Connecting Textual Patterns to Text Aboutness

Volume 1: thesis

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

Li Yuan ke

November 2011

Abstract

This thesis investigates the relationships between cohesion in texts and the meanings of the texts. I analyze hard news texts and the summaries written by competent readers for the texts in terms of nuclei (i.e. the combinations of Process and Medium) (Halliday, 1994), and examine the characteristics of the nuclei that recur in a text and are also considered as important to the meaning of the text. There are two main findings of my present study. The first is that when two or more recurrent nuclei in a text are considered as important to the meaning of the text, the relationship in the lead that holds between the nuclei are also thought of as important to the text's meaning. This finding provides evidence for the claim made by many linguists that the conjunctive relations that hold between the propositions in a passage are important to the meaning of the passage. The other main finding of my study is that when the nuclei that are considered as important to the meanings of the texts in which they occur are found in leads, they have a strong tendency to occur in primary or independent clauses; on the other hand, when the nuclei occurring in leads are however not considered as important to the meanings of the texts in which they occur, they tend to occur in secondary or embedded clauses. This finding indicates that the distribution of information in leads is not random. The propositions in a lead that are important to the meaning of the text in which they occur are often foregrounded in major types of clause, while those that are not thought of as important to the meaning of the text are often backgrounded in minor types of clause.

Acknowledgements

There are many people I want to thank, first and foremost, my parents. They for years have been funding me studying in UK, and gave me advices and assistance whenever I needed. I also want to thank all tutors in the School of English at the University of Liverpool who I have been working or once worked with, especially Geoff Thompson, Dr. Mike Scott and Pro. Michael Hoey. Geoff kindled my interests in learning functional grammar, answered my questions, and gave me papers that were difficult for me to get. Mike supervised me in the first two years of my PhD, guiding me through the initial stages, and Pro. Michael Hoey supervised me in the last two years. Michael enthused me about patterns of lexis in a lecture five years ago, and I have been working on this topic ever since. He very patiently read the drafts of my thesis, pointed out the problems that I needed to deal with, and advised me on the ways to solve the problems. Writing a mammoth PhD thesis is daunting to me, but Michael helped me to achieve it.

For my dear Dad and Mum

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 A short review of studies on cohesion and of aspects of cohesion that have not been explored sufficiently

The early 1970s saw the emergence of studies on cohesion, that is, studies that explore 'the way certain words or grammatical features of a sentence can connect that sentence to its predecessors (and successors) in a text' (Hoey, 1991: 3). Most of the early work on cohesion was concerned with categorizing cohesion and discussing the lexico-grammatical devices that realize different types of cohesion. Halliday and Hasan's (1976) *Cohesion in English*, which is widely considered as a landmark book on cohesion, discusses the nature of cohesion and the contribution of cohesion in making a text a meaningful product that does not consist of 'a collection of unrelated sentences' (Halliday and Hasan, 1976: 1), categorizes cohesion and surveys the different types of cohesive devices. Inspired by *Cohesion in English*, studies on cohesion burgeoned in the two decades that followed, but the studies that followed often did not research cohesion for its own sake but associated cohesion with particular characteristics of text. For instance, Hasan (1984), Halliday and Hasan (1985), and Parsons (1991a, 1991b) examine the contribution of patterns of lexis in text to the coherence of text. Hoey (1991) discusses the contributions of patterns of lexis in text to the coherence, organization and meaning of the text. Mann and Thompson (1988a, 1988b) discuss the contributions of the relations that hold between the clauses or larger parts of a text to the organization, coherence and meaning of the text.

There are, however, two aspects of cohesion that have not been explored sufficiently in previous studies on cohesion. Firstly, a number of previous works that discuss the relations between the clauses or larger parts of a text, such as Eggins (1994, 2004), Halliday and Hasan (1976), Halliday (1994), Halliday and Matthiessen (2004), Hoey (1983, 2001), Martin (1992), Martin and Rose (2003), Martin *et al* (2010), Mann and Thompson (1988a, 1988b), McCarthy (2002), Matthiessen and Thompson (1988), Matthiessen (2002), Quirk *et al* (1972, 1985), Thompson

(1996, 2004, 2005, 2010), and Winter (1977, 1994), have claimed that the relations that hold between the propositions contained in a text are important to the meaning of the text. But these works have provided relatively little empirical evidence in support of this view.

Secondly, there have been only a small number of studies on cohesion that have explored the impact of cohesion in a text on readers' interpretations of the text's meaning, though Kintsch and Keenan (1974), and Kintsch and Van Dijk (1978) are exceptions. These two studies analyze texts in terms of the propositions contained in them. A proposition is described by them as consisting of a process and one or more of its arguments. Both studies found that the propositions that share arguments with many other propositions in a text are recalled by readers much more often, and are therefore more important to the meaning of the text than, those that share arguments with few other propositions. A natural way to investigate recall is to ask readers to summarize what they have read, but few studies have examined summaries in this light. There has been research into the memory value of summaries and the methods that are often used by readers to write summaries. Reder and Anderson (1980), for example, asked some university students to read chapters taken from college text books and other university students to read abstracts that summarized the main points of the selected chapters. They found that the students that read the abstracts of the chapters recalled the main points of the chapters much better than those that read the chapters. Kintsch and Van Dijk (1983) noted that *deletion*, *generalization* and *construction* are often used by readers to summarize texts. The works of computational linguists that are concerned with automatic summarization of text often use summaries written by competent readers to evaluate the outputs produced by the summarizing schemes they developed (see Mani and Maybury, 1999). None of these studies have however examined summaries with a view to exploring the impact of cohesion in a text on readers' interpretations of the text`s meaning.

1.2 The aim of my present study

The aspects noted above that have not been explored sufficiently in previous studies on cohesion will be dealt with in my present study. The aim of my study is to investigate the relationships between cohesion in a text and the meaning of text. Most previous studies that worked on patterns of lexis in a text, such as Hasan (1984), Halliday and Hasan (1985), Parsons (1991a, 1991b) and Hoey (1991), have analyzed texts in terms of the repetitions of single elements. My present study, on the other hand, analyzes texts in terms of the repetitions of the combination of two elements, a Process and its Medium. I analyze texts in such a way because according to the Hallidayan approach, the Process is the centre of a clause, and the Medium is more closely related to the Process than any other type of participant roles (see Halliday (1994: 163-164), Halliday and Matthiessen (1999: 174)). More explanations are given below. A clause consists of a Process and of participant(s) being involved in the Process; and in addition there may be circumstance(s) of time, space, cause, manner, etc that qualify the Process in some way. Therefore, the Process is the centre of a clause, and both the participants involved in the Process and the circumstances relate to the Process in some ways. Of the participants involved in the Process, there is one type defined by Halliday as 'the participant through which the Process is actualized' (1994: 163). This type of participant is referred to by Halliday (1994: 163) as the Medium. The Medium is obligatory in all process types (the participant roles that act as the Medium in different process types will be given on pages 135 and 136). The Process and its Medium together form 'the nucleus of a clause' (Halliday, 1994: 164).

Analyzing the nuclei that recur in a text is in effect analyzing the propositions that recur in the text. The term proposition derives from logic, and, as already noted, is widely viewed by many authors, such as Brown and Yule (1983), Kintsch and Keenan (1974), and Kintsch and Van Dijk (1978), as consisting of a process and one or more of its arguments. The nucleus of a clause can be seen as a miniature representation of the proposition expressed by the clause.

My present study also analyzes the summaries written by competent readers to give what they think of as the gist of the texts that they read, with a view to identifying which of the recurrent nuclei in a text that they are summarizing seem to be important to the meaning of the text, and exploring their characteristics.

In short, my present study analyzes texts in a new way, and explores the characteristics of those cohesive features in texts that also occur in the summaries, with a view to investigating the impacts of cohesion in those texts on readers' interpretations of the texts' meaning. I hope these will give sufficient reasons for reading the thesis.

1.3 The organization of the thesis

The next chapter, chapter 2, reports on studies that have investigated the contributions of cohesion in a text to the coherence, organization and meaning of the text. The chapter begins with the works of Hasan, one of the co-authors of *Cohesion in English*, which discuss the contribution of patterns of lexis in stories written by young children to the coherence of those texts. I point out the limitations of her method used to analyze the texts, and discuss ways of refining her method. I then report on Parsons (1991a, 1991b) who tested a range of Hasan's claims. I then discuss Hoey (1991), who researched a type of cohesion that was not sufficiently dealt with in *Cohesion in English*, namely lexical cohesion, and discussed the contribution that patterns of lexis in a text make to the coherence, organization and meaning of the text. The chapter then examines Scott's studies on Keywords in texts. Keywords refer to words that occur with unusual frequency when compared with a reference corpus (Scott, 2010). Both Hoey (1991) and Scott's studies on Keywords show that the way lexis is patterned in a text is often indicative of what the text is about. Finally, the chapter discusses Hoey's later work (2001) and Thompson *et al*'s Rhetorical Structure Theory (1988a, 1988b) with regard to the

contribution that the relations that hold between clauses or larger parts of a text make to the coherence, organization and meaning of the text.

Chapters 3, 4 and 5 report on my present study, in which two questions given below are explored.

1. Which recurrent nuclei in a text are thought of by readers as important to the meaning of the text?
2. Do the recurrent nuclei in a text that are thought of by readers as important to the meaning of the text tend to appear in different types of clause contained in the lead from those that are not thought of as important to the meaning of the text?

Chapter 3 explains the method used to explore the first question. The method involves unpacking nominalizations. Previous studies on nominalization have not explored the appropriateness of unpacking nominalizations. Section 3.3.2 of the chapter contains the criteria that determine whether nominalizations in a text are suitable for unpacking. Section 3.2.6, similarly, gives the criteria that determine when to unpack nominalizations in summaries.

Chapter 4 reports the results of my investigation of the first research question and explores the characteristics of the recurrent nuclei in a text that are thought of by readers as important to the meaning of the text. The chapter begins with classifying the texts I analyzed into categories that indicate which recurrent nuclei have been thought of as important to the meaning of the texts in which they occur, and then discusses the characteristics of these nuclei. I examine the distribution in texts of the most frequent nuclei that are thought of as important to the meaning of the texts in which they occur. When two or more recurrent nuclei in a text are thought of as important to the meaning of the text, I investigate whether the relationship in the lead that links them is also considered as important to the meaning of the text. After that, I also examine the

probability of the presence in the leads of the recurrent nuclei in a text that are considered as important to the meaning of the text.

Chapter 5 explores the second research question. This chapter consists of two parts. The first part gives the method used to explore the question, and discusses the results and implications. The second part examines the nuclei that occur in secondary or embedded clauses contained in the leads but which are nevertheless thought of by readers as important to the meaning of the texts in which they occur, and explores the reasons why they are thought of as important, despite occurring in minor types of clause contained in the leads.

The final chapter brings the findings of my present study and those noted in the works of other authors together with a view to identifying better the relationships between the cohesion in a text and the meaning of the text. It also discusses information distribution in the lead and summarizes what previous studies and my own have found regarding the connection between the conjunctive relations that hold between the propositions contained in a text and the meaning of the text. Finally I note the paths that future research may take.

Chapter 2 Literature Background

2.1 Introduction

Inspired by Halliday and Hasan's (1976) *Cohesion in English*, studies on cohesion burgeoned. The studies that followed often associated cohesion to particular characteristics of text or in effect answer one or more of the following questions:

1 How does the presence of cohesion in a text contribute to the coherence of the text?

2 How does cohesion contribute to the meaning of a text?

3 How does cohesion contribute to the organization of a text?

This chapter discusses the works that have implications for answering one or more of the questions. We begin by discussing the works that answer the first question. The way in which the question is expressed implies that cohesion does contribute to the coherence of a text. Authors such as Brown and Yule (1983) and Widdowson (1978, 2004), however, contend that it is neither necessary nor sufficient for a text to have cohesive features to make it coherent. On the other hand, Hasan's, Parsons' and Hoey's studies reported below show that cohesion indeed plays an important part in making a text coherent.

2.2 Hasan (1984) and Halliday and Hasan (1985)

One of the authors of *Cohesion in English*, Hasan, again pioneered the studies of the relationships between cohesion in a text and the coherence of the text. She examines the relationships between the two in her 1984 paper and her 1985 book that she co-authored with

Halliday.

She uses 'cohesive chain' (1984, 1985) to explain the textual phenomenon that particular entities, concepts, events and qualities, etc are often recurrent at intervals in texts, running through texts vertically to indicate the continuities/relatedness among different parts of texts. A cohesive chain is formed when an item repeats earlier items and/or is repeated by subsequent items in texts. The semantic relations that enable items in texts to form chains are *co-reference*, *co-classification*, and *co-extension* (Halliday and Hasan, 1985: 84).

Co-reference is the meaning relation whereby every member of a chain refers to the same thing, event, or entity. This relation is typically realized by reference, such as the pronominals *he*, *she*, *it*, etc; or by the use of definite article *the* plus a noun that in context refers to the thing mentioned earlier (e.g. A *dog* is subsequently referred to as *the animal*); or by demonstratives *this* or *that* (Halliday and Hasan, 1985: 74). When these items are used to refer to the thing, event or entity mentioned earlier in the text, these items and their referents form an identity chain, a kind of chain that is made up of items expressing the same referent (Halliday and Hasan, 1985: 84). For instance, in Example 2.1 given below (reproduced from Halliday and Hasan, 1985: 72), the *girl* at the beginning of the text is subsequently referred to as *she* and *her*. *Girl*, *she* and *her* form an identity chain because the *girl* and each subsequent occurrence of *she* and *her* have the same situational identity of reference.

Example 2.1:

1. Once upon a time there was a little girl
2. and she went out for a walk
3. and she saw a lovely little teddybear
4. and so she took it home
5. and when she got home she washed it
6. and when she took it to bed with her she cuddled it
7. and she fell straight to sleep
8. and when she got up and combed it with a little wire brush the teddybear opened his eyes
9. and started to speak to her
10. and she had the teddybear for many many weeks and years
11. and so when the teddybear got dirty she used to wash it
12. and every time she brushed it it used to say some new words from a different country
13. and that's how she used to know how to speak English, Scottish, and all the rest.

Co-classification is the meaning relation in which the things that different members of a chain refer to belong to the same class, with each member referring to distinct member of this class (Halliday and Hasan, 1985: 74). Co-classification is different from co-reference in that items that stand in a co-reference relation have the same situational referent, whereas items that stand in a co-classification relation belong to the same class but do not have the same situational referent. Co-classification is normally 'realized either by substitution or ellipsis' (Halliday and Hasan, 1985: 74). For instance, in Example 2.2 given below (also reproduced from Halliday and Hasan, 1985: 74), the *'pen'* contained in the first utterance is a different object from the one contained in the second utterance (elliptical and included in *'yours'*, which is equivalent to *'your pen'*), but they belong to the same class.

Example 2.2:

-----'Can I borrow your pen?'

-----'Yes, but what happened to yours?'

Co-extension, on the other hand, is a third kind of meaning relation in which an item of a chain relates to the others by way of being 'in the same general field of meaning' (Halliday and Hasan, 1985: 74). This kind of meaning relation is realized by any of the following sense relations: synonymy, antonymy, hyponymy, meronymy and repetition[1]. In Example 2.1 *'went'* in sentence 2 stands in a co-extension relation with both *'walk'* that also occurs in this sentence and *'got'* (in the sense of *reached*) contained in sentence 5 because 'walking is a kind of going, and going is an important part of getting anywhere' (Halliday and Hasan, 1985: 85).

Co-classification and co-extension are the two types of meaning relation that enable items that are related to each other in such ways to form similarity chains. In other words, the members of a similarity chain are related to each other by co-classification and/or co-extension (Halliday and Hasan, 1985: 85). The fundamental difference between identity chains and similarity chains lies in that an identity chain is formed by items that co-refer to the same thing, event or whatever, whereas a similarity chain is made up of items that describe related or similar actions, events, objects or their attributes. To put it another way, sameness is the principle whereby identity chains are formed, whereas relatedness or similarity is the principle whereby similarity chains are constituted.

Table 2.1 below gives the chains contained in Example 2.1 that were identified by Halliday and Hasan (1985). The numbers in brackets show the numbers of tokens included in the chains. For example, identity chain (a) includes 17 tokens in total (including pronouns) that co-refer to *the*

[1] Henceforth, both Hasan (1984) and Hasan (1985, with Halliday) use 'repetition' to refer to distinct morphological forms of a lexical item (e.g. *suggest* and *suggestion*).

girl.

Table 2.1 Cohesive chains contained in Example 2.1 (adapted from Halliday and Hasan, 1985: 90)

Identity chains:	(a) girl (17)
	(b) teddy bear (14)
	(c) home (2)
Similarity chains:	(d) was got [in 11, = became]
	(e) went walk got [in 5, = reached]
	(f) lovely dirty
	(g) wash (2) comb brush
	(h) took had [in 10, = owned]
	(i) weeks years
	(j) many (2) some
	(k) new different
	(l) speak (2) say
	(m) took-to-bed fell-to-sleep got-up opened-eyes
	(n) words English Scottish all-the-rest
	(o) little (3)

So far I have discussed the distinctions between identity chains and similarity chains, and the meaning relations defining these two types of chain. I have also explained the lexico-grammatical devices that are typically used to realize each kind of meaning relation. The discussions above are summarized in Table 2.2 below (adapted from Halliday and Hasan, 1985: 82).

	Device	Typical tie relation
Grammatical Cohesive Devices	A: Reference 1. Pronominals 2. Demonstratives 3. Definite article 4. Comparatives	co-reference
	B: Substitution & Ellipsis 1. Nominal 2. Verbal 3. Clausal	co-classification
Lexical Cohesive Devices	A: General 1. Repetition 2. Synonymy 3. Antonymy 4. Hyponymy 5. Meronymy	co-classification or co-extension
	B: Instantial 1. Equivalence 2. Naming 3. Semblance	co-reference or co-classification

Table 2.2 Summary of relations that enable items to be put in a chain (adapted from Hasan, 1985: 82)

The table shows that Halliday and Hasan (1985) classified lexical cohesion into two major kinds: general and instantial. General relations refer to the relations that are well-established and acknowledged in the English language system. They are 'dissociated from a real context of utterance, and yet constitute a valid example of this meaning relation' (Hasan, 1984). In both Hasan (1984) and Halliday and Hasan (1985) general lexical cohesion was subcategorized into repetition (referring to distinct morphological forms of a lexical item), synonymy, antonymy, hyponymy and meronymy (part and whole relation). Examples of these relations, reproduced from Halliday and Hasan (1985), are given below.

i repetition e.g. leave, leaving, left
ii synonymy e.g. leave, depart
iii antonymy e.g. leave, arrive
iv hyponymy e.g. travel, leave (including co-hyponyms: leave, arrive)
v meronymy e.g. hand, finger (including co-meronyms: finger, thumb).

Classifying general lexical cohesion in such a way improves Halliday and Hasan's (1976) model, which classified lexical cohesion into two major types, reiteration and collocation. Reiteration was further classified into repetition (distinct morphological forms of a lexical item), synonym (or near-synonym), superordinate and general word (Halliday and Hasan, 1976: 288). It should be noted that 'collocation' is used by Halliday and Hasan (1976) in a way that is very different from that now adopted by many corpus linguists, i.e. the textual phenomenon that two words co-occur in a designated span with a probability that is more than would be expected by chance (Scott and Tribble, 2006). Halliday and Hasan (1976), instead, use it as a ragbag to include assorted lexical relations, but they do not further classify the relations subsumed in this category. Examples given by them (1976: 285-286) that are included in this category are *boy-girl*, *wet-dry*, *order-obey*, *poetry-literature-reader-writer-style*, *hair-comb-curl-wave*, *candle-flame-flicker*. Later Hasan (1984: 202) concedes that the way in which 'collocation' was defined in Halliday and Hasan (1976) was not suitable for consistent analysis. She makes the following statement in her 1984 paper.

> While I firmly believe that behind the notion of collocation is an intuitive reality, I have come to accept the fact that unless we can unpack the details of the relations involved in collocation in the Firthian sense, it is best to avoid the category in research. The problems of inter-subjective reliability cannot be ignored. (p. 195)

Therefore, she abandons 'collocation' and does not include it as a subcategory of general

lexical cohesion in her 1984 and 1985 works. But to compensate for the loss of this category, some relations that were previously handled in it, such as antonymy and meronymy, are made subcategories of general lexical cohesion in the two works.

Instantial relations are the other major type of lexical cohesion in her 1984 and 1985 works. They refer to the relations that are text-bound. 'Their validity is an artifact of the text itself, and does not extend to the system' (Hasan, 1984). In both Hasan (1984) and Halliday and Hasan (1985) instantial lexical cohesion is subcategorized into equivalence, naming and semblance. Examples of these three relations given by Hasan (1984: 202) are reproduced below. In each example the two items that are italicized stand in the relation that precedes the example.

i Equivalence: e.g. the *sailor* was their *daddy; you* be the *patient*, *I*'ll be the *doctor*
ii Naming: e.g. the *dog* was called *Toto*; they named the *dog Fluffy*
iii Semblance: e.g. the *deck* was like a *pool*; all my *pleasure* are like *yesterdays*

The addition of instantial lexical cohesion to her 1984 and 1985 works is another improvement of Halliday and Hasan's 1976 model. This kind of relation was not a category of lexical cohesion in their 1976 work. Making it a category of lexical cohesion in her later works is a big step-forward because instantial lexical cohesion is also thought of as an important type of device to create cohesive relation between two items in a text. Since then, the importance of instantial lexical cohesion in creating cohesive relation has been noted by many authors. McCarthy (1988) points out that three kinds of instantial relation, namely equivalence, opposition and inclusion, which correspond to general relations of synonymy, antonymy and hyponymy respectively, are important to account for the lexical cohesion that occurs in

speakers' utterances in his spoken data. Carter and McCarthy (1988) discuss the recognition of instantial relations in a text. They (1988: 204) make the following statements.

> Some commentators have recently suggested that cohesion is not so much a set of overt semantic markers which signal a coherent chain of meaning in a text, as a set of meaning *potentials* which can only be realized when the reader/listener taps his or her knowledge of the world and of how situations characteristically manifest themselves in our experience. In other words, experience organizes itself into a series of *schemata* or mental constructs, whereby we continuously reclassify phenomena. Only when we activate the appropriate schema for a given language situation or text can we then ultimately extract a coherent message from the flow of speech or writing. Experience builds on experience; we make links between vocabulary items in a discourse because we are predisposed to do so; we are more or less looking for them once the appropriate schema has been activated. It is only then that we can process the cohesive elements of the surface text (Original italics).

In their statements they stress the importance of schemata in recognizing instantial lexical cohesions. However, as noted in Hoey (2001: 122), schemata are not practical analytical tools. We need a method that enables us to identify the instantial relations that hold between two items in a text. Hoey (1983, 1991) provides one: the grammatical cohesion and/or general lexical cohesion that hold between other items occurring in the clauses or sentences in which the two items in question occur often aids the identification of the instantial lexical cohesion between the two items; signaling devices and parallelism may also occur to aid the identification (see the 'matching relation' that may hold between two clauses, as discussed in Hoey (1983, 2001) and section 2.7 below). This is illustrated with the example given below that was used by McCarthy (1988: 184) to illustrate instantial relation of equivalence. Two speakers, A and B, participated in a conversation.

A: So you want to meet Harry

B: Yes and I'm dying to see Bill too.

McCarthy (1988: 184) notes that providing that the intonation of *I'm dying to* that occurs in speaker B's utterance is non-prominent, it is largely synonymous with *you want to* that occurs in speaker A's utterance. There are clues in the conversation that lead us to make this interpretation. The adjunct *too* in the final position of speaker B's utterance signals that the clause *I'm dying to see Bill* that occurs in speaker B's utterance is 'matched' (Winter, 1974, 1986; Hoey, 1983) for similarity with speaker A's utterance. The instantial equivalence of *I'm dying to* and *you want to* is nestled in a textual context in which other pairs of items also form cohesive relations. *I* in speaker B's utterance is in a repetition relation with *you* that occurs in speaker A's utterance because they, in the context of the conversation, co-refer to speaker B. *See* in speaker B's utterance also stands in a general relation of synonymy with *meet* that occurs in speaker A's utterance. The repetitions of pairs of items in the two speakers' utterances create parallelism between the two utterances, shown in table 2.3 below.

A:	you	want	to	meet	Harry
B:	I	dying	to	see	Bill
constant	same referent (speaker B)	mental process	same preposition	repetition by synonymy	people
variable		different verbs			different persons

Table 2.3: The parallelism between speaker A's and B's utterances

The repetition relations that hold between pairs of items in the two speakers' utterances, the parallelism created by the repetitions and the signalling device *too* jointly indicate that a matching relation holds between the two speakers' utterances. This has an impact on our recognition of an instantial equivalence relation that holds between *I'm dying to* and *you want to* in the conversation.

Co-reference, co-classification and co-extension are the principles that enable items to form cohesive chains. They produce vertical unity in text (Halliday and Hasan, 1985:92) in that they permit items that are related to each other in such ways to be part of the same chain due to either referential equivalence (in the case of co-reference) or semantic relatedness/similarity (in the case of co-classification and co-extension). Chain formation is concerned with repetition among components of messages.

It is, on the other hand, important to look at how members of different chains connect in order to express complete messages. This leads to what Hasan refers to as the horizontal vector of unity in text, namely chain interaction (Halliday and Hasan, 1985: 91). The reason why chain interaction is important is given below:

> It is important to recall here that in constructing chains, we are concerned with components of messages. Our entire analysis has revolved around components rather than whole messages as such. On the other hand, it is only message as message that has any textual viability; and it is only at the rank of clause or above that a lexico-grammatical unit is contextually viable: it is only at this rank or above that a linguistic unit can encode a complete message. Although the chains go a long way towards building the foundation for coherence, they are not sufficient: we need to include

> some relations that are characteristic of those between the components of message. This is the relation that I refer to as CHAIN INTERACTION (original capitalization).
>
> (Halliday and Hasan, 1985: 91)

The relations that Hasan refers to as linking components of message and enabling members of different chains to interact are TRANSITIVITY relations. The minimum requirement of chain interaction is that '**at least two members of one chain should stand in the same relation to two members of another chain**' (Halliday and Hasan, 1985: 91, my emphasis). The interactions between the chains contained in Example 2.1 that were given in Halliday and Hasan (1985: 92) are reproduced in Figure 2.1 below. The chain labels are the same as those in Table 2.1. To save space, the chains do not include all their tokens but only those tokens that interact with tokens of other chains. Each interacting segment of the chain is boxed together to make the interaction display easier to follow. For example, the first and the second entries of *girl* interact with *went* and *got* contained in chain (e); the second and the third entries of *girl*, on the other hand, interact with two instances of *home* contained in chain (c). Each arrow registers an interaction between a (segment of a) chain and (segment of) another chain.

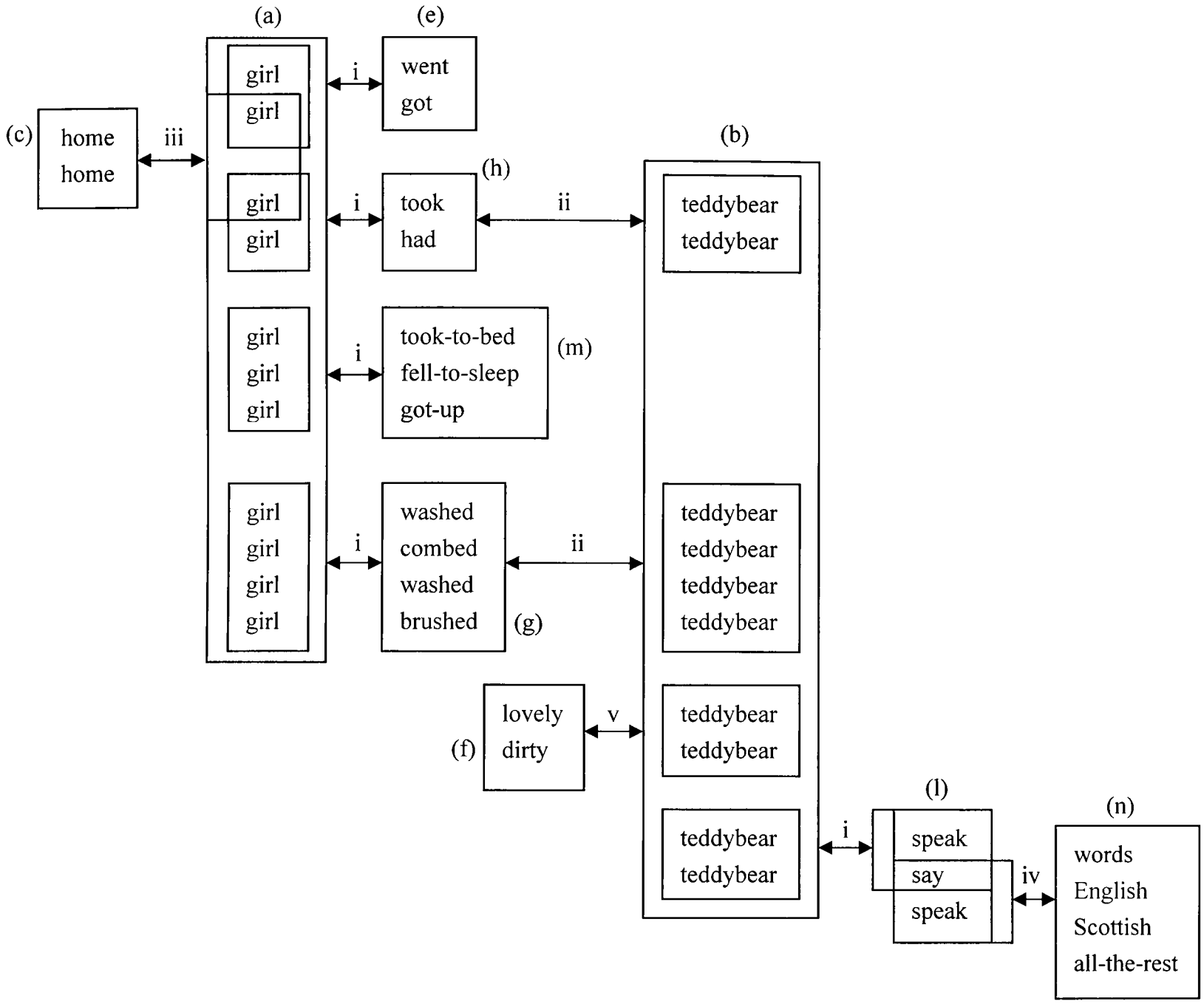


Figure 2.1 Chain interactions in Example 2.1 (reproduced from Halliday & Hasan, 1985: 92)

The Roman numbers on top of the arrows refers to the following relations (reproduced from Halliday and Hasan, 1985: 93):

i are in 'actor action relation' (for example, *girl went*)
ii are in 'action acted-upon relation' (for example, *took teddy bear*)
iii are in 'action location relation' (for example, *girl got home*)
iv are in 'saying text' relation (for example, *said words*)
v are in 'attribute attribuand' relation (for example, *lovely teddy bear*)

Chain formation and chain interaction are two complementary perspectives on textual unity.

While chain formation is concerned with the vertical unity among different members of an

individual chain, chain interaction examines the horizontal relations between two or more members of one chain and two or more members of another chain. These two vectors of textual unity are used by Hasan to examine the relationship between cohesive patterns in texts and texts' coherence. She concludes that the degree of coherence in a text (measured by readers' evaluations) is linked to the degree of interaction between cohesive chains (Hasan, 1984; Halliday and Hasan (1985)).

The degree of interaction between cohesive chains is measured by what she refers to as 'cohesive harmony' (Hasan, 1984; Halliday and Hasan (1985)). Relevant, peripheral, central and non-central tokens are used to measure the coherence of a text. Relevant tokens refer to the tokens that enter into chains. Peripheral tokens, on the other hand, refer to the tokens that do not belong to any chain. Central tokens refer to the tokens that enter into chains and interact with tokens of other chains. Non-central tokens, on the other hand, refer to the tokens that enter into chains but do not interact with tokens of any other chain. In two separate publications Hasan gives somewhat different accounts of cohesive harmony. In her 1984 paper (p218), she writes:

1. Any text will be seen as coherent, in which the central tokens form at least 50 percent of the total tokens. This percentage may be treated as a measure of their cohesive harmony;

2. Ranking by cohesive harmony will match the ranking of the texts on the cline of coherence by reference to informal reader/ listener reaction: cohesive harmony is the lexico-grammatical reflex of the semantic fact of coherence;

3. If two texts display no significant difference in their cohesive harmony, variation in their coherence will correlate with interactive gaps: all else being equal, the larger the number of such gaps, the less coherent the text would be:

4. The ratio of central tokens to peripheral tokens is associated with coherence: the higher the ratio of central to peripheral tokens, the more coherent the text will be, all else being assumed equal.

The first parameter was not included in her 1985 revision of cohesive harmony (given below). The second parameter was referred to as the underpinning principle of using different types of ratio to measure texts' coherence in her 1985 revision, but it was not one of the parameters of cohesive harmony in the later version. In addition, the ratio of central tokens to peripheral tokens contained in the fourth parameter was replaced by two other types of ratio in the 1985 revision. It is replaced by the ratio of peripheral tokens to relevant tokens contained in the first parameter in the 1985 revision, and by the ratio of central tokens to non-central tokens contained in the second parameter of this revision.

The parameters of cohesive harmony in Halliday and Hasan (1985: 93-94) are:

1. The lower the proportion of the peripheral tokens to the relevant ones, the more coherent the text is likely to be;
2. The higher the proportion of the central tokens to the non-central ones, the more coherent the text is likely to be;
3. The fewer the breaks in the picture of interaction, the more coherent the text.

Hasan's studies have two important implications for my research purpose. Firstly, it is more fruitful to examine chain interactions than consider individual chains in an isolated manner (i.e. consider each chain independently and do not consider the interactions of chains) because the second parameter of cohesive harmony in the 1985 version notes that central tokens are related to the coherence of text. Secondly, it is rewarding to consider the interactions between chains denoting processes and chains denoting entities. Figure 2.1 contains ten chains in total, five of which, namely chain (e), (h), (m), (g) and (l), denote processes. All these five chains interact

with chains that denote entities: chain (e) interacts with chain (a) (i.e. *girl*); chain (h) interacts with both chain (a) and chain (b) (i.e. *teddybear*); chain (m) interacts with chain (a); chain (g) interacts with both chain (a) and chain (b); chain (l) interacts with both chain (b) and chain (n) (i.e. *language*). These two implications lead me to examine in my present study the interactions between chains that denote processes and chains that denote entities.

2.3 The limitations of Hasan's method and improvements

Hasan used the method described above to analyze simple texts written by children aged between six and seven (Hasan, 1984). Her method, however, is not always capable of handling the sophisticated use of language in texts written by adults. This section discusses two limitations of her method and proposes improvements.

The first limitation is that some lexical repetitions that occur in my data cannot be handled by her categories of lexical cohesion (shown in Table 2.2 given earlier). An example is given in Example 2.3 below, which contains four sentences taken from text 1[2]. This text reports that no evidence was found to show that then Korean Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in the assassination of former Korean President Pak Chong-hui in 1979. The extract contains three occurrences of *assassination*, two occurrences of *killed*, and one instance of *slaying*. All these items are emboldened.

[2] My data are numbered and given in appendix A.

Example 2.3:

(1) The joint military-prosecution investigation team, which probed the **assassination** of President Pak Chong-hui by then Central Intelligence Agency Director Kim Chae-kyu in October 1979, had found no evidence that then Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in the **slaying**, a former team member testified recently.

(3) The joint investigation headquarters' announcement on Dec. 24, 1979, charging Chong with high treason was based not on any concrete evidence but on what he did after Kim **killed** Pak on the night of Oct. 26, 1979.

(5) Paek dismissed as "not true" then Defense Minister No Chae-hyon's statement on Dec. 13, 1979, that martial Law Commander Chong was arrested for interrogation because new evidence was discovered in the course of investigating the **assassination**.

(6) The joint probe team felt the need to question Chong because when Kim **killed** Pak he was nearby and Prosecutor Chong Kyong-sik, a member of the joint team, called on the martial law commander at his office three times to question him about his actions on the **assassination** day.

Killed contained in sentence 3 clearly repeats *assassination* contained in sentence 1, but the way in which these two items repeat cannot be handled by any of Hasan's subcategories of general lexical cohesion: *killed* is not a morphologically distinctive form of *assassination*; these two items do not stand in a synonymy or hyponymy relation because *killed* is a verb, whereas *assassination* is a noun; they are not antonyms because their meanings are apparently not opposite; neither do they stand in a part and whole relation. Similarly, the way in which *killed* in sentence 6 repeats *slaying* (as a noun) contained in sentence 1 cannot be handled in her model either.

My present study accounts for lexical repetitions with Hoey's (1991) categories, namely simple lexical repetition, complex lexical repetition, simple paraphrase and complex paraphrase. for

two reasons. Firstly, instances of lexical repetition that cannot be handled by Hasan's categories, such as the way in which *killed* repeats *assassination*, can be handled properly in 'complex paraphrase' category. The repetition between *killed* and *assassination* is accounted for in a triangular way in Hoey's approach, shown in Figure 2.2 below: a simple paraphrase of *killed* is *assassinated*, which is a complex repetition of *assassination*; these two links produce a complex paraphrase relation between *killed* and *assassination*.

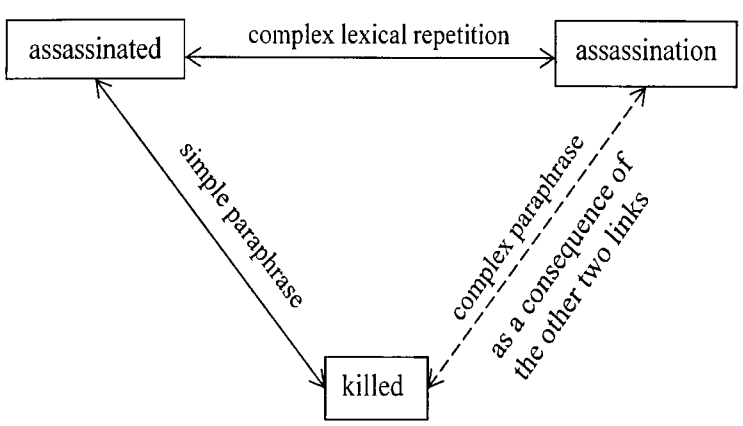


Figure 2.2: Complex paraphrase relation between *killed* and *assassination*

Secondly, Hoey's criteria of delimiting simple lexical repetition, complex lexical repetition, simple paraphrase and complex paraphrase have great value for my research purpose. I adapted his criteria to make them only relevant to *process.* Figures 3.4, 3.5, 3.6 and 3.7 given in chapter 3 provide the criteria that need to be satisfied so that process X stands in these repetition relations to process Y.

The second limitation of Hasan's method is that it cannot consider a nominalised process and the verb that repeats it together in chain interactions. Nominalizations are not likely to occur frequently in her data because increase in the use of nominalizations is typically found in children' writing when they reach age 9 to 10 (Painter *et al*, 2007; Martin, 1989: 31). Painter *et*

al (2007) writes:

> It is noteworthy that the impetus to make grammatically metaphorical meanings on a larger scale is provided by contact with disciplinary knowledge, such as that of history or science, and from an increase in the child's close engagement with the written mode in which such knowledge is mediated. Because of this, it is not until age nine to ten years, that we find a dramatic increase in the use of grammatical metaphors[3] in children's writing.

This limitation to Hasan's method affects texts written by adults. Example 2.4 given below illustrates this. This example contains four sentences taken from text 28 in my data, which reports on the arrest of Abdullah Ocalan, the leader of a Kurdish insurgency, in 1998. Two instances of *arrested* (one in sentence 1 and the other in sentence 2), *picked up* in sentence 5 and *arrest* (as a noun) in sentence 6 are emboldened. The items that co-refer to Abdullah Ocalan are italicized.

Example 2.4:

(1) *The leader of a 14-year-old Kurdish insurgency*, considered a terrorist by Turkey and its most-wanted criminal, has been **arrested** in Rome, setting up a major battle over *his* extradition with Italy.

(2) *Abdullah Ocalan* was **arrested** after stepping off a plane from Moscow, where *he* had sought asylum after recently fleeing *his* hideout in Syria, Turkish and Italian officials said Friday.

(5) Italian officials said *Ocalan* was **picked up** on arrival Thursday night at Leonardo da Vinci airport in Rome because of an outstanding Turkish warrant.

(6) "Once identified, the Italian government had the obligation to go ahead with *his* **arrest** given the serious charges against *him*," said Foreign Minister Lamberto Dini.

The emboldened items constitute a cohesive chain (referred to as chain (a) thereafter). The repetition relations that link these items are given in Table 2.4 below. The numbers in brackets

[3] Nominalization is an important type of grammatical metaphor (Halliday and Matthiessen, 1999).

are the ordinal sentence numbers in which the items occur.

	arrested (1)	arrested (2)	picked up (5)
arrested (2)	sr		
picked up (5)	sp	sp	
arrest (as a noun, in 6)	cr	cr	cp

The repetition relations that the symbols represent are given below:

sr: simple lexical repetition
cr: complex lexical repetition
sp: simple paraphrase
cp: complex paraphrase

Table 2.4: The repetition relations between the emboldened items in Example 2.4

The items put into italics also constitute another cohesive chain (referred to as chain (b) thereafter) because they co-refer to Abdullah Ocalan.

Chain (a) and (b) interact because two instances of *arrested,* and *picked up* contained in chain (a), stand in an 'action acted-upon' relation to *the leader of a 14-year-old Kurdish insurgency, Abdullah Ocalan* and *Ocalan* respectively contained in chain (b). Figure 2.3 below diagrammatically represents the interactions of these two chains in terms of Hasan's method. Both chains are put into rectangles. The three pairs of interacting items are linked by double-headed arrows. The Roman number 'ii' on top of the arrows refer to the second type of chain interacting relation (i.e. 'action acted-upon' relation) noted in Figure 2.1. The numbers in brackets are the ordinal sentence numbers.

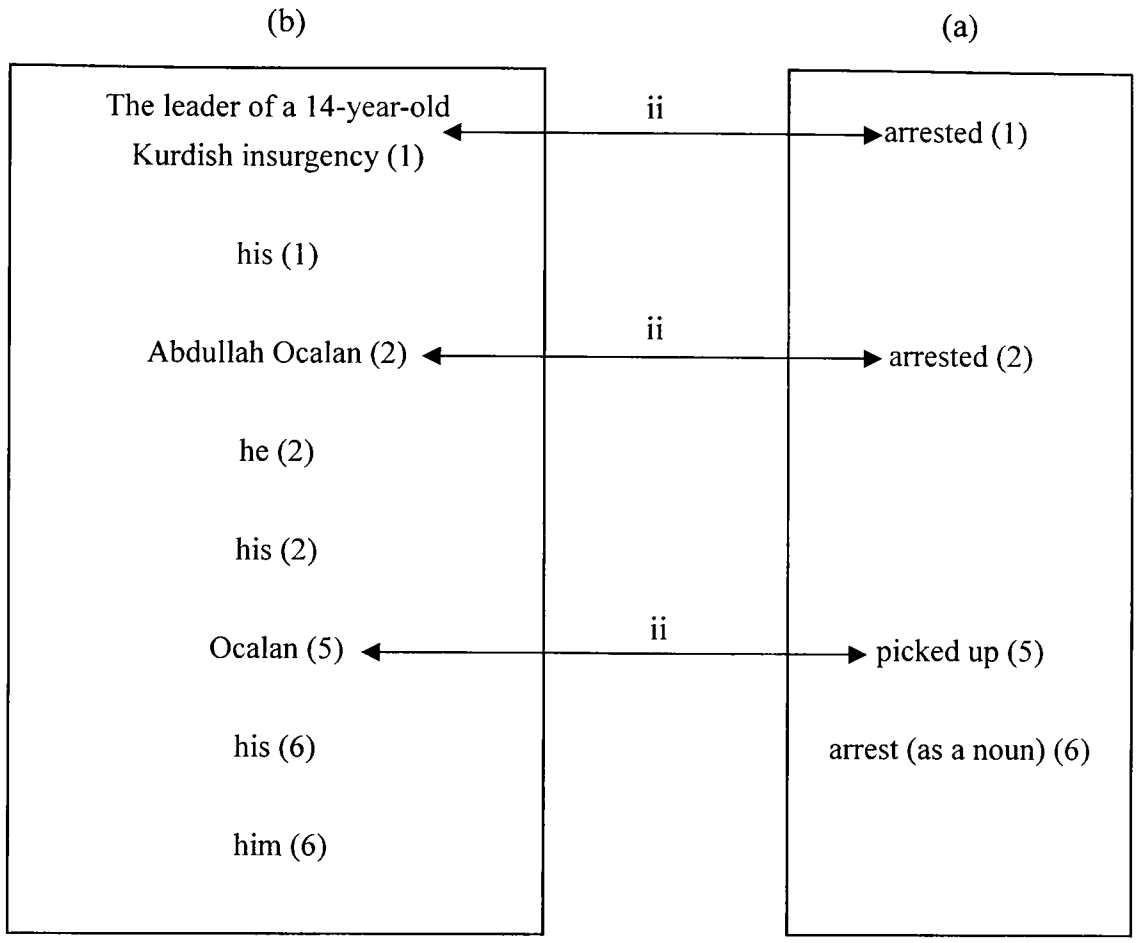


Figure 2.3: The interaction between chains (a) and (b) in Example 2.4

Arrest (as a noun) contained in chain (a) does not stand in an 'action acted-upon' relation to the parallel item in chain (b), namely *his* in sentence 6. These two items co-occur in the nominal group *his arrest* in which *arrest* is Thing and pre-modified by the Deictic element *his*. Therefore, in spite of the fact that the ideational meaning that *his arrest* refers to in the text is identical to *the leader of a 14-year-old Kurdish insurgency has been arrested, Abdullah Ocalan was arrested* and *Ocalan was picked up*, Hasan's method is not able to consider *arrest* (as a noun) along with the two instances of *arrested*, and *picked up* together in chain interactions.

Both Martin (1992) and my method (discussed in detail in chapter 3) deal with this limitation. Martin's method makes use of Halliday's logico-semantic relations of expansion, including

elaboration, extension and enhancement, explained as follows in Halliday (1994: 220).

Elaborating: one clause expands another by elaborating on it (or some portion of it): restating in 'i.e.' other words, specifying in greater detail, commenting, or exemplifying.

Extending: one clause expands another by extending beyond it: adding some new element, 'and, or' given an exception to it, or offering an alternative.

Enhancing: one clause expands another by embellishing around it: qualifying it with some 'so, yet, then' circumstantial feature of time, place, cause or condition.

Martin (1992) applies these three types of relation to clause, nominal group and verbal group, respectively. The following paragraphs give a synopsis of Martin's (1992) discussion of the realizations of these three types of relation in clause, nominal group and verbal group.

Elaboration is used to define *message parts* (i.e. the unit of analysis). In clause, verbal group and nominal group, items standing in elaboration relation are combined to form a single message part. In the discussion below, following Martin (1992), '=' indicates that the items placed on its two sides stand in an elaborating relation and together constitute a single message part.

