Listen. (normally) [12 cont.] I want you to compete to blowing, to blow the balloon.
S: [14] *Hý pow.* {= She asks us to blow.} [SS talk together loudly. The three volunteers start blowing up the balloons. M1's balloon bursts. T gives M1 a new balloon.]
T: [15] *Yah roong raang.* {= Blow Not too strongly.} [16] *Mý torng yý mahk.* {= It (the balloon) doesn't have to be very big.} [17] Don't blow too big. [M2's balloon reaches a reasonable size.] [18] *Por laaw.* {= That's enough.} [19] *Teu wý na.* {= Just hold it (the balloon).} (unclear) [M1's balloon bursts. T gives M1 a new balloon. M3's balloon reaches a reasonable size.] [20] *Dý mý?* {= Is it possible?} [M3's balloon bursts. T gives M3 a new balloon.] [21] *Ow mý.* {= Take a new one (balloon).} [M1's balloon reaches a reasonable size.] [22] OK, OK. *Kaa nee.* {= That's enough.} [23] *Mý torng payayam mahk.* {= You don't have to try very hard.} [24] *Dee-o ja taak.* {= It (the balloon)'ll break in a minute.} [25] *Yah peung bloy.* {= Don't let it (the balloon) go.} [M3's balloon reaches a reasonable size.] [26] OK. It's enough. [M3's balloon bursts.] [27] That's why it's enough. [T ignores M3 and turns to M1 and M2.] [28] Ah, *song kon mah nee.* {= You two, come here.} [M1 blows more air into his balloon.] [29] *Por laaw.* {= That's enough.} [30] *Ja taak see.* {= It (the balloon)'ll burst.} [31] OK, I want both of you to compete to let it (the balloon) go.
S: [32] *Bloy.* {= Let it (the balloon) go.}
T: [33] *Mý torng mut.* {= You don't have to tie it (the balloon).} [T undoes the knot in M2's balloon.] [34] I want you to let it (the balloon) go [35] and let's see which one (balloon) is, uh goes furthest. [36] I will count. [M2 still tries to sort out the knot in his balloon. M3 walks to the table to get a new balloon. M2's balloon bursts. M2 takes a new balloon, which also bursts. M3's balloon reaches a reasonable size.] [37] OK,

song kon. {= two people.} [T gestures for M1 and M3 to stand at the front of the class holding their balloons.] [38] OK, let's see which one (balloon) goes furthest. [39] I will count one, two, three, [40] and then you let it (the balloon) go, OK? [41] One, two, three. [M1 and M3 release their balloons which float to the ground. SS laugh.] [42] I just want to know how, uh why does the balloon fly like that. [SS talk together.] [43] *Fung ahjahrn na ka.* {= Listen to the teacher.} [44] **OK.** [45] What happened when your friends let the balloon go? [46] What happened?

S: [47] What happened?

T: [48] What happened?

S: [49] (unclear)

T: [50] Ah, the air is come out [51] and what happened when the air is come out?

S: [52] The balloon will go.

T: [53] OK, when the air is come out from the balloon, the balloon go, go ahead, right?, go forward, right? [54] Maybe it's not very forward [55] but it (the balloon) doesn't

T: turn around. [T makes circling motions with her hands.]

S: go around.

T: [56] OK, anyway, when you, when you let the balloon go out, [T draws a picture of a balloon on the board. She adds arrows showing air escaping from the balloon's neck.] and the balloon go forward, right? Actually, it (the balloon) go forward. [T draws an arrow showing the direction of movement of the balloon on the board.] [57] OK, why this happens? [58] Why it happens like this? [59] Why the balloon go forward like this?

S: [60] The balloon go forward because of Action reaction.

T: [61] Do you know why? (5.0) Why does the balloon go forward?

S: [62] The balloon go forward because of Reaction.

T: [63] What kind of reaction? (3.0) [Two SS get up to go to the toilet.] [64] *Tum mý torng bý doo-ay gun song kon nee.* {= Why do you two have to go together?} [Several other SS make sarcastic comments.] [65] *Pood nah klee-ad jing jing.* {= You really speak nastily.} [66] **OK.** [67] Can you answer my question? (3.0) [68] Why, why when you let it go, let the air out, the balloon go forward? (2.0) *Pror arai ka.* {= Because of what?}

S: [69] (unclear)

T: [70] Ah, because when the air go forward, uh go out, it (the air) push, right?, it (the air) push the balloon go forward, right? [71] **OK.** [72] And can you tell me what kinds of engine work like this? [73] What kinds of engine work like this? [T writes [74] "Engine" on the board.] [75] Do you know what is engine?

S: [76] Engine is *Kreu-ang yon.* {= Engine.}
T: [77] Engine *keu arai kha?* {= what is it?}
S: [78] Engine is *Kreu-ang yon.* {= Engine.}
T: [79] Engine is *Kreu-ang yon.* {= Engine.} [80] What kinds of engines that work like the balloon?
S: [81] (unclear)
T: [82] *Mý chý* energy *na ka.* {= It's not energy.} [83] Engine.
S: [84] Action reaction.
T: [85] Think about when, uh, what kind of engine, what kind of uh how to say. [86] **OK.** [87] Can you compare this (action reaction), with this technique with the engine? (3.0) [88] *Payayam torp dý mý pror wah gahn tee mee pror lom mun ork laa* balloon *mun bý kahng nah, nee plee-ab tee-ab gahp* engine *chunit ný, baap ný?* {= Can you try to answer because this action where the air goes out of the balloon and the balloon goes forward, this is comparable to what kind of engine, what type of engine?}
S: [89] (unclear)
T: [90] *Bok wah arai na?* {= What did you say?}
S: [91] (unclear) *raaw raaw krup.* {= quick quick.}
T: [92] Ah *mun chý kreu-ang yon baap arai ka?* {= what kind of engine does it use?} [93] What kind of engine?
S: [94] It uses Jet engine.
T: [95] It uses Jet engine, OK. [T writes [96] "Jet engine" on the board.] [97] Right? So it works just like jet engine. [98] And what kind of engine that you know? [99] What kind of engine that you know? (2.0) [100] *Mee eek mý nork jahk* {= Are there any others in addition to} jet engine? [101] *Pood teung tum ngahn baap nee,* {= We've been talking about action like this,} [102] *taa wah* {= but} let's talk about the engine. [103] What kind of engine that you know, that we use today? (4.0) [104] *Mee arai ka, kreu-ang yon, mee kreu-ang yon baap ný ka?* {= What is there, engines, what kind of engines?}
S: [105] There is Diesel engine.
S: [106] There is Benzene engine.
T: [107] There are Diesel, and benzene engines. [108] Benzene *tee mee krý torp keu* {= that someone answered is} petrol engine. [T writes [109] "Petrol engine" on the board.] [110] OK, petrol engine include gas engine, benzene, diesel, *gloom nee.* {= this group.}
S: [111] There is *Palung taan.* {= Alternative energy.}
T: [112] *Palung taan pen* engine *reu plow?* {= Is alternative energy an engine or not?}
S: [113] There is *Ý num na krup.* {= Steam.}
T: [114] Uh, *ý num keu arai?* {= steam is what?}
S: [115] There is Hydro, hydraulic engine.