In a clause, elaboration is realized through two kinds of structure. One is 'Process = Range: process' structure, such as *take = a walk*, because 'it involves just one meaning, which is realized through two lexical items, one elaborating the other' (Martin, 1992: 311). The other is 'a class of locative phrases which can be interpreted as subclassifying the Process whose ritual

destination they provide. These are associated with verbs of motion (especially *go* and *come*); their destination is realized through an unmodified nominal head (e.g. *come* = *to* ***tea***)' (Martin, 1992: 313, his example). More examples of this kind that are given in Martin (1992: 313) are reproduced below:

go/come =	to/from	work/school/class/university/theatre/daycare
	to/from	training/practice/football/hockey
	to/from	dinner/tea/lunch/breakfast/supper
	to/from	bed

In a verbal group, elaboration is also realized through two kinds of structure. One structure is Event = Particle. Phrasal verbs are handled in this situation because 'from the point of view of field only a single event is involved, and most phrasal verbs can be paraphrased with single lexical items (though these are typically less spoken in register)' (Martin, 1992: 312). An example given by Martin (1992: 312) is *look at. Look* stands in an elaborating relation with *at* because these two items together express a meaning that can be paraphrased by *examine*.

The other structure concerns *go* and *come* only. These two can be elaborated in verbal group complex involving specific types of coming and going with respect to leisure activities (Martin, 1992: 313). The examples given by Martin (1992: 313) are:

go/come = swimming, skiing, snorkeling, jogging, running, sailing,
handgliding, surfing, drinking etc.

In a nominal group, the elaborating relation is realized through the pre-modifications of Thing in three structures. The first structure is Classifier = Thing (e.g. *red* = *wine*), in which 'the Classifier functions to specify the relevant subclass of Thing' (Martin, 1992: 312). The second structure is possessive Deictic = Thing when 'the relationship between the participant realized through the Deictic and that realized through the Thing are in a part-whole relation rather than

an ownership one' (Martin, 1992: 314). An example of elaboration in nominal group given by Martin (1992: 326) is *her foot* because *her* and *foot* are in a part-whole relation. The third structure is the pre-nominal structure, including Pre-Deictic (e.g. *the top of = the ridge*), Pre-Numerative (e.g. *a herd of = buffalo*), Pre-Epithet (e.g. *the last of = the questions*), and Pre-Classifier (e.g. *this sort of = thing*) (Martin's examples). The reason why these four pre-nominal structures are treated in the same way as Classifier = Thing structure is explained by Martin as follows: 'like Classifier Thing structures, from the point of view of field these function simply as grammatical resources for isolating particular parts or classes of people, places and things' (Martin, 1992: 313-314).

Table 2.5 given below, adapted from Martin (1992: 326), summarizes the realizations of elaboration in clause, verbal group and nominal group.

	Multiple lexical items in elaborating relations	Examples
Clause	Process = Range: process	act = a role
	Process = ritual Location	go = to work
Verbal Group	Event = Particle	think = over
	Event = Leisure event	go = skating
Nominal Group	Classifier = Thing	skating = rink
	Pre-Deictic = Thing	the edge of = the rink
	Pre-Numerative = Thing	a pair of = skates
	Pre-Epithet = Thing	the largest of = the rinks
	Pre-Classifier = Thing	that kind of = rink
	Deictic (possessive) = Thing	her = foot

Table 2.5: Multiple items realizing elaborating relations in clause, verbal group and nominal group

Having defined message parts, Martin considers the relations between message parts in terms of the other two kinds of logico-semantic relation of expansion, namely extension and enhancement. The realizations of these two types of relations in a clause, a verbal group and a nominal group are described below. Following Martin, + indicates that the message parts on its two sides stand in an extension relation, whereas × indicates that they are in an enhancement relation.

Extension is realized in a clause through two structures: Process + Range: entities and Process + Medium. An example of the former given by Martin is *climb* + *mountain* and an example of the latter is *shoot* + *terrorist*.

All verbal group complexes are treated by Martin as involving extension (1992: 315). Martin (1992: 315) writes: 'all verbal group complexes will be treated as involving extension here, in order to bring out proportionalities of the following kind':

ELABORATION	EXTENSION	ENHANCEMENT
(Phrasal verb)	(verbal group complex)	(event × quality)
look into	keep looking	look carefully
run into	attempt to run	run quickly
see through	happen to see	see clearly
go over	promise to go	go reluctantly
etc.		

However, handling all verbal group complexes in such a way is somewhat indiscriminate and is not in line with Halliday (1994), who considers verbal group complexes involving 'phase' (Halliday, 1994: 279), such as *keep doing;* and *start/stop doing*, as elaboration. In these

instances, the first verbs (*keep, start/stop*) elaborate on the second verbs (*doing*) in terms of time---durative: *keep*, and inceptive/conclusive: *start/stop*. On the other hand, verbal group complexes involving 'modulation' (Halliday, 1994: 281), such as *happen to do* and *hesitate to do*, are analyzed in terms of enhancement because in these instances the first verbs (*happen* and *hesitate*) express circumstantial information to qualify the second verb (*do*): *happen to do* means 'do by chance'; *hesitate to do* means 'do reluctantly' (Halliday, 1994: 282).

In a nominal group, extension is realized through two structures. One is Epithet + Thing structure (Martin, 1992: 315), such as *ugly* + *car*. The other is possessive Deictic + Thing when the relation between the Deictic and Thing is ownership rather than composition (i.e. part and whole relation). Therefore, in *her dog*, *her* extends *dog* because the relation between *her* and *dog* is ownership. (On the other hand, as noted earlier, *her* elaborates on *feet* in *her feet* because the relation between *her* and *feet* is part and whole.)

In a clause, enhancement is realized by Process × Circumstance structure, in which the circumstantial elements qualify the processes in terms of Extent, Location, Cause, etc. The only exception is the adverbial realization of manner, such as *hungrily* in *eat* × *hungrily*. Martin treats adverbial realizations of manner as qualifying Processes in verbal groups rather than in clauses (1992: 316). In a nominal group, enhancement is realized through post-modification of Thing, namely Thing × Qualifier structure (e.g. *the story* × *about Trillian*).

The realizations of extension and enhancement in clause, verbal group and nominal group

described above are summarized in Table 2.6 given below (adapted from Martin, 1992: 317).

	Extension	Enhancement
Clause	Process + Medium (e.g. shoot + the deer) Process + Range: entity (e.g. play + the piano)	Process × Circumstance (e.g. shoot × in the field)
Verbal group	Event + Event (e.g. try to + shoot)	Event × Quality (e.g. shoot × carefully)
Nominal group	Epithet + Thing (e.g. new + car) Possessive Deictic + Thing (e.g. her + dog)	Thing × Qualifier (e.g. the car × in the park)

Table 2.6: Extension and enhancement between message parts in Martin (1992)

Having described the realizations of elaboration, extension and enhancement in clause, verbal group and nominal group, I will explain how Martin's method treats nominalised processes and the verbs repeating these nominalised processes together in chain interactions.

Martin (1992) defines the relations whereby chains interact not in terms of transitivity relations, but in terms of extension and enhancement. Martin's method interprets the interactions between chain (a) and (b) contained in Example 2.4 in the following way. Two instances of *arrested, picked up* and *arrest* (as a noun) contained in chain (a) stand in an extension relation to *the leader of a 14-year-old Kurdish insurgency*, *Abdullah Ocalan, Ocalan* and *his* (in sentence 6) respectively contained in chain (b). The first three pairs stand in such relation in clauses, whereas the last pair stands in such relation in a nominal group. Figure 2.4 below diagrammatically represents the interactions of chains (a) and (b) in terms of Martin's method. The interacting members of these two chains are linked by double headed arrows. The extension relation whereby these two chains interact is indicated by '+' on top of the arrows.

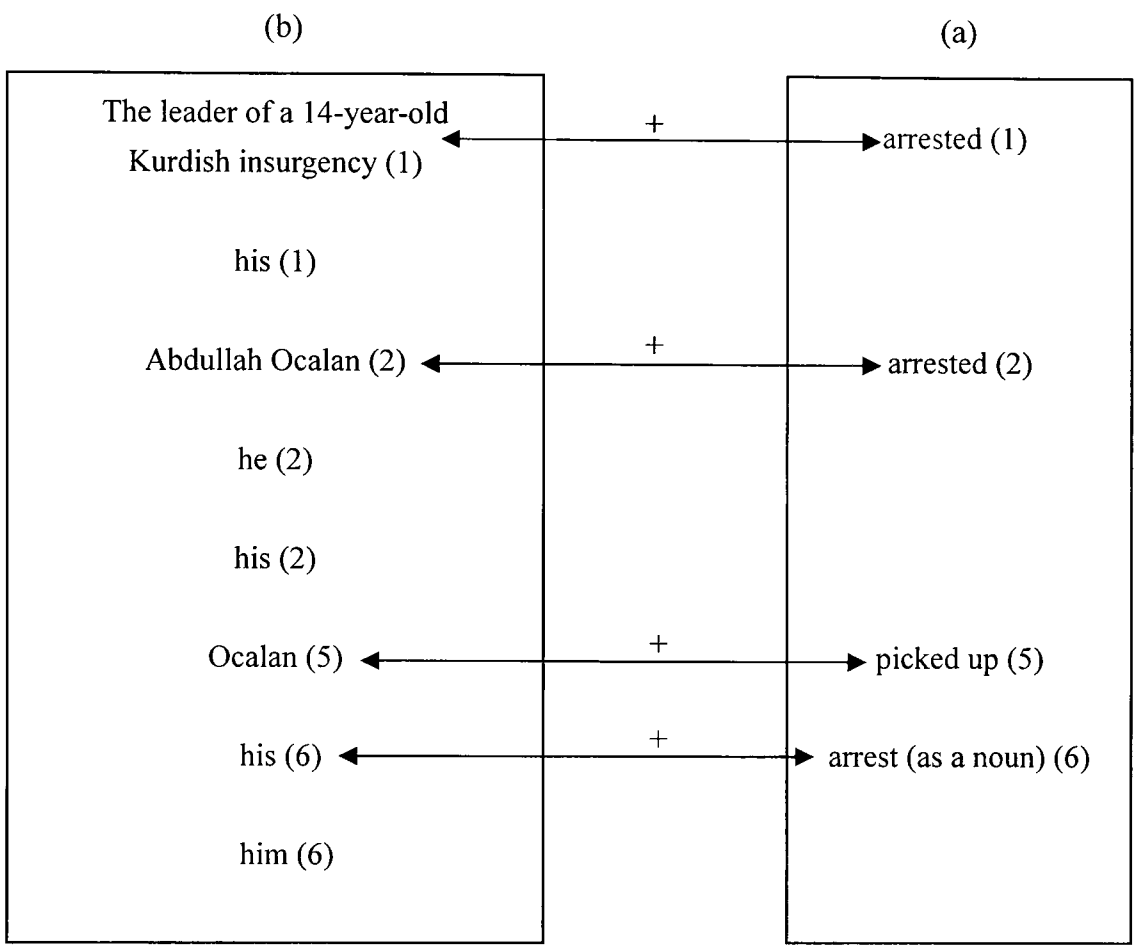


Figure 2.4: The interactions of chains (a) and (b) contained in Example 2.4 in terms of Martin's method

Martin (1992: 371) points out the main differences between his and Hasan's methods as follows:

> The main differences between nuclear relations[4] and cohesive harmony analysis is that Hasan's approach is based on lexical items rather than message parts and built up around TRANSITIVITY relations rather than extending and enhancing ones. This is because nuclear relations obtain at the level of discourse semantics in the model being developed here, with message parts realized by one or more lexical items and the elaboration, extension, enhancement analysis stratified with respect to TRANSITIVITY and group rank experiential grammar.

He also says that 'The major advantage of the stratified approach is related to the problem of grammatical metaphor. Being less tied to lexicogrammar, the nuclear relations analysis is freed

[4] Martin (1992) refers to 'nuclear relations' as the relations between message parts.

to recognize semantic continuities across a diversified range of realizations' (Martin, 1992: 372).

Martin's method however has two limitations. The first is that his method allows pairs of lexical items that stand in heterogeneous transitivity relations to interact. This limitation is illustrated by the short made-up text given in Example 2.5 below.

Example 2.5:

During the last few years, *the homeless* have been **harassing** other citizens in the city of San Francisco. But the situation has changed. *They* are now **harassed** by the police.

This passage contains two cohesive chains. One consists of *the homeless* (italicized) and the pronoun *they* (also italicized, referring to the homeless). The other chain consists of *harassing* (emboldened) and *harassed* (also emboldened). These two chains do **not** interact when interpreted by Hasan's method because *the homeless* is the Actor of *harassing*, whereas *They* is the Goal of *harassed*. On the other hand, these two chains interact when interpreted by Martin's method because *the homeless* stands in an extension relation to *harassing*, and *They* also stands in such relation to *harassed*.

The second limitation is that extension and enhancement with which Martin accounts for chain interactions are so abstract that they do no more than indicate that two or more items of one chain are either added to two or more items of another chain (in the case of extension) or qualified by two or more items of another chain (in the case of enhancement). This limitation is illustrated by Figures 2.5 and 2.6 given below. Figure 2.5 contains a part of Figure 2.1 given

earlier. In Figure 2.5 four instances of *girl* contained in chain (a) all stand in 'actor action relation' (marked by *i*) to two instances of *washed*, *combed* and *brushed* respectively contained in chain (g), all of which in turn stand in 'action acted-upon relation' (marked by *ii*) to four instances of *teddybear* respectively contained in chain (b).

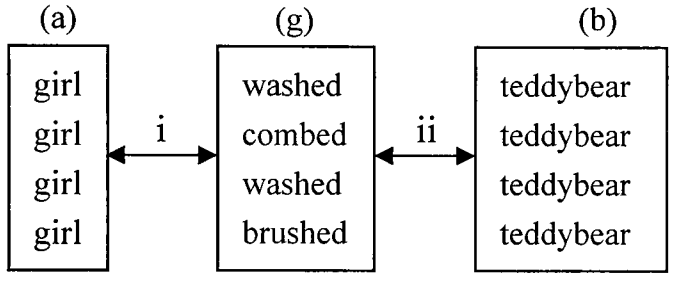


Figure 2.5: Hasan's way of accounting for the interactions of chain (g) and segments of chains (a) and (b) contained in Figure 2.1

Figure 2.6 below represents the interactions between chain (g) and the segments of chains (a) and (b) in Martin's method. The four instances of *girl* contained in chain (a) all stand in an extension relation (marked by +) to the four processes in chain (g) respectively, all of which in turn also stand in this relation to four instances of *teddybear* respectively contained in chain (b).

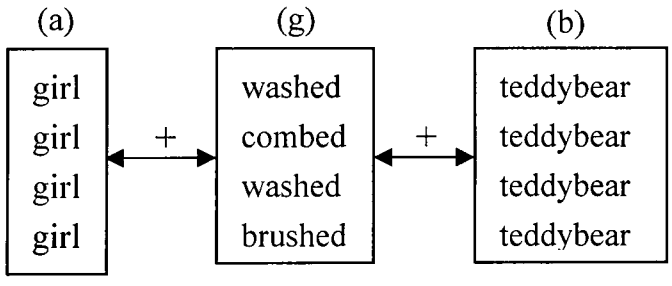


Figure 2.6: Martin's way of accounting for the interactions of chain (g) and segments of chains (a) and (b) contained in Figure 2.1

Figure 2.5 is more informative than Figure 2.6 about the semantic relations in which chain (g) interacts with the segments of chains (a) and (b). Figure 2.5 shows that the girl groomed the

teddybear in three ways (*washed*, *combed* and *brushed*). On the other hand, Figure 2.6 shows that both the girl and the teddybear are involved in these actions of grooming, but does not indicate the transitivity roles of the girl and the teddybear in relation to these actions.

My method involves unpacking nominalizations, which is an important type of grammatical metaphor (Halliday, 1994: 352). Grammatical metaphor means expressing a meaning through a lexico-grammatical form that originally evolved to express a different kind of meaning (Thompson, 2004: 223).

The starting point with which Halliday explains grammatical metaphors is the congruent relationship between semantic and grammatical categories. Congruent means that 'the relation between semantic and grammatical categories is natural' (Martin, 1993: 238). Semantic categories and their congruent realizations are shown in Figure 2.7 below (adapted from Halliday, 1999: 110).

Congruence in rank		Congruence in status (elements)	
semantic	*grammatical*	*semantic*	*grammatical*
sequence	clause complex	entity	noun (nominal group)
figure	clause	quality	adjective (in nom.gp.)
element	group/phrase	process	verb (verbal group)
		circumstance (1)	adverb (/adverbial gp.)
		circumstance (2)	prepositional phrase
		minor process	preposition
		relator	conjunction

Figure 2.7: Congruence between semantic and grammatical categories

A figure is a representation of experience in the form of a configuration that consists of a process, participant taking part in this process and associated circumstances (Halliday and Matthiessen, 1999: 52). It is realized congruently by a clause. A sequence occurs when two or more figures are related in some kind of logico-semantic relation, which can be classified into two types: expansion and projection. Expansion has been discussed earlier. Projection is further subcategorized into locution and idea; the former concerns verbal events, while the latter concerns mental events. A sequence is congruently realized by a clause complex, that is, a combination of two or more clauses.

Entities are congruently expressed by (or 'realized by' in Halliday's term) nouns, qualities are congruently expressed by adjectives, and processes are congruently expressed by verbs. Thus, in Example 2.4 both *arrested* and *picked up* express process meanings congruently.

However, if humans always expressed meanings in congruent ways, there would be no need to distinguish semantics and grammar in the first place (Martin, 1993: 238). When a semantic category is expressed incongruently, that is, when a semantic category is expressed by grammatical categories that congruently realize other semantic categories, grammatical metaphors occur.

A process, congruently worded as a verb, may be nominalised and thus expressed by a noun. In Example 2.4 *his arrest* is the nominalization of *he was arrested.* Unpacking nominalizations reverses nominalising, reconstruing semantic elements other than entities in terms of their

congruent grammatical categories. Unpacking nominalizations overcomes the second limitation of Hasan's method because it enables the cognate verb of a nominalised process and the verb that repeats the nominalised process to be considered together in chain interactions. This is illustrated by Figure 2.8 given below, which diagrammatically represents the interaction between chains (a) and (b) after *his arrest* was unpacked. We now have three instances of *arrested*, one of which is the result of the unpacking, and one instance of *pick up* contained in chain (a). These four processes stand in an 'action acted upon relation' to *the leader of a 14-year-old Kurdish insurgency*, *Abdullah Ocalan*, *he* (in sentence 6) and *Ocalan* respectively contained in chain (b).

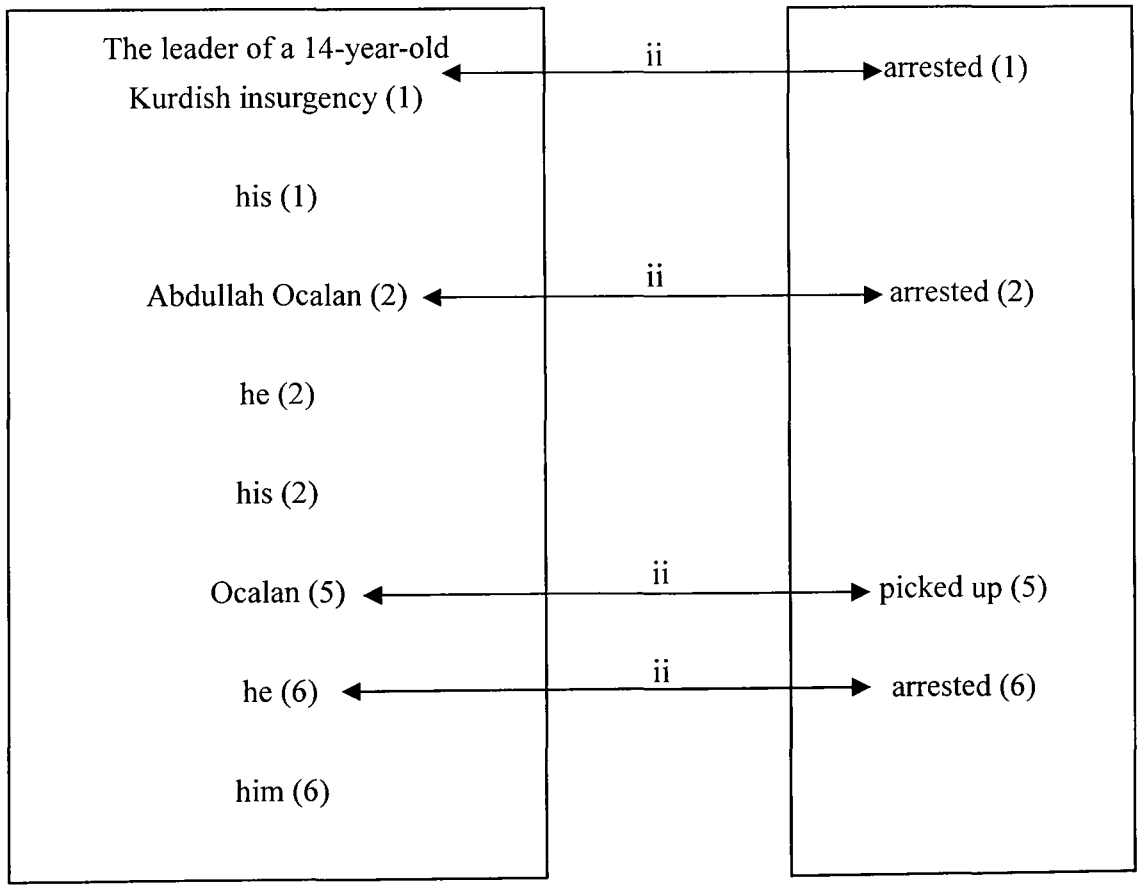


Figure 2.8: The interactions of chains (a) and (b) contained in Example 2.4 after *his arrest* was unpacked

2.4 Parsons (1991a, 1991b)

Hasan's study revealed an interesting relationship between cohesion and coherence, and attracted other researchers to work on this topic, one of whom was Parsons (1991a, 199b). He uses Halliday and Hasan's (1985) model to analyze eight scientific texts about the processes involved in the growing and harvesting of coffee that were written by Algerian post-graduate, civil or electrical engineering students, and eight scientific texts about the same topic written by native post-graduate students from the same discipline. He first identifies the identity and similarity chains in each text, and schematically represents chain interactions. After that he classifies lexical tokens into relevant tokens (the tokens that enter into a chain), peripheral tokens (the tokens that do not belong to any chain), central tokens (the tokens that enter into a chain and interact with tokens of other chains), and non-central tokens (the tokens that enter into chains but do not interact with tokens of any other chain), and calculates the number of each type of token in each text. He subsequently obtains the following four types of ratios for each text: relevant/peripheral, central/non-central, central/peripheral and central/total lexical tokens.

He also asked eight informants to evaluate the coherence of each of the 16 texts according to five categories: 'poor', 'below average', 'average', 'above average', and 'good'. The results are later given marks accordingly: 'poor' is given one point; 'below average' is given two points, etc, with one more point being added to each stage higher in the scale. His results are given in Table 2.7 below. The letters of the alphabet in the first column refer to the 16 texts, followed by the status of the writers. The last column, 'rank scores', grades the coherence of the texts based

on the results of the informant tests. The results are ordered in descending rank scores.

texts	ratios of relevant to peripheral tokens	ratios of central to non-central tokens	ratios of central to peripheral tokens	ratios of central tokens to total lexical tokens	rank scores
M native	3.0	1.1	1.6	0.40	38
A native	1.7	1.1	0.9	0.33	34
N native	1.1	0.6	0.4	0.19	34
K native	1.9	1.5	1.1	0.39	32
I native	2.9	0.9	1.3	0.34	28
H native	1.5	1.0	0.8	0.30	27
L overseas	3.0	1.2	1.6	0.40	27
O overseas	2.1	1.1	1.1	0.36	26
P native	1.6	0.4	0.5	0.19	25
J native	1.1	1.4	0.6	0.30	24
D overseas	2.0	1.0	1.0	0.33	21
F overseas	0.7	1.2	0.4	0.22	19
B overseas	3.0	2.0	2.0	0.50	18
E overseas	1.9	0.5	0.7	0.23	18
C overseas	1.5	0.5	0.5	0.20	9
G overseas	1.7	0.6	0.4	0.14	8

Table 2.7 The results of cohesive harmony in the 16 texts (adapted from Parsons, 1991b: 422)

The table shows some interesting results. Firstly, text B warrants special attention because the four types of ratios in this text are all the highest in the 16 texts but the coherence of this text is not judged to be high. This is explained below. The four types of ratio of this text are 3.0, 2.0, 2.0 and 0.50 respectively. It shares the largest first type of ratio with texts L and M. The other three types of ratio of this text are all the largest in the 16 texts. On the other hand, the rank score of this text is 18 and the mean value of it is 18/8=2.25 (8 refers to the number of informants taking part in the informant test). This suggests that the coherence of this text is considered to be 'below average'. It therefore stands out because of the discrepancy between its very high ratio scores and its low coherence and is a counter-example to Hasan's (1985) claims

in that, in contrast to what she predicts in her first two claims in her description of cohesive harmony, the ratio of relevant to peripheral tokens and the ratio of central to non-central tokens are very high in this text, but the coherence of the text is very low.

Secondly, Hasan's first claim regarding cohesive harmony noted in her 1984 work is also questionable. She claims that 'any text will be seen as coherent, in which the central tokens form at least 50 percent of the total tokens' (Hasan, 1984: 218). But in Parsons' results, text B is the only text that has a value of no less than 50 percent for this type of ratio, but its coherence is considered to be very low. On the other hand, text M has the highest rank score, 38, and therefore is, according to this criterion, considered to be most coherent, but the ratio of central tokens to total lexical tokens for this text is less than the minimum level (50%) noted by Hasan (1984).

In order to find out whether there is a consistent correlation between each of the four types of ratios in Table 2.7 and judgements of the coherence of texts, Parsons (1991a) calculated the correlation coefficients ('r') relating each of the four types of ratio to the rank scores. He found that only the last type of ratio (the ratio of central tokens to total lexical tokens) had a significant correlation to the coherence of text, whereas the other three types of ratios (relevant to peripheral tokens, central to non-central tokens and central to peripheral tokens) did not show significant correlation with judgments of the coherence of texts.

His results cast doubt on Hasan's (1985) first two claims in her description of cohesive

harmony, in which she hypothesizes that two types of ratio, namely relevant to peripheral tokens and central to non-central tokens, are correlated with the degree of coherence of texts, in that Parsons (1991a) failed to find a significant correlation between these two types of ratio and the coherence of texts in his results.

The ratio of central tokens to total lexical tokens is not noted in Hasan's 1984 or 1985 descriptions of cohesive harmony, but Parsons (1991a) found that this type of ratio correlates with the coherence of text. This leads him to notice the crucial role the central tokens play in contributing to the coherence of text. He then introduced another type of chain, namely significant chains, which refer to the chains that possess more central tokens than the average number of central tokens in the chains contained in the set of texts under examination. In the 16 texts that Parsons analyzed, the significant chains include at least 4 central tokens.

Parsons (1991a: 174-175) refers to the central tokens in significant chains as significant tokens. Significant tokens, by definition, must be found in significant chains; and they are by nature tokens that interact with tokens of other chains.

Having defined significant chains and significant tokens, Parsons (1991a) derives the ratios of significant tokens to total lexical tokens for the 16 texts. He finds that the correlation between this type of ratio and rank scores is better than that between the ratio of central tokens to total lexical tokens and rank scores. He says tentatively that 'it would seem that those texts which have central tokens organized into long chains are being perceived to be more coherent'

(Parsons, 1991b).

Parsons does not stop here. He moves further to consider the following three types of ratio for central tokens in chains including the number of central tokens at different thresholds (n stands for different thresholds).

Cn/TLT: the ratio of the number of central tokens found in chains including at least n central tokens to the number of total lexical tokens

Cn/PT: the ratio of the number of central tokens found in chains including at least n central tokens to the number of peripheral tokens

Cn/the remaining lexical tokens: the ratio of the number of central tokens found in chains including at least n central tokens to the number of the remaining lexical tokens (the number of the remaining lexical tokens is obtained by deducting the number of central tokens found in chains including at least n central tokens from the number of total lexical tokens)

Since significant chains in the 16 texts analyzed by Parsons are chains including at least 4 central tokens, he first examines the above three types of ratio for central tokens found in chains including at least 5 central tokens. Parsons finds that the three types of ratios for C5 (central tokens found in chains that include at least 5 central tokens) have better correlation to the coherence of text than those for C4 (central tokens found in chains that include at least 4 central tokens). Parsons (1991a: 204) interpreted the results as follows:

> It appears that the informants are perceiving those texts which have central tokens organized into long chains to be more coherent than those which have their central tokens organized mainly in short chains. It seems that chain interaction alone does not necessarily result in the most coherent texts, but that interaction which organizes the tokens into long chains is more likely to lead to coherent texts in which there are more

> occurrences when one is saying 'similar things about similar phenomena'.

After that, Parsons examines the above three types of ratio for central tokens found in chains including at least 6 central tokens (C6). He finds that the three types of ratios for C6 not only do not have better correlations to the coherence of text than C5 but also are too low for significant correlation at a safe level (Parsons, 1991a: 196). He points out that there seems to be 'a point beyond which informants did not favorably evaluate texts with long chains' (Parsons, 1991a: 196) and that 'it appears that a text can be too cohesive, and that beyond a certain point, an informant ceases to evaluate a text favorably, possibly because of the lack of lexical and structural variety. It is likely that there is a balance between a text being too cohesive and not being cohesive enough' (Parsons, 1991a: 197).

Parsons' findings have two important implications for my research purpose. The first implication is similar to the one derived from Hasan's research, that is, it is important to look at chain interactions because he shows that the ratio of central tokens to total lexical tokens correlates well with the coherence of text. The central tokens are by definition tokens that interact with tokens of other chains.

The second implication is that the examination of frequent occurrences of chain interactions is rewarding because he shows that the three types of ratio noted earlier for central tokens found in chains that contain at least five central tokens have better correlation with judgments of the coherence of texts than those for central tokens found in chains that contain at least four central

tokens, which in turn have better correlation with judgments of the coherence of texts than those for central tokens found in chains that contain at least three central tokens.

2.5 Hoey (1991)

Hoey (1991) is an important work because it discusses the contributions of patterns of lexis in a text to the coherence, organization, and meaning of the text.

Hoey establishes categories of repetition to account for different types of repetition relation that hold between pairs of lexical items in a text. Lexical repetitions are primarily handled in these four categories: simple lexical repetition, complex lexical repetition, simple paraphrase and complex paraphrase.

He defines simple lexical repetition as follows: 'simple lexical repetition occurs when a lexical item that has already occurred in a text is repeated with no greater alteration than is entirely explicable in terms of a closed grammatical paradigm' (1991: 53). The definition may be too succinct to be fully understood. A closed paradigm occurs when 'there is a limited set of choices dictated by grammatical considerations and to which no addition can be unilaterally made' (personal communication with Hoey). Thus, singular and plural variants of a noun form a closed paradigm, so do different variants of a lexical verb in the verb-form paradigm (i.e. the base form of a lexical verb, its third person singular form, its present continuous form, its past tense and past participle forms).

Complex lexical repetition occurs when 'two lexical items share a lexical morpheme. but are not formally identical (as defined in our discussion of simple repetition), or when they are formally identical, but have different grammatical functions' (Hoey, 1991: 55). The nominalizations of a lexical verb stand in this relation to the variants of the lexical verb that belong to the verb-form paradigm. For example, both *investigation(s)* and *investigator(s)* stand in a complex lexical repetition relation to each variant of *investigate* that belongs to the verb form paradigm. *Investigation(s)* is also a complex lexical repetition of *investigator(s).*

Simple paraphrase occurs 'whenever a lexical item may substitute for another in context without loss or gain in specificity and with no discernible change in meaning' (Hoey, 1991: 62). Thus, two lexical items that are similar in meaning in context stand in this relation. The complication is allowed, that is, a simple paraphrase is also established 'if an alteration needed to substitute one item for another in context is necessitated by a grammatical paradigm' (Hoey, 1991: 63). Thus, *probe* is a simple paraphrase of *investigate* in many contexts because when one is substituted for the other, no change in meaning of the stretch in which they occur is discernible. *Probe* is also a simple paraphrase of *investigated* because the alteration of *probe* to *probed* or *investigated* to *investigate* when they substitute for each other is conditioned solely by different choices from the verb form paradigm.

Complex paraphrase occurs 'when two lexical items are definable such that one of the items includes the other, although they share no lexical morphemes' (Hoey, 1991: 64). This definition is a very loose one. Hoey (1991) restricts the application of this relation to three types of

situation.

The first type of situation relates to antonymy. The definition of complex lexical repetition is such that it is able to pick up antonyms that share a lexical morpheme (e.g. happy/unhappy). It is, however, not able to pick up antonyms that do not share a lexical morpheme (e.g. cold/hot). Complex paraphrase is defined in such a way that it picks up antonyms that do not share a lexical morpheme. For instance, *cold* may be crudely defined as *far from hot*.

The second type of situation in which complex paraphrase may be recognized occurs when a third item is a complex repetition of one item and also a simple paraphrase (or an antonym) of the other. For example, *writer* is a complex paraphrase of *writings*, and *writer* is also a simple paraphrase of *author*. The presence of these two links produces a complex paraphrase relation between *writings* and *author*. The links between *writer*, *writings* and *author* are represented in Figure 2.9 below, reproduced from Hoey (1991: 65).

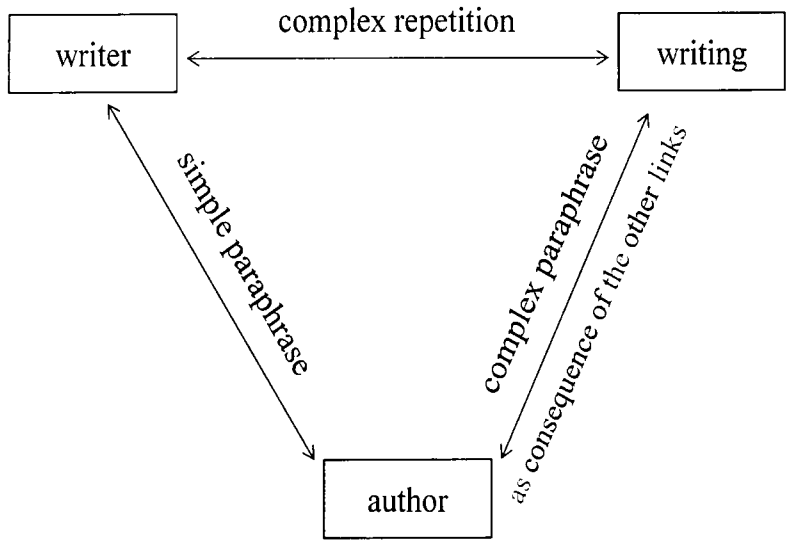


Figure 2.9: The links between *writer*, *writing* and *author*

The third type of situation is a complication of the second situation. It is shown in Figure 2.9 that *writer* functions as the mediator that enables the links between it and the other two items to establish a link between the other two. It is possible, though, that the mediator is not present in the text. When this occurs, a complex paraphrase will still be acknowledged between the other two provided that 'there is an item that is capable of paraphrasing exactly *in that context* one of the items and of repeating the other' (Hoey, 1991: 66, his emphasis).

Hyponymy is not handled in the above categories. It is treated in a very restrictive way in Hoey (1991). An item stands in this relation to a more general noun only if they have the same referent, whether the item occurs before or after the general noun. Hoey (1991: 70) says that 'the crucial feature is the comment referent...in the event of the 'link triangle' not operating, we will say that if, and only if, two items are interpreted as having identical referents they will be treated as repetitions, even though they are either unrelated as lexical items, or related as superordinate to hyponym'. Thus, when an item that has occurred earlier is followed by a more general noun that refers to it (e.g. in one of Hoey's texts *the animal* is used to refer to an aforementioned *bear*), a hyponymic relation is established between these two items. On the other hand, they do not stand in such a relation when the general noun does not refer to the aforementioned item. Hoey notes that had hyponymy not been restricted as such, an earlier mention of a general noun, such as 'thing', would have been in a repetition relation with a large number of different objects, both animated and non-animated by nature. Likewise, if 'people' is found early in the text, had the hyponymic relation not been restricted in such a way, it would

have been in a repetition relation with every mention of man and woman that followed, irrespective of profession, nationality, etc. 'Such a consequence would make nonsense of our objective, which has been to make sense of cohesion as an organizing feature of texts' (Hoey, 1991: 69).

In conjunction with defining the different types of lexical repetition, Hoey (1991) also delimits these relations by giving the criteria that need to be satisfied. The reason why Hoey takes the trouble to produce the criteria is that although different types of lexical repetitions are distinguished, and the definitions of them are explained, they are not sufficient to identify the boundaries of these relations. The criteria state clearly the conditions that need to be satisfied so that a pair of lexical items can be said to stand in a certain type of repetition relation. They have great value for cohesion analysis.

In addition to the lexical repetitions discussed above, grammatical cohesive devices, such as substitution, ellipsis, co-reference, and demonstrative pronouns (e.g. *this* and *that*), are also considered in Hoey's analyses, but are explained in less detail than are the lexical repetitions.

Having explained the different types of repetition, Hoey proceeds to examine exhaustively the repetitions between the lexical items in the sentence under consideration and those in each of the other sentences. A repetition is registered as a link. After that he measures the semantic relatedness of sentences in a text with the notion of a 'bond', which refers to above average degree of repetition between any two sentences in a text (Hoey, 1991: 91). The threshold for the

number of links that make up a bond is normally three. Hoey (1991: 133) finds that bonded sentences often read coherently when placed together no matter how distantly they are separate in a text. This is because parallelism occurs more or less overtly in a bonded pair and the links shared by the two sentences of the pair create the parallelism (Hoey, 1991). This supports Hasan's and Parsons' findings reported earlier in showing that in naturally occurring prose, the repetition of items across the sentences in a text plays an important part in enabling readers to perceive the text as coherent.

On the other hand, Hoey (1991: 134) also discusses pairs of bonded sentences that do not read coherently. Each pair is characterized by an excessive number of links shared by the two sentences that form the pair. This is in line with Parsons' finding noted earlier that excessive lexical repetitions shared by sentences have a negative effect on the coherence of a text. Therefore, there is a trade-off between the presence of lexical repetitions shared by sentences that enable readers to perceive a text as coherent and an excessive number of lexical repetitions that, on the other hand, have a negative effect on the coherence of the text.

The following points concerning the relationships between repetitions of items in a text and the coherence of the text are drawn from Hasan's, Parsons' and Hoey's studies.

1. The repetition of items shared by the sentences in a text contributes to the coherence of the text.
2. Excessive repetition of items shared by the sentences, however, has a negative impact on the coherence of the text.

Hoey (1991) and Hasan's and Parson's work reported earlier, however, do not discuss the contribution of conjunctive relations that hold between clauses, sentences, and parts contained in a text to the coherence of the text. We will discuss these in the last two sections of this chapter.

Hoey (1991) shows that the repetition of items shared by the sentences in a text also contributes to the organization and meaning of the text. His method enables us to identify topic opening and closing sentences on the one hand, and central and marginal sentences on the other. Topic opening and closing sentences mark the boundaries of a topic in a text that discusses two or more topics. The sentence that has a majority of its bonds with the sentences that come after it usually opens a topic, whereas the sentence that has a majority of its bonds with the sentences that come before it usually closes a topic. Berber Sardinha (1997, 2001) developed a program that implements Hoey's method to automatically identify the boundaries of topics in a text.

Central and marginal sentences are defined in terms of the number of bonds a sentence has with other sentences. The sentence that has bonds with many other sentences is referred to as a central sentence. On the other hand, the sentence that has few bonds with other sentences is referred to as a marginal sentence. Hoey (1991) notes that central and marginal sentences play different roles in a text. Central sentences are more important to the meaning of a text than marginal sentences. The juxtaposition of central sentences very often summarizes the original text if it is non-narrative (Hoey, 1991: 148). Marginal sentences, on the other hand, rarely contribute to the main theme of text (Hoey, 1991: 105), but typically function like sentence-

length conjuncts that relate earlier sentences to later sentences and signal the logical relationship between them (Hoey, 1991: 105).

A limitation of Hoey's method needs pointing out. His method gives priority to long sentences because long sentences have a higher probability of being selected and put in an abridgement than short sentences. A sentence that contains 20 lexical items is more likely to have three items repeated by three items contained in another sentence than a sentence that contains, say, only eight lexical items. The consequence is that short sentences that may contain information that is important to the gist of a text have a danger of being dropped out of the abridgement.

Hoey (1991) has an important implication for my research purpose. He notes that 'we are more likely to arrive at a satisfactory account of how cohesion works if we concentrate on the way repetition clusters in pairs of sentences' (1991: 20). His statement and Hasan's finding noted earlier that chain interactions are important to the coherence of a text indicate that it is more useful to examine the co-occurrence of lexical items that belong to different chains than to look at individual chains in an isolated manner. Examining the recurrences of the Process and Medium combination in my present study follows this implication.

Hoey's (1991) link-and-bond method was implemented automatically by a text processing system named Tele-pattan created by Benbrahim and Ahmad (1994). This system is capable of performing many of the applications of his method noted in Hoey (1991), including identifying the sentences that open and close a topic for segmenting a text that discusses two or more

topics, and central and marginal sentences in a text for producing summaries for the text.

2.6 Scott's *KeyWords* tool

Another author that is also interested in examining the relationship between repetitions in a text and the meaning of the text is Mike Scott. He is the author of *WordSmith*, a suite of lexical analysis software that was first published in 1996 and has been regularly updated ever since, now standing at version 5. The *KeyWords* tool in the suite enables one to identify the keywords in a text or corpus, that is, the words that occur with unusual frequency in the text or corpus in question when compared to a reference corpus (Scott, 2004, 2010; Scott and Tribble, 2006). The words that occur more often in a text or corpus in question than in a reference corpus are referred to as positive keywords; those that occur less often in the text or corpus under consideration than in the reference are referred to as negative keywords.

The *KeyWords* tool identifies the keywords of a text (or group of texts) on a mechanical basis by comparing the words and their frequencies in the text or corpus in question with those in a reference corpus. Scott and Tribble (2006) notes that for a word to be key, it must satisfy two conditions: (a) it must occur in a text or corpus at least as many times as the user has specified as a minimum frequency; (b) a statistical test of some suitable kind must show that its frequency in the text or corpus is outstanding when compared with its frequency in the reference corpus. The reference corpus functions as a 'filter' (Scott and Tribble, 2006). The definite article *the* is often the most frequent word in both the text in question and the reference corpus. But if its frequency as a percentage of the total number of running words is roughly the

same in the two places, it will not be outstanding, even if it is frequent. Therefore, it gets filtered out. After all the different word-types in the text under consideration have been processed to check whether each occurs in the text at least as frequently as the user specified as the minimum frequency and its frequency in the text is deemed as outstanding by a statistical test of some suitable kind when compared with a reference corpus, only those words that characterize the text remain. Scott (2010) notes that the *KeyWords* tool will usually throw up three kinds of words as 'key': proper nouns that are specific to a story depicted in the text, the words that indicate what the text is about, and the words that indicate the stylistic features of the text. For instance, the tool identifies 241 positive keywords in the chapter you are reading when compared to the British National Corpus (BNC). Table 2.8 given below shows the first 20 of these keywords, organized in descending order of keyness.

N	Keywords
1	TOKENS
2	HOEY
3	HASAN
4	TEXT
5	LEXICAL
6	CHAINS
7	CHAIN
8	RELATION
9	COHERENCE
10	ITEMS
11	COHESIVE
12	RELATIONS
13	COHESION
14	CENTRAL
15	SENTENCES
16	HALLIDAY
17	TEXTS
18	REPETITION
19	SENTENCE
20	REPETITIONS

Table 2.8: The top 20 positive keywords of this chapter when compared to the BNC

HOEY, HASAN, and HALLIDAY are names of authors that I refer to frequently in this chapter. The other keywords in the list given above indicate what this chapter is about. TOKENS is key (in the way defined by Scott) because its frequency is outstandingly high in the earlier sections that discuss Hasan's and Parsons' works, which involve examining a variety of ratios pertaining to different types of tokens. TEXT(S) are also key because this chapter reports on studies that discuss the characteristics of texts, concerning primarily the contributions that COHESION makes to COHERENCE, and also other characteristics of text. CHAIN(S) and COHESIVE are key because their frequencies are outstandingly high particularly in sections 2.2, 2.3 and 2.4 of this chapter: section 2.2 discusses Hasan's work on identifying the cohesive chains that occur in a text and examining the interactions of chains; section 2.3 discusses methods that allow for considering a nominalised process and the verbs that repeat it together in chain interactions; section 2.4 discusses Parsons' work, which is similar to Hasan's in many respects. REPETITON(S) are also key because their frequencies are outstandingly high in the sections that report Hasan's, Parsons', and Hoey's works because their works involve examining the repetitions of items in a text.

Hoey (1991) and Scott (1997, 2000, 2001, 2010) show that patterns of lexis in a text are often indicative of what the text is about. This view is supported by the works of computational linguists that employ patterns of lexis in a text for the purpose of text summarization, such as Benbrahim and Ahmad (1994), and Barzilay and Elhadad (1997).

An important point is evidenced by Hasan's, Parsons', Hoey's and Scott's work reported above,

that is, the repetition of items in a text contributes to the coherence, organization and meaning of the text. They, however, do not discuss whether another aspect of cohesion, namely the relations that hold between the clauses, sentences and larger chunks contained in a text, also contribute to these three aspects of characteristics of text. We will discuss this in the two sections that follow.

2.7 Hoey's (2001) *Textual Interaction*

Hoey's (2001) work, *Textual Interaction*, brings together his works about lexical signaling (1979, 1983, 1994), patterns of lexis in text (1991), and patterns of text organization (1983, 1986). *Textual Interaction* analyzes texts in terms of the relations that hold between the component parts (i.e. clauses, sentences and larger chunks). It is worth discussing this work because it discusses the contribution of the relations that hold between the parts of a text to the coherence, organization and meaning of the text as a whole.

Analyzing a text in terms of the relations that hold between the parts contained in it is a popular way of describing the organization of the text. Works about genres, such as Swales (1990), Martin (1997) and Ventola (1987), and Mann and Thompson's Rhetorical Structure Theory reported in the next and final section also analyze a text in such a way. But *Textual Interaction* accounts for the relations that hold between the parts of a text in terms of an interaction between the writer and the reader: the writer anticipates the questions that readers may ask while they read, and writes the material that follows in such a way that they answer the readers' questions. The interactions take place both on a small scale when writers answer the questions

that readers want answering on a sentence-by-sentence basis, and on a large scale when writers compose a part of a text in such a way that the sentences contained in that part collectively answer the question that readers may have generated after reading the previous part.

In the small scale interaction, the answer in a sentence that follows of the question that readers formulate after reading a previous sentence in effect spells out the relation that hold between the two sentences (Hoey, 1983). Hoey, following Winter (1974, 1986), classifies the relations that hold between a pair of clause or sentence into two broad classes, Sequence relations and Matching relations. Sequence relations occur when 'one clause/sentence answers a question such as 'What happened next?', 'What happened as a result?' and 'What do you infer from this?', all of which involve putting propositions in some order of priority in time, space or logic' (Hoey, 2001: 30). Typical Sequence relations are time sequence, cause-consequence, means-purpose, and premise-deduction. They are signalled by a variety of means. A fair number of lexical items, subordinators and sentence conjuncts inform readers of the existence of a sequence relation and what its nature is (Hoey, 1983, 2001). For instance, consider the lead of text 27 given below, which reports that the Police in a county of Hebei province issued fake warrants to detain villagers.

Police in a county of North China's Hebei province issued fake warrants to detain six villagers in a land dispute for bail money, the Beijing News reported on Wednesday.

In the lead the subordinator *'to'* (shorthand for 'in order to') signals to the readers that the clause that comes after it is about to give the purpose for the proposition expressed in the

clause that precedes it.

Matching relations, on the other hand, do not involve putting things in any order but compare statements with a view to examining the similarities and/or differences between them (Hoey, 2001: 31). They occur when a clause, sentence or group of sentences answer questions such as 'How does x compare with y?', 'How does x differ from y', 'What is an example of that?' (Hoey, 2001: 31). Typical Matching relations are contrast, similarity, exemplification, preview-detail, and exception. Like Sequencing relations, Matching relations are also frequently signalled by lexical items, subordinators and sentence conjuncts, but key signals for Matching relations are lexical repetition and parallelism (Hoey, 1983, 2001). Lexical repetitions, which embody the similarities between two statements, provide a context for highlighting the information contained in the later statement that replaces that given in the former statement (Winter, 1977, 1986; Hoey, 1983, 2001; Scott and Thompson, 2001).

The question that readers formulate after reading a sentence may not be answered in the sentence that immediately follows, but in a sentence that comes much later in a text. When this occurs, the link-and-bond method discussed in Hoey (1991) can be used to identify the sentence that answers the question that readers may ask and want answering. What readers need to do is to skim the discourse after the sentence that makes readers generate the question, and find the sentence that contains three or more lexical items that are in a repetition relation with three or more lexical items contained in that sentence.

In the large scale interaction the relations that hold between the parts contained in a text are similar to those that hold between clauses or sentences (Hoey, 2001: 56). *Textual Interaction* discussed five culturally popular patterns of text organization, namely the Problem-Solution pattern, the Goal-Achievement pattern, the Opportunity-Taking pattern, the Desire Arousal-Fulfilment pattern, and the Gap in Knowledge-Filling pattern. The Problem-Solution pattern is the prototype, and the other four patterns are similar to it in many respects. But each pattern has its own characteristics. Hoey (2001) dedicated considerable efforts to discussing the characteristics of each pattern.

The patterns embody 'a template of questions that both writer and reader know about and can refer to' (Hoey, 2001: 119). For instance, the Problem-Solution pattern can be diagrammatically represented in Figure 2.10 given below (adapted from Hoey, 2001). This pattern frequently consists of four stages in such a sequence: Situation-Problem-Response-Results/Evaluation. The questions that readers may ask to bring about the stages are given above them.

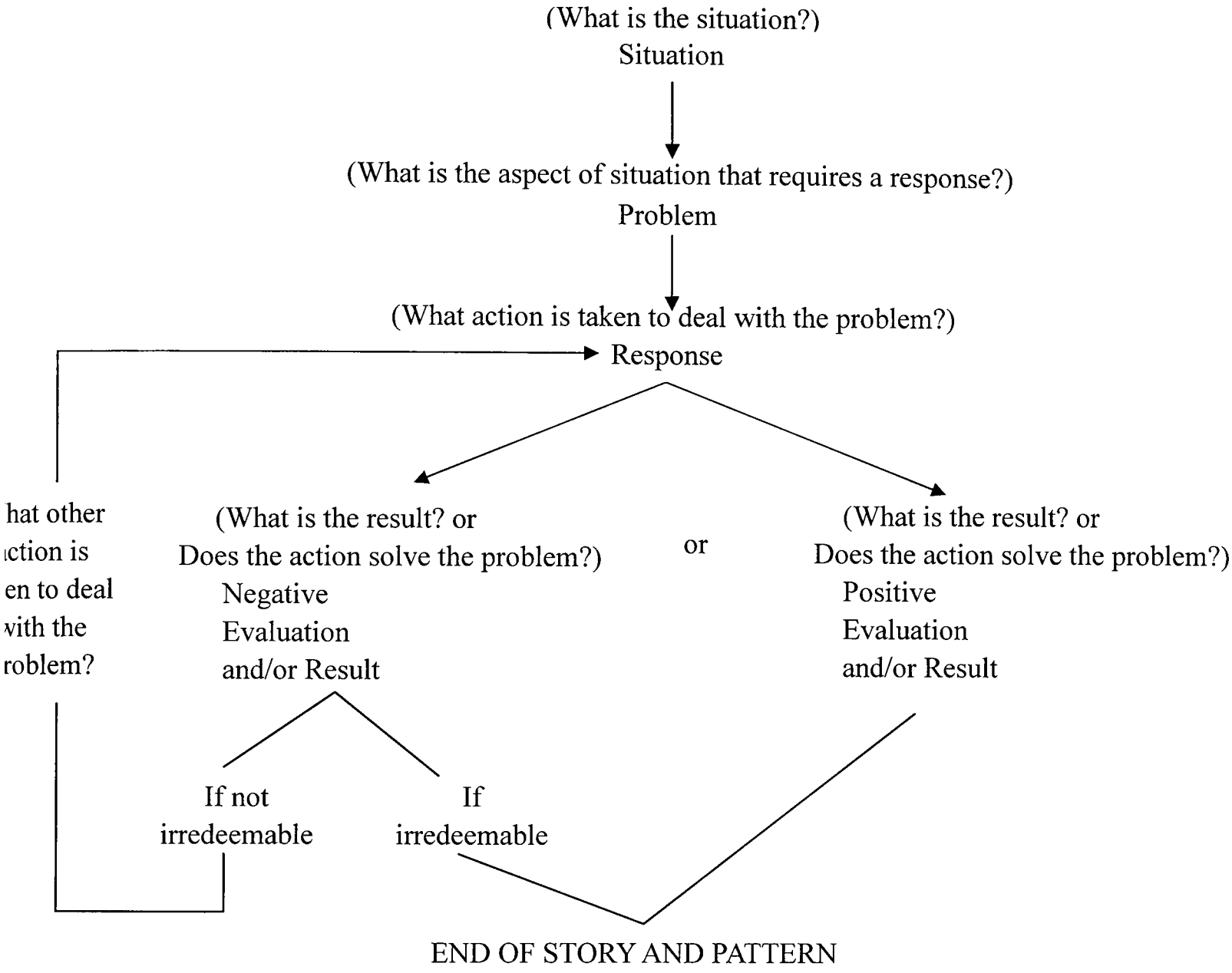


Figure 2.10: The Problem-Solution pattern and questions that bring about the stages of the pattern

It should be noted that the pattern recycles when the Result/Evaluation for the Response is negative but not irredeemable. On the other hand, if the Result/Evaluation for the Response is negative and irredeemable, it functions like a Positive Result/Evaluation to complete the pattern.

The text given below is used by Hoey to illustrate the pattern. As before, sentences are

numbered for ease of reference.

> (1) I was once a teacher of English language. (2) One day some students came to me unable to write their names. (3) I taught them text analysis. (4) Now they all write novels.

The text can be projected into a dialogue given below to spell out the relationships between the sentences.

> Sentence 1: I was once a teacher of English language.
> Question: What problem arose for you?
> Sentence 2: One day some students came to me unable to write their names.
> Question: What did you do about this?
> Sentence 3: I taught them text analysis.
> Question: What was the result?
> Sentence 4: Now they all write novels.
>
> (adapted from Hoey, 2001)

Textual Interaction indicates that the relations that hold between the parts of a text contribute to the organization and coherence of the text. The contribution of conjunctive relations to the organization of a text is also noted in Hoey (1983) and the works of many other linguists, such as Eggins (1994, 2004), Halliday and Hasan (1976), Halliday (1994), Halliday and Matthiessen (2004), Martin (1992), Martin and Rose (2003), Martin *et al* (2010), Mann and Thompson (1988a, 1988b), McCarthy (2002), Matthiessen and Thompson (1988), Matthiessen (2002), Quirk *et al* (1972, 1985), Thompson (1996, 2004, 2005, 2010), and Winter (1977, 1994). The contribution of conjunctive relations to the coherence of a text is also noted in Halliday and Hasan (1985), Hoey (1983, 1994), Martin (1992), Martin and Rose (2003), Mann and Thompson (1988a, 1988b), McCarthy (2002), Matthiessen and Thompson (1988), Matthiessen (2002), Sinclair (1993), Thompson (2005, 2010), and Winter (1977, 1994).

Given that the conjunctive relations that hold between the parts of a text are important in accounting for the coherence of the text, we may expect that they are also important to the meaning of the text in which they occur. Although this hypothesis is supported in theory by the linguists who discuss clause relations, no empirical evidence that validates this hypothesis is given in their works. Nor has the work of psychologists on the effects of conjunctive relations between events on the recall of the events, such as Black and Bern (1981), Bradshaw and Anderson (1982), Caron *et al* (1988), Keenan *et al* (1984), Myers *et al* (1987), and Maury and Teisserenc (2005), provided such empirical evidence because the concern of these studies has been the recall of propositions rather than the conjunctive relations that hold between them. We will examine the validity of this hypothesis in section 4.3.2 of chapter 4 and sections 5.5.1, 5.5.2, and 5.5.3(a) of chapter 5.

2.8 Mann and Thompson's (1988a, 1988b) Rhetorical Structure Theory

The last work reported in this chapter is Mann and Thompson's Rhetorical Structure Theory (RST). RST is 'a linguistically useful method for describing natural texts, characterizing their structure primarily in terms of relations that hold between parts of the text' (Mann and Thompson, 1988b:243). Once again, the size of a part may range from a clause, or a sentence, to a larger chunk.

Central to RST are the rhetorical relations that may hold between any two non-overlapping parts contained in a text. The types of rhetorical relations that occur frequently are given below (adapted from Mann and Thompson, 1988b: 250).

Antithesis
Background
Circumstance
Concession
Condition
Elaboration
Enablement
Evidence and Justify
Interpretation and Evaluation
Motivation
Relations of Cause
Restatement and Summary
Solutionhood

A rhetorical relation is referred to as a multi-nuclear relation when both parts are central to the writer's purpose. Only Contrast, Joint, List, and Sequence are multi-nuclear relations (Mann, 1999). On the other hand, a rhetorical relation is referred to as a nucleus-satellite relation when one part is more central to the writer's purpose than the other. All but Contrast, Joint, List, and Sequence are nucleus-satellite relations (Mann, 1999).

For illustration Mann and Thompson (1988a, 1988b) analyzed the text given below (henceforth referred to as the 'Syncom' text) in terms of the relations that hold between the parts contained in it. They firstly divided a text into units. Each unit is coded as a clause, except that a clausal subject and complement and a restrictive relative clause are not analyzed as a separate unit but as a constituent of the nominal group in which it occurs. The units contained in the 'Syncom' text are numbered for ease of reference. Unit 11 consists of two parts, 11a and 11b. They are separated by unit 12.

1. What if you're having to clean floppy drive heads too often?
2. Ask for Syncom diskettes, with burnished Ectype coating and dust-absorbing jacket liners.
3. As your floppy drive writes or reads,
4. a Syncom diskette is working four ways
5. to keep loose particles and dust from causing soft errors, drop-outs.
6. Cleaning agents on the burnished surface of the Ectype coating actually remove

build-up from the head,
7. while lubricating it at the same time.
8. A carbon additive drains away static electricity.
9. before it can attract dust or lint.
10. Strong binders hold the signal-carrying oxides tightly within the coating.
11a. And the non-woven jacket liner,
12. more than just wiping the surface,
11b. provides thousands of tiny pockets to keep what it collects.
13. To see which Syncom diskette will replace the ones you're using now,
14. send for our free 'Flexi-Finder' selection guide and the name of the supplier nearest you.
15. Syncom, Box 130, Mitchell, SD 57301, 800-843-9862; 605-996-8200.

The rhetorical structure of this text given in Mann and Thompson (1988a: 25) is reproduced in Figure 2.11 given below. In the figure the units that are pointed to by arrows function as nuclei, and those that initiate arrows function as satellites.

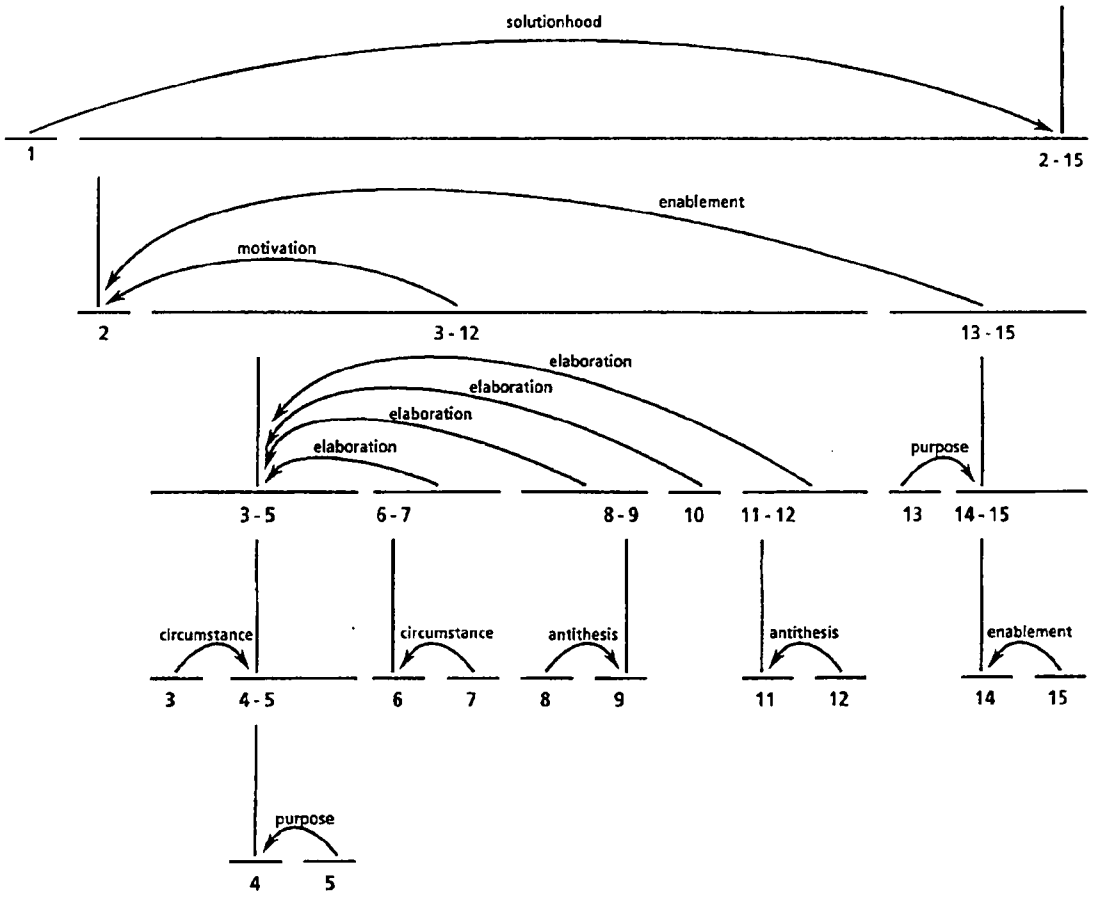


Figure 2.11: The rhetorical structure of the 'Syncom' text

The text is organized in terms of a Solutionhood relation, which is identical to Hoey's Problem-Solution pattern discussed earlier: Unit 1 presents a problem, and the rest of the text advertises a product that is capable of solving the problem. Unit 2 recommends the product (Syncom diskettes) and stands in a motivation relation to units 3-12 because the detailed descriptions in these units of the ways in which Syncom diskettes work are intended to increase the readers' desire to purchase the products. Unit 2 also stands in an enablement relation to units 13-15 because the means given in unit 15 enable the readers to get the services noted in unit 14, namely get a free 'Flexi-Finder' selection guide and the name of the supplier that is nearest to their locations, which in turn increases the readers' possibilities of asking for the products. Units 3-5 stand in an elaboration relation with units 6-7, units 8-9, unit 10, and units 11-12 respectively because each of these four parts elaborates on one of the four ways noted in units 3-5 in which the products work; on the other hand, unit 13 is the purpose of units 14-15. At the next level down a circumstance relation holds between unit 3 and units 4-5, and also between unit 6 and 7. On the other hand, an antithesis relation holds between units 8 and 9. Figure 2.11 shows that Mann and Thompson think of the nucleus of this pair as unit 9 and the satellite as unit 8. However, I would argue that it should be the other way round, that is, the nucleus should be unit 8, and the satellite should be unit 9, evidenced ironically by the statement made by Mann and Thompson pertaining to this pair: 'Unit 9 presents the 'thesis' satellite, the idea that static electricity attracts dust and dirt. By the use of 'before', the writer signals a lack of positive regard for this idea in favor of the nuclear 'antithesis', unit 8, which claims that the static electricity is drained away' (1988b: 262). An antithesis relation also holds between units 11 and 12. Unit 14 stands in an enablement relation with unit 15

because the readers will get the services noted in unit 14 if they contact the company by the means given in unit 15. At the bottom level unit 5 stands in a causal relation of purpose to unit 4.

Mann and Thompson (1988b: 270) note that the nuclei are more central than the satellites to the writer's purpose, and the nuclei are comprehensible independent of the satellites, but not vice versa (Mann and Thompson, 1988: 266). They also make the following statements:

> If units that only function as *satellites* and never as nuclei are deleted, we should still have a coherent text with a message resembling that of the original; it should be something like a synopsis of the original text. If, however, we delete all units that function as *nuclei* anywhere in the text, the result should be incoherent and the central message difficult or impossible to comprehend.
>
> (Mann and Thompson, 1988b: 267-268; original italics)

They illustrate the points noted in the above statements with the 'Syncom' text once again. The units that occur as nuclei at any level in the rhetorical structure of the text make up the following passage (reproduced from Mann and Thompson, 1988b: 268):

2. Ask for Syncom diskettes, with burnished Ectype coating and dust-absorbing jacket liners
4. a Syncom diskette is working four ways
6. Cleaning agents on the burnished surface of the Ectype coating actually remove build-up from the head
8. A carbon additive drains away static electricity
10. Strong binders hold the signal-carrying oxides tightly within the coating
11. And the non-woven jacket liner provides thousands of tiny pockets to keep what it collects
14. send for our free 'Flexi-Finder' selection guide and the name of the supplier nearest you

This passage reads coherently, which can be seen from the fact that it can also be characterized by the relationships that hold between the units. Unit 2 stands in a motivation relation to units

4-11, and also stands in an enablement relation to unit 14. The passage acts as a synopsis for the original text. Mann and Thompson (1988b: 268) noted that 'while this group of nuclear units lacks some cohesion and the grammar of clause combining is missing, we still have a reasonable idea of what the text is about. It tells us to buy Syncom diskettes and gives information motivating and enabling us to do so'.

On the other hand, the units that only function as satellites and never as nuclei make up the passage given below (also reproduced from Mann and Thompson, 1988b: 268):

1. What if you're having to clean floppy drive heads too often?
3. As your floppy drive writes or reads
5. to keep loose particles and dust from causing soft errors, drop-outs
7. while lubricating it at the same time
9. before it can attract dust or lint
12. more than just wiping the surface
13. To see which Syncom diskette will replace the ones you're using now

The juxtaposition of these satellite units do not form a coherent passage because the relations that hold between the units cannot be discerned. Mann and Thompson (1988b: 268) note that 'we cannot discern the purpose of the satellite-only text; it is incomprehensible and incoherent. Furthermore, the satellite-only text contains a number of non-sequiturs'.

Therefore, we may conclude that the relations that hold between the parts in a text contribute to the coherence of the text. A text in which the relationships that hold between the parts contained in it are discernable is more coherent than that in which the relationships that hold between its constituent parts cannot be discerned.

Marcu (1999) conducted experiments with a view to testing Mann and Thompson's claim that the nuclei are more central than the satellites to the meaning of the text in which they occur. He asked informants to judge the units contained in five scientific texts in terms of their degrees of importance to the meaning of the text in which they occur. The units were measured on a three-point scale: unimportant, moderately important and very important. He found that a majority of units in the five texts that are thought of by a majority of the judges as very important to the meaning of the text in which they occur functioned as nuclei at some level in the rhetorical structure of the text. The results of Marcu's experiments support Mann and Thompson's claim, but also suggest that the units that are thought of by a majority of the judges as very important to the meaning of the text in which they occur may not always function as nuclei at some level of the rhetorical structure of the text.

Mann and Thompson (1988a, 1988b)'s Rhetorical Structure Theory once again shows that the relations that hold between the parts of a text contribute to the coherence, organization and meaning of the text.

2.9 Conclusions

This chapter has discussed whether the cohesion that occurs in a text contributes to the coherence, organization, and meaning of the text. Hasan (1984), Halliday and Hasan (1985), Parsons (1991a, 1991b) and Hoey (1991) show that repetitions of items contained in a text contribute to the coherence of the text. Hoey (1991) and Scott's work on keywords show that patterns of lexis in a text are indicative of what the text is about. Hoey (1991) also shows that

patterns of lexis in a text play an important part in organizing the text. On the other hand, Hoey (2001) and Mann and Thompson (1988a, 1988b) show that another aspect of cohesion, namely the conjunctive relations that hold between the parts of a text, also contributes to the coherence, organization and meaning of the text.

In this chapter, I have not yet dealt with the following two issues. The first is that I have not given the criteria that need to be satisfied so that Process X stands in a repetition relation to Process Y. The second issue concerns nominalizations. I have proposed a method that involves unpacking nominalizations so that the cognate verb of a nominalised process and the verbs that repeat the nominalised process can be considered together in chain interactions. But I have not given the criteria that need to be satisfied to ensure that nominalizations are suitable for unpacking. These two issues will be dealt with in the next chapter.

Chapter 3 Research Question 1 and the Method

3.1 Introduction

Having discussed in the previous chapter the studies that discuss the contributions of cohesion in a text to the coherence, organization and meaning of the text, this and the two chapters that follow report on my present study. This chapter gives my first research question and explains the method that explores the question.

3.2 Research question 1 and the data

As noted in my Introduction chapter, my present study aims to investigate the relationships between cohesion in texts and the meaning of the texts. Previous studies working on patterns of lexis in texts, such as Hasan (1984), Halliday and Hasan (1985), Parsons (1991a, 1991b) and Hoey (1991), have analyzed texts in terms of the repetitions of single elements. My present study, on the other hand, analyzes texts in terms of the repetitions of nuclei. The nucleus of a clause is the combination of the Process and its Medium, which is 'the participant through which the Process is actualized' (Halliday, 1994: 163). I analyze texts in terms of repetitions of nuclei because, as noted in the Introduction chapter, the Process is the centre of a clause, the Medium is more closely related to the Process than any other type of participants, and the Process and its Medium together are 'the centre of gravity' of a clause (Halliday and Matthiessen, 1999: 156).

The first question my present study explores is **which recurrent nuclei in a text are thought of by readers as important to the meaning of the text?** In order to answer the question, we need to identify the nuclei that recur in a text. Section 3.3 below gives a method that identifies the repetition of nuclei in a text. After that we need to determine which recurrent nuclei in the

text we have analyzed are thought of by readers as important to the text's meaning. In order to do so, we need to ask readers to recall the gist of the texts they have read. A source from which we can obtain reliable summaries of texts is the Document Understanding Conferences (DUC). An introduction of the DUC is given on http://www-nlpir.nist.gov/projects/duc/intro.html, and cited below.

> Out of the initial workshop and the roadmapping effort has grown a continuing evaluation in the area of text summarization called the Document Understanding Conferences (DUC). Sponsored by the Advanced Research and Development Activity (ARDA), the conference series is run by the National Institute of Standards and Technology (NIST) to further progress in summarization and enable researchers to participate in large-scale experiments. NIST is an agency of the U.S. Commerce Department.

DUC annually invites research institutes around the globe that are interested in text summarization to participate in the conferences. The participants are asked to produce summaries for the texts given by DUC with the automatic summarization schemes they have developed. After that, the qualities of the summaries they produce will be evaluated in terms of the summaries written by information retrieval experts for those texts. My present study analyzed 80 hard news texts that were used in the conferences. These texts are numbered and given in Appendix A. Three summaries written by information retrieval experts employed by the DUC were given for each of 73 texts. On the other hand, four summaries also written by information retrieval experts were given for each of the other seven texts. The summaries of a text were written by different summarizers. Each summary is approximately 10 words in length.

3.3 The method used to explore research question 1

The recurrent nuclei in a text that are repeated in two or more summaries are thought of by readers as important to the text's meaning because they are echoed in a majority of summaries

(in the case of 73 texts, each of which has three summaries) or in at least half of the summaries (in the case of the other seven texts, each of which has four summaries). Therefore, research question 1 can be put alternatively as 'which recurrent nuclei in a text are repeated in two or more summaries of the text?'. Figure 3.1 given below outlines the method used to explore the question.

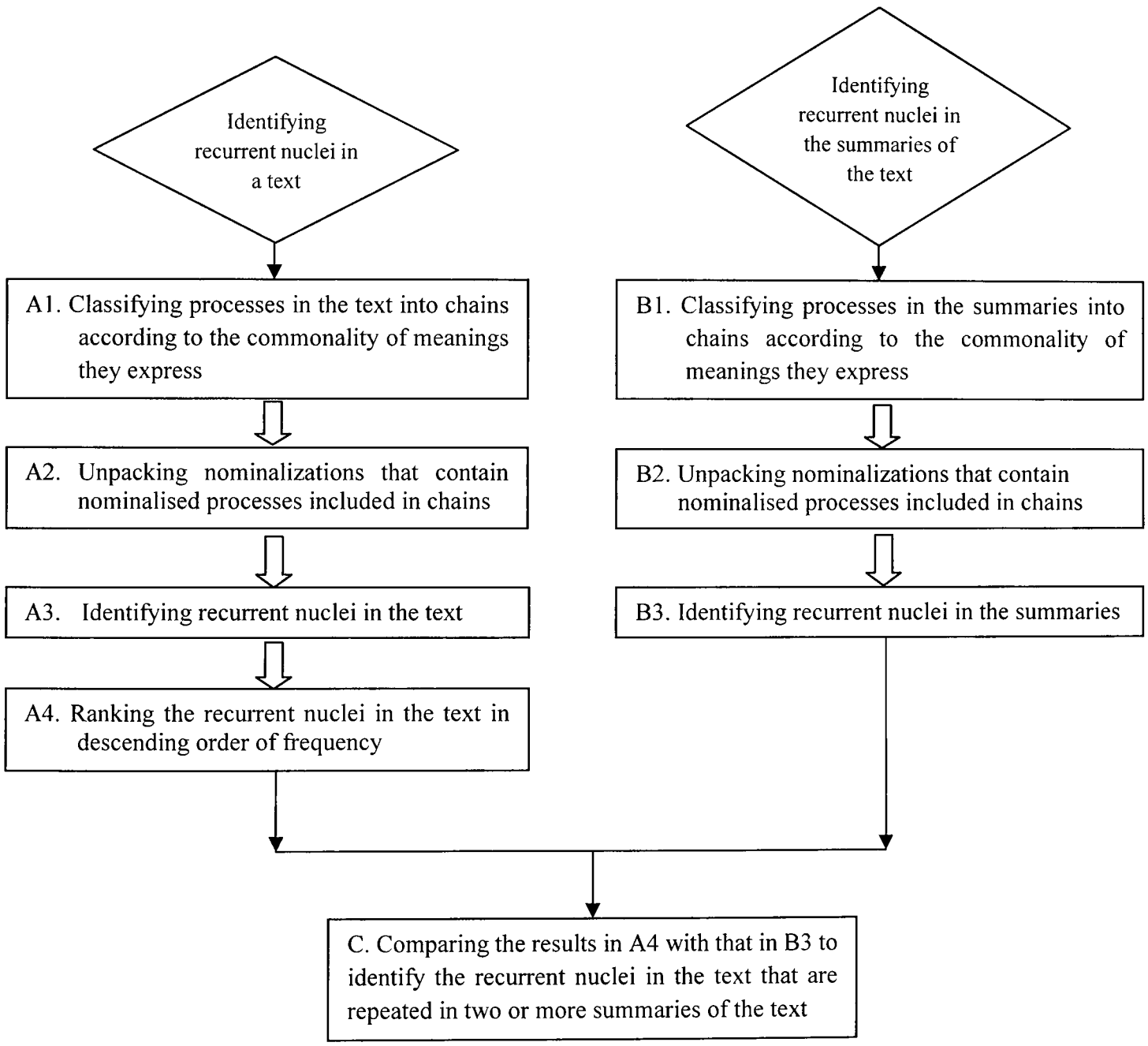


Figure 3.1 The method used to explore research question 1

My method is made up of three parts. The first part identifies recurrent nuclei in a text. In order to do so, we need to go through procedures A1, A2 and A3. A1 classifies the processes in a text into chains according to the commonality of meanings they express. The classifications make use of four types of lexical repetition relations noted in Hoey (1991), namely simple lexical repetition, complex lexical repetition, simple paraphrase and complex paraphrase. After that, A2 unpacks nominalizations that contain nominalised processes that have been included in chains. A3 identifies the repeated Medium(s) (if any) affected by the processes in each chain, and then identifies the recurrent nuclei in the text. Having done that, A4 ranks the recurrent nuclei in the text in descending order of frequency.

The second part of my method aims to identify recurrent nuclei in the summaries of the text. In order to do so, we need to go through procedures B1, B2 and B3. These three procedures are identical to A1, A2 and A3 respectively. The third and final part compares the results in A4 with the results in B3 in order to identify the recurrent nuclei in the text that are repeated in two or more summaries of the text.

The sections that follow explain the procedures in the figure.

3.3.1 Classifying processes in a text into chains according to the commonality of meanings they express

We begin with discussing step A1, which classifies processes in a text into chains according to the commonality of meanings they express. Processes are congruently expressed by verbs (see Figure 2.7 in chapter 2), but they may be incongruently expressed by nouns by means of nominalising (Halliday, 1994: 352; Eggins, 2004: 98-9; Martin and Rose, 2003: 104-5). Halliday (1998) classifies nominalising into four types given in Figure 3.2 below:

(1) verb: active (Actor) noun
'one who/ that which …-s'
e.g. make: maker;
do: doer

(2) verb: passive (Goal) noun
'that which is …-n'
e.g. be made: thing made
be done: thing done, deed

(3) verb: middle (Medium) noun
'…ing' (abstract)
e.g. make: making
do: doing, action

(4) adjective: noun of quality/degree
'being…; how…?'
e.g. big: size, greatness
deep: depth, deepness

Figure 3.2 Four types of nominalising (adapted from Halliday, 1998)

Nouns derived from types 1 and 2 describe 'things', those derived from type 3 describe 'processes', while those derived from type 4 describe 'qualities'. Halliday (1998) explains these four types of nominalising as follows:

> The nouns in (1) and (2) originate as concrete, or at least perceivable, 'things'. Type (1) is an entity, typically a person, identified as actor in, or causer of, a process; type (2) is an object coming into being as product or as outcome of a process. The nouns in (3) and (4), on the other hand, do not represent entities. Here some process itself (3), or else some quality (4), is being

> construed **as if it was** a 'thing': that is, as an ongoing, stable and hence, in (4) measurable phenomenon (original emboldening).

Example 3.1 below contains the first two sentences of text 58, which reports that the president of the U.N. war crimes tribunal referred to Yugoslavia as a "rogue state" after Yugoslavia refused to allow U.N. war crimes tribunal investigators to investigate alleged atrocities in Kosovo.

Example 3.1:

THE HAGUE, Netherlands (AP) _ 1 The president of the U.N. war crimes tribunal angrily branded Yugoslavia a "rogue state" Thursday after Belgrade refused to allow investigators to **probe** alleged atrocities in Kosovo.

2 Authorities in the Federal Republic of Yugoslavia confirmed they would not allow Louise Arbour, chief prosecutor of the Yugoslav war crimes tribunal, and a team of investigators to carry out *investigations* in the troubled region.

Probe (emboldened) in sentence 1 is a verb, expressing a process meaning congruently. *Investigations* (italicised) in sentence 2 is a noun derived from type 3 nominalising, incongruently expressing a process meaning that is in a repetition relation with *probe* contained in sentence 1. On the other hand, two instances of *investigators* (underlined) are derived from type 1 nominalising, expressing entities.

In procedure A1, nouns derived from type 3 nominalising were put into chains when they are repeated by other processes because they describe processes, while nouns derived from the other three types of nominalising were not put into any chain because they do not describe processes. Thus, in Example 3.1 *probe* (as a verb) and *investigations* were put into a single chain, but the two instances of *investigators* were not put into any chain.

The classification of processes in a text into chains is based on four types of repetition relation: simple (lexical) repetition, complex (lexical) repetition, simple paraphrase and complex paraphrase (Hoey, 1991). Process X and process Y were put into a single chain if X stands in any of these four types of relation to Y. Section 2.5 of Chapter 2 has discussed the four types of relation. In the paragraphs that follow I will show how the processes in a text are analyzed in terms of these relations.

Simple repetition occurs 'when a lexical item that has already occurred in a text is repeated with no greater alteration than is entirely explicable in terms of a closed grammatical paradigm' (Hoey, 1991: 53). For a lexical verb (e.g. *tell*), 'a closed grammatical paradigm' refers to its variants in the verb-form paradigm, including its base form (*tell*), its third person singular form (*tells*), its present in present form (*telling*), its past tense form (told), and its past participle form (*told*).

Example 3.2 below contains sentences 1, 5, 7 and 14 in text 73, which reports that China will start building a training center to train giant pandas born in captivity to live in the wild. *Train* in sentence 1, *trained* in sentence 5, two occurrences of *training* (one in sentence 7 and the other in sentence 14), and *teach* in sentence 14 are emboldened.

Example 3.2:

CHENGDU – (1) China will start building a center at the end of May to **train** giant pandas born in captivity to live in the wild, experts said on Wednesday.

(5) In the experimental zone, the giant pandas will be **trained** to reduce their dependency on humans.

(7) After five to 10 years **training** in the experimental zone, the giant pandas that perform well will be introduced into the "half-natural" zone.

(14) China started a giant panda **training** project in 1992 to **teach** the animals to live in the wild.

Trained in sentence 5 is a simple lexical repetition of *train* (as a verb) contained in sentence 1 because the difference between these two verbs is entirely explicable in terms of a different syntactic choice from the verb-form paradigm.

Complex lexical repetition occurs 'either when two lexical items share a lexical morpheme, but are not formally identical, or when they are formally identical, but have different grammatical functions' (Hoey, 1991: 55). In Example 3.2, both instances of *training* (as nouns) are complex lexical repetitions of *train* (as a verb) contained in sentence 1 because *training* (as a noun) and *train* (as a verb) share a lexical morpheme but are not formally identical.

Simple paraphrase occurs 'whenever a lexical item may substitute for another in context without loss or gain in specificity and with no discernible change in meaning' (Hoey, 1991: 62). In Example 3.2, *teach* contained in sentence 14 is a simple paraphrase of *train* (as a verb) contained in sentence 1 because these two verbs can substitute for each other in context.

A paraphrase link is still simple 'if an alteration needed to substitute one item for another in context is necessitated by a grammatical paradigm' (Hoey, 1991: 63). Thus, *teach* (in sentence 14) is a simple paraphrase of *trained* (in sentence 5) because the change of *teach* to *taught* when it substitutes for *trained* is necessitated by a grammatical paradigm.

The last type of repetition relation, namely complex paraphrase, occurs when 'two lexical items are definable such that one of the items includes the other, although they share no lexical morphemes' (Hoey, 1991: 64). As noted in section 2.5 of Chapter 2, Hoey (1991) restricts its application to three situations only.

The first situation relates to antonyms that do not share a lexical morpheme. Example 3.3 below contains the first two sentences in text 63, which reports on Pope Shnouda III's rejection of foreign interference in the scandal that Egyptian police tortured the Copts in a countryside village.

Example 3.3:

Cairo 11-6 (AFP) – (1) The Copts' Patriarch, Pope Shnouda III, affirmed today, Friday, his **rejection** of foreign interference in the affairs of the Copts of Egypt and stressed the solidarity between her Christians and Moslems after the police were accused for torturing Christians in a countryside village.

(2) Pope Shnouda said in a statement, "We do not **accept** foreign interference in our internal affairs, which we are resolving quietly with officials in our country."

If *his rejection of foreign interference in the affairs of the Copts of Egypt* contained in sentence 1 were unpacked to *he rejected foreign interference in the affairs of the Copts of Egypt, rejected* (underlined) in the result after unpacking would be an antonym of *accept* contained in sentence 2 because *rejected* may be paraphrased by *did not accept*.

The second situation in which complex paraphrase may be recognized occurs when an item (A) is a complex lexical repetition of another item (B) and also a simple paraphrase (or an antonym) of a third item (C). In such circumstance, Hoey acknowledges a complex paraphrase link between the second (B) and third items (C). In Example 3.2, a simple

paraphrase of *teach* contained in sentence 14 is *train* (as a verb), which is a complex repetition of *training* (as a noun). These two links establish a complex paraphrase relation between *teach* and *training* (as a noun). Figure 3.3 below shows the link triangle between *train* (as a verb), *training* (as a noun) and *teach*.

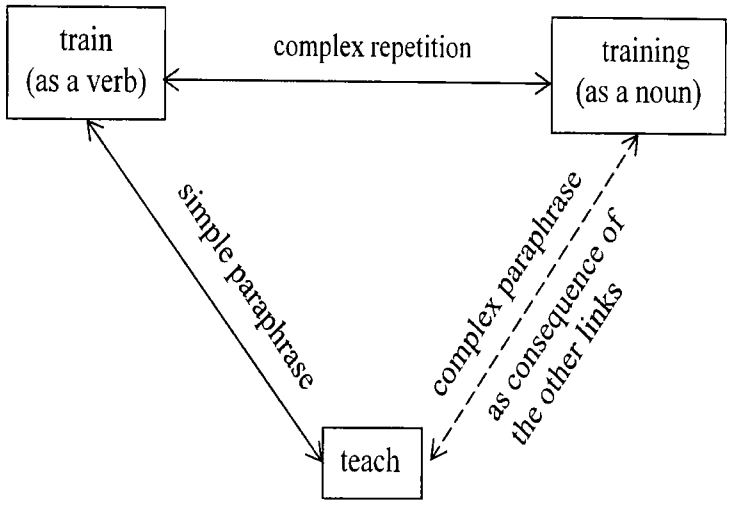


Figure 3.3: Link triangle with regard to *train*, *training* and *teach*

The third and final situation in which a complex paraphrase may be recognized between two items is an complication of the second situation. Figure 3.3 shows that *train* (as a verb) acts as the mediator to establish a repetition between *teach* and *training* (as a noun). It is possible that the mediator is not present in the text. In such situation, a complex paraphrase will still exist between the other two when there is an item 'that is capable of paraphrasing exactly *in that context* one of the items and of repeating the other' (Hoey, 1991: 66, his emphasis). In Example 3.3, *accept* is still a complex paraphrase of *rejection* though *rejected*, which is an antonym of *accept* and also a complex lexical repetition of *rejection*, does not occur in text 63.

In addition to defining the four types of repetition relation, Hoey (1991) also uses flow charts to describe the criteria that need to be satisfied so that a lexical item is treated as

repeating a previous item. I adapted his criteria so as to make them relevant to processes only. Figure 3.4 below contains questions leading to further flow charts. Figure 3.5 and Figure 3.6 give the criteria that need to be satisfied so that a process is considered as a simple or complex lexical repetition of a previous process in a text. Figure 3.7 gives the criteria for paraphrase.

The six criteria below are contained in these flow charts:

1(a) Is a paraphrase of one process also a paraphrase of the other process?

2(a) Does the paraphrase of one of the two processes include a process whose cognate verb is a paraphrase of the other process?

3(a) Is one process a simple paraphrase (or antonym) of a third process that is a complex lexical repetition of the other process?

4(a) Is it possible to paraphrase one of the two processes in such a way that the paraphrase includes a verb that is a simple lexical repetition of (a verb in the paraphrase of) the other process?

5(a) Do they share a common context, or are their contexts related in some way?

6(a) Does one of the two processes share a common context with a third process that is in a repetition relation with the other process?

They are adapted from the following six criteria contained in Hoey (1991) respectively

(1(a) is adapted from 1(b), 2(a) is adapted from 2(b), etc).

1(b) Is a paraphrase of one item a paraphrase of the other?

2(b) Are the items paraphrased in such a way that a paraphrase of one of the words may contain the other?

3(b) Is one of the items a simple paraphrase of a complex repetition of the other items?

4(b) Is it possible to paraphrase one item so that the paraphrase includes at least one lexical item in common with (the paraphrase of) the other item?

5(b) Do the items share a common context, or are their contexts related in some way?

6(b) Does one of the items share a context with an item that is in a repetition relation with the other item?