T: [116] Hydraulic engine *mee reu plow?* {= exists or not?} [T writes [117] "Hydrolic engine" on the board.]
S: [118] *Mee.* {= It (hydraulic engine) exists.}
T: [119] OK. **OK**. [120] Today we are going to talk about engines.

Extract E

T: [1] **Yes, OK,** [2] thank you. [3] For today, we will start with a game. OK. What we call twenty questions. [4] Have you ever played this game before? Twenty questions game. [T writes [5] "Twenty questions" on the board.] [6] Um, I will think of an object, OK?, [7] and you have to guess, you have to find out what I think of. [8] You can ask questions, [9] and within twenty questions if you can guess correctly, if you can find out what I think of, then you win. OK? [10] The questions can be yes-no questions or 'or' questions. [11] For example you can ask 'Can we eat it (the object)?' [12] I will answer just yes or no. [13] Or you can ask 'Is it (the object) big or small? Is it (the object) big or small?' [14] OK, then I may say big or small. [15] This is what we call twenty questions. [16] So are you ready? [17] **OK.** [18] Everyone today now I'm thinking of an object, an object, a thing. [19] You can ask questions. [20] I'm thinking of An object. (3.0) [21] You can ask Any yes-no questions or 'or' questions. (2.0) Yes, yes. [SS laugh.] [22] And I will count. [23] If within twenty questions you can get the answer, then you win the game. (3.0) [24] I'm thinking of an object. [25] Uh, I give you a clue. [26] It (the object) 's a machine. [27] I'm thinking of a machine. (8.0) [T writes [28] "I'm thinking of a machine" on the board.] [29] Yes, you can ask. [30] Within twenty questions if you guess correctly. (1.0) [31] You can ask, uh, you are not allowed to ask wh- questions [32] but you are allowed to ask yes-no questions or 'or' questions, questions with 'or'. [33] For example, is it (the object) big or small? [34] Uh, you can ask uh [One S sitting near the front looks as if he's about to ask a question] Yes? Any questions you can ask. [35] You can guess Within twenty questions. [36] Then I will count. [37] You try to ask in order to guess, to find out what I'm thinking of. [M1 raises his hand.] [38] Yes, Sombat?

M1: [39] Have it (the object) two wheels or four wheels?

T: [40] Has it (the object) two wheels or four wheels? [41] Some of these objects have wheels [42] but some of these objects don't have wheels. (7.0) [43] So I, I could not say whether the object has wheels

M2: [44] Do we Use the object every day or not?

T: [45] Pardon.

M2: [46] Do we Use the object every day or not?

T: [47] Do we use it (the object) every day or not? [48] For you, you don't. You don't use it (the object) every day.

M3: [49] Is it (the object) in a car?

T: [50] Is it?

M3: [51] Is it (the object) in a car?

T: [52] Is it (the object) used in a car? [53] I don't think it (the object) is used in a car. [54] So how many questions have you asked already? [55] You have asked Three questions, right?

S: [56] We have asked Two questions.

T: [57] You have asked One question from Sombat, [58] you have asked one question from you, [59] and you have asked one question from Gay. [60] You have asked Three questions already OK. [T makes marks on the board to indicate the number of questions.]

F1: [61] Can it (the object) carry things?

T: [62] Can it (the object) carry things? [63] Yes it (the object) can carry things. [64] It (the object) can help people to carry things. (7.0) [65] It (the object) can help people, it (the object) can help workers to carry things. (1.0) [66] Will you ask Any more questions? [67] You have asked only four questions. [T points to the marks on the board.] Just only four. [68] You have twenty questions to ask. [SS talk together.]

M2: [69] Is it (the object) used in supermarket?

T: [70] Is it (the object) used in supermarket, supermarket? [71] I haven't seen any (of the objects) in the supermarket. [72] In the future, maybe, in the future it (the object) will be used in the supermarket [73] but the object is not now used in the supermarket.

M4: [74] Is it (the object) used in the industrial?

T: [75] Is it (the object) used in industry? [76] Yes, it (the object) is used in factories. [SS talk together.] [77] Yes, it (the object) is used in industry, [78] it (the object) is used in factories.

SS: [79] (unclear)

T: (to these SS) [80] The answer is...

M1: Robot.

T: [81] The answer is Robot? [82] Yes, you are right. [83] So I'm thinking of a robot. [T writes [84] "a robot" on the board.]

Extract F

T: [1] What did you do this, last weekend? (1.0) [2] What did you do last weekend? After mid-term exam. (2.0) [3] Huh? *Lung* mid-term *tum arai, lung* mid-term *laaw*? {= After the mid-term, what did you do after the mid-term?} [SS talk together.]

S: [4] (unclear)

T: [5] You were Sleeping.

S: [6] I was Watching TV.

T: [7] You were Watching television.

S: [8] *Bý tee-o* RCA. {= I went out to RCA (an entertainment area for teenagers).}

T: [9] Uh, *bý tee-o RCA reu?* {= You went out to RCA, did you?} [10] *Bý tum arai RCA?* {= What did you go to RCA for?} [11] Did you watch, did you watch Star Wars Episode 1 yet?

SS: [12] No I didn't watch Star Wars.

T: [13] No you didn't watch Star Wars.

S: [14] Yes I watched Star Wars.

T: [15] Yes you watched Star Wars. [16] Who watched this film already? [T waves his hand in the air. Some SS put their hands up.] [17] One, two, only two, three, four watched Star Wars already. [18] **OK**, good. [19] For the one who haven't seen this film, go to the SALL and see this movie, OK? [20] But anyway, let's see this one (transparency). [T puts a transparency of C3PO on the transparency.] (5.0) [21] Do you know this robot? [22] For the one who have who has already watched this film, do you know this robot? [23] What is his (the robot's) name? [24] What is his (the robot's) name? [25] Do you know the robot's name?

M1: [26] The robot's name is Three-P-O.

T: [27] Huh?

M1: [28] The robot's name is Three-P-O.

T: [29] T- pardon?

M1: [30] The robot's name is Three-P-O.

T: [31] The robot's name is T-P-O? T-P-O? (2.0) [32] Huh? OK, his (the robot's) name is T-C-fourteen. [33] Is that (T-C-fourteen) right? OK. [34] What can he (T-C-fourteen) do? [35] Do you know what - what do you know he (T-C-fourteen) can do?

M1: [36] T-C-fourteen can Translate.

T: [37] Huh?

M1: [38] T-C-fourteen can Translate.