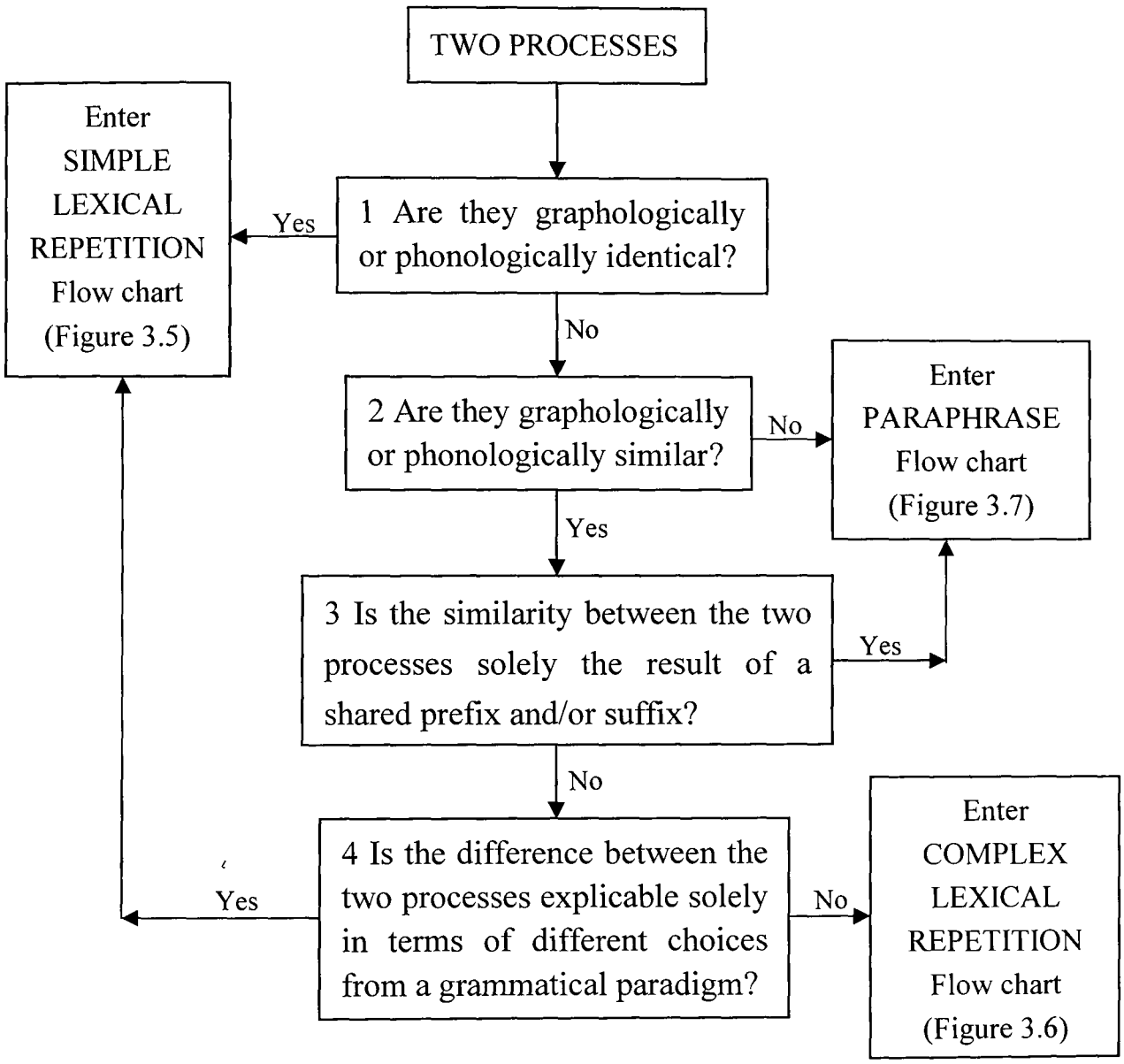


Figure 3.4: Questions leading to Figures 3.5, 3.6 and 3.7

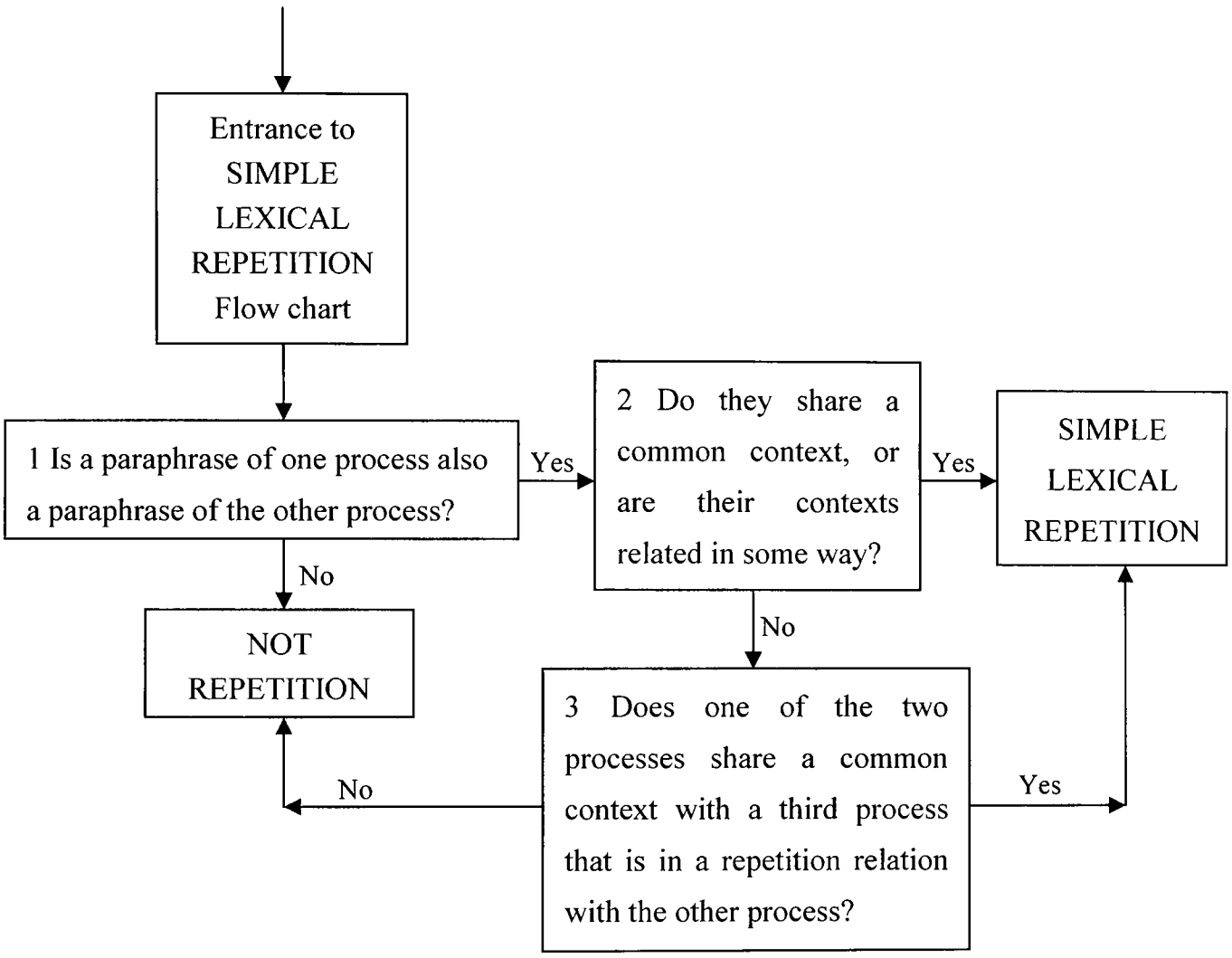


Figure 3.5: Simple lexical repetition

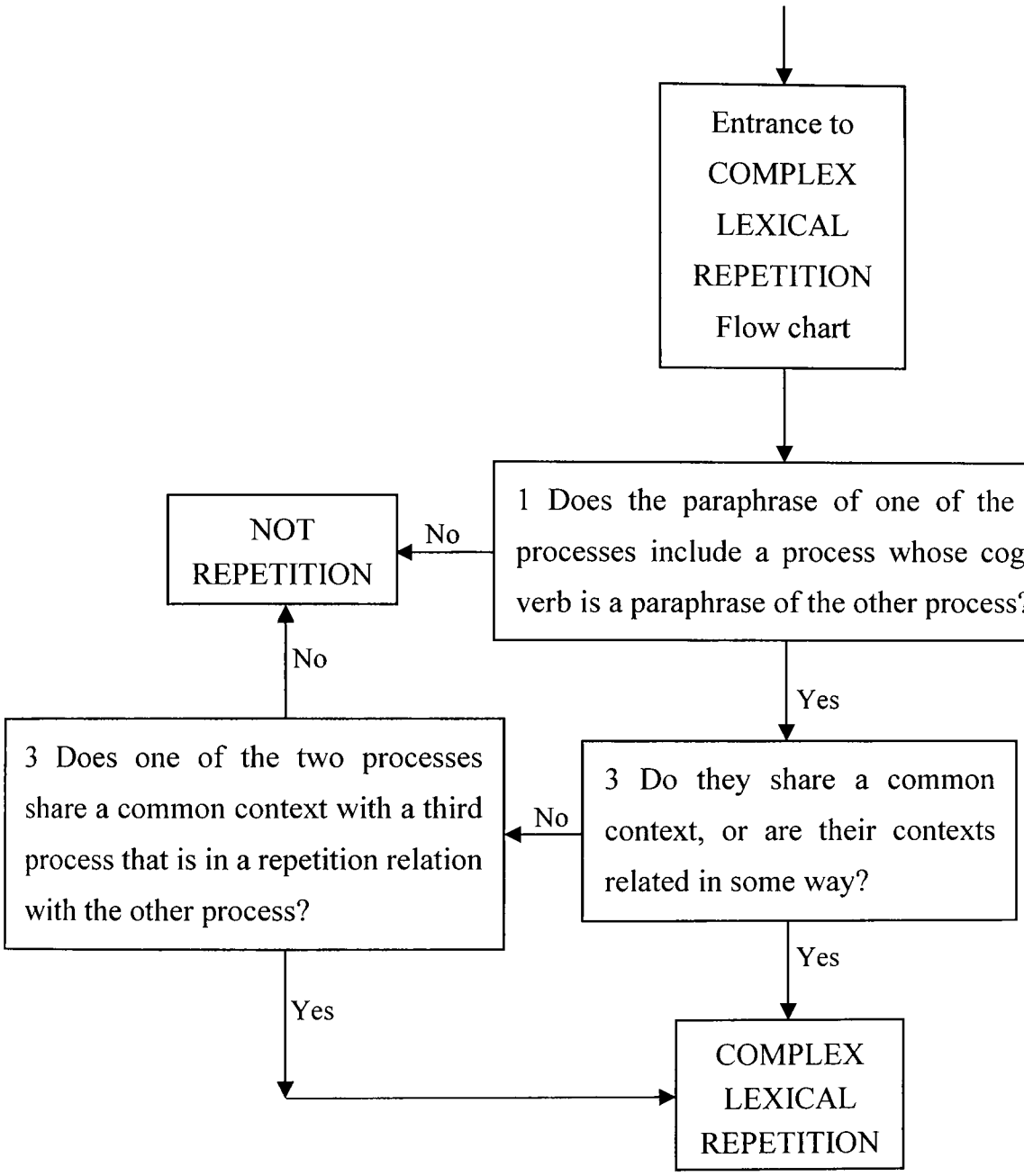


Figure 3.6: Complex lexical repetition

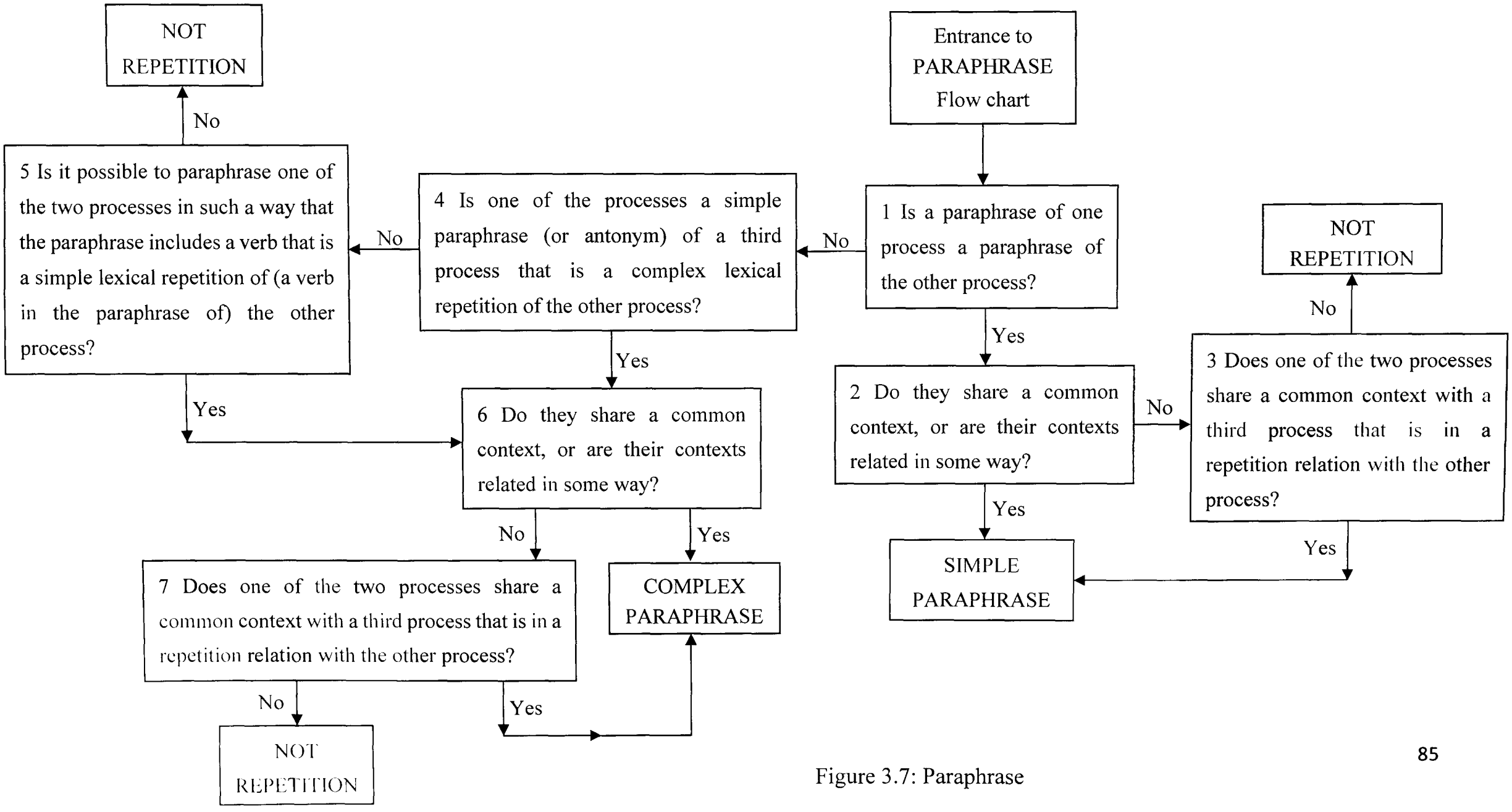


Figure 3.7: Paraphrase

The six criteria used in my present research are explained below. Criterion 1(a), namely 'Is a paraphrase of one process also a paraphrase of the other process?' is expressed in a triangular way with a view to facilitating the treatment of different tokens of the same type. Example 3.4 below contains sentences 4, 10, 12 and 13 of text 4. Two instances of *treatment* (one in sentence 4 and the other in sentence 12), and one instance each of *treating* in sentence 10 and of *treated* in sentence 13 are emboldened.

Example 3.4:

(4) When organizers of a national conference on homelessness wanted a city to illustrate how "mean-spirited" the nation's **treatment** of those without housing has become, they picked San Francisco.

(10) The center also criticized Atlanta, Chicago, New York and Tucson, Ariz., for "criminalizing" rather than **treating** homelessness.

(12) Protesters at the UN Plaza used skits, chants and banners to call on leaders around the nation to focus on long-term solutions to homelessness: affordable housing and better access to drug **treatment** and mental health care.

(13) "Homelessness is being **treated** like ... an animal-control problem," Max Biddel, a homeless advocate from Sacramento, told the crowd of 200 protesters gathered at United Nations Plaza.

The instance of *treatment* in sentence 4 and the one in sentence 12 do not satisfy criterion 1(a) because *treatment of* in sentence 4 can be paraphrased by *behaviour towards* (because no change in meaning of sentence 4 is discernible of after replacing *treatment of* with *behaviour towards*), but *behaviour* is not an acceptable paraphrase of the instance of *treatment* in sentence 12.

On the other hand, *treating* in sentence 10 and *treated* in sentence 13 satisfy this criterion because a paraphrase of the former would be *addressing*, which seems to be an acceptable paraphrase of the latter because no change in meaning of sentence 13 is discernible of after replacing *treated* with *addressed.*

Criterion 2(a), namely 'Does the paraphrase of one of the two processes include a process whose cognate verb is a paraphrase of the other process?', is contained in Figure 3.6. In Example 3.4, the instance of *treatment* in sentence 4 and *treating* in sentence 10 do not satisfy this criterion because, as noted earlier, a paraphrase of *treatment of* (in sentence 4) is *behaviour towards,* which includes a process (*behaviour*) whose cognate verb (*behave*) is not capable of paraphrasing *treating* (in sentence 10) because *behaving homelessness* sounds very odd.

On the other hand, in Example 3.2 the instance of *training* in sentence 7 and *train* (as a verb) in sentence 1 satisfy criterion 2(a) because a paraphrase of *training* (in sentence 7) is *teaching*, whose cognate verb (*teach*) is a paraphrase of *train* (in sentence 1).

Criterion 3(a) asks 'Is one of the processes a simple paraphrase of a third process that is a complex lexical repetition of the other process?'. It does not require the third process which helps us to recognize a complex paraphrase relation between the other two processes to be present in text. An example to illustrate this criterion is *rejection* and *accept* in Example 3.3 discussed earlier.

Criterion 4(a), namely 'Is it possible to paraphrase one of the two processes in such a way that the paraphrase includes a verb that is a simple lexical repetition of (a verb in the paraphrase of) the other process?', is contained in Figure 3.7. Example 3.5 below takes sentences 1, 3, 4 and 9 from text 4. *Arrested* in sentence 1, *firebombed* in sentence 3, *treatment* in sentence 4 and *harass* in sentence 9 are emboldened.

Example 3.5:

(1) SAN FRANCISCO _ In Milwaukee, homeless people can be **arrested** for sleeping on heating grates.

(3) In the farmland city of Jeffersonville, Ind., a homeless mother, father and their infant died last month when their shelter was **firebombed**.

(4) When organizers of a national conference on homelessness wanted a city to illustrate how "mean-spirited" the nation's **treatment** of those without housing has become, they picked San Francisco.

(9) In January, San Francisco was one of five cities named as being especially tough on the homeless, relying on the police to **harass** street people rather than employing social service programs to find health care, jobs and homes for them, said the report by the National Law Center on Homelessness and Poverty in Washington.

Any two of the emboldened processes satisfy criterion 4(a). For instance, in sentence 1 *arrested* can be paraphrased by *treated negatively by being taken to a police station.* In sentence 4 *treatment of* can be paraphrased by *'way of treating'*. The paraphrase of *arrested* includes *treated*, which is a simple lexical repetition of *treating* contained in the paraphrase of *treatment of.*

Criterion 5(a), namely 'Do they share a common context, or are their contexts related in some way?', encompasses three more delicate questions:

a. *Do they co-occur with lexical items denoting the same referent?*
b. *Do they co-occur with lexical items denoting referents that belong to the same class of thing?*

c. *Is there whole or partial parallelism between their contexts?*

Criterion 5(a) will be satisfied if the answer to any of these three questions is positive. These three questions are adapted from the three questions (d, e, and f given below) used by Hoey (1991: 57) when he explained his criterion *Do the items share a common context, or are their contexts related in some way*?:

d. *Do they have common or related contexts? or*

e. *Do the items share common relationships with neighbouring lexical items? or*

f. *Is there whole or partial parallelism between the contexts of the items?*

I do not use Hoey's first question (i.e. question d above) for two reasons. First of all, this question does not tell analysts clearly what to look for. *'Common or related contexts'* in this question is vague and needs further explanation, which is not provided in Hoey (1991). Secondly, this question does not help to explain the criterion (i.e. *'Do the items share a common context, or are their contexts related in some way?*) because it is too similar in expression to the criterion. This question seems to rephrase more than explain the criterion.

The rationale for my first two questions (questions a and b) is given below. Hoey (1991: 54) says that 'one of the factors involved in separating the different (though related) senses of common lexical items is that they have different collocational profiles'. The implication of Hoey's statement is that in order to determine whether two lexical items are

in a repetition relation, it is useful to consider whether they co-occur with lexical items denoting the same referent, and if not, whether they co-occur with lexical items denoting referents belonging to the same class of thing.

Criterion 5(a) is illustrated by the chain denoting 'negative treatment' in text 4. This chain consists of 16 processes given in the table contained in Appendix B. 13 processes in this chain co-occur with items referring to or exemplifying homeless people. The latter are italicised in Example 3.6 below, while the 13 processes are emboldened.

Example 3.6:

(1) SAN FRANCISCO _ In Milwaukee, *homeless people* can be **arrested** for sleeping on heating grates.

(3) In the farmland city of Jeffersonville, Ind., *a homeless mother, father and their infant* died last month when *their* shelter was **firebombed**.

(4) When organizers of a national conference on homelessness wanted a city to illustrate how "mean-spirited" the nation's **treatment** of *those without housing* has become, they picked San Francisco.

(5) About 100 housing advocates from around the country joined more than 100 San Francisco activists at the United Nations Plaza on Saturday to protest what they call a growing problem of civil rights abuses **against** *street people*.

(6) They called for an end to aggressive police policies which, they said, allow the **harassment** of *homeless people* or simply **push** *them* from neighborhood to neighborhood.

(9) In January, San Francisco was one of five cities named as being especially tough on *the homeless*, relying on the police to **harass** *street people* rather than employing social service programs to find health care, jobs and homes for *them*, said the report by the National Law Center on Homelessness and Poverty in Washington.

(16) In San Francisco, advocates for *the homeless* have been angered by numerous police moves **clearing** *street people* from public parks and plazas, and a series of proposals aimed at **controlling** panhandling and the use of shopping carts by *the homeless*.

(20) "They're **sweeping** *street people*, **closing** *their* parks, **confiscating** shopping carts.

(23) "We want to stop the **harassment** of *homeless people* on the streets," said Faith, who runs a homeless program in Columbus serving up to 15,000 *homeless people* each year.

Question (e) in Hoey's sub-list, namely 'Do the items share common relationships with neighbouring lexical items?', is not used at this point. It is reserved to procedure A3 in which repeated Mediums of Processes contained in chains will be identified. Mediums are participants standing in certain transitivity relations to processes; the participant roles that the Mediums are equivalent to in different types of clause are given in section 3.2.3 of this chapter. 'Common relationships' in Hoey's question is narrowed to transitivity relations in procedure A3.

My last question (question c), namely 'Is there whole or partial parallelism between their contexts?', is illustrated by the two partial parallelism structures, namely verb + Goal, and head noun + *of* prepositional phrase found in the contexts of the 13 emboldened processes contained in Example 3.6. Included in the first partial parallelism structure given below are *arrested* in sentence 1, *push* in sentence 6, *harass* in sentence 9, *clearing* in sentence 16 and *sweeping* in sentence 20. These five processes all appear in verb forms and co-occur with *homeless people* or an equivalent expression functioning as the Goal of the processes.

1) verb + goal

(1) *homeless people* can be **arrested**

(6) **push** *them* from neighbourhood to neighbourhood

(9) **harass** *street people*

(16) **clearing** *street people*

(20) **sweeping** *street people*

The other partial parallelism structure given below consists of *treatment* in sentence 4 and two instances of *harassment* (one in sentence 6 and the other in sentence 23). These three processes all function as head nouns in nominal groups, qualified by *of* + *homeless people* or *those without housing.*

2) the Head noun + *of* + *homeless people* (or its equivalent expression)

(4) mean-spirited the nation's **treatment** of *those without housing*

(6) the **harassment** of *homeless people*

(23) the **harassment** of *homeless people*

To the criterion *Do the items share a common context, or are their contexts related in some way?*, Hoey adds the following overriding criterion: 'if a lexical item appears for the third (or more) time in a text, it is only necessary to establish a context connection with *one* of the previous occurrences for the item to be treated as forming a repetition link with *all* previous occurrences' (Hoey, 1991: 57; original italics). The reason why Hoey adds

this overriding criterion is that 'in a long text, the contextual criterion can be difficult to operate, particularly if certain words are recurring with great frequency' (Hoey, 1991: 57). I also found that such an overriding criterion was necessary in my data. Criterion 6(a), namely 'Does one of the two processes share a common context with a third process that is in a repetition relation with the other process?' overrides criterion 5(a). A pair of processes, say A and B, will satisfy criterion 6(a) if A shares a common context with a process repeating B (e.g. C), or if the context of A is related in some way to that of C.

An example satisfying criterion 6(a) is *sweeps* (in sentence 17) and *harassment* (in sentence 23) contained in text 4 (These two sentences are given in Example 3.7 below). These two processes do not satisfy criterion 5(a) because none of the answers to the three questions subsumed in criterion 5(a) (i.e. question a, b or c) are positive. However, these processes satisfy the overriding criterion 6(a) because *clearing* (in sentence 16), which is in a repetition relation with *harassment*, shares a common context with *sweeps*: *In San Francisco* (underlined) contained in the context of *clearing* is in a repetition relation with *in the city* (also underlined) contained in the context of *sweeps*.

Example 3.7:

(16) In San Francisco, advocates for the homeless have been angered by numerous police moves **clearing** street people from public parks and plazas,

(17) Mayor Willie Brown has defended the **sweeps** as an improvement of quality of life in the city.

(23) "We want to stop the **harassment** of homeless people on the streets," said Faith, who runs a homeless program in Columbus serving up to 15,000 homeless people each year.

Neither *intolerance* in sentence 7 nor *being tough on* in sentence 9 given as Example 3.8 below is included in the chain denoting 'negative treatment'.

Example 3.8:

(7) "There is a growing **intolerance** of the homeless around the nation," said Bill Faith, board president of the National Coalition for the Homeless, which held its annual meeting in San Francisco last weekend.

(9) In January, San Francisco was one of five cities named as **being** especially **tough on** the homeless, relying on the police to harass street people rather than employing social service programs to find health care, jobs and homes for them, said the report by the National Law Center on Homelessness and Poverty in Washington.

These two processes appear to be similar to the 13 processes contained in Example 3.6 in that both *intolerance* and *being tough on* not only express negativities towards homeless people, as before, but also share common contexts with the 13 processes. The nominal group *the homeless* is found in the contexts of both *intolerance* and *being tough on*; the equivalent expressions or exemplifications of *the homeless* also occur in the contexts of the 13 processes.

In spite of the similarities described above, it is safe not to accept *intolerance* or *being tough on* into the chain denoting 'negative treatment' because *intolerance* is derived from a mental process verb (*tolerate*), telling that the authorities are reluctant to put up with homeless people; *being tough on* is in the form of a relational process, expressing negative attitude towards the homeless. Neither *intolerance of the homeless* nor *being tough on the homeless* tells that the homeless were physically ill-treated in any way.

3.3.2 Unpacking in text nominalizations that contain nominalised processes included in chains

Having classified processes in a text into chains according to the commonality of meanings they express, the next step (A2 in Figure 3.1) unpacks nominalizations that contain nominalised processes that have been included in chains. In this section, I shall answer three important questions regarding the unpacking of nominalizations, namely why nominalizations were unpacked in my present study, when should they be unpacked. and how they were unpacked.

Why nominalizations were unpacked in my present study

The first reason for unpacking nominalization in my study is that unpacking nominalizations enables the transitivity relations between the nominalised processes and the participants (if present) to be laid bare so that the Mediums of Processes can be identified in step A3.

This is illustrated by Example 3.9 given below, which contains the first two sentences of text 53. Both *arrest* (as a noun) in sentence 1 and *arrested* in sentence 2 are emboldened.

Example 3.9:

(1) The **arrest** of Gen. Augusto Pinochet shows the growing significance of international human-rights law, suggesting that officials accused of atrocities have fewer places to hide these days, even if they are carrying diplomatic passports, legal scholars say.

(2) Pinochet, who ruled Chile as a military dictator from 1973 to 1990, was granted amnesty in his homeland but was **arrested** on Friday in London at the request of Spanish authorities, who want him extradited to Spain.

In *the arrest of Gen. Augusto Pinochet* contained in sentence 1 the transitivity relation between *arrest* and *Gen. Augusto Pinochet* may be ambiguous when considered independently of the discourse after sentence 1 because *Gen. Augusto Pinochet* could be either the Goal (e.g. *Gen. Augusto Pinochet was arrested*) or the Actor (e.g. *Gen. Augusto Pinochet arrested human rights violators*).

After *the arrest of Gen. Augusto Pinochet* was unpacked (see below), the transitivity relation between *Gen. Augusto Pinochet* and the cognate verb of *arrest* (as a noun) (i.e. *arrested*) was rendered explicit: *Gen. Augusto Pinochet* was the Goal. The method of unpacking nominalizations will be given later in Figure 3.12. Sentence 1 after *the arrest of Gen. Augusto Pinochet* is unpacked is given below. *The arrest of Gen. Augusto Pinochet* is replaced by *The fact that Gen. Augusto Pinochet was arrested* (italicised). The remaining words in sentence 1 are unchanged.

(1) *The fact that Gen. Augusto Pinochet was arrested* shows the growing significance of international human-rights law, suggesting that officials accused of atrocities have fewer places to hide these days, even if they are carrying diplomatic passports, legal scholars say. (sentence 1 of text 53 after unpacking)

The second reason for unpacking nominalizations is that unpacking them enables the Mediums (of Processes) to be recovered so that they can be identified in step A3.

Construing processes as 'things' makes the mention of the participants involved in the process optional (Thompson, 2003) because after nominalising, the participants will become the modifiers of 'things' in nominalizations, and modifiers are optional in nominal groups.

This is illustrated by Example 3.10 given below, which contains sentences 1 and 16 of text 17 reporting on the kidnapping of three Britons and a New Zealand citizen in Chechnya.

Example 3.10:

(1) Unknown gunmen abducted three Britons and a New Zealand citizen in Russia's breakaway Chechnya region after a shootout with their bodyguards, a regional official said Sunday.

(16) **The kidnapping** came just two weeks after Chechen warlords released two British charity workers, ending the pair's 14 months of captivity in the turbulent republic in Russia's northern Caucasus region.

The kidnapping (emboldened) contained in sentence 16 recapitulates the abduction event described in sentence 1. The Medium (i.e. the participants that were kidnapped) is not present in this nominalization.

After *the kidnapping* was unpacked, the unmentioned Mediums, namely *three Britons and a New Zealand citizen*, were recovered. Sentence 16 after unpacking is given below. *The kidnapping* is replaced by *The incident of three Britons and a New Zealand citizen being kidnapped* (italicised). The remaining words in sentence 16 are unchanged.

(16) *The incident of three Britons and a New Zealand citizen being kidnapped* came just two weeks after Chechen warlords released two British charity workers, ending the pair's 14 months of captivity in the turbulent republic in Russia's northern Caucasus region. (sentence 16 of text 17 after unpacking)

When should nominalizations be unpacked

This question is concerned with the criteria that need to be satisfied to determine whether nominalizations are suitable for unpacking. To my best knowledge such criteria

do not exist in the current literature. I, therefore, made considerable effort to fill in this void.

My criteria for unpacking nominalizations in texts are given in Figures 3.9 and 3.11. It is important to point out that the nominalised processes judged by these criteria are those that have been put in chains in step A1. This means that only nouns derived from type 3 nominalising in Halliday's classification (shown in Figure 3.2 given earlier) are suitable to be judged by these criteria because only this type of noun denotes process. Nouns derived from the other three types of nominalising were not unpacked in my present study.

The nominalization in question firstly needs to be interrogated by the question contained in Figure 3.8, namely whether the nominalised process is derived from a verbal or a mental process?. When the answer to this question is negative, that is, when the nominalised process is derived from neither a verbal nor a mental process, we go to Figure 3.9. On the other hand, when the answer to this question is positive, that is, when the nominalised process is derived from either a verbal or a mental process, we go to Figure 3.11.

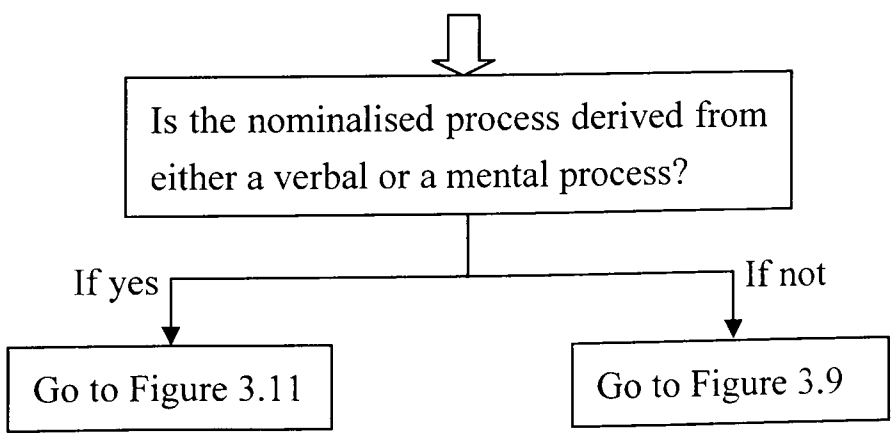


Figure 3.8: Directory to Figures 3.9 and 3.11

The remaining text of this section explains the criteria contained in the flowcharts represented as Figures 3.9 and 3.11.

The criteria for unpacking in a text nominalizations that contain nominalised processes derived from neither verbal nor mental processes

We begin with Figure 3.9 (given on p101-102), which gives the criteria for unpacking in a text nominalizations that contain nominalised processes derived from neither verbal nor mental processes.

Criterion 1 in Figure 3.9, namely 'Does the nominalised process have a cognate verb?', requires that the nominalised processes have cognate verbs in order to be suitable for unpacking because after nominalizations are unpacked, the nominalised processes are turned into their cognate verbs.

Criterion 2 requires that the nominalised process must function as Thing in the nominal group in which it immediately occurs. 'Thing' is the head noun of a nominal group that is not elliptical. 'Thing' is 'the semantic core of the nominal group' (Halliday and Matthiessen, 2004: 325).

In my data, it is not rare that nominalised processes function as the Classifiers of head nouns in nominal groups, such as *anti-**kidnapping** department*, *a giant panda **training** project*, *a **detention** center*, *the joint military-**prosecution investigation** team*, *the joint **probe** team*, *the **assassination** day*, etc (nominalised processes functioning as Classifiers

are emboldened). The last three given above are contained in Example 3.11 below, which takes sentences 1 and 6 from text 1 reporting that no evidence shows that then Korean Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in the assassination of former Korean President Pak Chong-hui in 1979. In Example 3.11 these three nominalizations are italicised, and the nominalised processes contained in these nominal groups are emboldened.

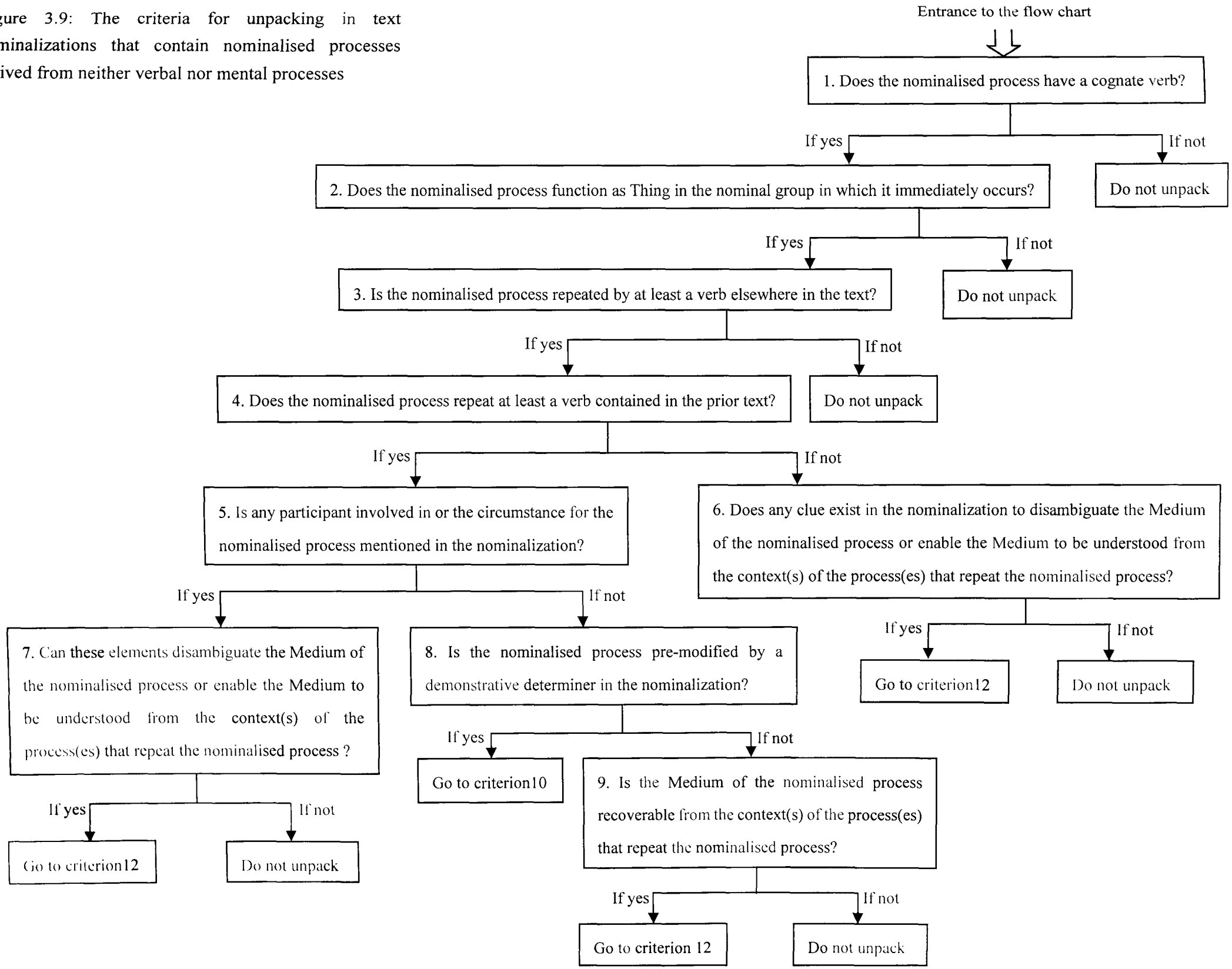
Example 3.11:

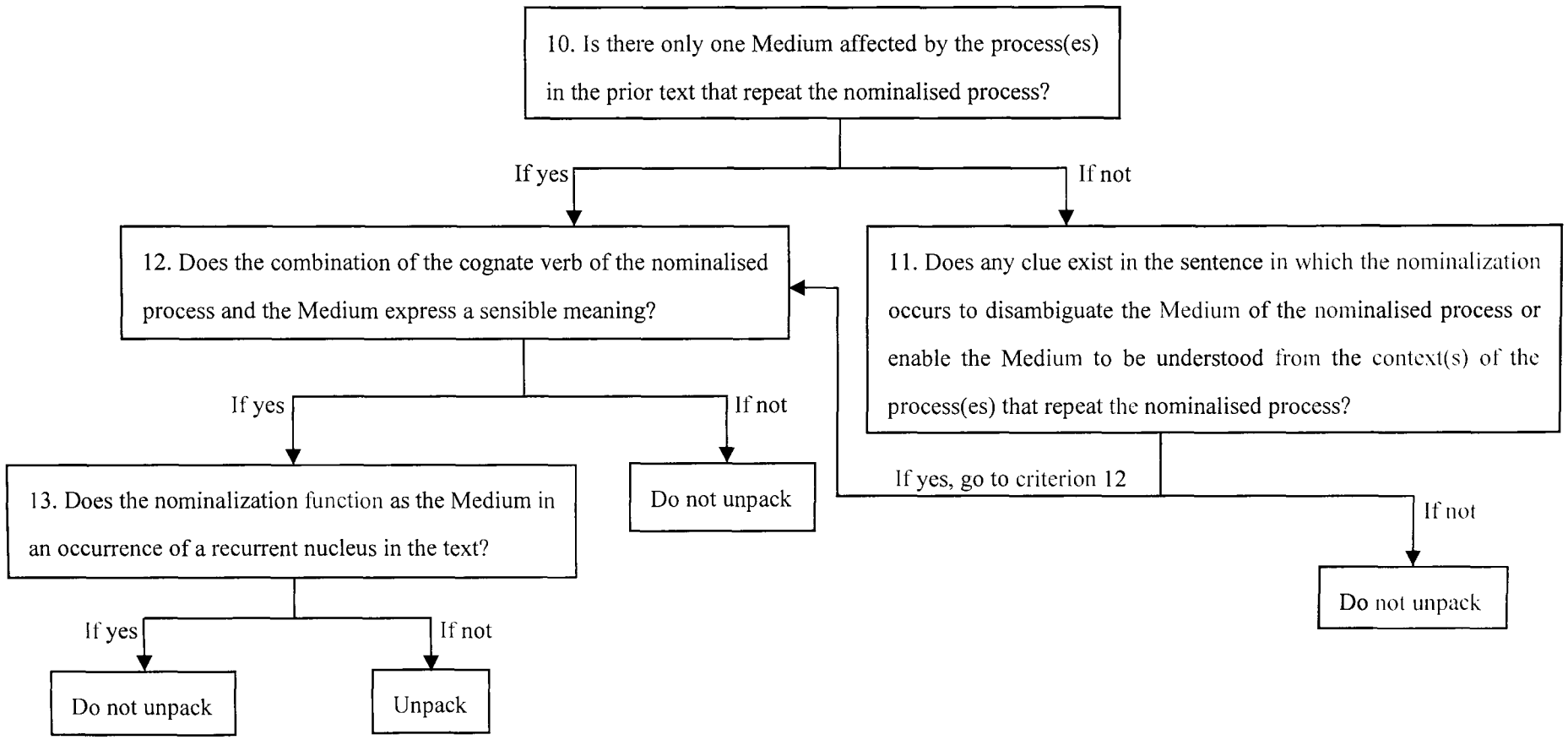
(1) *The joint military-**prosecution investigation** team*, which probed the assassination of President Pak Chong-hui by then Central Intelligence Agency Director Kim Chae-kyu in October 1979, had found no evidence that then Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in the slaying, a former team member testified recently.

(6) *The joint **probe** team* felt the need to question Chong because when Kim killed Pak he was nearby and Prosecutor Chong Kyong-sik, a member of the joint team, called on the martial law commander at his office three times to question him about his actions on *the **assassination** day*.

Example 3.11 also contains two instances of *assassination* (underlined). The one contained in sentence 1 functions as 'Thing' in this complex nominal group: *the assassination of President Pak Chong-hui by then Central intelligence Agency Director Kim Chae-kyu in October 1979*, post-modified by the prepositional phrase *of President Pak Chong-hui by then Central intelligence Agency Director Kim Chae-kyu in October 1979*. Thus, this instance satisfies criterion 2. On the other hand, the one contained in sentence 2 functions as Classifier pre-modifying 'Thing' (i.e. *day*) in the nominal group *the assassination day*. Therefore, this instance does not satisfy criterion 2.

Figure 3.9: The criteria for unpacking in text nominalizations that contain nominalised processes derived from neither verbal nor mental processes




10. Is there only one Medium affected by the process(es) in the prior text that repeat the nominalised process?
If yes
If not
12. Does the combination of the cognate verb of the nominalised process and the Medium express a sensible meaning?
11. Does any clue exist in the sentence in which the nominalization occurs to disambiguate the Medium of the nominalised process or enable the Medium to be understood from the context(s) of the process(es) that repeat the nominalised process?
If yes
If not
If yes, go to criterion 12
If not
13. Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the text?
Do not unpack
Do not unpack
If yes
If not
Do not unpack
Unpack

Criterion 2 allows for the situation that the nominal group in which the nominalised process immediately occurs is part of a prepositional phrase or an embedded clause post-modifying the head noun of a larger nominal group. This is illustrated by *his rejection of foreign interference in the affairs of the Copts of Egypt* (emboldened) contained in the lead of text 63 given in Example 3.12 below.

Example 3.12:

Cairo 11-6 (AFP) – 1 The Copts' Patriarch, Pope Shnouda III, affirmed today, Friday, **his rejection of foreign interference in the affairs of the Copts of Egypt** and stressed the solidarity between her Christians and Moslems after the police were accused for torturing Christians in a countryside village.

The analysis of *his rejection of foreign interference in the affairs of the Copts of Egypt* is given in Figure 3.10 below. This complex nominal group includes two nominalised processes: *rejection* and *interference*. *Interference* functions as Thing in Nominal Group A, namely *foreign interference in the affairs of the Copts of Egypt*. Nominal Group A is contained in the prepositional phrase *of foreign interference in the affairs of the Copts of Egypt*, which post-modifies *rejection* in the larger nominal group (Nominal Group B) *his rejection of foreign interference in the affairs of the Copts of Egypt*.

Both *rejection* and *interference* satisfy criterion 2. The nominal group in which *rejection* immediately occurs is Nominal Group B, and *rejection* is Thing in this nominal group. The nominal group in which *interference* immediately occurs is Nominal Group A. and *interference* is also Thing in this nominal group.

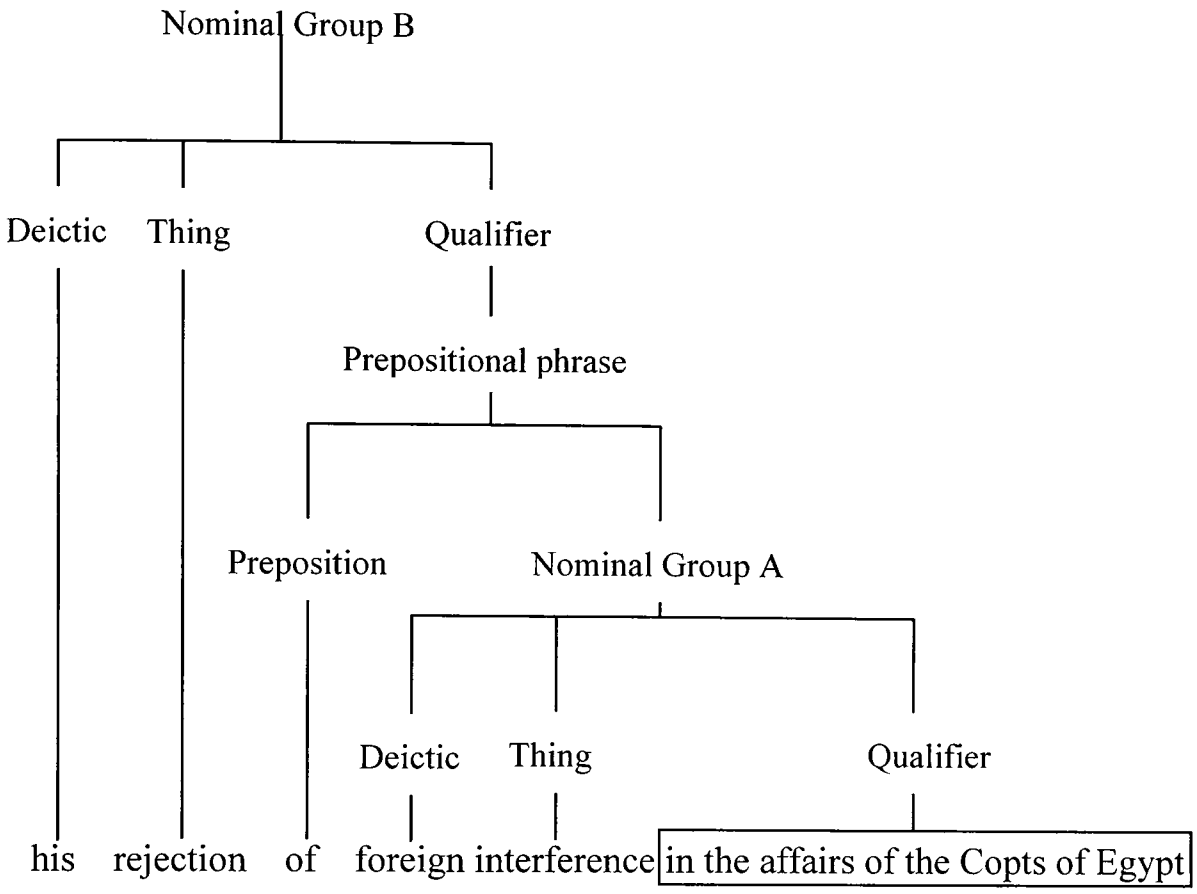


Figure 3.10: The analysis of *his rejection of foreign interference in the affairs of the Copts of Egypt*

Therefore, a complex nominal group may need to be unpacked more than once due to recursive post-modification. Recursive post-modification may in principle be repeated indefinitely, but it is fairly limited in naturally occurring text (Bloor & Bloor, 1995: 161), indicating that in naturally occurring text the actual number of steps taken to unpack nominalizations may be limited. In my analysis the largest number of steps taken to unpack nominalizations is two.

When criterion 2 is satisfied, we go to criterion 3, namely 'Is the nominalised process repeated by at least a verb elsewhere in the text?'. If the answer to this question is positive, unpacking nominalizations is safe because the meaning of the nominalised process is expressed as a verb elsewhere in the text. The nominalised process may be

repeated by a verb in terms of either complex lexical repetition defined in Figure 3.6 or complex paraphrase defined in Figure 3.7.

Following criterion 3, criterion 4 probes whether the nominalised process repeats at least a verb contained in the prior text (henceforth, 'the prior text' means the beginning of the text up to the point where the nominalization occurs).

Two routes lead from criterion 4. I begin by explaining what happens when the answer to criterion 4 is negative because the criteria relating to this situation are less complicated than those in the other situation.

When the answer to criterion 4 is negative, we go to criterion 6, namely 'Does any clue exist in the nominalization to disambiguate the Medium of the nominalised process or enable the Medium to be understood from the context(s) of the process(es) repeating the nominalised process?'. Criterion 6 is explained with reference to Example 3.13 given below, which includes five sentences, four of which, namely sentences 1, 2, 6, and 11, are drawn from text 3 which reports that a bereaved mother sought a national ban on cell phone use in cars. I added one sentence that is not present in text 3 in order to explain how criterion 6 deals with complicated situations. This made-up sentence appears after sentence 11. I give it a hypothetical ordinal sentence number, A. *Ban* contained in sentence 2 and the items repeating it in the other sentences are emboldened.

Example 3.13:

(1) Patricia Pena gave little thought to using a cellular telephone on the road before a distracted driver hit her car broadside, fatally injuring her 2-year-old daughter.

(2) Now she has devoted herself to pushing for a nationwide **ban** on cell phone use in cars, except in emergencies.

(6) No state **bans** the use of wireless phones in automobiles, and only California, Florida and Massachusetts have laws limiting cell phone use in moving vehicles, according to the National Conference of State Legislatures.

(11) Hilltown now **bans** certain cell phone uses for drivers.

(A) The **ban** on iPod use is also advocated because using iPod while driving is shown to be another major contributor to car accidents.

Suppose we are now considering the instance of *ban* in sentence 2. It occurs in this nominalization: *a nationwide ban on cell phone use in cars*. Had this instance of *ban* not been post-modified by *cell phone use in cars*, it would have been difficult to determine the Medium because two (different) Mediums are affected by the processes repeating this instance of *ban*. These two Mediums are *the use of wireless phones in automobiles* contained in sentence 6 (this Medium is repeated partially by *certain cell phone uses* contained in sentence 11) and *iPod use* contained in sentence A.

Criterion 6 requires that some clue exist in the nominalization to disambiguate the Medium of the nominalised process or enable the Medium to be understood from the contexts of the process(es) that repeat the nominalised process. The instance of *ban* in sentence 2 satisfies this criterion because, as just noted, the qualifier *on cell phone use in cars* serves to disambiguate the Medium.

Having discussed the situation when the answer to criterion 4 is negative, the other situation, that is, when the answer to criterion 4 is positive, is discussed below. In my data the nominalizations found in such a situation considerably outnumber those found in the situation when the answer to criterion 4 is negative.

When the answer to criterion 4 is positive, we arrive at criterion 5, namely 'Is any participant involved in or the circumstance for the nominalised process mentioned in the nominalization?'. An example satisfying this criterion is *the harassment of homeless people* contained in sentence 6 of text 4 because *homeless people*, which post-modifies the nominalised process (*harassment*), is a participant involved in the process. On the other hand, *the kidnapping* contained in sentence 16 of text 17 (given in Example 3.10) does not satisfy this criterion because the nominalised process, namely *kidnapping*, is not modified by any elements other than a definite article (*the)*, which describes neither the participant(s) involved in nor the circumstance(s) for the nominalised process.

Two different routes lead from criterion 5. When the answer to criterion 5 is positive, we go to criterion 7. On the other hand, when the answer is negative, we go to criterion 8.

Criterion 7 asks 'Can these elements (referring to the participant(s) involved in and/or circumstance(s) for the nominalised process) disambiguate the Medium of the nominalised process or enable the Medium to be understood from the context(s) of the process(es) that repeat the nominalised process?'. An example that satisfies this criterion is *The ban on iPod use* contained in sentence A of Example 3.13. The element post-modifying *ban* (i.e. *iPod use*) will be the Medium of the cognate verb of *ban* (as a noun) after this nominal group is unpacked.

When the answer to criterion 5 is negative, we go to criterion 8, namely 'Is the nominalised process pre-modified by a demonstrative determiner in the nominalization?'. Two routes lead out of criterion 8, which will be dealt with in turn.

When the answer to criterion 8 is positive, the demonstrative determiner pre-modifying the nominalised process signals that the nominalization recapitulates certain activities described in the prior text. (Note that criterion 4 must have been satisfied before reaching criterion 10 at this point, meaning that the nominalised process must repeat at least a verb contained in the prior text.) In such a situation, we go to criterion 10, namely 'Is there only one Medium affected by the process(es) in the prior text that repeat the nominalised process?'. Though the Medium of the nominalised process is not contained in the nominalization (because the answer to criterion 5 was negative earlier), when the answer to criterion 10 is positive, the demonstrative determiner pre-modifying the nominalised process signals that it is the same as that affected by the process(es) in the prior text repeating the nominalised process.

This is illustrated with Example 3.14 given below. Sentences 2, 3, 10 and 18 of text 17 are added to Example 3.10 to form Example 3.14. *Abducted* in sentence 1 and the processes in the other five sentences repeating *abducted* are emboldened. *Abducted*, *taken hostage* and *seized* are in verb form. The Mediums of these three verb forms are *three Britons and a New Zealand citizen*, *the four men* and *the four* respectively. These three Mediums, italicised, have the same situational referent.

Example 3.14:

(1) Unknown gunmen **abducted** *three Britons and a New Zealand citizen* in Russia's breakaway Chechnya region after a shootout with their bodyguards. a regional official said Sunday.

(2) *The four men*, who were in Chechnya to install a cellular phone system, were **taken hostage** Saturday after 20 men in camouflage attacked the house where they lived in the

Chechen capital Grozny, said Shadid Bargishev, head of the regional anti-**kidnapping** department.

(3) The British Foreign Office initially said *the four* were **seized** early Sunday morning, but later said the **kidnapping** had taken place at 4pm. Saturday.

(10) The **kidnapping** took place only 500meters from the anti-**kidnapping** department, Interfax said.

(16) The **kidnapping** came just two weeks after Chechen warlords released two British charity workers, ending the pair's 14 months of captivity in the turbulent republic in Russia's northern Caucasus region.

(18) Chechen President Aslan Maskhadov's government has been unable to stem **kidnappings** that have become a steady business for some criminal gangs.

Suppose we are now considering the instance of *the kidnapping* in sentence 16. The definite article *the* pre-modifying *kidnapping* signals that this nominalization recapitulates an abduction event described in the discourse before sentence 16. From the beginning of this text up to sentence 16, there is no other abduction event than the fact that three Britons and a New Zealand citizen were abducted in Chechnya. Therefore, the Mediums of the cognate verb of *kidnapping* (as a noun) after the three instances of *the kidnapping* (in sentences 3, 10, and 16) have been unpacked are *three Britons and a New Zealand citizen.*

Kidnappings contained in sentence 18 will not be discussed at this point because the nominalised process (*kidnappings*) is not pre-modified by any demonstrative determiner in the nominal group. Therefore, *kidnappings* does not belong to the situation we are now discussing. It will be discussed at a later point when the situation it belongs to, that is, when the answer to criterion 8 is negative, is discussed.

If the answer to criterion 10 is negative, there are different Mediums affected by the process(es) in the prior text that repeat the nominalised process. We are facing two difficulties when a nominalization occurring in such a situation is unpacked. Firstly, neither the participant(s) involved in nor the circumstance(s) for the nominalised process are mentioned in nominalization (as the answer to criterion 5 must be negative before reaching criterion 10). Thus it is not possible to use these elements to disambiguate the Medium of the nominalised process. Secondly, although the nominalised process is pre-modified by a demonstrative determiner, which signals that the nominalization recapitulates an activity described in the prior text, the negative answer to criterion 10 means that there is more than one Medium affected by the processes in the prior text that repeat the nominalised process.

Despite these two difficulties, some clues may exist in the sentences in which the nominalizations occur to disambiguate the Mediums of the nominalised processes or enable the Mediums to be understood from the contexts of the processes that repeat the nominalised processes. Criterion 11, namely 'Does any clue exist in the sentence in which the nominalization occurs to disambiguate the Medium of the nominalised process or enable the Medium to be understood from the context(s) of the process(es) that repeat the nominalised process?', allows for this possibility.

Criterion 11 is illustrated with Example 3.15 below, which contains sentences 1, 5 and 11 from text 35 which reports that US administration officials acknowledged that the

targets of the raid on Afghanistan were Osama bin Laden and his lieutenants. *Raid* in the sentence 1 and the items in the other two sentences repeating *raid* are emboldened. Some common contexts shared by *raid* (in sentence 1) and *attack* (in sentence 11) are underlined.

Example 3.15:

(1) The objective of last August's **raid** on Afghanistan was to kill Osama bin Laden and as many of his lieutenants as possible, administration officials said yesterday.

(5) But officials said the White House legal opinion drafted before the Afghan **strike** asserts that the president has authority to target the 'infrastructure' of terrorist groups that are **attacking** Americans.

(13) They (U.S. intelligence officials) said there were reports that at least one of his senior lieutenants may have died in the **attack**, and that total casualties in the complex were between 50 and 100.

Suppose we are considering *the attack* in sentence 13. The Medium of the cognate verb of *attack* (as a noun) is not mentioned in this nominalization. The definite article (*the*) signals that this nominalization recapitulates. It is, however, necessary to determine the activity that this nominalization recapitulates because both the attack on Afghanistan and the attack on the Americans are mentioned in the earlier discourse: the attack on Afghanistan is referred to in these two nominalizations: *last August's raid on Afghanistan* (in sentence 1) and *the Afghan strike* (in sentence 5); the attack on Americans is referred to in *attacking Americans* at the end of sentence 5.

Attack (in sentence 13) and *raid* (in sentence 1) share some common contexts: *his* and *lieutenants* contained in sentence 13 are a simple lexical repetition of *his* and *lieutenants*

respectively contained in sentence 1; *died* contained in sentence 13 is a complex paraphrase of *kill* contained in sentence 1. The common contexts of the two processes enable us to infer that *the attack* in sentence 13 refers to the attack on Afghanistan.

The above paragraphs discussed the situation when the answer to criterion 5 is negative but the answer to criterion 8 is positive. The situation when the answers to criterion 5 and criterion 8 are both negative is discussed below.

In such a situation, criterion 9, namely 'Is the Medium of the nominalised process recoverable from the context(s) of the processes that repeat the nominalised process?', is arrived at. The number of nominalizations occurring in this situation is very small in my data, one of which is *kidnappings* contained in sentence 18 of Example 3.14. The nominalised process (*kidnappings*) is not pre-modified by any demonstrative determiner, meaning that this nominalization does not recapitulate the kidnapping event mentioned earlier, namely three Britons and a New Zealand citizen were kidnapped in Chechnya. Furthermore, *kidnappings* has a plural marker *'-s'* at the end, indicating that it encompasses more than one kidnapping event. One of these events is indeed the kidnapping of three Britons and a New Zealand citizen in Chechnya, but others are not specified in the text. Therefore, the Medium of *kidnappings* that we are now discussing is not recoverable from the contexts of the processes repeating *kidnappings*. Thus, this nominalization does not satisfy criterion 9.

So far, I have explained criterion 1 to 11. The last two criteria will be explained below. Criterion 12, namely 'Does the combination of the cognate verb of the nominalised process and the Medium express a sensible meaning?', is arrived at either when the answer to criterion 10 is positive or when the answer to criterion 11 is positive.

Criterion 12 is explained with Example 3.16 given below, which contains the first three sentences in text 35. The two instances of *raid* contained in sentences 1 and 3 respectively and the items in sentence 2 repeating *raid* are emboldened.

Example 3.16:

(1) The objective of last August's **raid** on Afghanistan was to kill Osama bin Laden and as many of his lieutenants as possible, administration officials said yesterday.

(2) According to the officials, White House lawyers conducted a secret review in the months before the **attack** and concluded that **such operations** are legal under U.S. and international law.

(3) The officials said the **raid** was timed [...] when the Central Intelligence Agency believed he would be meeting with his chief operatives.

The two instances of *raid* and *attack* (as a noun) satisfy criterion 12 because the combinations of the cognate verbs of these nominalised processes and the Mediums: *raided Afghanistan*[5], and *attacked Afghanistan* express sensible meanings. On the other hand, *operations* contained in sentence 2 does not meet this criterion because the combination of its cognate verb and the Medium, namely *operated Afghanistan*. sounds odd. Neither would criterion 12 have been satisfied if *moves* or *acts* had been used

[5] After unpacking, the cognate verbs of these nominalised processes were written in past tense because the attack happened in the past.

instead of *operations* because the combinations of the cognate verbs of these two items and the Medium: *moved/acted Afghanistan* also do not express sensible meanings.

Criterion 13, namely 'Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the text?', deals with the situation in which a nominalization is embedded in and functions as (part of) the Medium of a recurrent nucleus.

This situation is illustrated with Example 3.17 given below, which contains sentences 1, 2, 6 and 11 from text 3. *Ban* (as a noun, emboldened) contained in sentence 2 is repeated by two instances of *bans* (as verbs, also emboldened) (one in sentence 6 and the other in sentence 11). *Using a cellular telephone* contained in sentence 1 is repeated by *cell phone use* contained in sentence 2, by *the use of wireless phones* and *cell phone use* contained in sentence 6, and by *cell phone uses* contained in sentence 11. These five are put into italics.

Example 3.17:

(1) Patricia Pena gave little thought to *using a cellular telephone* on the road before a distracted driver hit her car broadside, fatally injuring her 2-year-old daughter.

(2) Now she has devoted herself to pushing for a nationwide **ban** on *cell phone use* in cars, except in emergencies.

(6) No state **bans** *the use of wireless phones* in automobiles, and only California, Florida and Massachusetts have laws limiting *cell phone use* in moving vehicles, according to the National Conference of State Legislatures.

(11) Hilltown now **bans** certain *cell phone uses* for drivers.

Sentence 2 becomes as below after *a nationwide ban on cell phone use in cars* is unpacked. *A nationwide ban on cell phone use in cars* is unpacked to *legislation to ban*

cell phone use in cars nationwide (italicised). The remaining words in (original) sentence 2 are unchanged.

(2) Now she has devoted herself to pushing for *legislation to ban cell phone use in cars nationwide*, except in emergencies. (sentence 2 of text 3 after unpacking)

Ban cell phone use in sentence 2 after unpacking, and *bans the use of wireless phones* and *bans certain cell phone uses* contained in sentences 6 and 11 respectively together embody a recurrent nucleus (henceforth, this recurrent nucleus is referred to as *ban cell phone use*). In these three occurrences of this recurrent nucleus, *cell phone use, the use of wireless phones* and *certain cell phone uses* function as the Mediums respectively, expressing the same ideational meanings. These three Mediums satisfy criterion 13 and therefore are not suitable for unpacking.

Limiting cell phone use contained in sentence 6 is not an occurrence of the recurrent nucleus *ban cell phone use* discussed above because *limiting* is not in a repetition relation with *ban*: *limiting* means preventing something from becoming greater than a particular amount or degree (*Collins COBUILD English Dictionary for Advanced Learners*, third edition, 2001, p901), but not forbidding something entirely, whereas *ban* means prohibit or forbid entirely.

If a nominalization satisfying criterion 13 is in a repetition relation with a nucleus, this nominalization has not been added to the frequency of that nucleus because the nominalization was not suitable for unpacking. In Example 3.17 *cell phone use, the use*

of wireless phones and *certain cell phone uses* are all in a repetition relation with *using a cellular telephone* contained in sentence 1, but these three nominalizations were not added to the frequencies of *using a cellular telephone*.

The criteria for unpacking in a text nominalizations that contain nominalised processes derived from either verbal or mental processes

We now turn to Figure 3.11 (given on p117-118), which gives the criteria for unpacking in a text nominalizations that contain nominalised processes derived from either verbal or mental processes. The criteria in this figure are not identical to those in Figure 3.9 because when nominalizations that contain nominalised processes derived from verbal or mental processes are unpacked, we need to deal with issues ('projection' in particular) that do not arise when we unpack nominalizations that contain nominalised processes derived from neither verbal nor mental processes.

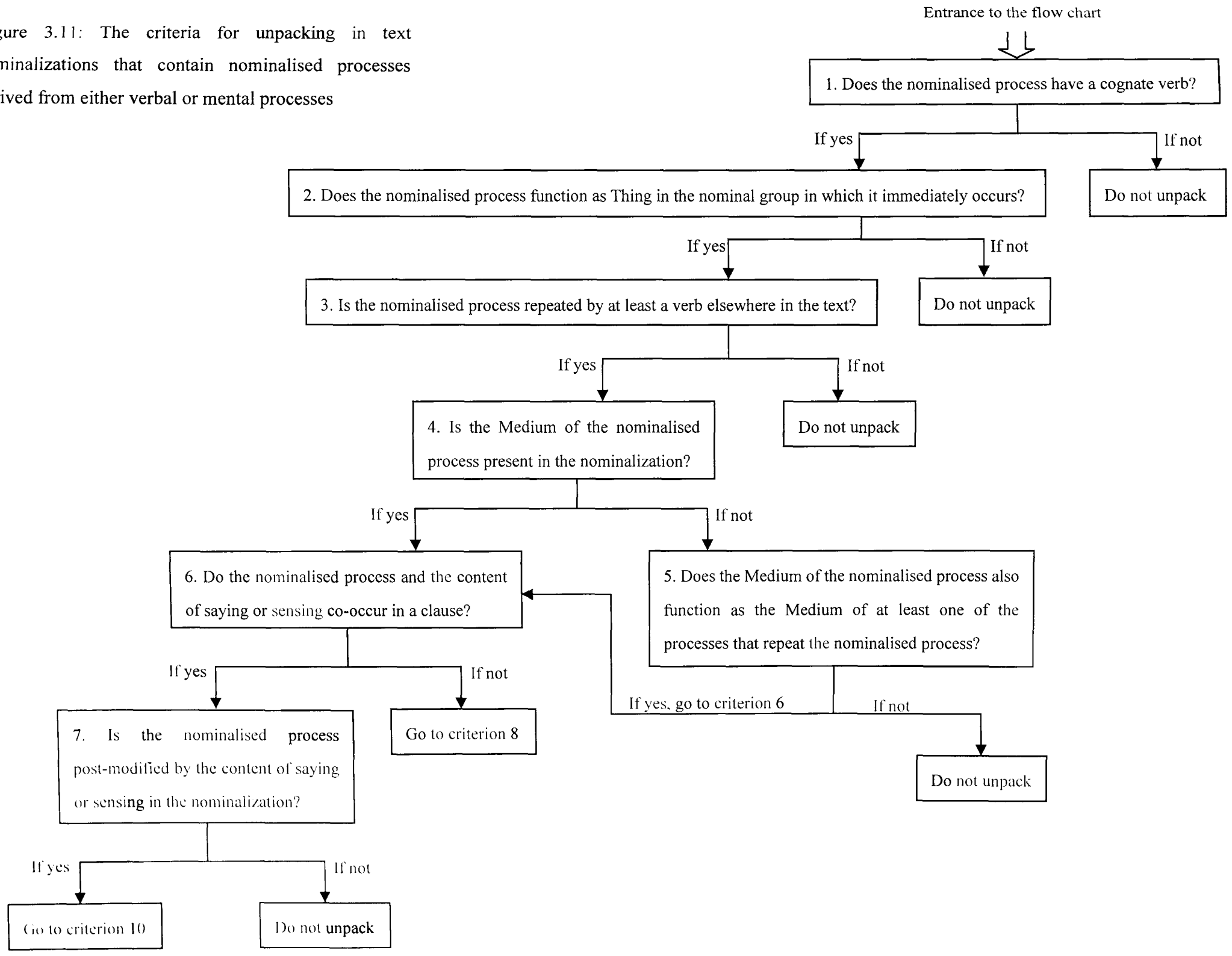
Criteria 1, 2, 3, and 10 in Figure 3.11 are also contained in Figure 3.9 and they were explained earlier. The following paragraphs explain criteria 4, 5, 6, 7, 8 and 9 in Figure 3.11.

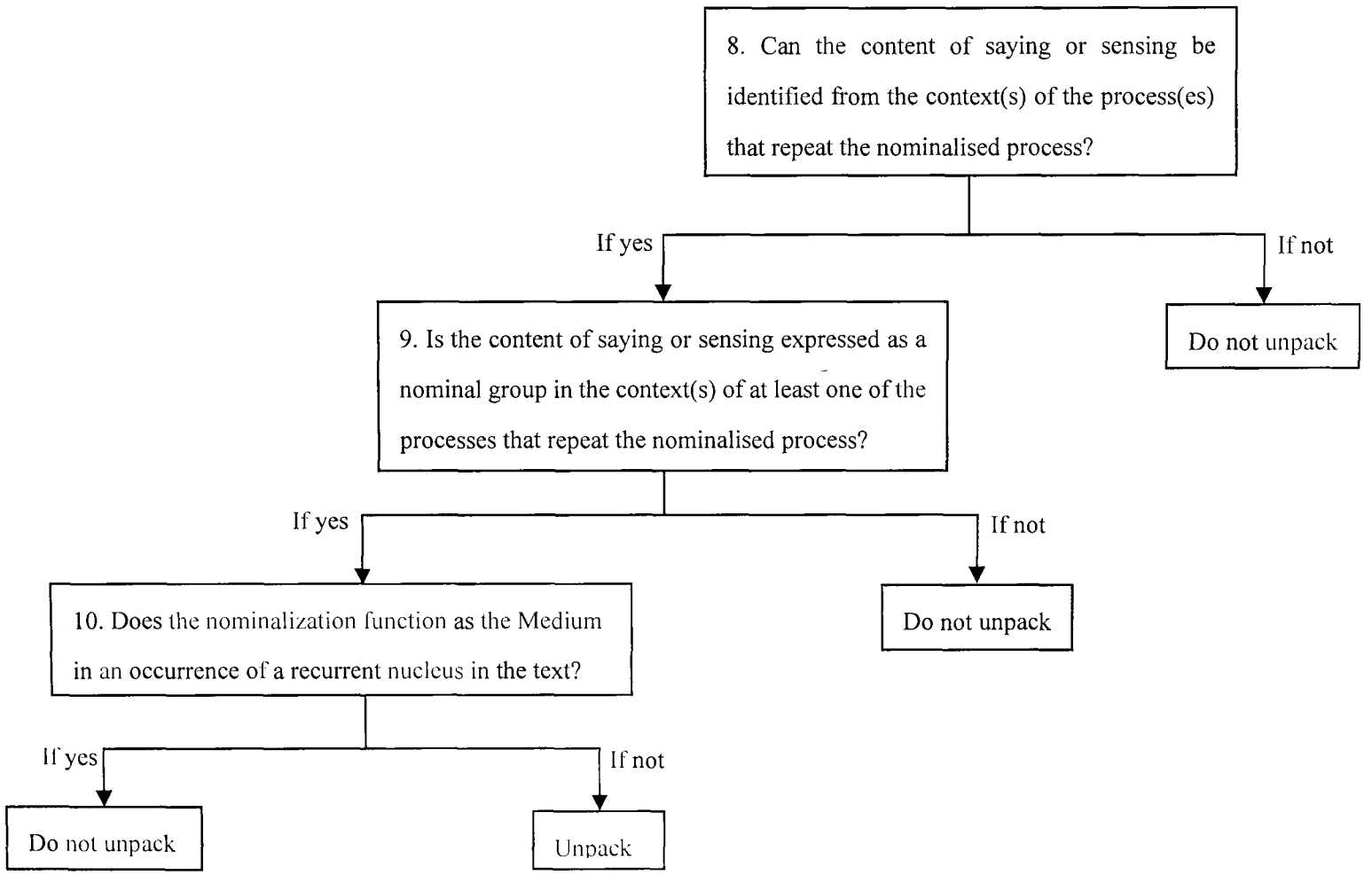
Criteria 4 and 5 are concerned with the Mediums of the nominalised processes derived from either verbal or mental processes. Halliday (1994: 165) defines the Mediums of these two types of process as follows:

The Medium is equivalent:

in mental process	to Senser
in verbal process	to Sayer (middle), Target (effective)

Figure 3.11: The criteria for unpacking in text nominalizations that contain nominalised processes derived from either verbal or mental processes




8. Can the content of saying or sensing be identified from the context(s) of the process(es) that repeat the nominalised process?
If yes
If not
Do not unpack
9. Is the content of saying or sensing expressed as a nominal group in the context(s) of at least one of the processes that repeat the nominalised process?
If yes
If not
Do not unpack
10. Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the text?
If yes
If not
Do not unpack
Unpack

PAGE

NUMBERING

AS ORIGINAL

The Mediums of mental processes are equivalent to a single participant role, namely the Sensers. The Medium of a verbal process is equivalent to the Sayer when the verbal clause is middle, and equivalent to the Target when the verbal clause is effective. Halliday (1994: 168) explains the voice system as follows: 'A clause with no feature of 'agency' is neither active nor passive but MIDDLE. One with agency is non-middle, or EFFECTIVE, in voice. An effective clause is then either active or passive: active if Agent/Subject, passive if Medium/Subject' (original capitalization).

Thompson (1994) classifies verbal processes accepting targets into two groups. The first group given below contains verbal processes meaning roughly 'say something bad about the target' (Thompson, 1994: 44):

abuse	charge	dismiss	reprove
accuse	condemn	disparage	revile
attack	criticize	insult	satirize
belittle	curse	lampoon	slam
bemoan	decry	libel	slander
besmirch	defame	malign	slate
bewail	denigrate	mock	upbraid
blame	denounce	pillory	vilify
castigate	deplore	rebuke	
censure	deprecate	reproach	

The second group given below contains verbal processes meaning roughly 'say something good about the target' (Thompson, 1994: 44):

acclaim	commend	endorse	flatter
applaud	compliment	eulogize	laud
bless	congratulate	extol	praise

The Mediums of verbal processes that accept Targets are not the Sayers but the Targets. Thus, in *the British "Sunday Telegraph" newspaper accused Egyptian police of crucifying and raping Copts,* the Medium of *accused* is not the Sayer (*the British "Sunday Telegraph" newspaper*) but the Target (*Egyptian police*) because *accused* is one of the verbal processes accepting Targets.

The Mediums of nominalised processes are identifiable when criterion 4 or 5 is met. When the answer to criterion 4 'Is the Medium of the nominalised process present in the nominalization?' is positive, the Medium is clearly identifiable because it modifies the nominalised process (as Thing, as required by criterion 2) in some way in the nominalization. For instance, *Ahmed's warning* (emboldened) contained in sentence 11 of text 32 satisfies criterion 4 because the Medium of *warning,* namely the Sayer *Ahmed*, is present in this nominalization:

(11) The State Department would not say whether her cables mentioned **Ahmed's warning**. (sentence 11 of text 32)

When the answer to criterion 4 is negative, we go to criterion 5, namely 'Does the Medium of the nominalised process also function as the Medium of at least one of the processes (including both nominalised processes and verbs) that repeat the nominalised process?'. When criterion 5 is satisfied, the Medium of the nominalised process is still identifiable because it is also the Medium of at least one of the processes repeating the nominalised process. This is illustrated with Example 3.18 given below, which contains sentences 1, 2, 8 and 11 from text 32. The four instances of *warning* contained in

Example 3.18 are emboldened. The nominalizations containing these four instances of *warning* are italicised.

Example 3.18:

NAIROBI, Kenya _ (1) Nine months before the attack on the American Embassy here, U.S. intelligence officials received *a detailed* ***warning*** *that Islamic radicals were plotting to blow up the building*, according to Kenyan and American officials.

(2) *The* ***warning*** forecast the Aug. 7 bombing in several particulars, the American officials said.

(8) Disclosure of *the* ***warning*** raises new questions about the State Department's protection of the embassies in Kenya and Tanzania, which did not meet the department's own minimum standards for security.

(11) The State Department would not say whether her cables mentioned *Ahmed's* ***warning***.

A detailed warning that Islamic radicals were plotting to blow up the building (in sentence 1) and two instances of *the warning* (one in sentence 2 and the other in sentence 8) satisfy criterion 5 because the Mediums of *warning* contained in these three nominalizations can be recovered from *Ahmed's warning* (in sentence 11).

When criterion 5 is satisfied, we go to criterion 6, namely 'Do the nominalised process and the content of saying or sensing co-occur in a clause?'. The contents of saying may be in the form of noun groups referred to as 'Verbiage' (Halliday, 1994: 141), or in the form of clauses (projected by the verbal processes) referred to as 'Locution' (Halliday, 1994: 255). Similarly, the contents of sensing may be in the form of noun groups referred to as 'Phenomenon' (Halliday, 1994:117), or in the form of clauses (projected by the mental processes) referred to as 'Idea' (Halliday, 1994: 255).

When the answer to the question that comprises criterion 6 is positive, we are dealing with intra-clausal encapsulations, i.e. nominalizations and the stretches of discourse they encapsulate co-occur in a single clause. Tang (2006) discusses this situation in detail. She refers to this situation as 'appositive packaging' (2006: 81) and identifies seven patterns in 'appositive packaging'. She refers to nominal groups that encapsulate as *'packet nouns'* (2006: 81). In the following examples taken from Tang (2006), the packet nouns are put into italics, and the stretches of discourse encapsulated by these packet nouns are the underlined *wh-/that-*/infinitive clause/gerundive[6] clause.

The seven patterns in 'appositive packaging' identified by Tang (2006) are:

i. packet noun + *BE that-*

The idea was <u>that all the major black political groups in South Africa would be represented</u>.

ii. packet noun + *BE wh-*/infinitive clause/gerundive clause

The temptation is <u>to hold on to what we have achieved and to refuse move forward</u>.

iii. *that-/wh-*/infinitive clause/gerundive clause + *BE* + packet noun

Warned panelist Dominique Moisi, associate director of the French Institute of International Relations in Paris: "<u>To think we can be an oasis of peace while the rest of Europe is a shambles</u> is *an illusion*."

iv. packet noun + *about* noun + *BE that-*

The point about Edna is <u>that she's an artist with tremendous technical understanding</u>.

[6] Henceforth, the term gerundive clause refers to 'ing' clause

v. **packet noun + *with* noun + *BE that-***

The trouble with TV these days is that there are not enough good comedy shows, things like Dad's Army and Fawlty Towers.

vi. **possessive + packet noun + *BE that-***

My complaint is that the bank does not pay interest on the interest that has been earned in previous years.

vii. **packet noun + *that-*/infinitive clause**

A characteristic formulation of this kind is *the statement* that "Weber showed the interdependence of ideas as social variable with other social variables".

In the first six patterns, the packet nouns and the *wh-/that-*/infinitive/gerundive clause they encapsulate occur on the two sides of copular 'be' in relational clauses. In pattern vii, on the other hand, the packet nouns and *that-*/infinitive clause co-occur in nominal groups in which the packet nouns are post-modified by the embedded *that-*/infinitive clause.

It does not seem appropriate to unpack nominalizations occurring in the first six patterns given above because doing so would involve replicating the discourse contained in the *wh-/that-*/infinitive clause/gerundive clause. The consequence would be that the nuclei contained in the *wh-/that-*/infinitive clause/gerundive clause would be counted twice after the nominalizations containing packet nouns were unpacked. When nominalizations occurring in these six patterns are interrogated by criterion 7, namely 'Is

the nominalised process post-modified by the content of saying or sensing in the nominalization?', the answers are negative, making them unsuitable for unpacking.

When nominalizations occurring in pattern vii are interrogated by criterion 7, the answer is positive. These nominalizations are referred to as embedded projections by Halliday (1994: 263). Halliday and Matthiessen (2004: 468-469) say that 'nouns that project belong to clearly defined classes, verbal process nouns (locutions) and mental process nouns (ideas); they correspond rather closely to, and in many instances are derived from, the verbs used in the projecting clause, especially the reporting ones'. Halliday and Matthiessen (2004: 468) consider nominal groups with embedded projections as nominalised expressions of their agnate clauses in which the cognate verbs of nominalised processes function as Processes. For instance, they consider this nominal group: *the/their assertion that Caesar was ambitious* as a nominalised expression of *they asserted that Caesar was ambitious*. Halliday and Matthiessen (2004: 468) make the following statements:

> 'the nominal group construction with an embedded projection clause is agnate with a clause nexus of projection: the nominal group is a metaphorical, nominalised version of the clause nexus; and the noun *assertion* serving as Head or Thing is in fact a nominalised variant of the verb *assert* serving as Process in the agnate clause. The congruent Sayer may be left out in the nominal group; or it may be represented either as the Deictic (*their assertion that...*) or as a Qualifier (*the assertion by the government that...*).'

Having discussed the situation when criterion 6 is met, we turn to the situation when criterion 6 is not met, where we are dealing with inter-clausal encapsulations, namely

nominalizations, and the stretches of discourse they encapsulate are contained in different clauses.

Francis (1994) studies this situation in detail. She uses the term 'labels' to refer to nominal groups that encapsulate and 'lexicalizations' to refer to the stretches of discourse that are encapsulated by these nominal groups. Labels may function either cataphorically or anaphorically. 'Where label precedes its lexicalization, it will be termed an advance label; where it follows its lexicalization, it will be called a retrospective label' (Francis, 1994: 83). Francis finds that retrospective labelling nouns are far more common than advance labelling nouns (Francis, 1994: 89). She classifies four major types of label: 'illocutionary' nouns, 'language activity' nouns, 'mental process' nouns and 'text' nouns. 'Illocutionary nouns' and 'mental process' nouns encompass a large number of nominalizations derived from verbal processes and mental processes respectively. These two types of nouns are therefore our major concern here.

Illocutionary nouns are 'nominalizations of verbal processes, usually acts of communication; they typically have cognate illocutionary verbs' (Francis, 1994: 23). Head nouns of this type found in *The Times* corpus are given below.

accusation, admission, advice, affront, allegation, announcement, answer, appeal, argument, assertion, charge, claim, comment, complaint, compliment, conclusion, contention, criticism, decision, (level of) denial, disclosure, excuse, explanation, indication, objection, observation, pledge, point, prediction, projection, proposal, proposition, protestation, reassurance, recognition, recommendation, rejection, remark, reminder, reply, report, request, response, revelation, statement, suggestion, warning.

Mental process nouns, on the other hand, refer to 'cognitive states and processes and the results thereof. They include nominalizations of mental-process verbs of the type that are used to project ideas, e.g. *think* and *believe,* but not all of them have cognate verbs' (Francis, 1994: 92). Head nouns of this type found in *The Times* corpus are given below.

analysis, assessment, assumption, attitude, belief, concept, conviction, doctrine, doubt, finding, hypothesis, idea, insight, interpretation, knowledge, misconception, notion, opinion, philosophy, position, principle, rationale, reading, suspicion, theory, thesis, thinking, thought, (point of) view, vision.

One important point made by Francis (1994) is that anaphoric encapsulation does not necessarily refer to a clearly delimited or identifiable stretch of discourse. She (1994: 88) writes:

> It is not always possible to decide where the initial boundary of a label's referent lies. This may be explained in terms of the intrinsic cohesive function of retrospective labels: they are used, like the anaphoric *this*, to tell the reader to section off in his or her mind what has gone before. The precise extent of the stretch to be sectioned off may not matter: it is the shift in direction signalled by the label and its immediate environment which is of crucial importance for the development of the discourse. It could even be argued that referential indistinctiveness of this kind may be used strategically by the writer to creative or persuasive effect, perhaps providing scope for different interpretations, or blurring the lines of specious or spurious arguments.

Francis's point is supported by my data because I found that the stretches of discourse encapsulated by nominalizations do not always have clearly delimited boundaries. When nominalizations that encapsulate stretches of discourse having no clearly delimited boundaries are interrogated by criterion 8, namely 'Can the content of saying or sensing

be identified from the context(s) of the processes that repeat the nominalised process?'. the criterion will not be met. In such a situation, it is not appropriate to unpack nominalizations because the contents of saying or sensing cannot be precisely recovered.

In Example 3.18 the two instances of *the warning* (one in sentence 2 and the other in sentence 8) and *Ahmed's warning* in sentence 11 all satisfy criterion 8 because the stretch of discourse encapsulated by these three nominalizations can be recovered from the embedded clause post-modifying *warning* contained in sentence 1, namely *that Islamic radicals were plotting to blow up the building*.

When criterion 8 is met, we go to criterion 9, namely 'Is the content of saying or sensing expressed as a nominal group in the context(s) of at least one of the process that repeat the nominalised process?'. The rationale of this criterion is given below. When nominalizations encapsulate, the encapsulated stretch of discourse often contain more than one nucleus. In Example 3.18 *that Islamic radicals were plotting to blow up the building* encapsulated by the two instances of *the warning* and *Ahmed's warning* contains two nuclei: *Islamic radicals were plotting* and *blow up the building*. Had these three nominalizations been unpacked, these two nuclei would have been duplicated three times.

With criterion 9 therefore I restrict the scope of the recovered contents of saying or sensing to nominal groups, making it inappropriate to unpack nominalizations when the

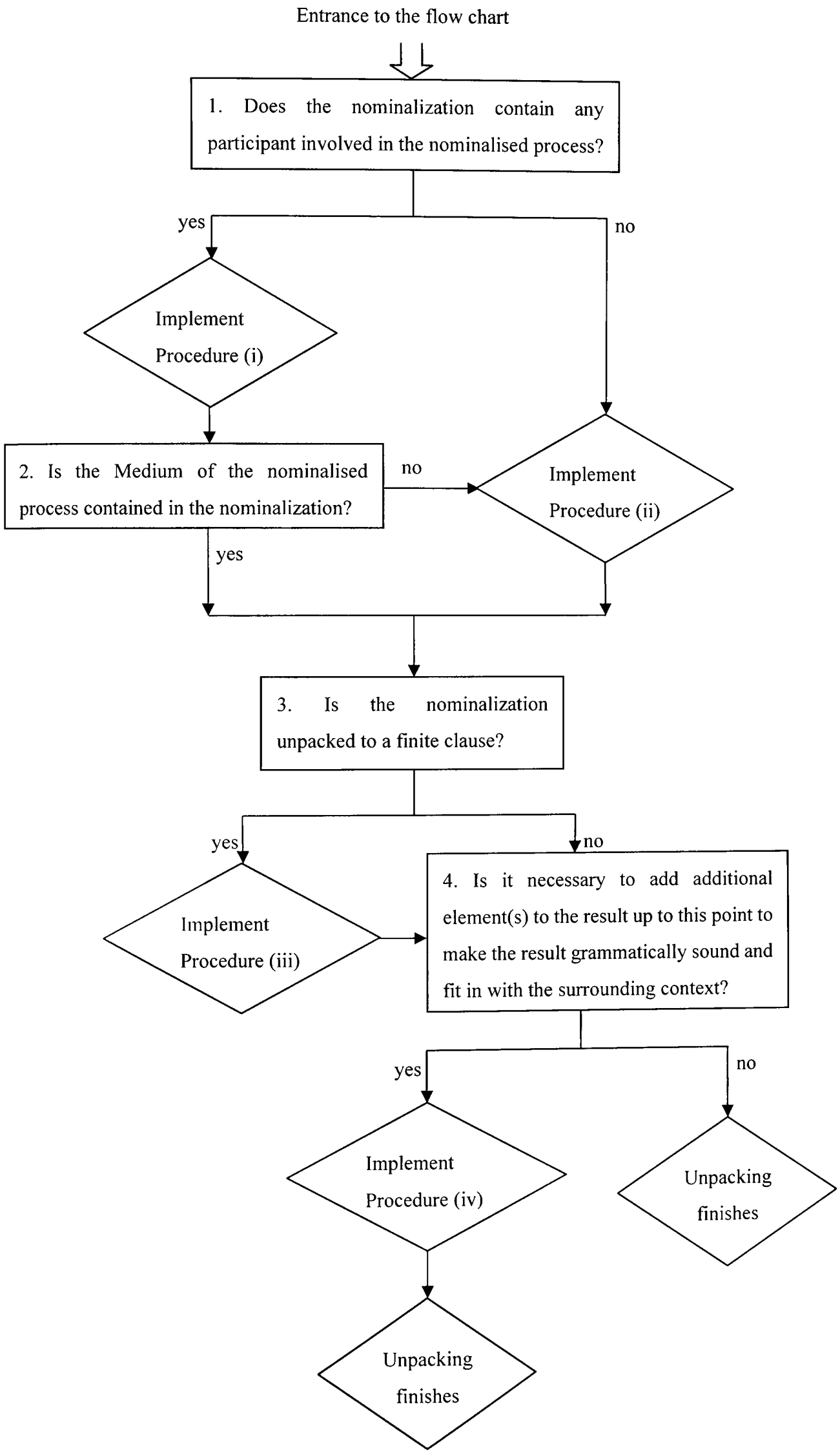
recovered contents are expressed in any unit larger than nominal group (i.e. a single clause, clause complex, paragraph, section, etc). Therefore, none of the three nominalizations described above satisfy criterion 9 because *Islamic radicals were plotting to blow up the building* is expressed in a clause complex.

Criterion 9 does not require that the content of saying or sensing be expressed as nominal groups in the context of every process repeating the nominalised process under consideration. It is sufficient to satisfy this criterion when the content of saying or sensing is expressed as a nominal group in the context of **one** of these processes.

How nominalizations were unpacked

Having considered the criteria for unpacking nominalizations in text, we turn to the last question, namely how nominalizations were unpacked. The steps of unpacking nominalizations are given in Figure 3.12 below.

Figure 3.12: The steps of unpacking nominalizations



Question 1 begins the flowchart, and it asks 'Does the nominalization contain any participant involved in the nominalised process?'. When the answer is positive. procedure (i), namely determining the transitivity relation(s) between the nominalised process and the participant(s), is implemented.

A variety of knowledge, including knowledge of the world and the happenings in the world that we remember, may be drawn on, and the context(s) of the verb(s) (if present) that repeat the nominalised process can be referred to for determining the transitivity relations between the nominalised processes and the participants contained in the nominalizations we analyze.

For instance, when we unpack the nominal group *the attack on the American Embassy here* (emboldened) contained in sentence 1 of text 32, if the analyst knew that the American Embassy in Nairobi was attacked in 1998 beforehand, he/she would bring his/her knowledge of this event to interpret the transitivity relation between *attack* (as a noun) and *the American Embassy here*.

Nine months before **the attack on the American Embassy here**, U.S. intelligence officials received a detailed warning that Islamic radicals were plotting to blow up the building, according to Kenyan and American officials. (sentence 1 of text 32)

We can also refer to the context(s) of the verb(s) that repeat *attack* contained in the nominalization *the attack on the American Embassy here*, such as the context of *blow up*. in which *the building*, which repeats *the American Embassy here*. is the Goal of *blow up*. Therefore, *the American Embassy here* should be interpreted as the Goal of *attack* (as a noun) so that sentence 1 reads coherently.

After implementing procedure (i), we consider question 2, which asks 'Is the Medium of the nominalised process contained in the nominalization?'. When *the attack on the American Embassy here* is interrogated by this question, the answer is positive because the Medium of *attack*, namely *the American Embassy here*, is contained in this nominalization.

When the answer to question 1 or 2 is negative, procedure (ii), namely recovering the Medium of the nominalised process, is implemented. In order to do so, we need to refer to the contexts of the processes (including both nominalised processes and verbs) that repeat the nominalised process. For instance, when *the attack* contained in sentence 11 of text 37 (given in Example 3.20 below) is unpacked, the Medium of *attack* (as a noun) needs to be recovered because it is not contained in the nominalization.

Example 3.20:

(1) The militant Palestinian movement Islamic Holy War said Saturday that it **attacked the Jerusalem market** on Friday, which prompted arrests by the Palestinian Authority overnight.

(11) The Palestinian cabinet, in a statement issued after a meeting on Friday night, accused "foreign forces" of standing behind **the attack.**

The definite article (*the*) pre-modifying *attack* indicates that *the attack* recapitulates an assault described in the discourse before sentence 11. In the text, no assault by Islamic Holy War other than its attack on the Jerusalem market has been mentioned from the beginning up to sentence 11, meaning that *the attack* recapitulates the event reported in sentence 1 that *the militant Palestinian movement Islamic Holy War attacked the Jerusalem market on Friday*. In this clause *the Jerusalem market* is the Goal (also the

Medium) of *attacked*. Therefore, after *the attack* is unpacked the Medium of the cognate verb of *attack* (as a noun) is *the Jerusalem market*.

After implementing procedure (ii), or when the answer to question 2 is positive, we consider question 3, namely 'Is the nominalization unpacked to a finite clause?'. If the answer to this question is positive, procedure (iii), namely recovering the tense of the cognate verb of the nominalised process, is implemented. In order to do so, we need to refer to the context(s) of the verb(s) that repeat the nominalised process. This is illustrated by Example 3.9 given earlier (reproduced below for convenience of discussion). When *the arrest of Gen. Augusto Pinochet* (italicised) contained in sentence 1 was unpacked to a finite clause, it was necessary to recover the tense of the cognate verb of *arrest* (as a noun, emboldened). In order to do so, we needed to refer to the context(s) of the verb(s) that repeat *arrest* (as a noun), such as the context of *arrested* in sentence 2, namely *Pinochet* [...] *was arrested*, in which the tense of *arrested* is past in present. Therefore, after *the arrest of Gen. Augusto Pinochet* was unpacked the tense of the cognate verb of *arrest* (as a noun) was also past in present: *the arrest of Gen. Augusto Pinochet* was unpacked to *Gen. Augusto Pinochet was arrested.*

Example 3.9:

(1) *The **arrest** of Gen. Augusto Pinochet* shows the growing significance of international human-rights law, suggesting that officials accused of atrocities have fewer places to hide these days, even if they are carrying diplomatic passports, legal scholars say.

(2) *Pinochet*, who ruled Chile as a military dictator from 1973 to 1990, was granted amnesty in his homeland but *was **arrested*** on Friday in London at the request of Spanish authorities, who want him extradited to Spain.

After implementing procedure (iii), or when the answer to question 3 is negative, we consider question 4, namely 'Is it necessary to add additional element(s) to the result up to this point to make the result grammatically sound and fit in with the surrounding context?'. If the answer to this question is positive, procedure (iv) is implemented to add additional elements. If the answer is negative, unpacking finishes.

What happens when the answer to question 4 is positive is discussed below. Heyvaert (2003) points out that 'some nominalizations [...] cannot be associated with a good clausal agnate unless one adds lexicogrammatical items that do not function in the nominalised structure'. In my analysis the items that are added to make the unpacked results grammatically sound and fit in with the surrounding contexts are frequently (a) participant roles other than the Mediums involved in the nominalised processes but not contained in nominalizations, and/ or (b) fact nouns of 'case' type (Halliday, 1994: 267), such as *fact*, *case*, and *accident*. In order to recover the participant roles other than the Mediums, it is necessary to refer to the context(s) of the process(es) (including both nominalised processes and verbs) that repeat the nominalised process. A fact noun of 'case' type is written as the head noun and post-modified by the clause after nominalization is unpacked so that the derived nominal group acts as an element in the ranking clause that contains the derived nominal group.

Procedure (iv) is illustrated below. *The arrest of Gen. Augusto Pinochet* was unpacked to *Gen. Augusto Pinochet was arrested* after procedure (iii) was implemented. *Gen. Augusto Pinochet was arrested* was grammatically sound but did not fit in with the

surrounding context. To achieve this, *the fact*, which is a fact noun of 'case' type, was added so that the nominal group *the fact that Gen. Augusto Pinochet was arrested* acted as the subject of the ranking clause that contains it, namely *The fact that Gen. Augusto Pinochet was arrested shows the growing significance of international human-rights law.*

To conclude, this section has discussed three important questions regarding the unpacking of nominalizations in text, namely why, when and how nominalizations are unpacked. The next section explains step A3 in Figure 3.1, namely identifying recurrent nuclei in text.

3.3.3 Identifying recurrent nuclei in text

Once we have unpacked any nominalizations, step A3 (of Figure 3.1) identifies recurrent nuclei in text.

The nucleus of a clause is the combination of the Process and the Medium, which is the participant 'through which the process is actualized' (Halliday, 1994: 163). The nucleus is the centre of gravity of a clause, around which the other elements are organized (Halliday and Matthiessen, 1999: 156): Agent represents the nucleus as being engendered by some external source; Circumstance qualifies the nucleus in terms of when, where, why or how it takes place.

The Medium is equivalent:

in material process	to Actor (middle), Goal (effective)
in behavioural process	to Behaver

in mental process	to Senser
in verbal process	to Sayer (middle), Target (effective)
in attributive process	to Carrier
in identifying process	to Identified
in existential process	to Existent

(reproduced from Halliday, 1994: 165)

In behavioural, mental, relational attributive, relational identifying and existential processes, the Medium is equivalent to a single participant role. On the other hand, in both material and verbal processes the Medium is equivalent to two participant roles, depending on whether the clause is middle or effective in voice (see p120).

The steps of identifying recurrent nuclei in text are given below. To begin with, the Mediums of the processes contained in the chain one starts with[7], excluding the nominalised process(es) contained in the nominalizations that are not suitable for unpacking, are identified. After that, any repeated Medium(s) are identified. Two Mediums are in a repetition relation when either criterion (a) or (b) given below is satisfied:

(a) They co-refer to the same 'thing'; or
(b) The 'things' they refer to belong to the same class

After the two steps described above, the recurrent nucleus/nuclei involving the processes contained in the chain under consideration and the repeated Medium(s) are identified. These two steps are then applied to the remaining chains in the text in order to identify the remaining recurrent nuclei in the text.

[7] It does not matter which chain one starts with.

The steps given above are illustrated by identifying the recurrent nuclei in text 4. I start with identifying any recurrent nuclei involving the processes contained in the chain denoting 'negative treatment' (because a majority of processes contained in this chain have been discussed in section 3.3.1). *The nation's* ***treatment*** *of those without housing* contained in sentence 4, *the* ***harassment*** *of homeless people* contained in sentence 6, *the* ***sweeps*** contained in sentence 17, and *the* ***harassment*** *of homeless people on the streets* contained in sentence 23 have already been unpacked. Sentences 4, 6, 17 and 23 given below embody the results after these four nominalizations are unpacked. On the other hand, *arrest* (as a noun) contained in sentence 14 was not suitable for unpacking because criterion 9 in Figure 3.9 was not satisfied: the Medium of *arrest* (as a noun), i.e. anyone who read the banner, cannot be recovered from the contexts of the processes repeating *arrest* (as a noun). Therefore, sentence 14 is unchanged. In the sentences given below the processes contained in the chain denoting 'negative treatment' are emboldened, and the Mediums affected by these processes (excluding *arrest* contained in sentence 14) are underlined.

(1) In Milwaukee, homeless people can be **arrested** for sleeping on heating grates.

(3) In the farmland city of Jeffersonville, Ind., a homeless mother, father and their infant died last month when their shelter was **firebombed**.

(4) When organizers of a national conference on homelessness wanted a city to illustrate how mean-spirited the nation has become to **treat** those without housing, they picked San Francisco.

(5) About 100 housing advocates from around the country joined more than 100 San Francisco activists at the United Nations Plaza on Saturday to protest what they call a growing problem of civil rights abuses **against** street people.

(6) They called for an end to aggressive police policies which, they said, allow the police to **harass** homeless people or simply **push** them from neighbourhood to neighbourhood.

(9) In January, San Francisco was one of five cities named as being especially tough on the homeless, relying on the police to **harass** street people rather than employing social service programs to find health care, jobs and homes for them, said the report by the National Law Center on Homelessness and Poverty in Washington.

(10) The center also criticized Atlanta, Chicago, New York and Tucson, Ariz., for **"criminalizing"** rather than treating homelessness.

(14) He stood in front of a row of shopping carts, decorated with banners reading, "Warning: push this cart and risk **arrest**."

(16) In San Francisco, advocates for the homeless have been angered by numerous police moves **clearing** street people from public parks and plazas, and a series of proposals aimed at **controlling** panhandling and the use of shopping carts by the homeless.

(17) Mayor Willie Brown has defended the move to **sweep** homeless people as an improvement of quality of life in the city.

(20) They are **sweeping** street people, **closing** their parks, **confiscating** shopping carts.

(23) "We want to stop the police from **harassing** homeless people on the streets" said Faith, who runs a homeless program in Columbus serving up to 15,000 homeless people each year.

Four instances of *homeless people* (in sentences 1, 6, 17 and 23 respectively); four instances of *street people* (in sentences 5, 9, 16 and 20 respectively); *those without housing* in sentence 4, and *them* in sentence 6 are in repetition relations with each other because they refer to the same class of people. For convenience, I refer to this repeated Medium as 'homeless people'.