T: [39] Yes, T-C-fourteen can translate. [40] He T-C-fourteen acts as a translator, right? Translator and as a (1.0) servant, servant, right? (2.0) [41] Let's see another one (transparency). [T puts a transparency of R2D2 on the

OHP.] [42] How about this one (transparency)? [43] What is his (the robot's) name?

M1: [44] The robot's name is R-two-D-two.

T: [45] Huh?

M1: [46] The robot's name is R-two-D-two.

T: [47] The robot's name is R-two-D-two, R-two-D-two. [48] What is his (R-two-D-two's) duty?

M1: [49] R-two-D-two's duty is Repair.

T: [50] R-two-D-two's duty is To repair, right. [51] R-two-D-two's duty is To repair what? [SS talk together.]

M1: [51] R-two-D-two's duty is to repair Spaceship.

T: [52] R-two-D-two's duty is to repair Spaceship, OK. [53] His (R-two-D-two's) duty is to repair a spaceship. (6.0) [54] One more. [T puts a transparency of a skinless C3PO on the OHP.] [55] How about this one (transparency)? Huh? [56] How about this one (transparency)? [57] What is his (the robot's) name?

S: [58] The robot's name is E-T.

T: [59] Huh? The robot's name is E-T? [60] The robot's name is C-three-P-O, right? C-three-P-O. [61] What is his (C-three-P-O's) function? [62] What is his (C-three-P-O's) duty? (3.0) [63] What is his (C-three-P-O's) duty, do you know?

S: [64] C-three-P-O's duty is Servant.

T: [65] Huh?

S: [66] C-three-P-O's duty is Servant.

T: [67] C-three-P-O's duty is Servant. [68] He (C-three-P-O) acts as a servant, right? OK. [T puts a new transparency on the OHP. It contains the 3 pictures given in Hutchinson and Waters (1984: 40). Picture A is of a robot arm on a submersible, but is very murky and unclear on the OHP. Picture B is a bit clearer and shows a man operating robot hands in a nuclear laboratory. Picture C is quite clear and shows robots welding a car body on a factory assembly line.] [69] That is about robots. [70] There are three pictures here. [71] Can you see? OK. [71] What do you think each one is doing? [72] Let's see this one (Picture A) first. [T puts a pencil on the transparency pointing to Picture A.] [73] What do you think this one is doing? (14.0) [74] Can you see? Clearly? [75] Is it (Picture A) clear?

SS: [76] No Picture A is not clear.

T: [77] Never mind. [78] Let's see this (Picture B) first, this. [T moves pencil to point at Picture B.] (3.0) [79] How about this (Picture B)? [80] What do you see in this picture (Picture B)? (9.0) (unclear) [81] Do you see Anything else? Huh?

S: [82] (unclear)

T: [83] You see *Kreu-ang* uranium. {= Machine involving uranium.} (2.0) [84] Where is the man? [85] Where is the man? [86] Where is the man? [87]

The man is In the toilet? [88] Where is the man? Where? [89] The man is At the back inside this

M2: safety room.

T: safety room [T moves the pencil to the man in the picture.] or operating room, right?, so to control these hands, right?, to handle something outside this room. [90] This means that he (the man) cannot stay outside this room, right? [91] And he (the man) has to be inside this room to control the mechanical hands, right? [92] How about this one (Picture C)? [T moves the pencil to point at Picture C.]

S: [93] Picture C is Car.

T: [94] Picture C is Car?

S: [95] Picture C is Car factory.

T: [96] Picture C is Welding? Huh? [97] What is the picture about? [98] The picture is about Car welding, right? Car welding? [99] The picture is about Welding which part of the car, which part of the car?

S: [100] The picture is about welding car Body.

T: [101] The picture is about The body, right. The body of the car. Welding the car body, right? Welding the car body. [102] How about this one (Picture A)? [T moves pencil to point at Picture A.] [103] What is in here (the submersible)? [104] What, what is this, this, this part (the submersible)? (4.0) [105] Why do we have to use this mechanical hand? [106] Why don't we use human hands instead of this mechanical one?

M2: [107] Because it's high pressure.

T: [108] Because it's high pressure?

M2: [109] Yes because it's high pressure.

T: [110] It's high pressure In this condition? [111] Maybe this condition is too dangerous, right? Right? Too dangerous for humans to be around here. [112] So here, maybe this one (hand) [T points to mechanical hand on the transparency.] may contain radioactive material, radioactive material, right? [113] Do you understand radioactive material?

S: [114] (unclear)

T: [115] Radioactive material is *Wusadoo, watook tee mee sahn gumatapahp rungsee.* {= Material which is radioactively poisonous.} [116] **OK**, **now**, [117] it's your turn to answer these questions. [T puts a new transparency on the OHP. The transparency contains the following: [118] "Will robots replace people at work? [119] Some people say that they (robots) create more jobs, [120] some people say that they (robots) create less jobs. [121] Who do you agree with?"] [122] Will robots replace people at work?

S: [123] Yes robots will replace people at work.

T: [124] Yes robots will replace people at work. [125] Who agree yes? [Most SS raise their hands.]

Extract H

[T writes [1] "Robot" on the board. T writes [2] "= motor machine" on the board.]

T: [3] OK, what else that you think? [4] Somebody said that it (robot)'s a motor machine.

S: [5] Robot's a Toy.

T: [6] It (robot) 's a toy. [T writes [7] "= Toy" on the board]

S: [8] Robot's a Toy.

T: [9] It (robot) 's a toy? [10] What else? (2.0) What robot is. [11] *Keu arai* {= What is it (robot)} Wacharin?

F1: [12] (unclear)

T: [13] Ah Wacharin *torp wah yahng ngý?* {= answered what?}

F1: [14] Wacharin answered Use battery.

T: [15] OK, you think robot use battery. [T writes [16] "= use battery" on the board.] [17] Right? **OK**. (3.0) [18] From your (3.0) from your description of robots, now let's see this picture. [T puts 4 pictures on the OHP. Picture A is a piece of factory machinery. Picture B is a tank. Picture C is a lunar module. Picture D is a humanoid science-fiction robot.] [19] Can you tell me which one is robot? (2.0) A, B ...

SS: [20] D is a robot.

T: [21] D is a robot. [22] From your opinion, right?, this one (Picture D) is robot. [T points at Picture D.] [23] How about this one (Picture B) ? [T points at Picture B.]

SS: [24] Picture B is a Tank.

T: [25] Tank *keu arai?* {= what is it?}

SS: [26] Tank is *Rot tung.* {= Tank.}

M1: [27] Tank is *Tung num.* {= Water tank.}

T: [28] It (tank) 's a kind of vehicle, OK. [T points at Picture C.] [29] A lunar module. Lunar module *keu arai?* {= what is it?}

SS: [30] Lunar module is *Yahn awahgaht.* {= Spaceship.}

T: [T points at Picture C.] [31] OK, also this one (Picture C) is a vehicle, right? OK. (unclear) (4.0) [T puts a reading passage on the OHP.] [32] **Now**, [33] read this passage ...