The Medium of *criminalizing* in sentence 10, namely *homelessness*, is not in a repetition relation with 'homeless people' because *homelessness* refers to a social phenomenon, whereas 'homeless people' refers to a class of people.

The remaining Mediums are *their shelter* contained in sentence 3, *panhandling and the use of shopping carts by the homeless* in sentence 16, *their parks* and *shopping carts* in sentence 20, in which there is no repeated Medium. Therefore, 'homeless people' is the only repeated Medium affected by the processes in the chain denoting 'negative treatment'.

Thus, we identify the recurrent nucleus *mistreat homeless people*, which occurs ten times in text 4. The occurrences of this recurrent nucleus (emboldened) are given below.

mistreat homeless people (10)

(1) In Milwaukee, **homeless people** can be **arrested** for sleeping on heating grates.

(4) When organizers of a national conference on homelessness wanted a city to illustrate how mean-spiritedly the nation is **treating those without housing**, they picked San Francisco.

(5) About 100 housing advocates from around the country joined more than 100 San Francisco activists at the United Nations Plaza on Saturday to protest what they call a growing problem of civil rights abuses **against street people**.

(6) They called for an end to aggressive police policies which, they said, allow the police to **harass homeless people** or simply **push them** from neighbourhood to neighbourhood.

(9) In January, San Francisco was one of five cities named as being especially tough on the homeless, relying on the police to **harass street people** rather than employing social service programs to find health care, jobs and homes for them, said the report by the National Law Center on Homelessness and Poverty in Washington.

(16) In San Francisco, advocates for the homeless have been angered by numerous police moves **clearing street people** from public parks and plazas, and a series of proposals aimed at controlling panhandling and the use of shopping carts by the homeless.

(17) Mayor Willie Brown has defended the move to **sweep homeless people** as an improvement of quality of life in the city.

(20) They are **sweeping street people**, closing their parks, confiscating shopping carts.

(23) "We want to stop the police from **harassing homeless people** on the streets" said Faith, who runs a homeless program in Columbus serving up to 15,000 homeless people each year.

The application of the two steps described above to the remaining chains in text 4 leads us to identify six more recurrent nuclei (given below) in this text. The headings represent recurrent nuclei. The numbers in brackets next to the headings show the frequencies of the recurrent nuclei. The sentences containing the occurrences of these recurrent nuclei (emboldened) are given under the headings.

housing advocates said (6)

(5) About 100 housing advocates from around the country joined more than 100 San Francisco activists at the United Nations Plaza on Saturday to (Ø **about 100 housing advocates** and more than 100 San Francisco activists) **protest** what **they**[8] **call** a growing problem of civil rights abuses against street people.

(6) **They called for** an end to aggressive police policies which, **they said**, allow the harassment of homeless people or simply push them from neighborhood to neighborhood.

[8] *They* in sentence 5 and the two instances of *they* in sentence 6 refer to both 'about 100 housing advocates' and 'more than 100 San Francisco activists'. Therefore, *they call*, *they called for*, and *they said* are classified as occurrences of both recurrent nuclei: *housing advocates said* and *San Francisco activists said*.

(13) **Max Biddel**[9], a homeless advocate from Sacramento, **told** the crowd of 200 protesters gathered at United Nations Plaza.

(15) "All people have the right to exist in this country, even if it means sleeping in tents or sleeping bags," he (**Max Biddel) told** the crowd.

San Francisco activists said (4)

(5) About 100 housing advocates from around the country joined more than 100 San Francisco activists at the United Nations Plaza on Saturday to (Ø about 100 housing advocates and **more than 100 San Francisco activists) protest** what **they call** a growing problem of civil rights abuses against street people.

(6) **They called for** an end to aggressive police policies which, **they said**, allow the harassment of homeless people or simply push them from neighborhood to neighborhood.

Bill Faith said (4)

(7) "There is a growing intolerance of *the homeless* around the nation," **said Bill Faith**, board president of the National Coalition for the Homeless, which held its annual meeting in San Francisco last weekend.

(8) "Part of the reason the coalition wanted to come to San Francisco is because we have heard for years about the civil rights abuses that go on here are more egregious than most cities' in the country," **said Faith**, who is from Columbus, Ohio.

(23) "We want to stop the harassment of homeless people on the streets," **said Faith**, who runs a homeless program in Columbus serving up to 15,000 homeless people each year.

(26) People want real options for improving their lives," he **(Bill Faith) said**.

---

[9] Both *'Max Biddel'* in sentence 13 and *he* (referring to *'Max Biddel'*) in sentence 15 repeats *'housing advocates'* because *'Max Biddel'* is said to be *'a homeless advocate'* in sentence 13. *'Max Biddel'* is, therefore, a specific exemplification of this group of people.

treat homelessness (3)

(10) The center also criticized Atlanta, Chicago, New York and Tucson, Ariz., for "criminalizing" rather than **treating homelessness**.

(13) "**Homelessness is being treated** like ... an animal-control problem," Max Biddel, a homeless advocate from Sacramento, told the crowd of 200 protesters gathered at United Nations Plaza.

(22) Coalition members called for a federal effort to **deal with homelessness** by creating jobs, building affordable housing and making health care accessible.

served homeless people (2)

(18) Homeless people are facing similar problems around the nation _ from New York City with its massive homeless problem, to Jeffersonville, a town of 29,000 in which aid agencies **served 1,300 homeless people** last year.

(23) "We want to stop the harassment of homeless people on the streets," said Faith, who runs a homeless program in Columbus **serving** up to **15,000 homeless people** each year.

homeless people died (2)

(2) In Atlanta, more than 60 **homeless people died** on the streets last year.

(3) In the farmland city of Jeffersonville, Ind., **a homeless mother, father and their infant**[10] **died** last month when their shelter was firebombed.

Housing advocates said, *San Francisco activists said* and *Bill Faith said* are the three verbal recurrent nuclei in text 4. The verbal processes contained in the occurrences of these three recurrent nuclei, namely *protest* and *call* in sentence 5, *called for* in sentence

[10] 'A homeless mother, father and their infant' is an exemplification of 'homeless people'.

6, two instances of *told* (one in sentence 13 and the other in sentence 15), and five instances of *said*, belong to a single chain concerning verbal processes of saying. Also included in this chain are *said* in sentence 9, *criticized* in sentence 10, and *called for* in sentence 22.

In sentence 9 the Medium of *said* is the Sayer, namely *the report by the National Law Center on Homelessness and Poverty in Washington*. In sentence 10 the Medium of *criticized* is the Target, namely *Atlanta, Chicago, New York and Tucson, Ariz* (because *criticised* is one of the verbal processes accepting Targets). In sentence 22 the Medium of *called for* is *coalition members*. None of these three Mediums is in a repetition relation with the Medium involved in the three verbal recurrent nuclei noted above.

3.3.4 Ranking the recurrent nuclei in text in descending order of frequency

Once we have identified the recurrent nuclei in text, step A4 (Figure 3.1) ranks them in descending order of frequency.

In text 4 the most frequent nucleus is 'mistreat homeless people'. The least frequent recurrent nuclei are 'served homeless people' and 'homeless people died', both occurring twice. Table 3.2 below ranks the frequencies of the recurrent nuclei in this text in descending order.

The recurrent nuclei in text 4	**Frequencies**
mistreat homeless people	10

housing advocates said	6
San Francisco activists said	4
Bill Faith said	4
treat homelessness	3
served homeless people	2
homeless people died	2

Table 3.1: The ranking of the recurrent nuclei contained in text 4 in descending order of frequency

So far, steps A1, A2, A3 and A4 in Figure 3.1 have been explained. After these four steps, the recurrent nuclei in a text have been identified and ranked in descending order of frequency. We now deal with steps B1, B2, B3, which analyze summaries in terms of recurrent nuclei.

3.3.5 Classifying processes in summaries into chains according to the commonality of meanings they express

Step B1 is analogous to A1, but in B1 we classify the processes in summaries into chains according to the commonality of meaning they express. As with step A1, nouns in summaries that are derived from type 3 nominalising (shown in Figure 3.2) are put into chains in step B1 when they are repeated by other processes contained in the summaries, while nouns derived from the other types of nominalising are not put into any chain. This is illustrated with reference to the summaries of text 4 given below (alphabetised for convenience of reference).

a) San Francisco worst of many offenders **against** homeless.

b) National Law Center on Homelessness and Poverty alleges **mistreatment** of street people.

c) Advocacy coalition criticizes national **mistreament** of homeless.

Offenders contained in summary (a) is derived from type 1 nominalising, denoting an entity rather than a process. Therefore, it was not put into any chain (and consequently was not unpacked). On the other hand, the two instances of *mistreatment* (emboldened) (in summaries b and c) are derived from type 3 nominalising, denoting processes.

A method is needed for determining whether process A in a summary is in a repetition relation with process B contained in another summary. My method makes use of Hoey's 'link triangle' (Hoey, 1991) given in Figure 3.13 below (reproduced from Hoey, 1991: 65).

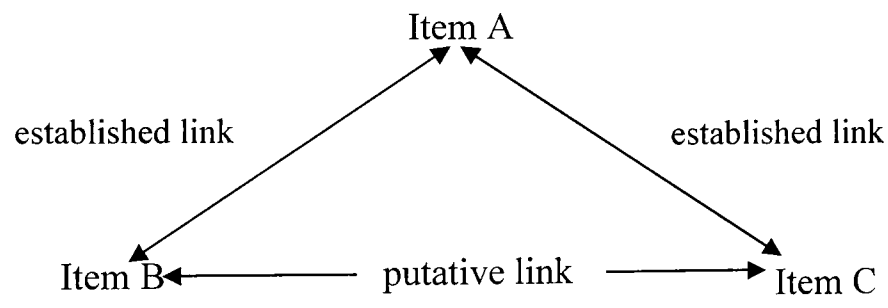


Figure 3.13: Link triangle

What this diagram says is that 'wherever links have been established between one item and any two other items in a text (by whatever means), there is also a link between the two other items' (Hoey, 1991: 65).

In Hoey (1991) 'link triangle' was mainly used to account for complex paraphrase relation between two items in a text. 'Link triangle' is extended here to determine

whether process X in a summary is in a repetition relation with process Y contained in another summary. Process X will be in a repetition relation with process Y when either condition a or b given below is satisfied.

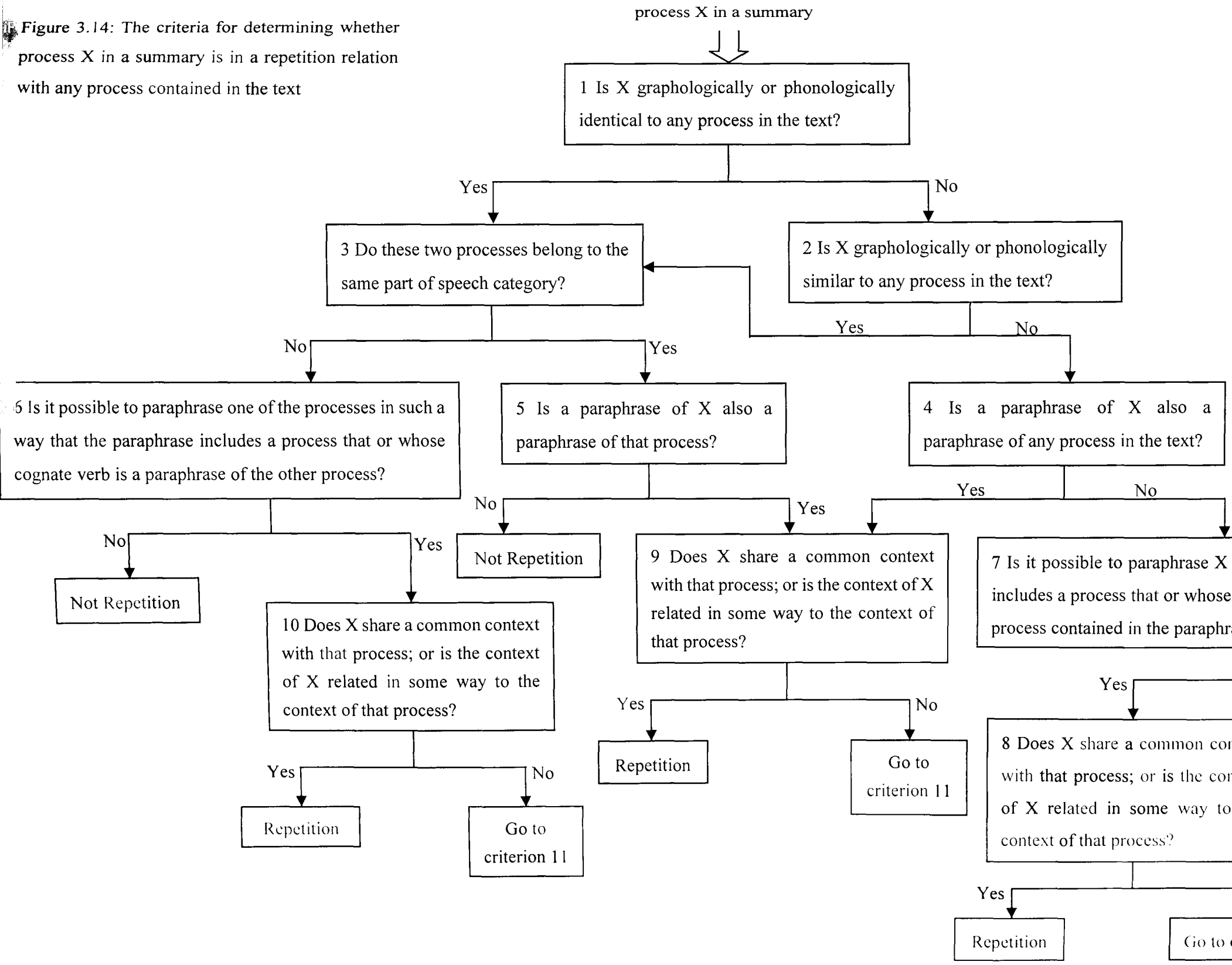
a. Process X is in a repetition relation with process Z contained in the text, and process Y is also in a repetition relation with process Z.
b. Process X is in a repetition relation with process W contained in the text, process Y is in a repetition relation with process Z also contained in the text, and W is in a repetition relation with Z.

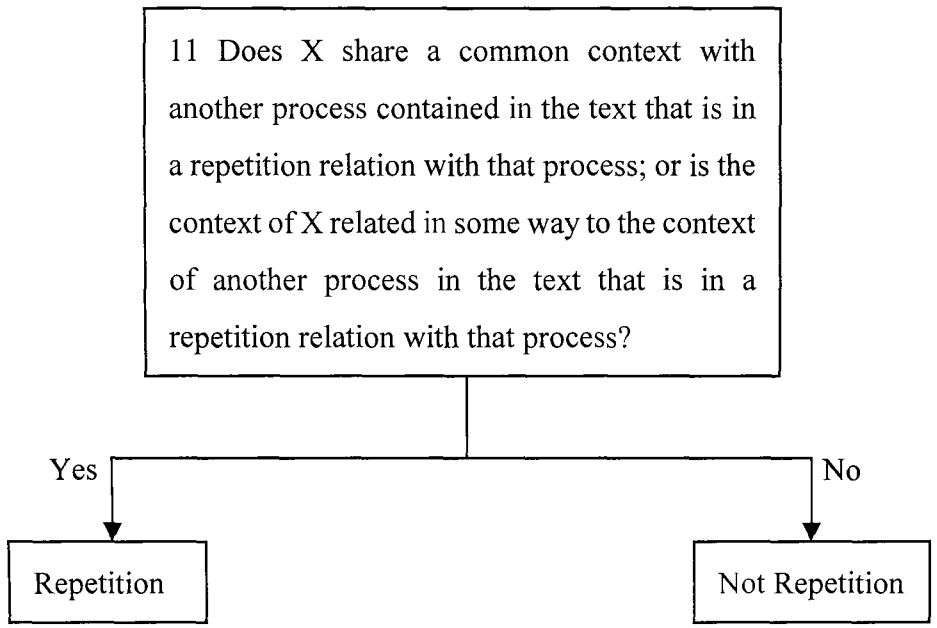
The criteria for determining whether a process in a summary is in a repetition relation with any process contained in the text are given in Figure 3.14 (on p147-148).

Let us consider whether *against* contained in summary (a) is in a repetition relation with *mistreatment* contained in summary (b). In order to do so, both items need to be examined by the criteria in Figure 3.14.

Against in summary (a) is considered first. When it is examined by criterion 1, namely 'Is X graphologically or phonologically identical to any process in the text?', the answer is positive because *against* (in summary (a)) is graphologically identical to *against* contained in sentence 5 of text 4. Therefore, we go to criterion 3, namely 'Do these two processes belong to the same part of speech category?'. The answer to this criterion is positive because both instances of *against* are prepositions. Therefore, we go to criterion 5, namely 'Is a paraphrase of X also a paraphrase of that process?'. Prepositions can be considered as minor processes (Halliday, 1994: 212) because their meanings are similar to those expressed by some processes, e.g. near/adjoining (the house), without/not wearing (a hat), about/concerning (the trial) (examples are taken from Halliday, 1994).

Figure 3.14: The criteria for determining whether process X in a summary is in a repetition relation with any process contained in the text




11 Does X share a common context with another process contained in the text that is in a repetition relation with that process; or is the context of X related in some way to the context of another process in the text that is in a repetition relation with that process?
Yes
No
Repetition
Not Repetition

Therfore, if *against* in summary (a) is replaced by *mistreating*, the thrust of this summary is not altered, albeit *against homeless* is less specific than *mistreating homeless.* It seems therefore appropriate to accept *mistreating* as a paraphrase of *against* in summary (a). Similarly, we may also accept *mistreating* as a paraphrase of *against* contained in sentence 5 of text 4. Once criterion 5 is satisied, we go to criterion 9, which asks 'Does X share a common context with that process; or is the context of X related in some way to the context of that process?'. Criterion 9 will be satisfied if the answer to any of the three questions given below is positive:

a. *Do they co-occur with lexical items denoting the same referent?*
b. *Do they co-occur with lexical items denoting referents belonging to the same class of thing?*
c. *Is there whole or partial parallelism between their contexts?*

For the pair we are considering, the answer to question (b) is positive because *homeless* contained in the context of *against* in summary (a) refers to the same class of people as that referred to by *street people* in the context of *against* in sentence 5. Therefore, *against* in summary (a) is in a repetition relation with *against* contained in sentence 5 of text 4.

We now consider the instance of *mistreatment* in summary (b). When it is judged by criterion 1, the answer is negative because it is not graphologically or phonologically identical to any process contained in text 4. Therefore, we go to criterion 2, namely 'Is X graphologically or phonologically similar to any process in the text?'. The answer is positive because it is graphologically similar to four processes in text 4: *treatment* in sentence 4, *treating* in sentence 10, *treatment* in sentence 12, and *treated* in sentence 13.

The first pair, namely *mistreatment* in summary (b) and *treatment* in sentence 4, is considered below. The satisfaction of criterion 2 earlier leads us to criterion 3, namely 'Do these two processes belong to the same part of speech category?'. The answer to this criterion is positive because both *mistreatment* and *treatment* are nouns. Therefore, we go to criterion 5, which asks 'Is a paraphrase of X also a paraphrase of that process?'. The answer to this criterion is also positive because a paraphrase of *mistreatment of* in summary (b) is *negative behaviour towards*, which is also a near-paraphrase of *"mean-spirited"...treatment of* contained in sentence 4. Therefore, we go to criterion 9, namely 'Does X share a common context with that process; or is the context of X related in some way to the context of that process?'. This criterion is also satisfied because the answer to question (b) is positive: *street people* in the context of *mistreatment* in summary (b) refers to the same class of people as that referred to by *those without housing* in the context of *treatment* in sentence 4. Therefore, *mistreatment* in summary (b) is in a repetition relation with *treatment* contained in sentence 4.

Now we have seen that *against* in summary (a) is in a repetition relation with *against* contained in sentence 5, and that *mistreatment* in summary (b) is in a repetition relation with *treatment* contained in sentence 4. The repetition between *against* in sentence 5 and *treatment* in sentence 4[11] establishes a repetition relation between *against* in summary (a) and *mistreatment* contained in summary (b) because condition (b) given earlier on page 146 is met.

[11] Both *against* in sentence 5 and *treatment* in sentence 4 belong to the chain denoting *'negative treatment'*.

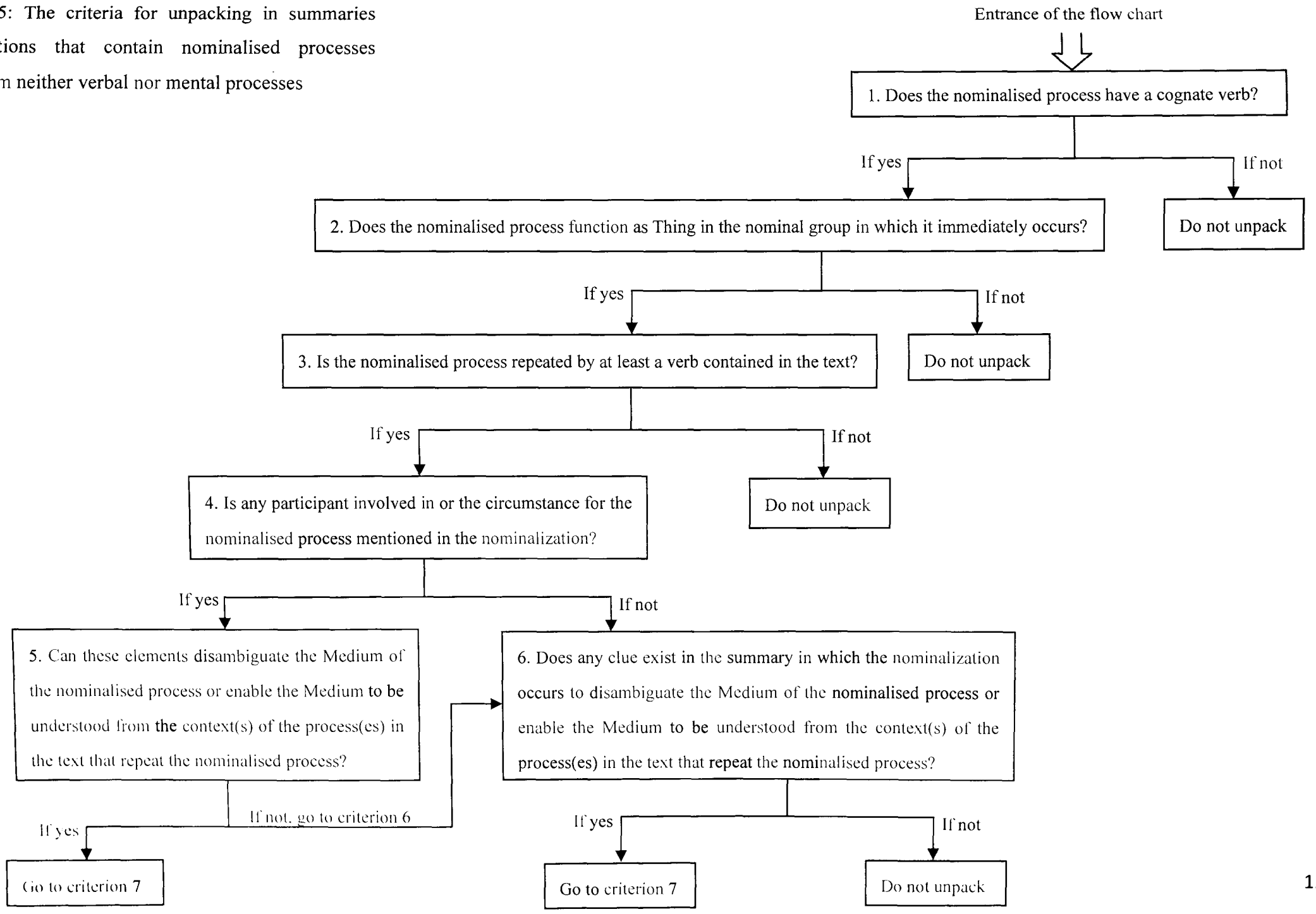
When the instance of *mistreatment* contained in summary (c) is examined in terms of the criteria in Figure 3.14, we find that it is also in a repetition relation with *treatment* contained in sentence 4. Therefore, *against* in summary (a), *mistreatment* contained in summaries (b) and (c) are in a repetition relation with each other.

The summaries of text 4 contain another process chain, which concerns processes of saying. Members of this chain are *alleges* in summary (b) and *criticizes* in summary (c).

3.3.6 Unpacking in summaries nominalizations that contain nominalised processes included in chains

We now turn to step B2 (of Figure 3.1), which unpacks in summaries nominalizations containing nominalised processes that have been included in chains. We also need criteria to determine when nominalizations in summaries are suitable for unpacking. Figure 3.15 (given on p152-153) gives the criteria for unpacking in summaries nominalizations that contain nominalised processes derived from neither verbal nor mental processes, while Figure 3.16 (given on p154-155) gives the criteria for unpacking in summaries nominalizations that contain nominalised processes derived from either verbal or mental processes. The criteria in Figure 3.15 are similar to those in Figure 3.9. and the criteria in Figure 3.16 are similar to those in Figure 3.11 so that nominalizations contained in summaries were unpacked on equal terms to those contained in the original text.

Figure 3.15: The criteria for unpacking in summaries nominalizations that contain nominalised processes derived from neither verbal nor mental processes



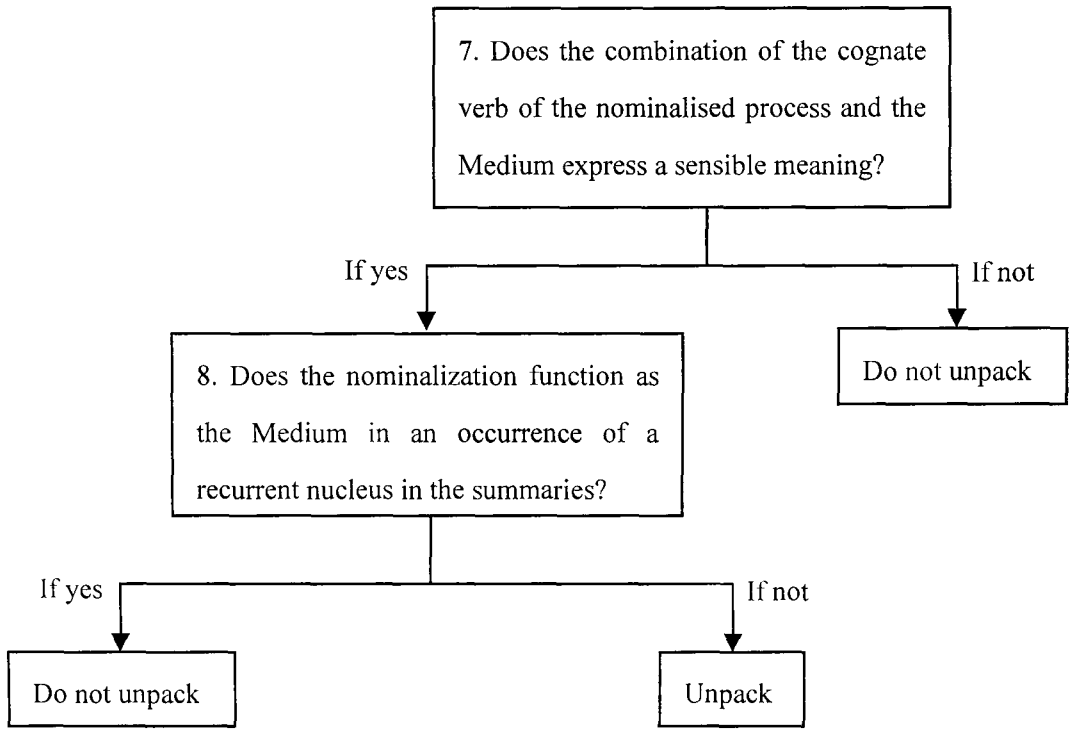
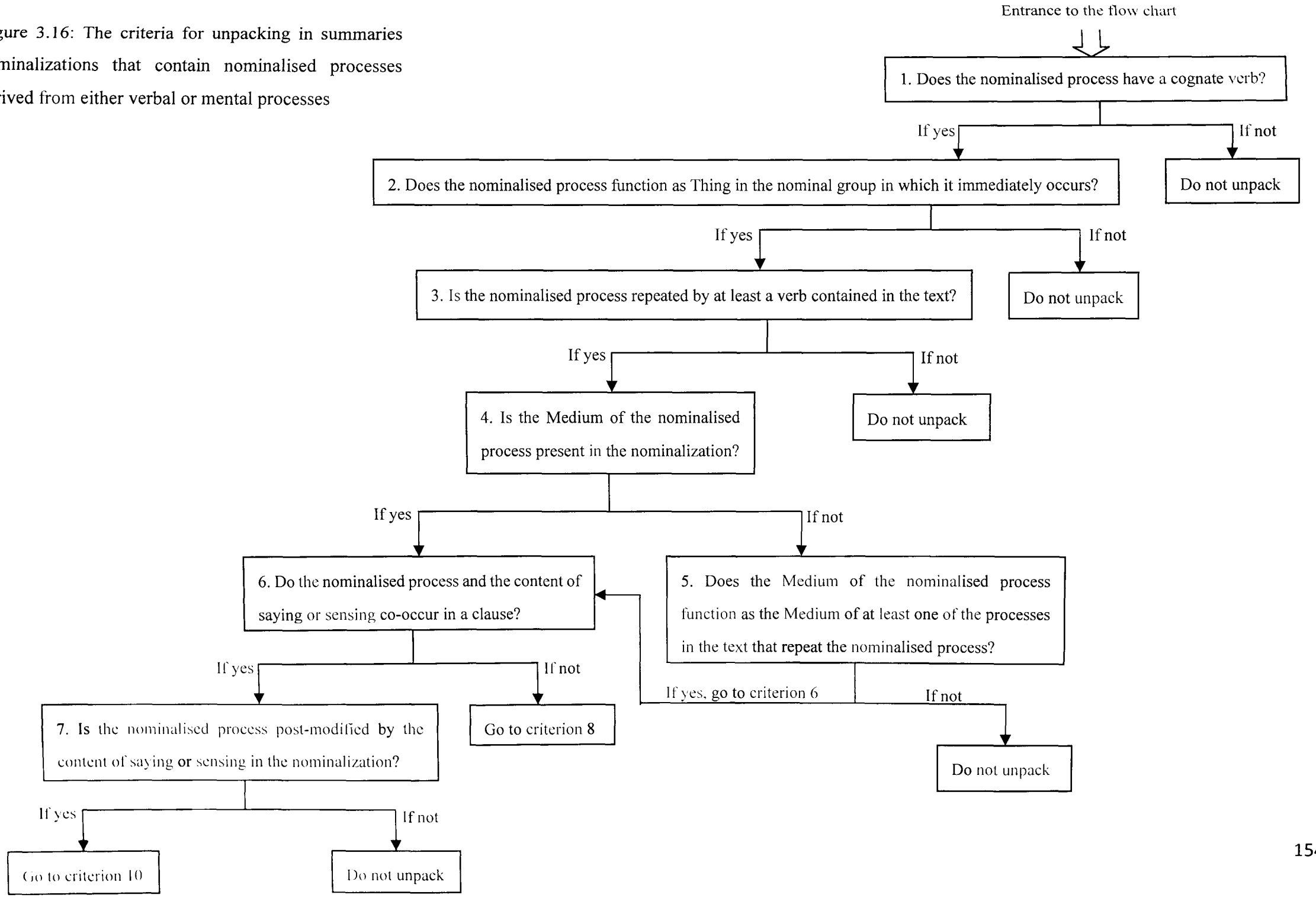
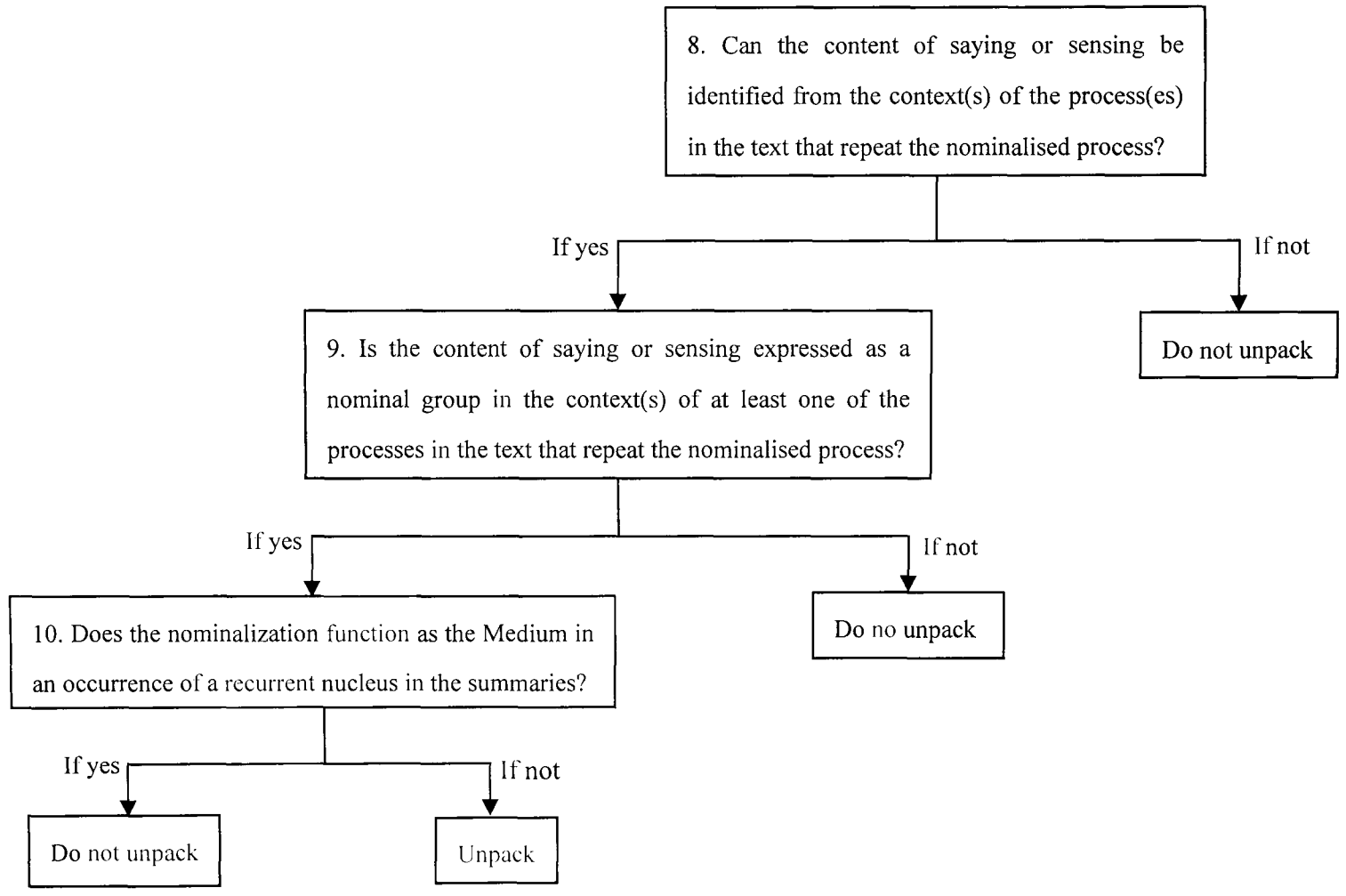

7. Does the combination of the cognate verb of the nominalised process and the Medium express a sensible meaning?
If yes
If not
8. Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the summaries?
Do not unpack
If yes
If not
Do not unpack
Unpack

Figure 3.16: The criteria for unpacking in summaries nominalizations that contain nominalised processes derived from either verbal or mental processes




8. Can the content of saying or sensing be identified from the context(s) of the process(es) in the text that repeat the nominalised process?
If yes
If not
Do not unpack
9. Is the content of saying or sensing expressed as a nominal group in the context(s) of at least one of the processes in the text that repeat the nominalised process?
If yes
If not
Do no unpack
10. Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the summaries?
If yes
If not
Do not unpack
Unpack

Both *mistreatment of street people* contained in summary (b) and *national mistreatment of homeless* contained in summary (c) are examined in terms of the criteria in Figure 3.15 so as to determine whether they are suitable for unpacking.

Mistreatment of street people is considered first. Criterion 1 in Figure 3.15, namely 'Does the nominalised process have a cognate verb?', is satisfied because *mistreatment* has a cognate verb (*mistreat*). Therefore, we go to criterion 2, namely 'Does the nominalised process function as Thing in the nominal group in which it immediately occurs?'. This criterion is also satisfied because *mistreatment* functions as Thing in *mistreatment of street people*, leading us to criterion 3, which asks 'Is the nominalised process repeated by at least a verb contained in the text?'. The answer to this criterion is positive because *mistreatment* is repeated by the verbs contained in the chain denoting 'negative treatment' in the text, such as *firebombed* in sentence 3, *push* in sentence 6 and *harass* in sentence 9, etc. After that, we go to criterion 4, namely 'Is any participant involved in, or the circumstance for, the nominalised process present in the nominalization?'. The answer to this criterion is also positive because the post-modifier *street people* refers to a participant involved in the process *mistreatment*. After that, we go to criterion 5, namely 'Can these elements disambiguate the Medium of the nominalised process or enable the Medium to be understood from the context(s) of the process(es) in the text that repeat the nominalised process?'. The transitivity relation between *mistreatment* and *street people* can be determined by referring to the contexts of the verbs in the text that repeat *mistreatment*, such as the contexts of *arrested* in sentence 1, *push* in sentence 6, *harass* in sentence 9, *clearing* in sentence 16, and

sweeping in sentence 20, in which *homeless people* or its equivalent expressions (or exemplification) function as the Goals of these verbs. Therefore, *street people* should be interpreted as the Goal (i.e. the Medium) of *mistreatment* when we unpack *mistreatment of street people*. Once criterion 5 is satisfied, we go to criterion 7, namely 'Does the combination of the cognate verb of the nominalised process and the Medium express a sensible meaning?'. This criterion is also satisfied because the combination of the cognate verb of *mistreatment* and the Medium, namely *mistreat street people*, expresses a sensible meaning. Finally, the answer to criterion 8, namely 'Does the nominalization function as the Medium in an occurrence of a recurrent nucleus in the summaries' is negative. Therefore, *mistreatment of street people* in summary (b) is suitable for unpacking.

When *national mistreatment of homeless* in summary (c) is examined in terms of the criteria in Figure 3.15, we find that it is also suitable for unapcking.

How nominalizations in summaries are unpacked

The steps of unpacking nominalizations in summaries are the same as those of unpacking nominalizations in text, as given earlier in Figure 3.12. The procedures in this Figure are given in the first column of Table 3.2 below. The second column of the table gives the contexts that need to be referred to when the procedures are implemented to unpack nominalizations in summaries.

The procedures that are implemented	The contexts that need to be referred to
When procedure (i) is implemented to determine the transitivity relation(s) between the nominalised process and the participant(s) in the nominalization	Refer to the context(s) of the verb(s) in the text that repeat the nominalised process
When procedure (ii) is implemented to recover the Medium of the nominalised process	Refer to the context(s) of the process(es) (including both nominalised process(es) and verb(s)) in the text that repeat the nominalised process
When procedure (iii) is implemented to recover the tense of the cognate verb of the nominalised process	Refer to the context(s) of the verb(s) in the text that repeat the nominalised process
When procedure (iv) is implemented to add the participant role(s) that are involved in the nominalised process but are not contained in the nominalization	Refer to the context(s) of the process(es) (including both nominalised process(es) and verb(s)) in the text that repeat the nominalised process

Table 3.2: The contexts that need to be referred to when the procedures in Figure 3.12 are implemented to unpack nominalizations in summaries

3.3.7 Identifying recurrent nuclei in summaries

Once we have unpacked nominalizations in summaries, we arrive at step B3 (of Figure 3.1) to identify recurrent nuclei in summaries.

The procedures to identify recurrent nuclei in summaries are the same as those for identifying recurrent nuclei in text. Step B3 is illustrated below by identifying any recurrent nucleus in the summaries of text 4. After *mistreatment of street people* and *national mistreatment of homeless* are unpacked, summaries (b) and (c) are rewritten as follows. Summary (a) is not changed.

summary (a): San Francisco worst of many offenders **against** *homeless*.

summary (b) after *mistreatment of street people* is unpacked: National Law Center on Homelessness and Poverty alleges that *street people* are **mistreated**.

summary (c) after *national mistreatment of homeless* is unpacked: Advocacy coalition criticizes the fact that *the homeless* are **mistreated** in the nation.

Let us consider whether any recurrent nucleus involving *against* and two instances of *mistreated* (emboldened) is contained in the summaries. In order to do so, firstly we need to identify the Mediums of these three processes respectively: in summary (a) the Medium of *against* is *homeless* (italicised); in summary (b) the Medium of *mistreated* is *street people* (also italicised); and in summary (c) the Medium of *mistreated* is *the homeless* (also italicised). After that, we consider whether any repeated Mediums exist. The three Mediums just noted co-refer to the same class of people. Therefore, a recurrent nucleus referred to as *mistreat homeless people* is identified.

After that, the two steps noted above are applied to the other chain, which consists of *alleges* in summary (b) and *criticizes* in summary (c). The Medium of *alleges* is the Sayer, namely *National Law Center on Homelessness and Poverty*. The Medium of *criticizes* is not the Sayer but the Target, namely *the fact that the homeless are mistreated in the nation*. These two Mediums are not in a repetition relation because *National Law Center on Homelessness and Poverty* is an organization, whereas *the fact that the homeless are mistreated in the nation* is a proposition.

Therefore, *mistreat homeless people* is the only recurrent nucleus in the summaries of text 4.

3.3.8 Comparing the results in A4 with that in B3 to identify the recurrent nuclei in text that are repeated in two or more summaries of the text

Having identified both recurrent nuclei in text and recurrent nuclei in the summaries of the text, step C (of Figure 3.1) compares these two results with a view to identifying the recurrent nuclei in text that are repeated in two or more summaries of the text.

For instance, the recurrent nuclei in text 4 were ranked in descending order of frequency in Table 3.1. On the other hand, the summaries of this text only contain one recurrent nucleus, namely *mistreat homeless people*. The comparison of the two results shows that only the most frequent nucleus in this text, namely *mistreat homeless people*, is repeated in two or more summaries.

This chapter has explained the method used to identify recurrent nuclei in a text that are important to the meaning of the text. The two chapters that follow will discuss their characteristics.

Chapter 4 Findings with Regard to Research Question 1

4.1 Introduction

Having analyzed my data following the methodology described in chapter 3, this chapter gives the results of research question 1, namely which recurrent nuclei in a text are repeated in two or more summaries of the text, discusses the findings, and explores the characteristics of the recurrent nuclei in texts that are important to the texts' meanings.

4.2 The results with regard to research question 1

The texts I analyzed were classified into the categories given below.

Category 1 contains texts in which only the most frequent recurrent nucleus is repeated in two or more summaries of the text. It is useful to further divide it into three subcategories. Subcategory 1(a) contains texts in which only the recurrent nucleus that is more frequent than any other recurrent nucleus is repeated in two or more summaries of the text. This subcategory contains 13 texts, one of which is text 4 we have discussed in chapter 3. In this text, only the recurrent nucleus *mistreat homeless people*, which is more frequent than any other nucleus, is repeated in two or more summaries of the text.

Subcategories 1(b) and 1(c), on the other hand, contain texts in which two or more recurrent nuclei are jointly most frequent. If only one of the jointly most frequent nuclei is repeated in two or more summaries, the text is assigned to subcategory 1(b). This subcategory contains seven texts, one of which is text 18. In this text *New Zealand children have suffered* and *gather information*, both occurring three times, are more frequent than any other nucleus. But only *New Zealand children have suffered* is repeated in two or more summaries of the text. If, on the

other hand, two or more of the jointly most frequent nuclei are repeated in two or more summaries of the text, the text is assigned to subcategory 1(c). Only one text, text 21, belongs to this subcategory.

Category 2 contains texts in which not only the most frequent nucleus but also one or more of the less frequent nucleus are repeated in two or more summaries of the text. This category can be also divided into two subcategories. Subcategory 2(a) is derived after adding 'one or more of the less frequent recurrent nuclei' to category 1(a). In other words, subcategory 2(a) contains texts in which not only the single most frequent nucleus but also one or more of the less frequent recurrent nuclei are repeated in two or more summaries of the text. This subcategory contains 18 texts, one of which is text 27. In this text *detain villagers*, occurring six times, is more frequent than any other nucleus. Both this recurrent nucleus and *issued fake warrants* (which occurs three times) are repeated in two or more summaries of the text.

Similarly, subcategory 2(b) is derived after adding 'one or more of the less frequent recurrent nuclei' to categories 1(b) and 1(c) together. In other words, subcategory 2(b) contains texts in which not only one or more of the jointly most frequent nuclei but also one or more of the less frequent recurrent nuclei are repeated in at least two summaries of the text. This subcategory contains five texts, one of which is text 41. In this text *increase interest rates* and *Chinese economy grows*, both occurring four times, are more frequent than any other nucleus. However, *Chinese economy grows* is not repeated in two or more summaries of the text. On the other hand, *increase interest rates* and *control investment,* which occurs three times in the text, are repeated in two or more summaries of the text.

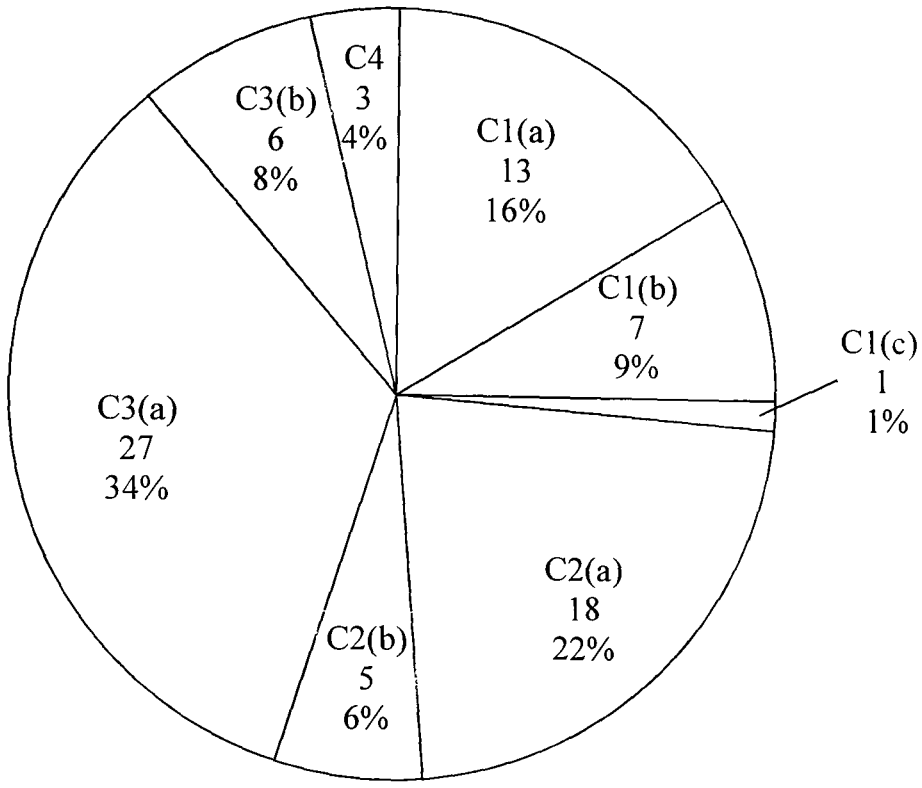
Category 3 contains texts in which the most frequent recurrent nucleus is not repeated but one or more of the less frequent recurrent nuclei is repeated in at least two summaries of the text. This category can be also divided into two subcategories. If only one of the less frequent recurrent nuclei in a text is repeated in two or more summaries, the text is classified into subcategory 3(a). This subcategory contains 27 texts, one of which is text 49. In this text the most frequent nucleus *teaching evolution* (occurring four times) is not repeated in the summaries of the text, but *removed evolution* (occurring three times) is repeated in two summaries. If, on the other hand, two or more of the less frequent recurrent nuclei in a text are repeated in at least two summaries of the text, the text will be classified into subcategory 3(b). This subcategory contains six texts, one of which is text 73. In this text the two jointly most frequent nuclei, *giant pandas live* and *Zhang Zhihe said*, both occurring six times, are not repeated in the summaries of the text, but two less frequent recurrent nuclei, *train giant pandas* (occurring five times) and *build a center* (occurring twice) are repeated in at least two summaries of the text.