Extract I

T: [1] **OK**, [2] are you ready?
SS: [3] Yes we're ready.
SS: [4] No we're not ready.
T: [5] No? You're not ready?
F1: [6] Yes, yes, yes we're ready.
T: [7] Um, OK, today I show you a picture. [T holds up a heart-shaped picture of a water pump.] [8] Can you tell me what it is? In the heart. [9] What is it (the picture)? [10] What is it (the picture)? This one.
SS: [11] The picture is A pump.
T: [12] The picture is A pump, yes, a pump. [13] What kind of pump is it (the picture)? This one. [14] What kind of pump is it (the picture)? [15] What kind of pump is it (the picture)?
S: [16] The picture is Water pump.
S: [17] The picture is *Pum num.* {= Water pump.}
T: [18] The picture is Water pump. [19] In English, not in Thai. [20] *Pum num roo juk laaw.* {= You already know (the Thai words for) water pump.} [21] So in this picture you see a pump, a pump. [T writes [22] "a pump" on the board.] [23] P - U - M - P. [24] What kind of pump is it (the pump)?
SS: [25] The pump is a Water pump.
T: [26] The pump is a Water pump. [T writes [27] "a water pump" on the board.] [26 cont.] A water pump. [28] What is it used for, the water pump here? [29] What is it's the water pump's function? For a water pump?
M1: [30] The function is Pumping water up to the (unclear)
T: [31] OK, you mean for pumping water, right? Pumping water, right? [T writes [32] "pumping water" on the board.] [31 cont.] For pumping water, for forcing water, right?, [T writes [33] "forcing water" on the board.] [31 cont.] forcing water to (1.0) flow, right?, to flow. [T writes [34] "to flow" on the board.] [35] There are so many kinds of pumps. [36] Apart from a water pump, are there any other pumps you know? (2.0) [37] Apart from a water pump, you still have many kinds of pump. [38] What are they (other kinds of pump)? (3.0) [39] **OK**, [40] you ride a bicycle, [41] then you have a flat tyre, OK? [42] You ride a bicycle, [43] then you have a flat tyre. [44] You must use a pump, right?, to blow up air into the wheel of your bicycle. [45] Is that (the pump you use) a water pump? [46] No it's not a water pump. [47] What is it (that pump)? (2.0) [48] Yes, what is it (that pump), Bangpot?
M2: [49] That pump is an Air pump.
T: [50] An air pump, yes, you have an air pump. [T writes [51] "an air pump" on the board.] [50 cont.] Air pump. [52] What is it used for, an air pump? [53] Is an air pump used For pumping water? For forcing water? [54] No an air pump is not used for pumping water. [55] An air pump is used For pumping ...

S: air.

T: [56] An air pump is used for pumping Air, yes, into something, right?, into your tyre if you have a flat tyre. [57] *Roo juk* flat tyre *mý?* {= Do you know 'flat tyre'?} [58] *Keu arai?* {= What is it (flat tyre)?} [59] When you ride a bicycle then you have a flat tyre, what happens to your tyre? [60] *Keu arai?* {= What is it (flat tyre)?} [61] *Yahng baan.* {= Flat tyre is A flat tyre.} [62] OK. In our body, do you think we have a pump in our body?

SS: [63] Yes we have a pump in our body.

T: [64] Yes we have a pump in our body? [65] Mmm-hmm. I agree with you, yes. [66] We know that, OK, in our body actually there are two pumps in our body, two pumps side by side. [67] There are two pumps side by side in our body. [68] What organ is the two pumps?

F2: [69] The organ is the Heart.

T: [70] The organ is Our heart, our heart. [71] Can you draw a heart? [T gives a marker pen to F2.] Yes, yes. [F2 draws a [72] " " on the board.] [73] Yes, OK. You write a word for your heart. [74] You write a word for your heart. Heart. [75] How do you spell heart? [F2 writes [76] "Heart" on the board.] [77] H - E - A - R - T. [78] Is it (H - E - A - R - T) right? Right or wrong? Right or wrong?

SS: [79] H - E - A - R - T is Right.

T: [80] H - E - A - R - T is Right, yes. (to F2) [81] Thank you. [F2 sits down.] (to all SS) (unclear) [82] **OK.**

Extract J

T: [T writes [1] "Pumping Systems" on the board.] [2] **OK**. [3] I will let you think about pumping systems, pumping systems. [4] Think about, think about the examples that have pumping systems, right? [5] **OK**. [6] I will check your attendance first, [7] but while I'm checking your attendance, think about the example, OK? [T checks attendance.] [8] **OK**, [9] what are examples of, of what? Examples of anything which has pumping system. (2.0)

S: [10] Answer *Pahsah unggrit reu krup?* {=In English?}

T: [11] Mmm? What are examples of anything which has The pumping system.

M1: [12] Answer in English?

T: [13] What do you mean? [14] Do you know, do you understand the pumping system?

M1: [15] Yes I understand the pumping system.

T: [16] It (pumping system) 's not *kreu-ang soop num.* {= water pump.} [17] It (pumping system) is a system which has a pump, pump, water pump, right?, water pump.

S: [18] Air pump is an example which has a pumping system.

T: [19] Air pump is an example which has a pumping system. (6.0) [20] Is Anything else an example of a pumping system? [21] Water pump is an example which has a pumping system. [T writes [22] "water pump" on the board.]

S: [23] Well pump is an example which has a pumping system.

T: [24] What is Any example which has pumping system.

S: [25] Hydraulic pump is an example which has a pumping system.

T: [26] Hydraulic pump is an example which has a pumping system.

S: [27] Well pump is an example which has a pumping system.

T: [28] Huh?

S: [29] Well pump is an example which has a pumping system.

T: [30] Well pump is an example which has a pumping system.

M2: [31] *Meu-a chow don pum.* {= This morning I was raped.} [SS laugh.]

T: [32] **OK**. [33] Think about inside yourself. Inside your body.

M2: [34] *Hoo-a jý.* {= Heart.} is an example which has a pumping system

T: [35] The heart, right, the heart, the human heart is an example which has a pumping system.

S: [36] (unclear)

T: [37] The human heart is an example which has a pumping system. [T writes [38] "human heart" on the board.] [39] Do you think that your heart has this pumping system?

SS: [40] Yes our heart has this pumping system.

Extract K

T: [1] **OK**, [2] if you don't have any questions about the experiment, we're going to start Unit six. [3] Unit six is about ...

S: pump.

T: [4] Pump. What is it? Pump.

SS: [5] Pump is *Pum.* {= Pump.}

T: [6] Pump is *Pum.* {= Pump.} [7] It (pump)'s a pump. [8] What do we use it for, the pump? [9] What do we use pump for? [SS talk together.]

S: [10] (unclear)

S: [11] We use pump for *Soop num.* {= Draw water.}

T: [12] (nodding) To, we use it (pump) for moving fluid, right? [13] *Row chý sumrup tum mý ka, pum poo-ak kong laaw reu num tahng tahng.* {= We use it for what reason? For pumping fluids and kinds of water.} [14] But, but this unit is not just about pump. [15] You are, uh I play a tape [16] and you are going to hear the sound. [17] What sound is it (the sound)? [T plays a tape of a heart beating.] [18] All right, (unclear) what sound is it (the sound)?

SS: [19] The sound is a Heart.

SS: [20] The sound is *Hoo-a jý.* {= Heart.}

T: [21] The sound is a Heart.

F: [22] The sound is Heart rate.

T: [23] The sound is Heartbeat.

SS: [24] The sound is Heartbeat.

SS: [25] The sound is Beat.

T: [26] Like the name of (unclear). [27] **OK**. [28] Do you know what the heart does when it beats? [29] Do you know what the heart does when it beats? (4.0) [T holds up her hand and opens and closes it.]

SS: [30] (unclear)

S: [31] The heart Pumps.

T: [32] What the heart does when it beats. [33] It (the heart) (2.0).

S: pump

T: [34] It (the heart) contracts [T closes her hand.] and [T opens her hand.]

SS: *klai.* {= unfurls.}

T: [35] It (the heart) contracts and expands right?

SS: expands

T: [36] So, uh, *waylah hoo-a jý see ka mun* contracts [T closes her hand.] *keu mun tum arai ka?* {= when the heart contracts, what does it do?}

SS: [37] The heart *Beep too-a.* {= Contracts.}

T: [38] The heart Uh, get smaller. [39] *Mun hot too-a long,* {= It (the heart) gets smaller} [40] *laa mun go* {= and it (the heart)} expand [T opens her hand.] [41] *Mun kayai.* {= It (the heart) gets bigger.}

SS: [42] The heart *Kayai.* {= Gets bigger.}

T: [43] *Tee taa la waylah mun beep laa kayai,* {= Every time it (the heart) gets smaller and bigger,} what happens?

S: [44] *Leu-at lý.* {= Blood flows.}

SS: [45] (unclear)

T: [46] Uh, right, it (the heart) creates difference in pressure, difference in pressure. [47] **All right, OK.** [48] In this handout you are going to read ....

Extract L

T: [1] OK. (3.0) [2] I have some question for you. [3] *Fung na.* {= Listen.} [4] Do you know which part of body that is most important? (2.0) [5] Which organ? Which organ that is most important to you?

M1: [6] Eye, eye is the most important organ.

T: [7] Eye is the most important organ.

M2: [8] Ear is the most important organ.

T: [9] Ear is the most important organ.

M2: [10] Mouth is the most important organ.

F1: [11] Brain is the most important organ.

M3: [12] Heart is the most important organ.

T: [T writes [13] "heart" on the board.] [14] Heart is the most important.

M: [15] Brain is the most important organ.

T: [16] Why the heart is the most important to you? (3.0) [17] Why? Why is it (the heart) important to you? (3.0) [18] What is the function of the heart? [19] What does it (the heart) do?

M4: [20] The heart Inject.

T: [21] Huh?

M4: [22] The heart Inject.

T: [23] The heart Inject what?

M5: [24] The heart Inject blood.

T: [25] What does heart do? [T writes [26] "What does the heart do?" on the board.] [27] What does the heart do? (7.0)

S: [28] The heart *Doot chit.* {= Draw and inject.}

T: [29] The heart *Doot chit.* {= Draw and inject.}

S: [30] The heart Pump.

T: [31] Ah.

SS: [32] The heart Pump, pump.

T: [T writes [33] "pump" on the board.] [34] So it (the heart) works like a pump to pump blood, right?

Appendix B Participant data for extract G

Teacher's journal for extract G

I did quite well this time even though there were some careless mistakes. I had more confidence and less nervousness. When I prepared the lesson, a question came up to my mind. What should I do to make the lesson more interesting when eliciting responses from the students for the first 5-10 minutes of the session? I went through the handout and found the answer to my question. The topic of the lesson was "Robots". So why didn't I find something about robots which was very interesting and "in the news". I chose a film entitled "Star Wars Episode 1" since I knew that there were absolutely robots in it. The students saw a scene from the film and answered questions about what they had seen in it. It seemed very interesting and enjoyable for the students to watch a film rather than look at the pictures in handouts. This technique was effective and helped increase the students' motivation. The only problem was that the students sometimes couldn't understand my questions when I asked them in English. They didn't know the meaning of some words like "naked" and "replace". I solved this problem by using gesture and related words when giving explanations of some unknown words. If it didn't work, I would eventually explain them in Thai as I didn't want to waste time.

Tapescript of student interview for extract G

The following is a translation from Thai of the interview with two students concerning extract G. The interview was conducted while playing a video of extract G. The video was paused at the points indicated to allow the students to be interviewed. In the tapescript, R is the researcher, and S1 and S2 are the two students.

After viewing the extract up to T-unit 4

R: The teacher said that she would let you watch a video, didn't she?
S1: Yes.
R: At this point, what did you think the video would be about?
S1: A technical topic.
S2: Something like a video about science or something like that.
R: So you thought it wouldn't be interesting.
S2: Yes.
R: Before this lesson, did you know you would be learning about robots?
SS: I didn't know.

After viewing up to T-unit 7

R: After you'd seen a small extract like this, did you know it was Star Wars?

S1: Yes, I knew.

R: Had you already seen the film?

S2: Yes.

R: Were you surprised?

S1: Yes, I was surprised that the teacher wanted us to watch this video. It didn't seem related to anything we were likely to learn.

S2: I thought it was going to be a listening exercise where we would watch the video and the teacher would check how much we had understood.

After viewing up to T-unit 17

R: The teacher asked you about things happening in the film. Why do you think she was asking about this?

S1: To check whether we could understand details, to check our understanding.

R: Earlier you said that you thought the teacher would use the video to check your listening ability. Were you more certain about that at this point?

S2: I thought she was still checking our listening.

After viewing up to T-unit 20

R: The teacher asked about robots at this stage.

S2: Yes.

R: Did you still think the teacher was checking your listening?

S1: I thought she had changed.

S2: The teacher said straight out that this was about robots, so I knew the lesson was going to be about robots.

R: How did you know?

S2: The teacher emphasised robots. Before she was just asking questions, but now she was stressing robots, so I knew it would be about this.

R: A moment ago, the teacher asked, "Who is this person?" and a student answered "Mother". Why didn't you think the lesson would be about mothers?

S2: No.

R: So why did you think it would be about robots?

S2: At first, it's because, well, I've already seen this film and there is, well, the teacher asked us about this film and I've already seen it and there's robots in most of the film.