Category 4 contains texts in which none of the recurrent nuclei are repeated in two or more summaries of the text. There are three texts in this category. Category 5 contains texts that do not contain any recurrent nuclei. However this is an empty category since none of the texts I analyzed belong to this category. Similarly, category 6 would in theory contain (original) texts of which the summaries do not contain any recurrent nuclei. However, as with category 5, none of the texts I analyzed belong to this category.

Figure 4.1 below diagrammatically represents the proportions of the different categories in my results. The number of texts contained in the categories is given under the category labels. The

ratio in each category is the ratio of the number of texts in that category to 80 (i.e. as a proportion of the total number of texts I analyzed). Henceforth, C stands for category.

Figure 4.1: The proportions of different categories in my results



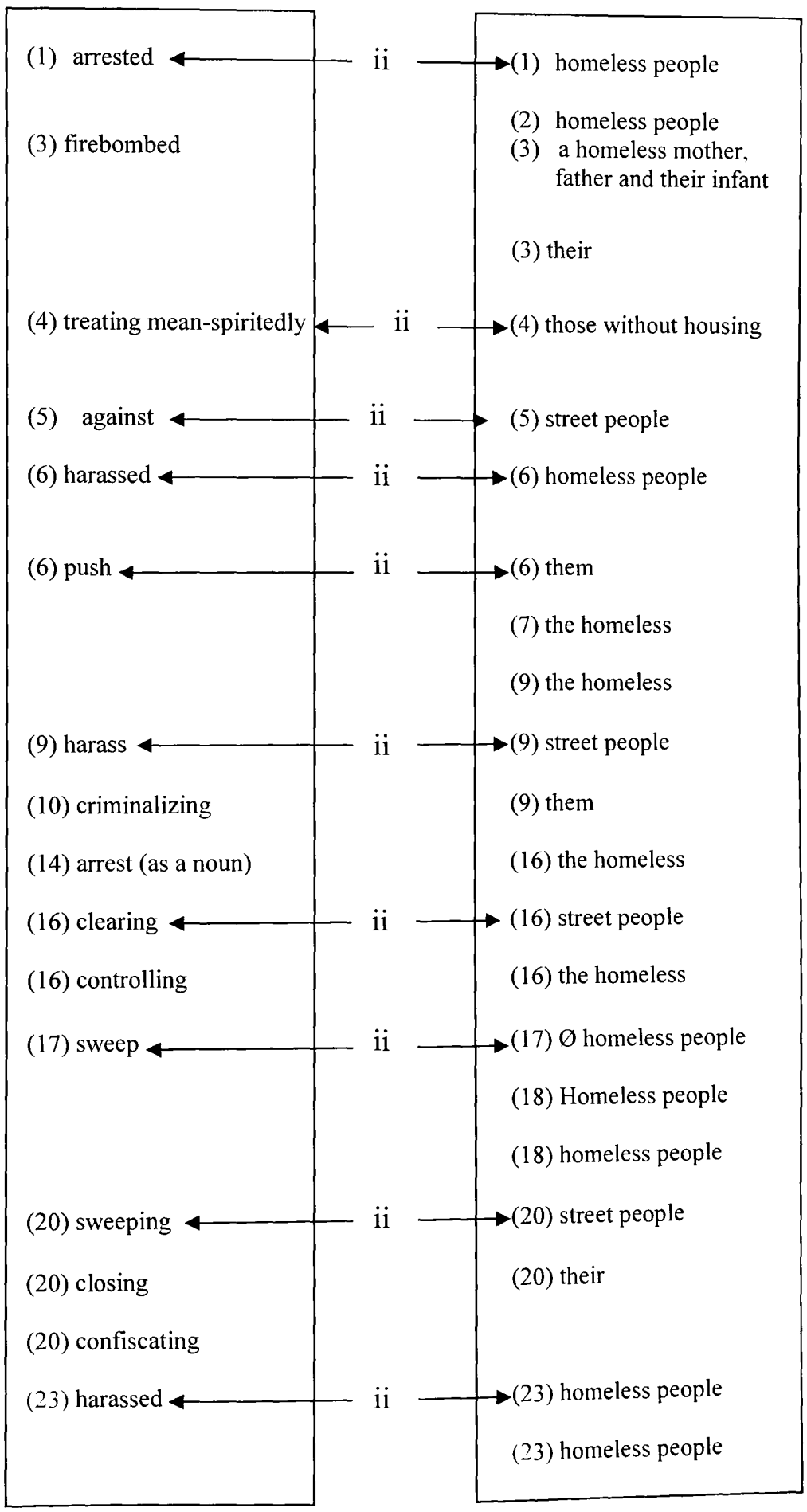
If we look at the results given in Figure 4.1 more closely, we can make four observations. The first is that C1(a), C1(b), C1(c), C2(a), C2(b), C3(a) and C3(b) together contain 77 texts, accounting for approximately 96% of my data. This result shows that more often than not one or more recurrent nuclei in a hard news text are repeated in two or more summaries of that text, indicating that one or more of the chain interactions in a hard news text between processes that are similar in meaning and their Mediums is often important to the text's meanings. This is explained below. A recurrent nucleus consists of interacting members of two chains that interact in Process-Medium relation. For instance, in text 4 the most frequent nucleus *mistreat homeless people* consists of interacting members of two chains that interact in Process-Medium

relation (see Figure 4.2 given on p166). One chain concerns 'negative treatment' and the other contains *homeless people* and its equivalent expressions and exemplification.

Figure 4.2 represents the interactions of the two chains after four nominalizations, namely *the nation's* ***treatment*** *of those without housing* in sentence 4, *the* ***harassment*** *of homeless people* in sentence 6, *the* ***sweeps*** in sentence 17, and *the* ***harassment*** *of homeless people on the streets* in sentence 23, have been unpacked. In these nominal groups the nominalised processes belonging to the chain that concerns 'negative treatment' are emboldened. After they have been unpacked, the cognate verbs of the emboldened nominalised processes interact with *homeless people* or its equivalent expressions. The sentences in which the items occur are given in brackets. In sentence 17 the Medium of *sweeps* (as a noun) is not present in the nominal group *the sweeps*, but can be inferred as *homeless/street people* from either *clearing street people* in sentence 16 or *sweeping street people* in sentence 20. This Medium is preceded by Ø symbol to show that it is elliptical and recovered. The interacting members of the two chains are linked by double headed arrows. The 'ii' symbol stands for Process-Medium relation.

The recurrent nucleus *mistreat homeless people* in the text is repeated in all summaries of the text, meaning that the interaction of the two chains given in Figure 4.2 is important to the text's meaning.

Figure 4.2: The interactions between the chain that concerns 'negative treatment' and the chain that contains *homeless people* and its equivalent expressions and exemplification



The second observation to be drawn from Figure 4.1 comes from comparing the combined result of the subcategories of C1 and C2 with that of the subcategories of C2 and C3. Both C1 and C2 involve the most frequent recurrent nucleus in a text being repeated in two or more summaries of the text. Therefore, the combined result of the five subcategories comprising C1 and C2 shows that the most frequent nuclei in hard news texts have approximately a 54% probability of being important to the texts' meanings. On the other hand, the combined result of the four subcategories comprising C2 and C3 shows that less frequent recurrent nuclei in hard news texts have approximately a 70% probability of being important to the texts' meanings. The comparison of the two ratios shows that in hard news texts less frequent recurrent nuclei have a higher probability than the most frequent nuclei of being important to the texts' meanings.

The implication of this perhaps surprising finding is that the frequency hierarchies of recurrent nuclei in hard news texts do not equate to the degree of importance of recurrent nuclei to the texts' meanings. The frequency hierarchy of recurrent nuclei in a hard news text arranges recurrent nuclei in descending order of frequency. The most frequent nucleus in the text is naturally at the top of the hierarchy, representing the meaning that is most frequently referred to in the text, followed by the recurrent nuclei occurring less frequently in the text. It would have been tempting to assume that the more frequently a meaning is referred to in a text, the more important it is to the text's meanings. Had this assumption been true the following two statements would have been valid.

1) The most frequent nucleus is always more important than any other less frequent nucleus to the text's meanings.
2) The more frequently a nucleus occurs in a text, the more important it is to the text's meanings.

Statement 1 is not valid because the second observation of Figure 4.1 shows that in hard news texts less frequent recurrent nuclei have a higher probability than the most frequent nuclei of being important to the texts' meanings.

Statement 2 is also not valid because in each text contained in C3(a) and C3(b), the most frequent nucleus is not but one or more of the less frequent recurrent nuclei are thought of (by readers) as being important to the text's meanings.

An example in which neither statement 1 nor 2 is valid is text 72 given below in full (this text is contained in category 3b). Henceforth, sentences are numbered for convenience of discussion.

YUXIAN, Shanxi – (1) At least **10 people died** and *one is missing* after a coal mine gas explosion Tuesday in north China's Shanxi Province, a spokesperson for work safety authorities said Wednesday.

(2) The accident occurred at around 4:50 p.m. Tuesday when 41 miners were working in the shaft of Chentong Coal Industry Co., Ltd., in Yuxian County.

(3) Twenty-seven people escaped from the scene unhurt and 14 were trapped underground, said Liu Xianyun, of the work safety committee of Yangquan City, which administers Yuxian county.

(4) Three of the trapped had been rescued, while **10 died** and *one is missing*, Liu said at a press conference.

(5) The company was licensed with an annual production capacity of 600,000 tonnes.

(6) Rescue[12] work is continuing, and work safety authorities are investigating the cause of the accident.

This text contains three recurrent nuclei: *a spokesperson for work safety authorities said* (underlined), *10 people died* (emboldened) and *one is missing* (italicised). *A spokesperson for*

[12] *Rescue* was not suitable for unpacking because criterion 2 in Figure 3.9 is not satisfied, i.e. *rescue* doesn't function as Thing (i.e. the head noun) in the nominal group *rescue work*.

work safety authorities said occurs three times, while both *10 people died* and *one is missing* occur twice.

The summaries of this text, alphabetised, are given below. In the summaries the nuclei that repeat the recurrent nucleus *10 people died* contained in the text are emboldened, while the nuclei that repeat the recurrent nucleus *one is missing* contained in the text are italicised.

a) At least **10 killed** and *one was missing* in an explosion.
b) **10 died** and *one was missing* in a coal mine explosion.
c) **10 died** and *one was missing* in mine explosion in north China.

The most frequent recurrent nucleus *a spokesperson for work safety authorities said* in the text is not repeated in any summary. On the other hand, both *10 people died* and *one is missing* are repeated in each summary.

The converse implication of the result reported above is that the recurrent nuclei that are not frequent in hard news texts may be important to the texts' meanings. Categories 2 and 3 together contain 32 recurrent nuclei that are important to the meanings of the texts in which they occur but occur only twice in their texts. Two such recurrent nuclei are *10 people died* and *one is missing* in text 72 discussed above.

The third observation concerns category 4. This category contains only approximately 4% of my data, suggesting that it is rare for none of the recurrent nuclei in a hard news text to be important to the text's meanings.

The final observation concerns categories 5 and 6. The absence of texts in category 5 indicates that repetitions are characteristic of hard news texts. White (1997) says that in hard news texts, the headlines and the leads together constitute the 'nuclei' and that different parts

of the remaining discourse after the leads can be seen as the satellites of the nuclei, specifying the nuclei in a variety of relations, such as elaboration, contextualization, explanation, appraisal and justification, etc. He (1997: 420) also adds that 'the need for each satellite to establish a direct link back to the headline/lead nucleus means that there is **significant redundancy or repetition** in the body of the news story as the original point of impact is elaborated or restated by individual satellites' (my emboldening). White's statements suggest that in hard news texts information contained in the headline/lead nuclei are often repeated in the remaining paragraphs after the leads.

The absence of texts in category 6 indicates that it is very rare for the summaries of a hard news text not to contain any recurrent nucleus. This is not surprising because researchers working on automatic text summarizations found that very important information and unimportant information tend to be agreed on by a majority of judges, whereas moderately important information is a matter of disagreement more often than are very important information and unimportant information (Marcu, 1999). The implications are that the recurrent nuclei in a text that are repeated in 50% of a group of summaries of the text may be considered as moderately important to the text's meaning, and that those that are repeated in a majority of summaries may be considered as important to the text's meanings.

4.3 The characteristics of the recurrent nuclei in texts that are important to the texts' meanings

Having discussed the results of my first research question and their implications, this section looks more closely into texts contained in the categories with a view to exploring the characteristics of the recurrent nuclei in texts that are thought of by readers as important to the texts' meanings.

4.3.1 Category 1

We begin with category 1. Subcategory 1(a) contains 13 texts in which there is a single nucleus that is most frequent and is repeated in two or more summaries of the text in which it occurs. Subcategories 1(b) and 1 (c) together contain nine jointly most frequent nuclei that are repeated in two or more summaries of the texts in which they occur. These 22 recurrent nuclei are distributed in the (original) texts in four ways. For ease of description, each reference in this section to the most frequent nucleus includes jointly most frequent nucleus.

The first way is that at least one occurrence of the most frequent nucleus in a text that is repeated in two or more summaries of the text is present in the lead and the remaining occurrences recur at intervals in the sentences after the lead. 14 of the most frequent nuclei that are repeated in two or more summaries of the texts in which they occur are distributed in such a way. 12 have single occurrences in the leads and the other two occur twice in the leads.

For instance, the most frequent nucleus in text 4, *mistreat homeless people*, is distributed in such a way. As shown in Figure 4.2, the first occurrence *homeless people can be arrested* is present in the lead; the remaining occurrences recur at intervals in the remaining sentences after the lead.

The 14 most frequent nuclei that are distributed in such a way occur in 14 hard news texts, accounting for approximately 18% of my data. Therefore, the repetitions in the sentences after the lead of nuclei in the lead that are important to a text's meaning may be a popular means of writing hard news texts. White also points this out in his (1997) paper, saying that 'both event stories and issues reports are marked by repetitions of the original point of

newsworthy impact as set out in the headline/lead. These repetitions occur in a pulse-like rhythm[13] as the text unfolds' (1997: 412).

The second way in which the most frequent nucleus in a text that is repeated in two or more summaries of the text is distributed is that it does not occur in the lead, but recurs at intervals in the sentences after the lead.

The second way differs from the first way only in that none of the occurrences of the most frequent nucleus in a text that is replicated in the text's summaries are found in the lead. Only the most frequent nucleus in text 3, namely *ban cell phone use*, is distributed in such a way. This recurrent nucleus occurs in sentences 2, 4, 6, 11, 16, 24, 25, 29 and 33. These sentences and the lead of this text are given below. The occurrences embodying *ban cell phone use* are emboldened. The results after *a nationwide ban on cell phone use in cars* in sentence 2, *an outright ban* in sentence 24, and *outright bans* in sentence 33 have been unpacked are given in square brackets under the original sentences. On the other hand, *cell phone use in cars* in sentence 2 after *a nationwide ban on cell phone use in cars* has been unpacked, *the use of wireless phones in automobiles* in sentence 6, *certain cell phone uses* in sentence 11, *car phone use* in sentence 16, *cell phone use* in sentence 24 after *an outright ban* has been unpacked, *use of hand-held phones* in sentence 29, and *cell phone use* in sentence 33 after *outright bans* has been unpacked are not suitable for unpacking because when these nominalizations are interrogated by criterion 13 in Figure 3.9, the answers are positive, that is, these nominalizations all function as the Mediums of the recurrent nucleus *ban cell phone use*.

(1) Patricia Pena gave little thought to using a cellular telephone on the road before a distracted driver hit her car broadside, fatally injuring her 2-year-old daughter.

[13] The term 'a pulse-like rhythm' in White's statement is inherently difficult to pin down precisely. Roughly, it means that the point of newsworthy impact set out in the headline/lead recur at intervals in the sentences after the lead.

(2) Now she has devoted herself to pushing for a nationwide ban on cell phone use in cars, except in emergencies.

[Sentence 2 after *a nationwide ban on cell phone use in cars* has been unpacked: Now she has devoted herself to pushing for a legislation to **ban cell phone use** in cars nationwide, except in emergencies.]

(4) Lawmakers in states and cities across the nation are considering legislation to **ban drivers from using their cell phones** in cars.

(6) No state **bans the use of wireless phones** in automobiles, and only California, Florida and Massachusetts have laws limiting cell phone use in moving vehicles, according to the National Conference of State Legislatures.

(11) Hilltown now **bans certain cell phone uses** for drivers.

(16) Proposals to **ban car phone use** have been introduced in Massachusetts and in Nebraska, where talking on a cell phone would be considered contributory negligence in auto accidents.

(24) "Educating drivers about how to use their cell phones responsibly will do more good than an outright ban", said Lisa Ihde, spokeswoman for the Cellular Communications Industry Association, a Washington-based industry group representing wireless carriers and manufacturers.

[Sentence 24 after *an outright ban* has been unpacked: "Educating drivers about how to use their cell phones responsibly will do more good than **banning cell phone use** outright", said Lisa Ihde, spokeswoman for the Cellular Communications Industry Association, a Washington-based industry group representing wireless carriers and manufacturers.]

(25) "We don't want to see any legislation that would **deter drivers from using their phones** in their cars," she said.

(29) Such measures have failed to get out of committee in many states, including New York, where bills **prohibiting use of hand-held phones** while driving have languished in committee since 1996.

(33) Opponents of outright bans believe the benefits of car phones -- the ability to quickly report drunken drivers, accidents and breakdowns -- far outweigh any drawbacks.

[Sentence 33 after *outright bans* has been unpacked: Opponents of **banning cell phone use** outright believe the benefits of car phones -- the ability to quickly report drunken drivers, accidents and breakdowns -- far outweigh any drawbacks.]

The summaries of this text are given below. The results after *bans on cell phone usage* in summary (a) and *ban on cell phone use in cars* in summary (b) are unpacked are given in square brackets under the original summaries.

The summaries of text 3 are:

a) More legislators nationwide considering bans on cell phone usage.

[Summary (a) after *bans on cell phone usage* is unpacked: More legislators nationwide considering proposals to **ban cell phone usage**]

b) Bereft mother seeks ban on cell phone use in cars.

[Summary (b) after *ban on cell phone use in cars* is unpacked: Bereft mother seeks to **ban cell phone use** in cars]

c) Mother against cell phone driving.

Once the nominal group *bans on cell phone usage* in the first summary is unpacked to *[proposals] to ban cell phone usage*, we can see that the recurrent nucleus *ban cell phone use* in the text is repeated in this summary. Similarly, once the nominal group *ban on cell phone use in cars* in the second summary is unpacked to *ban cell phone use in cars*, we can see that the recurrent nucleus noted above in the text is also repeated in this summary.

It is, however, difficult to determine whether the recurrent nucleus noted above in the text is repeated in the last summary because though *against* is in a repetition relation with *ban*, the meanings of *cell phone driving* can be interpreted in more than one way: it could mean, among other things, using a cell phone while driving (if so, the recurrent nucleus *ban cell phone use* in the text is repeated in the last summary), or driving with a cell phone in hand but

not using it, or even driving with the help of a cell phone (in the latter two cases, the recurrent nucleus *ban cell phone use* in the text is not repeated in the last summary). For safety reason, my analysis did not consider *ban cell phone use* in the text to be repeated in the last summary.

None of the occurrences of *ban cell phone use* is present in the lead. The implication is that the information that is important to the meaning of a hard news text may not be present in the lead. Both Bell (1991: 183) and Van Dijk (1986; 1988: 53) say that the leads can be seen as abstracts that summarize the gist of hard news texts. This must be regarded as a statement of typicality not of absolutes, that is, the leads typically, but not inevitably, summarize the gist of hard news texts.

The third way of distribution in the text of the most frequent nucleus in a text that is replicated in the text's summaries is that one of its occurrences is present in the lead and at least half of its total occurrences are present in a single sentence after the lead. The most frequent nuclei in texts 5, 8, 14, and 16 that are repeated in two or more summaries of these texts respectively are distributed in such a way. An example is the most frequent nucleus in text 8, *killed people*, which occurs eight times in this text. The first occurrence *killed at least 82 people* (emboldened) is present in the lead. Another four occurrences (also emboldened) are present in sentence 9. These two sentences are given below.

(1) Typhoon Babs headed toward southern China Friday after battering the central and northern Philippines with heavy winds and rain that **killed at least 82 people**, flattened crops, and forced more than 100,000 people to flee their homes.

(9) The fatalities included **four people hit** by falling trees, **three girls buried** by landslides in central Iloilo province, a man **who died** of shock, and **a fireman crushed** after his van overturned on a slippery road, Red Cross officials said.

Hit by falling trees, *buried by landslides*, *died of shock* and *crushed* in sentence 9 are the different ways in which people were killed.

The fourth and final way of distribution in the text of the most frequent nucleus in a text that is repeated in two or more summaries of the text is that it occurs only twice in the text, with one of the occurrences being present in the lead.

Yilmaz's government fell in text 19, *evacuate 185 villagers* and *shut down a highway* in text 21 are distributed in such a way. Text 21 in full is given below for illustration. This text contains three recurrent nuclei: *evacuate 185 villagers* (emboldened), *shut down a highway* (italicised) and *Chen Tianqi said* (underlined). Each of these recurrent nuclei occurs only twice.

(1) A gas leak caused by a truck accident forced **185 villagers to evacuate** and *shut down a highway* in northwest China's Gansu Province, a local newspaper reported on Wednesday.

(2) The accident happened around 7:45 a.m. on Tuesday in Qingquan Township, Yumen City, when a truck carrying 21 tons of natural gas veered off the highway and subsided into a four-meter-deep roadside ditch.

(3) The driver and another person in the truck were killed. 4 Fire fighters extinguished a blaze caused by the truck's oil tank.

(5) Officials **evacuated 185 residents** from a village about 100 meters from the accident and *closed a section of the highway* from Yumen to Jiayuguan City, which lies to the west.

(6) Chen Tianqi, vice mayor of Yumen, who headed the rescue effort, said the accident was caused by driver fatigue.

(7) Chen said the wreck had been transported to safer ground to let out the remaining gas, and normal traffic would resume on Wednesday night.

(8) The truck was owned by a transportation company in Miquan City, in central-north Xinjiang.

The summaries of this text are given below. Both *evacuation* and *road closure* contained in summaries (a) and (c) are suitable for unpacking. The results after these two nominalizations

have been unpacked are given in square brackets under the original summaries. On the other hand, *gas leak* contained in all the three summaries is not suitable for unpacking because criterion 3 in Figure 3.15 is not satisfied, that is, *leak* (as a noun) is not repeated by at least a verb in the text.

a) Gas leak in truck accident forces evacuation and road closure in Gansu.
[Summary (a) after *evacuation* and *road closure* have been unpacked: Gas leak in truck accident forces **185 people to be evacuated** and *road to be closed* in Gansu.]

b) Gas leak forced **185 to be evacuated**. *A highway was shut down.*

c) Gas leak forced evacuation and led to road closure in Gansu.
[Summary (c) after *evacuation* and *road closure* have been unpacked: Gas leak forced **185 people to be evacuated** and led *road to be closed* in Gansu.]

The recurrent nucleus *evacuate 185 villagers* in the text is repeated by *185 to be evacuated* in summary (b), and the recurrent nucleus *shut down a highway* in the text is repeated by *A highway was shut down* in the same summary.

Once *evacuation* in summaries (a) and (c) is unpacked, we can see that the recurrent nucleus *evacuate 185 villagers* in the text is also repeated in these two summaries. Similarly, once *road closure* contained in these two summaries has been unpacked, we can see that the recurrent nucleus *shut down a highway* in the text is also repeated in these two summaries.

Therefore, both *evacuate 185 villagers* and *shut down a highway* are important to the meanings of the text. On the other hand, the other recurrent nucleus in the text, namely *Chen Tianqi said*, which has a frequency that is equal to the two recurrent nuclei noted above, is not repeated in any summary, meaning that it is not important to the text's meanings.

Text 19 contains four recurrent nuclei, all of which occur only twice. Only *Yilmaz's government fell* is repeated in two or more summaries of this text.

Texts 19 and 21 show the fact that recurrent nuclei have the same frequency in a hard news text does not mean they are equally important to the text' meanings, supporting the conclusion drawn from Figure 4.1 that the frequency hierarchies of recurrent nuclei in hard news texts do not equate to their degrees of importance to the texts' meanings.

To sum up, the present section has discussed four ways of distribution in my data of the most frequent recurrent nuclei that are replicated in the summaries of the texts in which they occur. Of the four ways of distribution, the first way is most common because a majority of the most frequent nuclei that are important to their texts' meaning are distributed in such a way. All but the second way of distribution involve the presence in the leads of the most frequent nuclei that are important to the meaning of the texts in which they occur. Section 4.3.4 of this chapter investigates the probability of the presence in the leads of nuclei that are important to texts' meaning.

4.3.2 Categories 1(c), 2 and 3(b)

Categories 1(c), 2 and 3(b) are considered together because they all involve two or more recurrent nuclei in a text being repeated in two or more summaries of the text. Henceforth the initial occurrence of a recurrent nucleus in a text refers to its occurrence that is contained in the lead. The initial occurrence in a text of two or more recurrent nuclei that are repeated in two or more summaries of the text are linked in four ways given below in the leads of 23 texts, accounting for approximately 77% of the number of texts contained in the three categories.

Linkage 1

Linkage 1 occurs when the initial occurrence in a text of a recurrent nucleus that is repeated in two or more summaries is (part of) the content of saying of the initial occurrence of a verbal recurrent nucleus (or less often is (part of) the content of sensing of the initial occurrence of a mental recurrent nucleus) that is also repeated in two or more summaries.

This type of linkage involves a verbal recurrent nucleus or less often a mental recurrent nucleus in a text being repeated in two or more summaries of the text. It occurs in the lead of texts 24, 25, 32, 36, 37, 38, 39, 42 and 75. This type of linkage is often realized by projection. An example is the lead of text 25 given below, which reports on a warning by doctors that using a mobile phone in a thunderstorm is life-threatening. *Using a mobile phone* (emboldened), *kill people* (italicised) and *doctors warned* (underlined) are the initial occurrences of three recurrent nuclei in the text that are repeated in two or more summaries.

Using a mobile phone during a thunderstorm could *kill people*, doctors warned today. (the lead of text 25)

In this lead both *using a mobile phone* and *kill people* occur in the clause complex that is projected by *doctors warned.* In such a way, the first two nuclei are linked to the third. The projected clause complex gives the content of the doctors' warning. Therefore, both *using a mobile phone* and *kill people* are part of the content of the warning.

Linkage 2

The second type of linkage occurs when the initial occurrences of two recurrent nuclei in a text that are repeated in two or more summaries of the text are linked in causal relations (purpose, reason and consequence).

The linkage concerning purpose is expressed by *to* (shorthand for *in order to*) in the lead of texts 27 and 73, by *in an effort to* in the lead of text 41, by *the objective of ...was...* in the lead of text 35, by *in hopes of* in the lead of text 74. An example is the lead of text 27 given below. This text reports that the police in Hebei province, China, issued faked warrants to detain six villagers for bail money.

BEIJING – Police in a county of North China's Hebei province **issued fake warrants** to *detain six villagers* in a land dispute for bail money, the Beijing News reported on Wednesday.

Issued fake warrants (emboldened) and *detain six villagers* (italicised) are the initial occurrences in this text of two recurrent nuclei that are repeated in two or more summaries of the text. *Detain six villagers* is linked to *issued fake warrants* by *to* (shorthand for *in order to*). Therefore, *detain six villagers* is expressed as the purpose of *issued fake warrants*.

The linkage concerning reason occurs in the lead of text 26 only, expressed by *due to*.

TAIXING - The man who was detained after **a knife attack on children** at an east China kindergarten Thursday carried out **the attack** due to *personal humiliations*, police said Friday. (The lead of text 26)

A knife attack on children and *the attack* (both are emboldened) are the first two occurrences in the text of a recurrent nucleus that is repeated in two or more summaries of the text: *personal humiliations* (italicised) is the initial occurrence in the text of another recurrent nucleus that is also repeated in two or more summaries. *Personal humiliations* is linked to *the attack* by *due to*. Therefore, *personal humiliations* is expressed as the reason why the man attacked children with a knife.

The linkage concerning consequence is expressed by *and* (followed by implicit *therefore*) in the lead of text 76 given below.

Chinese police opened fire and *wounded four protesters* "in self-defence" last Sunday in a Tibetan area of Sichuan province, the Xinhua news agency says.

In the lead *Chinese police opened fire* (emboldened) and *wounded four protesters* (italicised) are the initial occurrence in the text of two recurrent nuclei that are repeated in two or more summaries of the text. *'Therefore'* seems to be implicit after the coordinator *'and'* that link them because the shooting of the police resulted in the injuries of the protesters.

The connector *and* can be also used in other ways. When the initial occurrences in a text of two or more recurrent nuclei that are repeated in two or more summaries of the text are linked by this connector and a temporal relation hold between them, the instance is classified into linkage 3 given below. On the other hand, when the meanings of the nuclei add to each other but no circumstantial relation (e.g. causal or temporal relation) holds between them, the instance is classified into the fourth and final linkage.

Linkage 3

The third type of linkage occurs when the initial occurrences in a text of two recurrent nuclei that are repeated in two or more summaries are linked in temporal relations. This linkage is expressed by *and* (followed by implicit *then*) in the leads of texts 23 and 77 respectively, by *after* in the lead of text 40, and by *while* (meaning *at the same time*) in the lead of text 44.

An example is the lead of text 23 given below. This text reports on the early returning of many Mozambicans to their villages after the breakout of a flood.

MAPUTO, Mozambique (AP) – Not fearing more flooding and a possible cholera outbreak in Mozambique, restless **flood victims were abandoning** aid camps and (O *flood victims were*)[14] *heading* back to their devastated villages, officials said.

[14] The subject and finite element of this nucleus, namely *flood victims were*, are bracketed and preceded by O to indicate that they were recovered from the preceding context.

Flood victims were abandoning (emboldened) and *flood victims were heading* (italicised) are the initial occurrences in the text of two recurrent nuclei that are repeated in two or more summaries of the text. The temporal adjunct *'then'* is implicit after *'and'* (underlined) that links them because it is not possible for the flood victims to head back to their homes if they had not have left the aid camps.

Linkage 4

The fourth and final type of linkage occurs when the initial occurrences in a text of two or more recurrent nuclei that are repeated in two or more summaries of the text are coordinated but no circumstantial relation holds between them. This type of linkage is expressed by the coordinator *'and'* in the leads of texts 21, 34 and 72. An example is the lead of text 72 given below.

YUXIAN, Shanxi – At least **10 people died** and *one is missing* after a coal mine gas explosion Tuesday in north China's Shanxi Province, a spokesperson for work safety authorities said on Wednesday.

In the lead *10 people died* (emboldened) and *one is missing* (italicised) are the initial occurrences in the text of two recurrent nuclei that are repeated in two or more summaries of the text. These two nuclei are coordinated by the connector *'and'* but no circumstantial relation holds between them.

Replication in the summaries of the linkages that occur in the leads

Having considered the relationships that link the initial occurrences in a text of two or more recurrent nuclei that are repeated in two or more summaries of the text, I examined whether these linkages are echoed in the summaries of the texts in which they occur. I found that the linkages that occur in the leads of 21 texts are echoed (albeit not necessarily with the same

wording) in two or more summaries, accounting for approximately 91% of the texts where the leads contain the four types of linkage noted above. An example is text 27 of which the lead is reproduced below for convenience of discussion. More examples are discussed in sections 5.5.1, 5.5.2 and 5.5.3(a) of chapter 5.

BEIJING – Police in a county of North China's Hebei province **issued fake warrants** to *detain six villagers* in a land dispute for bail money, the Beijing News reported on Wednesday. (the lead of text 27)

The summaries of this text are given below. *Issued fake warrants* in the lead is repeated by *issued fake warrants* (emboldened) contained in summaries (a) and (c), but is not repeated in summary (b) because *issued* in the lead is not in a repetition relation with *used* contained in this summary. *Detain six villagers* in the lead is repeated by *detain six villagers* (italicised) contained in all three summaries.

the summaries of text 27 are:

a) Police **issued fake warrants** to *detain six villagers* for money.
b) Police used fake warrants to *detain six villagers* for bail.
c) Police in Hebei **issued fake warrants** to *detain six villagers*.

The relationship in the lead that links *detain six villagers* to *issued fake warrants* is echoed in summaries (a) and (c). In these summaries *to* is short for *in order to*.

The replication in summaries of the linkages that occur in the leads shows that when two or more recurrent nuclei in a hard news text are repeated in two or more summaries of the text, the relationship in the lead between these nuclei is also often repeated in the summaries. The implication is that when two or more propositions in a text are thought of by readers as important to the text's meanings, the relationship in the lead that links these propositions is also thought of as important to the text's meanings. This provides empirical evidence for the

claim made by many linguists, such as Eggins (1994, 2004), Halliday and Hasan (1976), Halliday (1994), Halliday and Matthiessen (2004), Hoey (1983, 2001), Martin (1992), Martin and Rose (2003), Martin *et al* (2010), Mann and Thompson (1988a, 1988b), McCarthy (2002), Matthiessen and Thompson (1988), Matthiessen (2002), Quirk *et al* (1972, 1985), Thompson (1996, 2004, 2005, 2010), and Winter (1977, 1994), that the relations that hold between the propositions in a text are important to the meaning of the text.

4.3.3 Subcategory 3(a)

We now consider subcategory 3(a). This subcategory contains 27 less frequent recurrent nuclei that are important to the meaning of the texts in which they occur. 24 such recurrent nuclei have a single occurrence in the lead, another one has two occurrences in the lead, while the other two, namely *replace Hubble's gyroscopes* in text 48 and *remove evolution* in text 49, do not occur in the leads of their texts. Therefore *replace Hubble's gyroscopes* and *remove evolution* are worth examining closely.

Text 48 reports that mechanical problems have threatened to shut down the astronomical observations of the Hubble Space Telescope, making the repair mission happen earlier than planned. In this text *replace Hubble's gyroscopes* (emboldened) occurs three times, found in sentences 5, 25 and 26 given below. The result after *the repairs* in sentence 5 has been unpacked is given under sentence 5.

(5) Instead, the failure of the gyros would cause the craft to go into an automatic "safe mode" until the repairs are made.

[Sentence 5 after *the repairs* has been unpacked: Instead, the failure of the gyros would cause the craft to go into an automatic "safe mode" until **the gyros were repaired**]

(25) NASA already had planned for a June 2000, Hubble servicing mission that would include **replacing all of the gyroscopes**, along with a computer and some other instruments.

(26) But Leckrone said he and others at the Goddard Spaceflight Center have proposed that an emergency mission be flown this October to **change out the ailing gyros**, and that the rest of the serving be done on a mission in 2001.

The summaries of the text are as follows:

a) Problems may stop Hubble astronomical observations; NASA may accelerate repair mission[15].
b) Shuttle launch planned earlier to **replace Hubble gyroscopes**.
c) NASA may **replace Hubble's** ailing **gyroscopes** six months early.

The recurrent nucleus *replace Hubble's gyroscopes* in the text is important to the text's meanings because it is repeated by *replace Hubble gyroscopes* (emboldened) contained in summary (b), and by *replace Hubble's...gyroscopes* (also emboldened) contained in summary (c). But *replace Hubble's gyroscopes* does not occur in the lead of text 48, given below:

Mechanical problems have threatened to shut down the astronomical observations of the Hubble Space Telescope, NASA officials told Congress on Wednesday.

The implication is that information that is important to the meanings of a hard news text may not be contained in the lead.

Text 49 criticizes the State Board of Education of Kansas for removing evolution from the curriculum. In this text *remove evolution* (emboldened) occurs three times, found in sentences 4, 9 and 14 given below.

(4) The State Board of Education has **removed evolution** from the list of matters that public school biology students must know.

(9) **Voting it away** is akin to removing math from physics.

[15] In summary (a) *repair mission* is not suitable for unpacking because criterion 2 in Figure 3.15 is not satisfied, that is, *repair* does not function as Thing (i.e. the head noun) in this nominal group.

(14) Unable to force their contrary pseudo-science of creationism into public curricula _ the courts quickly saw through that dodge for slipping sectarian religion into public classrooms _the fundamentalists are settling for **killing evolution off** by political fiat.

The summaries of the text are given below:

a) Kansas **removes evolution** as required biology study.
b) Kansas **removes evolution** as something public school students must know.
c) Board decision on evolution criticized. Blamed on activist reactionary groups.

The recurrent nucleus *remove evolution* in the text is important to the text's meanings because it is repeated by *removes evolution* (emboldened) in summaries (a) and (b). But this recurrent nucleus does not occur in the lead of text 49, given below:

Stop the world. Kansas wants to get off. In fact, it already has one foot dragging.

Once again, text 49 shows that information that is important to the meaning of a hard news text may not be contained in the lead.

4.3.4 The probability of presence in the leads of the recurrent nuclei that are important to texts' meanings

In section 4.3.1 we have seen that all but the second way of distribution in texts of the most frequent nuclei that are important to their texts' meanings involve their presence in the leads. The discussions in section 4.3.2 imply that when two or more recurrent nuclei in a text are thought of as important to the meanings of the text, each of these nuclei often occur in the lead, and in section 4.3.3 we have seen that approximately 93% of the less frequent recurrent nuclei that are important to their texts' meanings are, once again, present in the leads. These observations led me to find out the probabilities of presence in the leads of the recurrent nuclei that are important to the meaning of the texts in which they occur.

Table 4.1 below represents the probabilities in the different categories. In order to save space, 'important recurrent nuclei' in the table and the paragraph following the table is the shorthand for 'the recurrent nuclei in texts that are important to the meaning of the texts in which they occur', as defined by their occurrence in two or more summaries.

Category	the number of important recurrent nuclei	the number of important recurrent nuclei that do not occur in the leads	the number of important recurrent nuclei that occur only once in the leads	the number of important recurrent nuclei that occur twice in the leads	The proportion of important recurrent nuclei that occur at least once in the leads
C1(a)	13	1	10	2	92%
C1(b)	7	0	7	0	100%
C1(c)	2	0	2	0	100%
C2(a)	43	2	37	4	95%
C2(b)	11	0	11	0	100%
C3(a)	27	2	24	1	93%
C3(b)	13	0	13	0	100%
The above seven subcategories together	116	5	104	7	96%

Table 4.1: The probabilities of presence in the leads of the nuclei that are important to their texts' meanings

We can make two observations on the table. Firstly, the last column shows that in all seven subcategories the important recurrent nuclei are often present in the leads. Secondly, the seven subcategories comprising categories 1, 2 and 3 contain 116 important recurrent nuclei, only five of which do not occur in the leads. This is in line with the claim made by Bell (1991: 183) and Van Dijk (1986; 1988: 53) that the leads very often summarize the gist of hard news texts because if the leads indeed often summarize the gist of hard news texts, they would be expected to contain information that is important to the texts' meanings. This result also shows that the sentences after the lead seldom add additional information that is important to the meaning of a hard news text.

4.3.5 Category 4

Having considered categories 1, 2 and 3 in terms of the characteristics of recurrent nuclei that are important to the meaning of the texts in which they occur, let us consider the final category, category 4, in which none of the recurrent nuclei in a text are thought of as important to the text's meaning.

The summaries of the texts in this category were assessed with a view to determining whether it is the qualities of the summaries that resulted in the fact that none of the recurrent nuclei in each text contained in this category are important to the text's meaning.

Nine mature native speakers of English evaluated the summaries of the three texts in this category, namely texts 78, 79 and 80, on a five-point scale: 1 means very poor, 2 means poor, 3 means acceptable, 4 means good, 5 means very good. The informants read the three texts, and then judged their summaries. Each informant judged only one summary of each text. Therefore, each summary was assessed by three informants. The mean scores for the summaries of the three texts are given in Table 4.2 below.

Summary	Mean score
78(a)	4.3
78(b)	4
78(c)	4
79(a)	4.3
79(b)	3
79(c)	4
80(a)	4.3

80(b)	4.3
80(c)	4.3

Table 4.2: The mean scores assigned to the summaries of texts 78, 79 and 80

The results show that all but summary 79(b) were thought good or better, and summary 79(b) was thought acceptable, indicating that it is not the qualities of the summaries that result in the fact that none of the recurrent nuclei in each text contained in category 4 are thought of as important to the text's meaning. Therefore, we turn our attention to the three original texts, the summaries of which were assessed.

Text 78 is as follows:

BEIJING – (1) A new circular by Fujian's provincial education department on Tuesday has targeted academic plagiarism by college teachers, amid increasing worries over the practice.

(2) College teachers in Fujian may also be dismissed if they spread misinformation against the country's laws and regulations to mislead students, the circular said.

(3) An increasing number of teachers in universities in China are turning to the Internet or other academics' research to advance their own careers.

(4) Shen Yang, a professor at Wuhan University who released a research paper in 2009, said the country lacks an effective thesis supervision system and the convenience brought by the Internet drives the booming ghostwriting market.

(5) His study shows there were more than 1.1 million full-time teachers in universities and colleges across the country in 2007.

(6) They had to publish more than half a million theses within two years in nearly 1,800 important periodicals to keep their positions.

(7) Other banned practices include teachers abusing their power for personal benefit and teachers acting fraudulently on student enrolment, assessment and exams.

(8) The circular also emphasized that teachers will lose out on promotion opportunities and pay rises if they are irresponsible in students' safety or induce students to participate in any "illegal or superstitious activities".

(9) It said teachers were not allowed to use "physical punishment on students or insult them".

(10) Violators will have any academic award and honor canceled, and will not be able to apply for new research projects for specified periods.

This text contains only one recurrent nucleus: *the circular said*, which occurs three times, namely *the circular said* in sentence 2, *the circular emphasized* in sentence 8, and *it* (referring to *the circular*) *said* in sentence 9.

The summaries of the text are:

a) Circular published to punish errant teachers in Fujian.
b) Fujian published new circular to punish misbehaving teachers.
c) New circular by Fujian aims to punish misbehaving teachers.

The recurrent nucleus *the circular said* in the text is not important to the text's meanings because it is not repeated in any summary.

Both Van Dijk and Kintsch (1983: 190) and Van Dijk (1988: 116) point out that deletion, generalization and construction are frequently used in summarization. Van Dijk (1988: 116) notes:

> Deletion applies to local information that is not further used, as a presupposition, for the interpretation of the rest of the text. Generalization occurs when similar properties are relevant for different actors or situations or when a given property can be applied to different members of a set. Construction requires the combination of several partial acts or events into an overall macroact or macroevent.

The construction strategy was apparently used by my summarizers to write summaries for text 78.

College teachers in Fujian may also be dismissed in sentence 2, *teachers will lose out on promotion opportunities and pay rises* in sentence 8, and the entire sentence 10 together are summarized by *punish teachers* contained in all three summaries.

Academic plagiarism by college teachers in sentence 1, *they (referring to college teachers) spread misinformation against the country's laws and regulations to mislead students* in sentence 2, *abusing their power for personal benefit and teachers acting fraudulently on student enrolment, assessment and exams* in sentence 7, *they* (also referring to *college teachers*) *are irresponsible in students' safety or induce students to participate in any "illegal or superstitious activities* in sentence 8, and *use "physical punishment on students or insult them"* in sentence 9 together are very succinctly summarized by *errant* in summary (a) and by *misbehaving* in summaries (b) and (c).

The other two texts in this category show that the information that is important to a text's meanings may not be recurrent in the text. This is illustrated with text 80 given below in full.

(1) Sixty-one police were injured in last Friday's riot in Lhasa, six of them seriously, *said Qiangba Pungcog, chairman of the Tibet Autonomous Regional Government*, at a press briefing in Beijing on Monday.

(2) Rioters **attacked** public security **personnel** in an extremely cruel manner while they were maintaining order and refrained from using weapons, *said the chairman.*

(3) "For instance, the rioters **beat a** patrolling police **officer** until he got into coma, and rioters cut out a piece of flesh, as big as a fist, from his buttock," *he said.*

(4) Some rioters **stoned** patrol **officers** around 11:00 a.m. on March 14 at the Ramogia Monastery in Lhasa. 5 They then gathered around the Bargor street, chanting "independence" slogans and going on a destruction rampage.

(6) The violence accelerated quickly. 7 The mob set fire to shops, vehicles and pedestrians that passed by. 8 Schools, banks, hospitals and communication facilities were also their targets, *Qiangba Pungcog said.*

(9) *He stressed* that throughout the incident, both public security personnel and armed police showed great professional spirit and adherence to law. 10 "They didn't carry or use any lethal weapons," *he added.*

This text contains three recurrent nuclei, *Qiangba Pungcog said* (italicised, occurring six times), *attacked officers* (emboldened, occurring three times), and *use weapons* (underlined, occurring twice).

The summaries of text 80 are given below:

a) 61 police were injured, six seriously, in Lhasa riot.
b) 61 police were injured in Lhasa riot, six of them seriously.
c) 61 police were injured, six seriously due to rioters attack.

None of the three recurrent nuclei in the text can be regarded as important to the text's meaning because none of them are repeated in two or more summaries.

On the other hand, *sixty-one police were injured*, which occurs in the lead, is important to the text's meaning because it is repeated by *61 police were injured* in all three summaries. But *sixty-one police were injured* is not recurrent in the text.

4.4 Conclusions

This chapter has given my results with regard to my first research question, namely which recurrent nuclei in a text are thought of by readers as important to the text's meaning. The results of the question are given in Figure 4.1, which classifies the replication in summaries of nuclei that recur in texts into different categories so that we can make a generalization of the results. The figure shows that C3(a) contains a larger number of texts than any other category. Therefore, in approximately one third of my data the most frequent nucleus is not replicated but one of the less frequent recurrent nuclei is replicated in the summaries of a text. C2(a) contains the second largest number of my data. Therefore, in approximately one fifth of my data not only the nucleus that is most frequent but also one or more of the less frequent nuclei are replicated in the summaries of their text. C1(a) is the third largest category, containing approximately 16% of my data. Therefore, in less than one fifth of my data, only

the recurrent nucleus that is more frequent than any other nuclei is replicated in the summaries of its text. Each of the other categories contains less than 10% of my data.

Two main findings emerge from my results given in Figure 4.1. The first finding is that more often than not one or more recurrent nuclei in a hard news text are thought of by readers as important to the meaning of the text. The implication of this finding is that one or more of the chain interactions in a hard news text between processes that are similar in meaning and their Mediums are often important to the text's meanings. Hasan's study reported in Hasan (1984) and Halliday and Hasan (1985) concerning the relationship between cohesion in texts and their coherence show that the interactions of cohesive chains in a text contribute to the coherence of the text. My finding noted above shows that the interactions of chains between processes that are similar in meaning and their Mediums also contribute to the meanings of the text.

The second finding that emerges from the observation of the results given in Figure 4.1 is that the frequency hierarchy of the recurrent nuclei in a hard news text does not equate to the degree of importance of the nuclei to the text's meaning. This is evidenced by the texts contained in C3(a) in particular, in which the most frequent nucleus is not replicated but one of the less frequent recurrent nuclei is replicated in the summaries of the text.