After viewing up to T-unit 38

R: At this point were you certain that the lesson would be about robots?
SS: Certain.
R: One hundred per cent?
SS: One hundred per cent.
R: But what aspect of robots did you think it would be about? ... Did you think it would be about robots in the future?
S1: It would probably be about ... what. At this point, the teacher said that, well, I thought that robots from this point on would be about how to use robots and the teacher would talk about the future, something like that, that the teacher would talk about methods, would recommend ways that robots can be used, something like that.
S2: Well, I thought she would talk about how robots are used, especially robots in the present, maybe the future.

After viewing up to T-unit 59

R: At this point, do you understand the teacher's question?
SS: I understand.
R: And why do you think she asked this?
S1: I thought she was afraid we wouldn't understand.
R: So you didn't change your mind back to listening comprehension?
S1: No, she asked us ... at first it was about our listening comprehension, but now the teacher asked, she was checking our understanding. She asked whether we could understand from listening but I didn't think the teacher would emphasise listening comprehension now, but listening was one part of the lesson but not an important part.
R: Why do you think it wasn't important?
S1: Well, the teacher didn't teach it for a long time and now the lesson was different and focuses on the differences between robots and people and how they are different, because the questions and answers were in the video such as the boy doesn't have enough money to buy, but this is background for us, so that we think and we also listen, whether we can understand or not.

After viewing up to T-unit 63

R: At this point, what were you thinking?
S1: It's about robot work for certain.
S2: What things robots can do.
S1: What are the uses of robots.

After viewing up to T-unit 70

R: At this point, the teacher is talking about the kinds of work that robots can do, isn't she?

SS: Yes.

R: And did you think that the whole lesson would be about the kinds of work that robots can do?

S2: It seemed likely to be like that. The teacher keeps on asking, how do they work and what kinds of work can they do.

After viewing up to T-unit 74

R: Do you understand this question?

S2: What question?

R: [In English] "Will robots replace people at work?"

S2: Robots replacing people at work.

R: And do you think it was strange? Because she'd just been talking about the uses of robots, and now she was talking about robots replacing people. Do you think the relationship between the two points is clear?

S1: Clear enough. She's asking us to compare people and robots.

R: But before she asked this, had you thought about comparing people and robots?

SS: Not yet.

R: You'd only been thinking about work that robots can do?

SS: Yes.

R: OK. All of this is how the teacher introduced the topic of the lesson.

S1: Yes.

R: Do you think she introduced the topic in an appropriate way?

S2: When I think about how she introduced the topic, it was appropriate, because it provided appropriate background.

S1: She made us interested to learn in this lesson.

R: And do you think she gave enough background to be able to compare robots and people easily?

S1: That depends, because what the teacher was doing was asking for our opinions more and asking us to speak out. It depends on each student and the background knowledge that each student has.

R: Do you think that the teacher should have introduced the topic more than this?

S2: She asked for our opinions so she could see whether we understood, but another thing she did in the lesson was strange. It wasn't learning any theory or language point, but it was learning together. She made us think for ourselves and use our own thoughts in this lesson.

R: OK. That's it. Thank you for helping.

Appendix C Summary of analysis identifying discontinuities and topic entities

Extract A

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	box	OK	-	toy	toy	box
2		box	box			
3						guess
4	toy	jigsaw	toy			toy
5						
6						
7						
8						
9						
10						
11						
12						
13						
14	balloon	balloon			volunteer	balloon
15						
16						
17						
18						
19						
20						
21	balloon	blow up		balloon	blow up	balloon
22						
23	volunteer					
24	balloon					
25						
26						
27						
28	blow up					
29						
30						
31	balloon -				volunteer	
32						
33	volunteer	volunteer				
34						
35						
36		who			balloon	
37		you	you			
38						
39	you	enough	-			
40						
41	OK	OK	Kittipong			
42	Kittipong	Kittipong		name		
43						
44						
45						
46						
47	balloon			balloon	balloon	
48						
49	balloon	air	I			
50						
51	(wave motions)		balloon			balloon
52						
53						
54						
55	balloon					
56						
57	balloon					
58						
59						
60						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
61				balloon	balloon	
62						
63	balloon	air		balloon		
64						
65	balloon					
66						
67			balloon		balloon	
68						
69						
70	forward	forward		forward		
71						
72						
73						balloon
74	balloon					
75	neck				neck	
76						
77	neck					
78						
79	neck					
80						
81				neck		
82		pressure				
83						
84	balloon					
85						
86						
87			balloon			
88						
89	pressure				pressure	
90						
91						pressure
92						
93						
94						
95	directions	pressure		directions		
96						
97						
98						
99						
100	balloon				balloon	
101		you				
102			you			
103						
104	technician	afraid			technician	
105						
106						
107	balloon					balloon
108						
109	count		balloon	balloon		
110						
111	forward	you			forward	
112						
113						
114	forward					
115						
116						
117	Kittipong	Kittipong	Kittipong		blow	
118		neck	neck			
119						
120			what			
121	escape					
122						
123		neck			escape	balloon
124			escape	escape		
125						
126	neck					
127						
128						
129	air	escape	escape			

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
130	air	escape	escape	escape	escape	balloon
131	forward		balloon	balloon	forward	
132						
133						
134						
135						
136	blow	Kittipong	Kittipong		small balloon	big balloon
137		I				
138		Mongkut	you			
139						
140						
141		you	you			
142						
143		thank you	-			
144						
145	balloon	small balloon	balloon		big balloon	
146						
147	balloon			balloon		
148	farther					
149						
150						
151	try	big balloon	big balloon	balloon	big balloon	
152	faster					
153						
154						
155		big balloon				
156	big balloon					
157	try	we				
158						
159	one two three	one two three	one two three			
160	what	what	-			
161						
162	blow	we	you			big balloon
163	prove		we			
164						
165	OK	OK	you			
166	blow	now				
167		you	you			
168						
169						
170						
171	you		you			
172						
173	count					
174						
175	farther	what	-			
176		big balloon	big balloon			
177						
178						
179	big balloon					
180	question					
181			you			
182						
183						
184	you	OK	you			
185	pressure	neck	air	air	action reaction	air
186		air				
187						
188		air				
189	air					
190						
191						
192						
193	action reaction	reaction	action reaction			action reaction
194						
195						
196	air			balloon	big balloon	
197	farther	big balloon	big balloon			big balloon
198						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
199	farther	big balloon	big balloon	balloon	big balloon	big balloon
200						
201						
202	big balloon					
203						
204						
205	air					
206						
207	pressure	aeroplane	pressure	action reaction	action reaction	action reaction
208	vehicle		aeroplane			
209						
210						
211						
212	pilot					
213						
214						
215						
216						
217	aeroplane				action reaction	
218						
219						
220						
221						
222	aeroplane					
223						
224						
225						
226	action reaction	rocket	action reaction			
227						
228	action reaction					
229						
230						
231						
232						
233						
234	rocket					
235						
236	jet engine					
237	engine	today	today			

Extract B

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	we	so	-	balloon	balloon	balloon
2		balloon	balloon			
3						
4	balloon					
5						
6						
7	balloon					
8						
9						
10	blow					
11						
12	balloon					
13						
14						
15	big					
16						
17						
18						
19	I	volunteer	I	volunteer	volunteer	volunteer
20						
21	volunteer				volunteer	
22		volunteer	Rubi		balloon	

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
23	volunteer	volunteer	Rubi	volunteer	volunteer	volunteer
24		Rubi				
25						
26	Rubi					
27	neck	tie	blow		balloon	
28						
29		hold	-			
30		see	-			
31		stop	-			
32		thank you	-			
33	balloon	air	balloon	balloon		balloon
34						
35						
36						
37						
38						
39						
40	air					
41						
42						
43						
44	air					
45						
46						
47						
48						
49	pressure					
50						
51						
52	balloon	OK	-			
53						
54	pressure	pressure	balloon			
55						
56						
57						
58	open end	neck	neck	balloon		neck
59						
60						
61						
62						
63						
64						
65	balloon	balloon	balloon			balloon
66						
67						
68						
69	balloon		balloon			
70						
71						
72	direction	direction				
73						
74						
75						
76	balloon					balloon
77						
78						
79						
80						
81		air				
82	we	we	we	release		release
83		too quick	-			
84	you	you	you			
85						
86		you	you			
87						
88						
89		thank you	you			
90	rubber	this	this	rubber		
91						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
92	rubber	this	this	rubber	balloon	rubber
93						
94	reaction	reaction	reaction	reaction	reaction	reaction
95						
96						
97						
98						
99						
100						
101	balloon	action reaction	balloon		action reaction	
102						
103	engine	jet	engine		engine	action reaction
104						
105	jet					
106						
107						
108						
109		jet	action reaction	action reaction	engine	
110						
111	rocket	rocket				
112						
113						
114						
115						
116						
117						
118						
119						
120						
121						
122						
123						
124	aeroplane		aeroplane		engine	engine
125						
126						
127						
128						
129						
130						
131	engine	engine				
132						
133	engine					
134						
135	bahngfý	bahngfý	you			bahngfý
136						
137						
138						
139	bahngfý		bahngfý			
140						
141						
142						
143	car	car	car	car	engine	car
144						
145						
146		quarter mile				
147						
148						
149		car	engine			
150						
151						
152	engine	diesel				
153						
154						
155						
156						
157						
158						
159						
160				engine		engine

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
161	engine					
162						
163						
164	diesel		engine			
165						
166						
167		diesel				
168						
169			-			
170	train			engine	engine	engine
171						
172						
173						
174						
175						
176	steam		train			
177						
178						
179						
180						
181						
182	petrol			petrol		
183						
184						
185						
186	petrol engine	now				

Extract C

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	we	OK	-			
2		we	we			
3						
4						
5						
6	balloon					
7						
8		balloon	balloon	balloon	balloon	
9						
10						balloon
11						
12	balloon					
13						
14						
15	OK					
16						
17		balloon	balloon		balloon	
18	balloon					
19						
20						
21		we	we	balloon		
22		I	volunteer			
23	volunteer				volunteer	
24		come on	volunteer			volunteer
25						
26		thank you	-			
27						
28	neck	balloon				
29			balloon	balloon	balloon	balloon
30	balloon	why				
31						
32	OK	OK	-			
33						
34	groups	four	groups	groups	groups	groups
35						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
36	groups	four	groups	groups	groups	groups
37						
38	reason					
39	OK	why			balloon	balloon
40						
41					engine	
42	question				fly	
43						
44		engine	engine	engine		
45						
46	engine					
47						
48						
49	Duangdeaun	Duangdeaun	-			
50	Ekchalor	Ekchalor	-			
51	where				engine	
52	engine					
53						
54						
55		engine	engine			
56						
57						engine
58	engine					
59						
60						
61		electric	electric		electric	
62				engine		
63						
64	rot tý nah					
65						
66						
67	pump					
68		what else	engine		engine	
69						
70	robot					
71						
72						
73		why	balloon		balloon	balloon
74	groups		engine	groups	engine	
75		groups	groups		groups	groups
76						

Extract D

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	OK	OK	-			
2						
3			volunteer			
4	volunteer					
5						
6		three volunteers			volunteer	
7			volunteer	volunteer		
8	volunteer					volunteer
9						
10	do					
11						
12		I				
13						
14			I			
15	balloon	not strongly				
16		not big			balloon	
17				balloon		balloon
18		enough	enough			
19		hold				
20		possible	-			
21		that	balloon			

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
22		that	-			
23			balloon			balloon
24		in a minute	-			
25			-			
26		that	enough			
27						
28			two people			
29		two people	balloon			
30	balloon					
31		I	balloon	balloon		
32						
33						
34						
35						
36						
37		I	balloon		balloon	balloon
38						
39						
40						
41						
42						
43		listen	listen			
44	OK	OK				
45						
46						
47	friends					
48						
49						
50						air
51				balloon		
52	air					
53						
54	forward	air	balloon			balloon
55						
56						
57						
58						
59	balloon					
60						
61						
62						
63	reaction					
64	you two	why	two people			
65		speak	speak			
66	OK		-			
67						
68						
69	balloon		balloon			air
70						
71						
72	OK					
73	engine					
74		engine				
75						
76						
77	engine		engine			
78						
79				engine		
80						
81						
82						
83	engine		engine		engine	engine
84						
85		action reaction	action reaction			
86	engine					
87	OK	OK				
88						
89		you	engine			
90	engine					

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
91	engine	engine	jet engine	engine	engine	engine
92						
93	jet engine			engine		
94						
95						
96						
97						
98						
99	engine	engine	petrol engine			
100						
101						
102						
103						
104						
105						
106						
107						
108						
109						
110						
111			alternative energy			
112						
113			hydrolic engine			
114						
115						
116						
117						
118						
119	we	OK	-			
120		today	we			

Extract E

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	today	yes	-	-	game	question
2		thank you	-			
3		you	you			
4	20 questions					
5						
6	question				question	
7						
8						
9						
10						
11						
12						
13						
14						
15						
16	you				question	machine
17	OK	I	question	question		
18	machine					
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						question
30						
31						
32						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
33	machine	I	question	question	object ask	question
34						
35						
36						
37						
38	wheels	Sombat	-		object	object
39		object	object	object		
40						
41						
42						
43						
44	every day	object				
45						
46						
47						
48						
49	car					
50						
51						
52						
53						
54	question	object	question	question	question	question
55						
56						
57						
58						
59						
60						
61	carry		object	object	carry	carry
62						
63						
64						
65						
66	question				object	question
67	ask	object	object			
68						
69	supermarket			object		object
70						
71						
72						
73						
74	industry			object		
75						
76						
77						
78						
79						
80	robot	answer	answer	robot	robot	robot
81						
82		you	robot			
83		robot				
84						

Extract F

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	last weekend	RCA	Star Wars	RCA	RCA	Star Wars
2						
3						
4						
5						
6						
7						
8						
9						
10	RCA					

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
11						
12						
13	Star Wars	RCA	Star Wars			
14						
15					Star Wars	Star Wars
16	Star Wars	who		Star Wars		
17						
18	OK	OK	Star Wars			
19	Star Wars	one				
20						
21						
22						
23						
24						
25						
26						robot
27	robot					
28						
29				robot	TC14	
30		3PO	TC14			
31						
32						
33						
34						
35						
36						TC14
37	TC14					
38						
39						
40						
41	transparency	we		R2D2		
42						
43						
44	robot					
45			R2D2			
46					R2D2	
47		R2D2				
48						R2D2
49	R2D2					
50						
51						
52	repair					
53				R2D2		
54						
55	one more					
56		how about				
57						
58	ET					
59						
60						
61						
62			C3PO		C3PO	
63		servant				
64						
65	C3PO			C3PO		C3PO
66						
67						
68						
69						
70	robot	that	robot			
71		3 pictures	3 pictures			robot
72	you					
73	you					
74	Picture A					
75		you	you	picture	picture	
76						Picture A
77	Picture A					
78						
79		you	Picture B			Picture B

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
80						
81	Picture B	you	Picture B	picture	picture	Picture B
82						
83						
84	machine					
85						
86						
87						
88						
89	man	where	man	man	man	man
90						
91						
92						
93						
94						
95		car				
96	Picture C					
97						
98			Picture C		car	Picture C
99	car welding	car welding		picture		
100						
101						
102	welding					
103						
104						Picture A
105	submersible	what				
106						
107						
108						
109		why				
110	hands		hands		pressure	hands
111						
112				hands		
113						
114		this condition				
115	radioactive material					
116						
117						
118	OK	OK	-			
119						
120						
121	robot				robot	
122		some people	robot	robot		robot
123						
124	robot					
125						

Extract G

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1		OK	-			
2	video				video	
3						
4	video	Star Wars	you	film		Star Wars
5						
6	Star Wars				Star Wars	
7						
8						
9	boy	where			boy	
10						
11			boy	boy		boy
12						
13	storm	why			boy	
14						
15						
16	woman	robot	boy	woman	boy	R2D2

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
17	woman	robot	boy	woman	boy	R2D2
18	robot			robot	C3PO	
19						
20						
21	robot			R2D2		
22						
23						
24	C3PO					
25						
26	robot	R2D2	R2D2		R2D2	
27						
28						
29						
30						
31						
32						
33	R2D2			R2D2	R2D2	
34						
35						
36						
37	OK	these	robot	-	robot	robot
38	robot			robot		
39	robot	robot				
40						
41						
42						
43						
44						
45						
46	robot					
47						
48	robot					
49						
50						
51						
52						
53						
54	robot					
55	seen	who	Star Wars	seen	seen	robot
56		-				
57		who	robot			
58						
59						
60	robot	robot		robot	robot	
61						
62						
63						
64	robot					
65						
66						
67						
68	robot					
69						
70						
71	robot					
72	question	question	question			
73						
74						

Extract H

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	robot	robot	-	robot	robot	robot
2		motor machine	-			
3	toy	robot	robot			
4						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
5	toy	robot	robot	robot	robot	robot
6						
7						
8						
9						
10	battery					
11						
12						
13						
14						
15						
16						
17	robot	OK	tank	robot	robot	
18		Picture D				
19	robot					
20						
21				tank		
22						
23						
24	Picture B	tank			tank	tank
25						
26	tank					
27						
28						
29						
30	lunar module	lunar module	lunar module		lunar module	lunar module
31						
32		this one	Picture C			
33	passage	now	-			
34		read	-			

Extract I

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	OK	OK	-	ready	ready	ready
2	ready	you	picture			
3						
4						
5						
6						
7	picture	pump			picture	picture
8						
9						
10						
11						
12						
13	pump			pump	water pump	water pump
14						
15						
16						
17						
18						
19		water pump	water pump			
20						
21						
22						
23						
24	water pump					
25						
26						
27						
28	pumping water			pump		
29						
30						
31						
32						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
33	pumping water	water pump	water pump	pump	water pump	water pump
34						
35	pump	pump	pump		pump	
36						
37						
38						
39	OK	OK	you		bicycle	bicycle
40	bicycle	air pump				
41						
42						
43						
44						
45					air pump	air pump
46						
47	air pump					
48						
49						
50						
51						
52	air pump				flat tyre	
53						
54						
55						
56						
57	flat tyre	you				flat tyre
58						
59						
60						
61						
62	body	body	heart	pump	pump	pump
63						
64						
65						
66	pump			heart	heart	heart
67						
68						
69						
70						
71	heart					
72						
73	heart	heart				
74						
75	heart					
76						
77						
78	heart					
79						
80						
81	you	thank you	you			
82	OK	OK	-			

Extract J

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	pumping system	pumping system	-	pumping system	pumping system	pumping system
2	pumping system	OK	pumping system			
3		I				
4						
5	attendance				attendance	attendance
6						
7						
8	OK	what			pumping system	pumping system
9	pumping system	English	pumping system		answer	
10						
11					pumping system	pumping system
12						

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
13	pumping system	you	pumping system	pumping system	pumping system	pumping system
14						
15	pumping system					
16						
17						
18	air pump	water pump				
19						
20	water pump					
21						
22						
23						
24	pumping system					
25						
26						
27						
28						
29						
30						
31		this morning	-			
32	OK	OK	-			
33	heart	heart	heart		heart	
34						
35						
36						
37						
38						
39	heart					
40						

Extract K

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	you	OK	-	pump	pump	pump
2		pump	pump			
3	pump					
4						
5	pump					
6						
7						
8	pump					
9						
10						
11						
12						
13						
14	you	this unit		-	pump	
15		heart	sound	sound	heart	sound
16						
17	sound					
18						
19						
20						
21						
22						
23						
24						
25						
26						
27	OK	OK	-			heart
28	heart	heart	heart	heart		
29						
30						
31						
32						
33						
34	heart			heart		

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
35	heart	heart	heart	heart	heart	heart
36	heart					
37						
38						
39						
40						
41						
42						
43	heart					
44						
45						
46						
47	All right	All right	-			

Extract L

	Sinclair/Coulthard	Theme-rheme	Given-new	Hoey (1991)	Topic-based (relations)	Topic-based (associations)
1	question	OK	-	-	heart	most important organ
2		I	I			
3	listen	listen	most important organ	most important organ		
4	most important organ	heart				
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16	heart					-
17						
18	heart	heart	heart	heart		heart
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						