Having discussed in section 4.2 the results given in Figure 4.1 with regard to the first research question, the section that follows, namely section 4.3, explores the characteristics of the recurrent nuclei that are important to the meanings of the texts in which they occur. Subsection 4.3.1 has discussed four ways of distribution in my data of the most frequent nuclei that are important to the meanings of the texts in which they occur. A majority of the

most frequent nuclei considered as important to the meanings of their texts are distributed in such a way: at least one occurrence is present in the lead and the remaining occurrences recur after the lead at intervals. White (1997) characterizes hard news texts in terms of repetitions in the paragraphs after the leads of the points of newsworthy impact set out in the headlines and leads. Therefore, repeating in the paragraphs after the leads the propositions in leads that are important to the meaning of the texts in which they occur is a popular way of writing hard news texts.

In subsection 4.3.2, we have seen that when two or more recurrent nuclei in a text are replicated in the summaries of the text, the relationship in the lead that links them is also often replicated in the summaries. This provides empirical evidence for the claim made by many linguists, namely the relations holding between the propositions in a text are important to the meaning of the text. The finding also shows that replication in summaries of the linkages that occur in leads is a strategy that is often used by readers to write summaries for hard news texts. We will consider in sections 5.5.1, 5.5.2 and 5.5.3(a) of the next chapter more instances in which readers use this strategy to write summaries for hard news texts.

Subsection 4.3.3 has examined two recurrent nuclei that do not occur in the leads but are however important to the meanings of the texts in which they occur. The discussions of the two nuclei show that information that is important to the meaning of a hard news text may not be contained in the lead.

Subsection 4.3.4 has explored the probabilities of presence in the leads of recurrent nuclei that are important to the meaning of the texts in which they occur. Two findings emerged. The first is that no matter which category defined in section 4.2 a text belongs to, the

recurrent nuclei that are important to the meanings of the texts in which they occur have a strong tendency to occur in the leads. This is in line with the claim made by Bell (1991: 183) and Van Dijk (1988: 53) that the leads very often summarize the gist of hard news texts. The second finding is that the sentences after the lead seldom add additional information that is important to the meaning of a hard news text.

Section 4.4 has discussed the three texts contained in category 4. The discussion of text 78 shows that the summary writers used the construction strategy to summarize the gist of this text. On the other hand, the discussion of the other two texts show that information that is important to a text's meaning may not be recurrent in the text.

The next chapter, chapter 5, extends the study reported in section 4.3.4 of this chapter. In chapter 5, we will consider the distribution in the leads of the recurrent nuclei that are considered as important to the meanings of the texts in which they occur.

Chapter 5 Research Question 2 and Findings

5.1 Introduction

This chapter extends the study reported in section 4.3.4 of the previous chapter. It consists of two parts. The first part, consisting of sections 5.2, 5.3 and 5.4, investigates whether the occurrences in the leads of recurrent nuclei that are important to the meanings of the texts in which they occur tend to be found in certain type(s) of clause. The second part, section 5.5, examines the occurrences in the leads of the recurrent nuclei that are important to the meanings of the texts in which they occur but are found in minor types of clause.

5.2 The definitions of type A and type B nuclei

In the course of my analysis I noticed that some nuclei contained in the leads were thought of by readers as important to the meanings of the texts in which they occur, while others also contained in the leads were not, leading me to make a distinction between type A and type B nuclei.

The two types of nuclei are explained with reference to the lead of text 60 given below, which reports on the impacts of Typhoon Babs on some areas in Asia.

TAIPEI, Taiwan (AP) _ Typhoon Babs **brought torrential rains** and landslides to Taiwan and lashed Hong Kong with strong winds Sunday after *killing at least 156 people* in the Philippines and leaving hundreds of thousands homeless.

In the lead *brought torrential rains* (emboldened) and *killing at least 156 people* (italicised) are the occurrences in the lead of two recurrent nuclei in the text. Henceforth, the occurrence in the lead of a recurrent nucleus in a text is referred to as its initial occurrence.

The summaries of the text are given below. The nuclei contained in the summaries are given in the square brackets under the summaries.

a) Typhoon Babs floods Taiwan, lashes Hong Kong, and heads toward China.

[floods Taiwan, lashes Hong Kong, Typhoon Babs heads]

b) Babs brought Taiwan torrential rains, Hong Kong winds.

[brought torrential rains, brought winds]

c) Babs brings strong winds and torrential rains to Taiwan and Hong Kong.

[brings strong winds, brings torrential rains]

Brought torrential rains appearing in the lead is repeated by *brought torrential rains* contained in summary (b), and by *brings torrential rains* contained in summary (c), indicating that *brought torrential rains* in the lead is important to the text's meaning.

On the other hand, *killing at least 156 people* also appearing in the lead is not repeated in any summary, indicating that it is not important to the text's meaning.

I refer to *brought torrential rains* as the initial occurrence of a type A nucleus in the text, and to *killing at least 156 people* as the initial occurrence of a type B nucleus in the text. The definitions of the two types of nuclei are given below.

A type-A nucleus is a recurrent nucleus in a hard news text that occurs at least once in the lead and is repeated in two or more summaries of the text. This type of nucleus not only is present in the lead, which represents the text's 'anchor point' or 'textual centre of gravity' (White, 1997: 406), but also is important to the meaning of the text in which it occurs due to its replications in the summaries.

A type B nucleus, on the other hand, is a recurrent nucleus in a hard news text that also occurs at least once in the lead but is repeated in less than two summaries of the text. Though also present in the leads, this type of nucleus would not appear to be thought of as important to the meaning of the text in which it occurs because it is repeated in less than half of the summaries of the text.

The procedures for identifying type A and type B nuclei in a hard news text are shown in Figure 5.1 below.

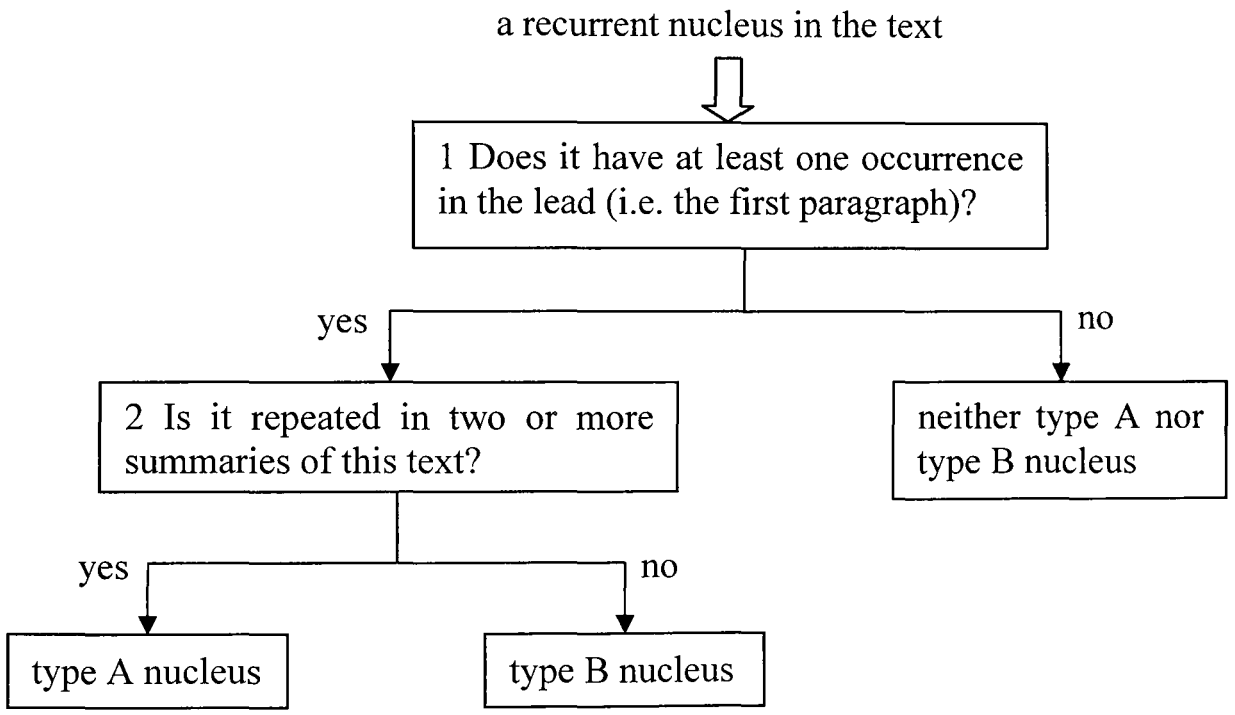


Figure 5.1: The procedures for identifying type A and type B nuclei in a hard news text

5.3 Research question 2 and the method of exploring the question

Having identified type A and type B nuclei in each of the 80 texts I analyzed, I further classified these two types of nuclei according to process types (Halliday, 1994). Each nucleus was characterized in terms of the process type to which the nucleus belonged. For instance, in text 60 *brought torrential rains* is a type A material nucleus because *brought* is a material process; *killed people* is a type B material nucleus because *killed* is also a material process.

Table 5.1 below gives the number of the initial occurrences of type A and type B nuclei contained in the leads of my 80 texts, classified according to process.

	the number of the initial occurrences of type A nuclei contained in the leads of my 80 texts	the percentages of type A nuclei in the leads, classified according to process	the number of the initial occurrences of type B nuclei contained in the leads of my 80 texts	the percentages of type B nuclei in the leads, classified according to process
material	86	74%	43	43%
verbal	15	13%	48	48%
mental	9	8%	3	3%
relational	3	3%	4	4%
causative	3	3%	0	0
existential	0	0	1	1%
in total	116	100%	99	100%

Table 5.1: The number of the initial occurrences of type A and type B nuclei in the 80 leads, classified according to process

We can make three observations on table 5.1. Firstly, the nuclei contained in the leads are mainly material and verbal nuclei. The leads do not often contain nuclei other than these two types.

Secondly, approximately 74% of the initial occurrences of type A nuclei are material, indicating that the nuclei contained in the leads that are thought of by readers as important to the meanings of the texts in which they occur have a strong tendency to be material nuclei.

Thirdly, there are 63 verbal nuclei in the leads, 48 of them (approximately 76%) are not thought of as important to the meanings of the texts in which they occur, indicating that more often than not readers are interested in the news events, rather than the sources expressed by the verbal nuclei from which the news events were obtained.

Given that the initial occurrences of type A nuclei have a strong tendency to be material nuclei, and that one of the two major types of type B nuclei in the leads are material nuclei, it is therefore worth investigating **whether the initial occurrences of type A and type B material nuclei tend to appear in different types of clause?**

Four types of clause, namely primary, secondary, independent and embedded clauses, are considered. A primary clause is the main clause when two clauses stand in a hypotactic relation, i.e. an unequal relation 'between a dependent element and its dominant, the element on which it is dependent' (Halliday and Matthiessen, 2004: 374-375). A secondary clause is the clause that is subordinated to the primary clause. For instance, the lead of text 6 given below consists of two clauses in a hypotactic relation. *'The Palestinians on Tuesday inaugurated Gaza International Airport, their first gateway to the world, with cheers, tears and an outpouring of patriotism'* is the primary clause, while *'taking a major step toward Statehood'* is the secondary clause.

DAHANIEH, Gaza Strip (AP) _ Taking a major step toward Statehood, the Palestinians on Tuesday inaugurated Gaza International Airport, their first gateway to the world, with cheers, tears and an outpouring of patriotism.

On the other hand, when two or more clauses stand in a paratactic relation, each of these clauses is an independent clause. Parataxis is 'the relation between two like elements of equal status' (Halliday and Matthiessen, 2004: 374-375). For instance, the lead of text 5 given below consists of two clauses in a paratactic relation: *'A knife-wielding villager stabbed two women to death'* and *'injured seven other women and children in Northwest China's Shaanxi province on Monday'*. These two clauses are coordinated by *and*. Both are independent clauses.

BEIJING – A knife-wielding villager stabbed two women to death and injured seven other women and children in Northwest China's Shaanxi province on Monday.

An embedded clause post-modifies the head noun in a nominal group. 'Embedding is a semogenic mechanism whereby a clause or phrase comes to function as a constituent within the structure of a group, which itself is a constituent of a clause' (Halliday and Matthiessen, 2004: 426). For instance, in the lead of text 8 given below the stretch included in double square brackets is an embedded clause complex post-modifying *'heavy winds and rain'*.

Typhoon Babs headed toward southern China Friday after battering the central and northern Philippines with heavy winds and rain [[that killed at least 82 people, flattened crops, and forced more than 100,000 people to flee their homes]].

Having explained the four types of clause, I will explain the procedures I followed in my analysis. I firstly classified the leads of my 80 texts into two groups according to the criteria given in Figure 5.2 below.

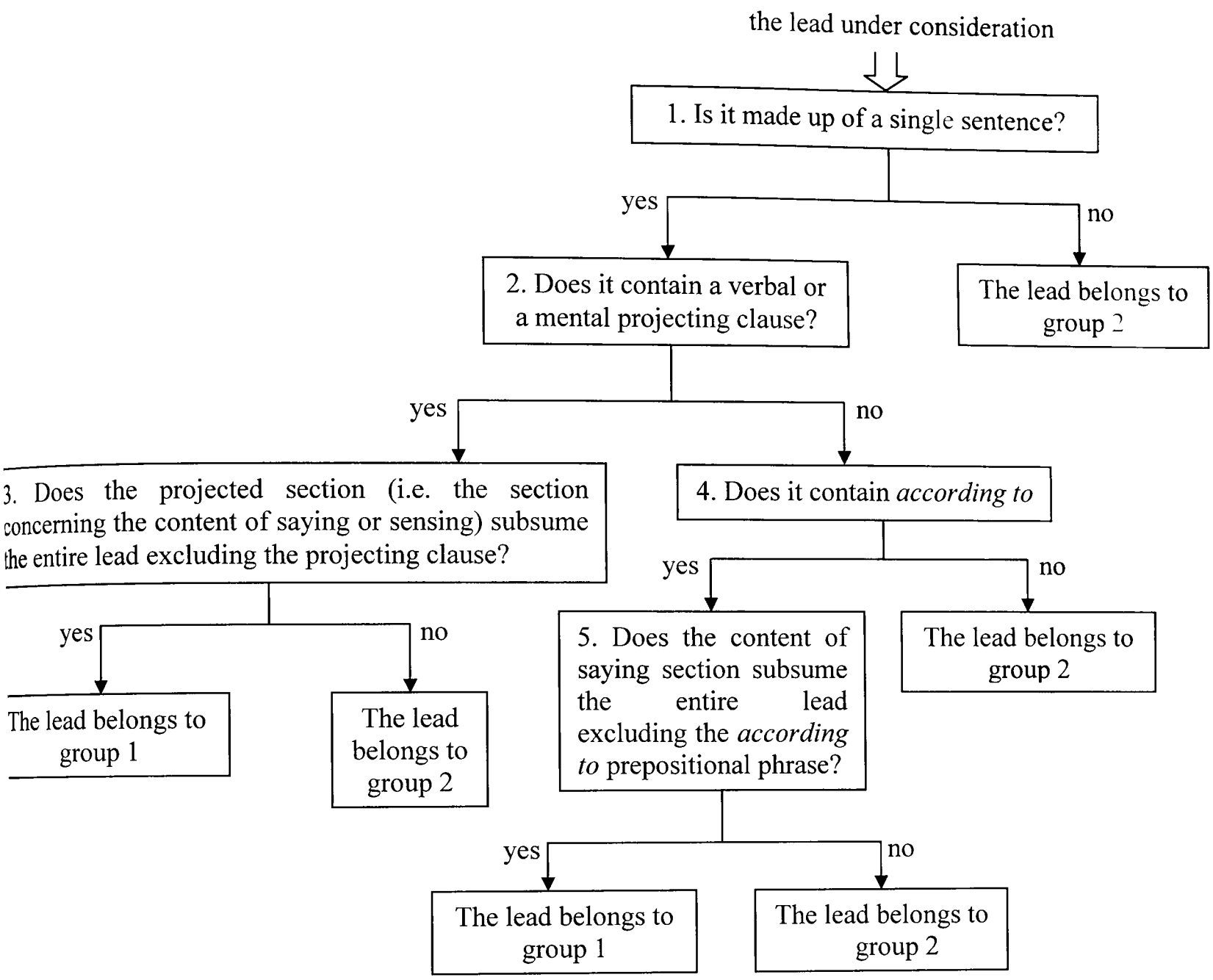


Figure 5.2 The criteria for classifying the 80 leads

Group 1 contains 42 leads. 34 of these leads contain a verbal projecting clause in the final position, e.g.

China will start building a center at the end of May to train giant pandas born in captivity to live in the wild, experts said on Wednesday. (the lead of text 73)

On the other hand, five of the leads contain a verbal projecting clause in the initial position. An example is the lead of text 10 given below.

An informed source in Bern confirmed today, Thursday, that Barazan Al-Takriti, Iraqi President Saddam Hussein's half-brother and former Ambassador to the United Nations, is returning to Baghdad at the end of this month.

Each of the other three leads contains an *according to* prepositional phrase in the final position[15], e.g.

NAIROBI, Kenya _ Nine months before the attack on the American Embassy here, U.S. intelligence officials received a detailed warning that Islamic radicals were plotting to blow up the building, according to Kenyan and American officials. (the lead of text 32)

Group 2 contains 38 leads. 24 of these leads do not contain any verbal clauses or *according to* prepositional phrases. An example is the lead of text 8 given earlier.

Another seven of the leads contain a verbal clause that does not project, e.g.

MARIJA BISTRICA, Croatia (AP) _ Pope John Paul II on Saturday honored Croatia's World War II cardinal, a hero to Roman Catholics, but long a symbol of divisions in the Balkans. (the lead of text 7)

In this lead *'honored'* is a verbal process, but it does not project.

Another five of the leads contain a verbal projecting clause but the projected clause does not subsume the entire lead excluding the verbal projecting clause, e.g.

The militant Palestinian movement Islamic Holy War said Saturday that it attacked the Jerusalem market on Friday, which prompted arrests by the Palestinian Authority overnight. (the lead of text 37)

In this lead the verbal clause *'The militant Palestinian movement Islamic Holy War said Saturday'* projects a *that* clause, namely *that it attacked the Jerusalem market on Friday*. But this *that* clause does not subsume the stretch after the comma, namely *which prompted arrests by the Palestinian Authority overnight*.

Each of the other two leads contains more than one sentence, e.g.

[15] This is in line with Hoey and O'Donnell's (2007) finding that when *according to* occurs in the leads, it has a strong tendency to occur at the end.

Zhao Zhenshang in central China's Henan province has been jailed for ten years, accused of "murdering" his neighbor Zhao Zuohai. However, the "victim" reappears recently, u-turning an old-time case that lays bare custody torture by police officers and sends public confidence on local authorities into a tailspin. (the lead of text 29)

Having classified the leads into the two groups, I recorded the types of clause in which the initial occurrences of type A and type B material nuclei are present when **the 'content of saying' sections** in the leads belonging to group 1 were analyzed at the most general level.

The 34 leads that contain verbal projecting clauses in the final position are referred to by Halliday and Matthiessen (2004: 465) as 'free indirect speech'. Halliday and Matthiessen (2004: 465) say that in free indirect speech the projected sections stand in paratactic (equal) relations with the verbal projecting clauses. Therefore, the two authors analyze the projected sections and the verbal projecting clauses as two independent clauses[16] (Halliday and Matthiessen, 2004: 466). Consequently the initial occurrence of every type A or type B material nucleus that is found in a non-embedded clause appears as part of an independent clause. Figure 5.3 below illustrates this.

projected section (independent clause complex)	projecting clause (independent clause)
China will start **building a center** at the end of May to **train giant pandas** born in captivity to (∅ *giant pandas*) *live* in the wild,	experts said on Wednesday.

Figure 5.3: First level of analysis of the lead of text 73

Text 73 contains two type A material nuclei, the initial occurrences of which are *'building a center'* and *'train giant pandas'* (both are emboldened), and one type B material nucleus, the

[16] Henceforth, each reference in this chapter to a clause includes reference to a clause complex.

initial occurrence of which is *'giant pandas live'* (italicised) (The Medium of *'giant pandas live'*, namely *'giant pandas'*, needs inferring from the preceding clause.)

Doing the analysis in this way is not, however, informative about the type of clause in which the initial occurrences of type A and type B material nuclei tend to occur. A second level of analysis is therefore necessary. Figure 5.4 shows such an analysis applied to the lead of text 73.

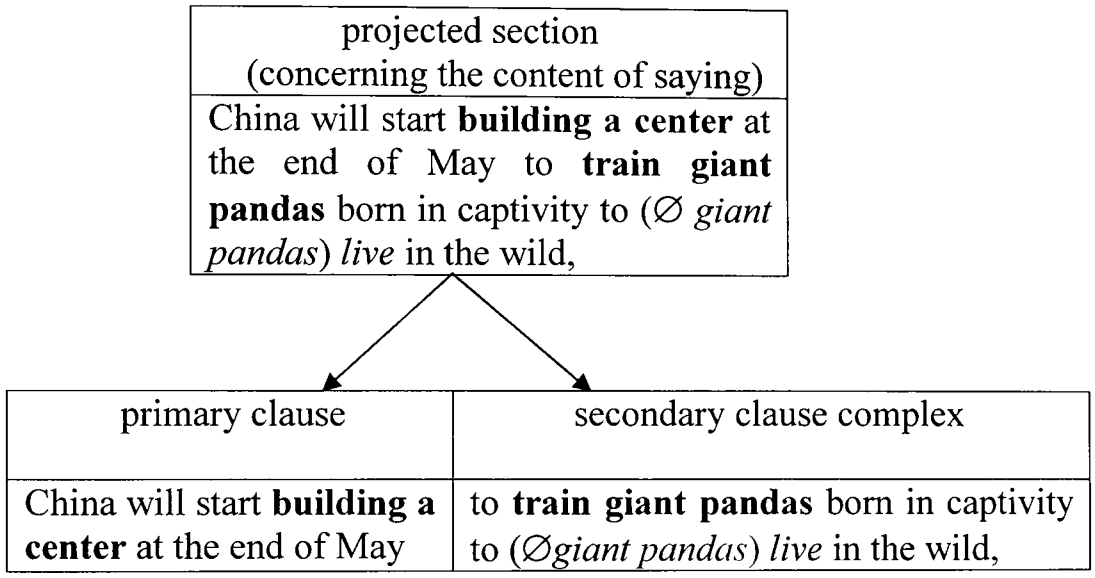


Figure 5.4 A second level of analysis of the lead of text 73

Figure 5.4 shows that *building a center* is present in the primary clause, whereas both *train giant pandas* and *giant pandas live* are present in the secondary clause complex.

The five leads that contain verbal projecting clauses in the initial position are referred to by Halliday and Matthiessen as 'indirect speech' (2004: 466). In indirect speech, the verbal projecting clause is the primary clause, and the projected section is the secondary clause (or clause complex). Consequently the initial occurrence of every type A or type B material nucleus that is found in a non-embedded clause appears as part of a secondary clause. Figure 5.5 below illustrates this.

projecting clause (primary clause)	projected section (secondary clause complex)
Secretary-General Kofi Annan said Wednesday	that **he may travel** to Libya next week in hopes of **closing a deal** [[to *try two Libyan suspects* in the Pan Am Lockerbie bombing]].

Figure 5.5: First level analysis of the lead of text 74

Text 74 contains two type A material nuclei, the initial occurrences of which are *he may travel* and *closing a deal* (both are emboldened), and one type B material nucleus, the initial occurrence of which is *try two Libyan suspects* (italicised). Figure 5.5 shows that after the lead of text 74 is analyzed at the most general level, both *he may travel* and *closing a deal* are present in the secondary clause complex, and *try two Libyan suspects* appears as part of the embedded clause (put in double square brackets) that post-modifies *deal*.

Doing the analysis in this way is, however, also not informative of the types of clause in which the initial occurrences of type A material nuclei and the initial occurrences of type B material nuclei tend to occur. A second level of analysis is therefore also necessary. Figure 5.6 below shows such an analysis applied to the lead of text 74.

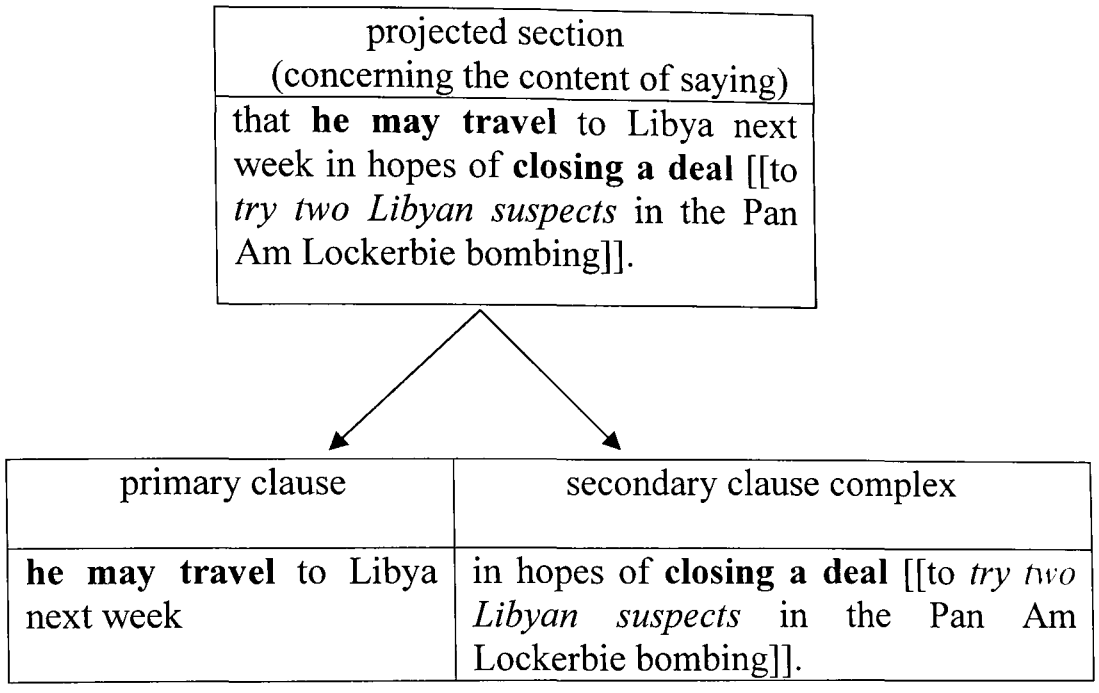


Figure 5.6 A second level analysis of the lead of text 74

Figure 5.6 shows that after the projected section of the lead is analyzed, *he may travel* is present in the primary clause, *closing a deal* is present in the secondary clause complex, while *try two Libyan suspects* is present in the embedded clause (marked as [[…]]).

According to X also serves to attribute locutions to their sources. In each of the three leads that contain *according to X,* the content of saying section (i.e. the entire lead excluding *according to X*) was analyzed at the most general level so as to be in line with the two sets of analyses described above. For instance, the 'content of saying' section in the lead of text 32 given below, namely *Nine months before the attack on the American Embassy here, U.S. intelligence officials received a detailed warning that Islamic radicals were plotting to blow up the building*, is an independent clause complex, with *warning* being post-modified by an embedded clause complex (marked as [[…]]).

NAIROBI, Kenya _ Nine months before **the attack on the American Embassy here**, U.S. intelligence officials *received a* detailed *warning* [[that Islamic radicals were plotting to **blow up the building**]], according to Kenyan and American officials.

This text contains two type A material nuclei: *attacked the American Embassy in Nairobi* and *received a warning*. The former has two occurrences (emboldened) in the lead, *the attack on the American Embassy here* and *blow up the building*. The latter has only one occurrence in the lead, namely *received a warning* (italicised). Both *the attack on the American Embassy here* and *received a warning* are present in the independent clause complex, while *blow up the building* is present in the embedded clause complex.

In group 2 the 36 leads that consist of single sentences were analyzed at the most general level. Figure 5.7 below illustrates such an analysis with the lead of text 37.

primary clause complex	secondary clause
The militant Palestinian movement Islamic Holy War said Saturday that it **attacked the Jerusalem market** on Friday,	which prompted *arrests by the Palestinian Authority overnight*.

Figure 5.7: First level analysis of the lead of text 37

This text contains one type A material nucleus, the initial occurrence of which is *attacked the Jerusalem market* (emboldened), and one type B material nucleus, the initial occurrence of which is *arrests by the Palestinian Authority overnight* (italicised). *Attacked the Jerusalem market* is present in the primary clause complex, whereas *arrests by the Palestinian Authority overnight* is present in the secondary clause.

In each of the two leads that consist of more than one sentence, the sentence that contains the initial occurrence of a type A or type B material nucleus was analyzed at the most general level. Such an analysis is illustrated with the lead of text 29 below.

(1) {**Zhao Zhenshang** in central China's Henan province **has been jailed** for ten years}. <accused of *"murdering" his neighbor Zhao Zuohai*>. (2) {However, the "victim" reappears

recently}, <u-turning an old-time case [[that lays bare custody torture by police officers and sends public confidence on local authorities into a tailspin]]>.

This text contains three type A material nuclei, the initial occurrences of which are *Zhao Zhenshang has been jailed* (emboldened), *"murdering" his neighbor Zhao Zuohai* (italicised), and *the "victim" reappears* (underlined), and one type B material nucleus, the initial occurrence of which is *custody torture by police officers* (underlined by a wavy line).

Both sentences were analyzed at the most general level. The primary clause is marked as {...}, the secondary clause is marked as <...>, and the embedded clause is marked as [[...]]. In sentence 1 *Zhao Zhenshang has been jailed* is present in the primary clause, while *"murdering" his neighbor Zhao Zuohai* is present in the secondary clause complex. In sentence 2 *the "victim" reappears* is present in the primary clause, while *custody torture by police officers* is present in the embedded clause complex that post-modifies *an old-time case*.

V/VP[17] followed by '-ing'/infinitive structure (e.g. *start building a center*) was analyzed following Francis *et al* (1996). When the first verb stands in a 'phase' relation with the verb contained in the '-ing' or infinitive structure, the two verbs constitute a complex verb group, meaning that they occur in a single clause. Francis *et al* (1996: xxi) noted that 'when two verbs are in phase, they together form a complex verb group. This means that the actions or states expressed by the two verbs cannot be separated from each other'. For instance, *'China will start building a center at the end of May'* contained in the lead of text 73 was analyzed as a single clause (see Figure 5.4 given earlier) because *'start'* stands in a phase relation with the *'-ing'* form of a verb when the meaning of *start* belongs to the 'start' and 'stop' group

[17] VP means a verb plus a particle.

(verbs in this group are concerned with starting, stopping, or continuing doing an action) (Francis *et al*, 1996: 81).

On the other hand, when the first verb stands in a 'verb with object' relation with the verb contained in the '-ing' or infinitive structure, the two verbs occur in two clauses. Francis *et al* (1996: 93) note that when two verbs stand in a 'verb with object' relation, they occur in two clauses because they 'express two closely-related but separate actions or states'. The '-ing' or infinitive clause is treated as the object of the first verb.

We can illustrate how the analysis works with the lead of text 22 given below. Text 22 contains one type A material nucleus, the initial occurrence of which is e*vacuating 330,000 people* (emboldened), and one type B material nucleus, the initial occurrence of which is *safeguard cities* (italicised).

China was **evacuating 330,000 people** Friday from land along the raging Yangtze River [[that officials were preparing to sacrifice to flooding to *safeguard cities* downstream]].

This lead is an independent clause complex, with *'land along the raging Yangtze River'* post-modified by an embedded clause complex (as before marked as [[...]]). *Evacuating 330,000 people* is part of the independent clause complex. On the other hand, *safeguard cities* is contained in the embedded clause complex.

'Preparing to sacrifice' contained in the lead is an instance of V followed by infinitive structure. Francis *et al* (1996) consider that *prepare* stands in a 'verb with object' relation with the verb contained in the to-infinitive when the meaning of *prepare* belongs to the 'promise' group (verbs in this group are concerned with being committed to a future action).

The meaning of *preparing* contained in '*preparing to sacrifice*' belongs to this group. meaning that *preparing* and *sacrifice* are present in two clauses.

Francis *et al* (1996) also consider that two verbs occur in two clauses when they stand in a 'verb with adjunct' relation. But instances of this kind are not found in the 80 leads I analyzed, and therefore we do not need to consider this possibility further.

5.4 The results with regard to research question 2

After analyzing the initial occurrences of type A and type B material nuclei in the ways described above, the results are given in tables 5.2 and 5.3 below.

Table 5.2 shows the distributions of the initial occurrences of type A and type B material nuclei in the 'content of saying' section of the leads in group 1. Henceforth, P means primary clause, I means independent clause, S means secondary clause and E means embedded clause. As before, each reference to a clause includes reference to a clause complex.

	P+I	S+E
the initial occurrences of type A material nuclei	39	14
the initial occurrences of type B material nuclei	7	12

Table 5.2 The distributions of the initial occurrences of type A and type B material nuclei in the 'content of saying' section of the leads in group 1

We can make four observations on the table. Firstly, the table shows that the initial occurrences of type A material nuclei found in primary and independent clauses are more common than those found in secondary and embedded clauses. Primary and independent

clauses are major clause types because they are neither subordinated to any other clause nor 'relegated' to modifying elements in nominal groups; on the other hand, secondary and embedded clauses are minor clause types in that secondary clauses are subordinated to primary clauses, and embedded clauses are 'relegated' to being modifying elements in nominal groups. Since type A nuclei are important to the meaning of the texts in which they occur, the implication of this result is that information that is important to a text's meaning is not only often elevated to the leads (Bell, 1991; White, 1997; Van Dijk, 1988), but also tends to be foregrounded in major types of clause contained in the 'content of saying' section of the leads in group 1.

The second observation arises out of the results in the third row of the table. The leads in group 1 contain 19 initial occurrences of type B material nuclei in total, 12 of which (approximately 63%) are present in secondary and embedded clauses. Since type B nuclei are not important to the meaning of the texts in which they occur, this result indicates that the information in the 'content of saying' section that is not important to a text's meaning tends to be backgrounded in minor types of clause.

The third observation arises out of the results in the second column of the table. Approximately 85% of the initial occurrences of material recurrent nuclei found in primary and independent clauses are occurrences of type A nuclei, whereas only approximately 15% of the initial occurrences of material recurrent nuclei found in these two types of clause are occurrences of type B nuclei. The implication is that readers seem to have a strong tendency to consider information given in major types of clause in the 'content of saying' section of leads as important to the meanings of the texts they appear in.

The fourth observation concerns the results given in the third column of the table. The number of initial occurrences of type A material nuclei found in secondary and embedded clauses is roughly even with that of initial occurrences of type B material nuclei found in these two types of clause. This is unexpected because I anticipate that readers may consider information given in minor types of clause as unimportant to the meanings of a text. Therefore, the 14 initial occurrences of type A material nuclei found in secondary or embedded clauses are marked and deserve close examination (see section 5.5 of this chapter).

Table 5.3 given below shows the distributions of the initial occurrences of type A and type B material nuclei in the leads in group 2.

	P+I	S+E
the initial occurrences of type A material nuclei	24	9
the initial occurrences of type B material nuclei	3	21

Table 5.3 The distributions of the initial occurrences of type A and type B material nuclei in the leads in group 2

We can analyze the results in this table in the same way as those in table 5.2. Table 5.3 shows that the initial occurrences of type A material nuclei found in primary and independent clauses are more than twice as frequent as those found in secondary and embedded clauses. The implication is that information that is important to a text's meaning tends to be foregrounded in major types of clause in the leads in group 2.

The third row of this table shows that the initial occurrences of type B material nuclei found in secondary and embedded clauses are seven times as frequent as those found in primary and independent clauses, indicating that information that is not important to a text's meaning has a strong tendency to be backgrounded in minor types of clause in the leads in group 2.

The second column of the table shows that approximately 89% of the initial occurrences of material recurrent nuclei found in primary or independent clauses are occurrences of type A nuclei, whereas only approximately 11% are occurrences of type B nuclei. The implication is that readers seem to have a strong tendency to consider information given in major types of clause in the leads in group 2 as important to the meaning of the texts in which they occur.

The third column of the table shows that 70% of the initial occurrences of material recurrent nuclei found in secondary or embedded clauses are occurrences of type B nuclei, while only 30% are occurrences of type A nuclei. The implication is that readers tend to consider the information given in minor types of clause in the leads in group 2 as less important to the meaning of the texts in which they occur.

Tables 5.2 and 5.3 are similar in three respects. Firstly, the second row of table 5.2 shows that approximately 74% of the initial occurrences of type A material nuclei are present in major types of clause in the 'content of saying' section of the leads in group 1. Similarly, the second row of table 5.3 shows that approximately 73% of the initial occurrences of type A material nuclei are also present in major types of clause in the leads in group 2. These two results show that the initial occurrences of type A material nuclei, which embody information that is important to the meaning of the texts in which they occur, have a strong tendency to be present in major types of clause.

Secondly, the third row of table 5.2 shows that approximately 63% of the initial occurrences of type B material nuclei are present in minor types of clause in the 'content of saying' section of the leads in group 1. Similarly, the third row of table 5.3 shows that 87.5% of the

initial occurrences of type B material nuclei are also present in minor types of clause in the leads in group 2. These two results show that the initial occurrences of type B material nuclei, which embody information that is not important to the meaning of the texts in which they occur, tend to be present in minor types of clause.

Thirdly, the second column of table 5.2 shows that approximately 85% of the initial occurrences of material recurrent nuclei found in primary or independent clauses in the 'content of saying' sections of the leads in group 1 are occurrences of type A nuclei. Similarly, the second column of table 5.3 shows that approximately 89% of the initial occurrences of material recurrent nuclei found in these two types of clause in the leads in group 2 are also occurrences of type A nuclei. These two results show that the initial occurrences of material recurrent nuclei that are found in primary or independent clauses have a strong tendency to be considered by readers as important to the meaning of the texts in which they occur.

The discussions above indicate that the 'content of saying' section of the leads in group 1 resembles the complete leads in group 2. A great majority of the leads in group 1 contain verbal clauses or prepositional adjuncts (frequently *according to*) in the final position (cf. Hoey & O'Donnell's (2007); Hoey & O'Donnell (2008)). These verbal clauses and prepositional adjuncts express attributions, indicating that the journalists are reporting on what someone else said. As pointed out by Thompson (1994: 79), had these verbal clauses and prepositional adjuncts been left out, the 'content of saying' section would look like news given directly by the journalist. Putting attributions at the end enables the news events being reported in 'the content of saying' section to be foregrounded, while incidentally reminding the reader that this is a report.

To say that the 'content of saying' section of the leads in group 1 resembles the complete leads in group 2 does not mean that the verbal clauses or prepositional adjuncts expressing attributions in the leads in group 1 are redundant. Through the use of reporting structures, journalists present their descriptions of the reported events as second-hand (Thompson, 1994: 151), telling the reader that the news events described in the 'content of saying' sections originated from other sources than the journalists themselves. This enables the journalists to avoid taking responsibility when the reports challenge influential person(s) or institution(s), or enhance the trustworthiness of the reported events when the sources are authoritative.

5.5 Examining the 'maverick' cases

Table 5.2 contains 14 initial occurrences of type A material nuclei found in secondary or embedded clauses, and table 5.3 contains 9 initial occurrences of this type of nucleus that are likewise found in these two types of clause. It is worth looking at these 23 initial occurrences of type A material nuclei closely because they were considered by my summary writers as important to the meaning of the texts in which they appear, despite occurring in minor types of clause. In what follows the patterns in which these 23 initial occurrences are found are discussed.

5.5.1 The seven initial occurrences of type A material nuclei that express causes

There are seven initial occurrences of type A material nuclei that express causes. They are *train giant pandas* in lead 1, *detain six villagers* in lead 2, *closing a deal* in lead 3, *suffocate his 17-month-old son* in lead 4, *raping a woman* and *killing a woman* in lead 5, and *rein in booming investment* in lead 6. In the leads in which they occur, each of these nuclei occurs in the secondary clause (or clause complex) that is linked to the preceding primary clause in a causal relation.

The six leads noted above are given below, and numbered for ease of reference. The texts in which the leads occur and the groups they belong to are indicated in the brackets following the leads. The summaries of the texts in which the leads occur are given under the leads. Henceforth, in the 'content of saying' section of the leads in group 1 and in the complete leads in group 2, different types of clause are marked according to the following scheme: the primary clause (or clause complex) is marked as {...}, the secondary clause (or clause complex) is marked as <...>, the independent clause (or clause complex) is marked as (...), and the embedded clause (or clause complex) is marked as [[...]]. Both the initial occurrence of the recurrent nucleus under consideration and the items repeating it in the summaries are emboldened.

The first nucleus we consider is *train giant pandas* in lead 1.

1) {China will start building a center at the end of May} *to* <**train giant pandas** [[born in captivity]] to live in the wild>, experts said on Wednesday. (the lead of text 73; group 1)

   The summaries of text 73 are:

   a) China to **train giant pandas** to survive the wild.
   b) China will build a center *to* **train pandas** to survive.
   c) A training center built in China *to* **train pandas**.

The 'content of saying' section of the lead consists of a primary clause: *China will start building a center at the end of May*, and a secondary clause complex: *train giant pandas born in captivity to live in the wild. 'Train giant pandas'* is present in the secondary clause complex.

The proposition expressed by the primary clause contained in the 'content of saying' section of the lead is summarized by *China will build a center* in summary 1(b), and by *A training center built in China* in summary 1(c). The proposition *train giant pandas* contained in the

secondary clause complex is summarized by *train giant pandas* in summary 1(a), by *train pandas* in the last two summaries. It is worth noting that the semantic relation in the lead that links *train giant pandas* to the preceding proposition expressed by the primary clause is echoed in summaries 1(b) and 1(c): in the lead these two propositions are linked by *'to'* (shorthand for *'in order to'*); in the last two summaries the propositions that repeat *train giant pandas* are also linked by this connector to the preceding propositions that summarize the proposition expressed by the primary clause in the 'content of saying' section of the lead.

The next two nuclei we consider are *detain six villagers* in lead 2 and *closing a deal* in lead 3.

2) {Police in a county of North China's Hebei province issued fake warrants} *to* <**detain six villagers** in a land dispute for bail money>, the Beijing News reported on Wednesday. (the lead of text 27; group 1)

The summaries of text 27 are:

a) Police issued fake warrants *to* **detain six villagers** for money.
b) Police used fake warrants *to* **detain six villagers** for bail.
c) Police in Hebei issued fake warrants *to* **detain six villagers**.

The 'content of saying' section of lead 2 consists of two clauses in a hypotactic relation. The primary clause is *Police in a county of North China's Hebei province issued fake warrants.* The secondary clause is *detain six villagers in a land dispute for bail money. 'Detain six villagers'* is present in the secondary clause.

The proposition expressed by the primary clause in the 'content of saying' section of the lead is summarized by *Police issued fake warrants* in summary 2(a), by *Police used fake warrants* in summary 2(b), and by *Police in Hebei issued fake warrants* in summary 2(c). The proposition *detain six villagers* contained in the secondary clause complex of the lead is repeated word for word in all three summaries.

In the lead the proposition *detain six villagers* is linked to the preceding proposition expressed by the primary clause by *'to'* (shorthand for *'in order to'*), which functions again as a causal connector expressing purpose. The summaries also link these two propositions with this connector.

3) Secretary-General Kofi Annan said Wednesday that {he may travel to Libya next week} *in hopes of* <**closing a deal** [[to try two Libyan suspects in the Pan Am Lockerbie bombing]]>. (the lead of text 74; group 1)

The summaries of text 74 are:

a) UN chief may visit Libya *to* **close Lockerbie trial deal**.
b) U.N. Secretary-General considering travel to Libya *to* **close Lockerbie trial deal**.
c) UN Chief Kofi Annan may visit Libya *hoping to* **close Lockerbie trial**.

When a clause follows a preposition (or a prepositional phrase) that functions as a conjunction, the clause can be analyzed as a rank-shifted (i.e. embedded) clause (Quirk *et al*, 1972), or as a ranking (i.e. non-embedded) clause standing in a relation to the other clause that the preposition (or the prepositional phrase) connects it to (Halliday, 1994; Halliday and Matthiessen, 2004). My present study follows Hallidayan approach.

Therefore, the 'content of saying' section of lead 3 consists of a primary clause: *he may travel to Libya next week*, and a secondary clause complex: *closing a deal to try two Libyan suspects in the Pan Am Lockerbie bombing*. *'Closing a deal'* is present in the secondary clause complex.

The proposition expressed by the primary clause in the 'content of saying' section of the lead is summarized by *UN chief may visit Libya* in summary 3(a), by *(U.N. Secretary-General) travel to Libya* in summary 3(b), and by *UN Chief Kofi Annan may visit Libya* in summary 3(c). The proposition *closing a deal* contained in the secondary clause complex of

the lead is repeated by *close Lockerbie trial deal* in summaries 3(a) and 3(b), and by *close Lockerbie trial* in summary 3(c).

In lead 3 the proposition *closing a deal* is linked to the preceding proposition *he may travel to Libya next week* by *in hopes of*, which also functions as a causal connector expressing purpose. In the first two summaries the propositions that summarize these two propositions are once again linked by *'to'* (shorthand for *'in order to'*), again functioning as a causal connector expressing purpose. In the last summary *'hoping to'* is a paraphrase of *'in hopes of'* contained in the lead. Therefore, in this summary *close Lockerbie trial* is also expressed as the purpose of UN Chief Kofi Annan's visit to Libya.

In the three examples discussed above, we can see that the relation in the lead that links the proposition expressed by the nucleus under consideration to the proposition expressed by the preceding primary clause is echoed in two or more summaries of the text in which the nucleus occurs. We will find out in the examples to be discussed below whether the relations in the leads that link the other four nuclei to the propositions expressed by the primary clauses in (the 'content of saying' section of) the leads are also echoed in two or more summaries of the texts in which they occur.

We next consider *suffocate his 17-month-old son* in lead 4.

4) {A young father has been detained} *for* <allegedly attempting to **suffocate his 17-month-old son** in a fit of anger after a quarrel with the baby's mother last week>, Chinese-language newspapers reported yesterday. (the lead of text 14; group 1)

The summaries of text 14 are:

a) Ma detained *for* failed attempt to **kill his son**.
b) Ma detained *for* attempting to **kill his son**.
c) Father detained *for* attempting to **suffocate his son**.

The 'content of saying' section of lead 4 consists of two clauses in a hypotactic relation. The primary clause is *A young father has been detained.* The secondary clause is *allegedly attempting to suffocate his 17-month-old son in a fit of anger after a quarrel with the baby's mother last week. Suffocate his 17-month-old son* is present in the secondary clause.

The proposition expressed by the primary clause in the 'content of saying' section of the lead is summarized by *Ma detained* in the first two summaries, and by *Father detained* in the last summary. The proposition *suffocate his 17-month-old son* is repeated by *kill his son* in the first two summaries, and by *suffocate his son* in the last summary.

In the lead the proposition *suffocate his 17-month-old son* is linked to the preceding proposition expressed by the primary clause by *'for'*, which also functions as a causal connector expressing reason. The three summaries also link these two propositions by this connector.

The next two nuclei we consider are *raping a woman* and *killing a woman* in lead 5.

5) {POLICE detained a security guard} *for* <allegedly raping (∅ a woman) and **killing a woman** in a factory in Bao'an District on June 3>, the Daily Sunshine reported yesterday. (the lead of text 77; group 1)

The summaries of text 77 are:

a) Security guard detained *for* the alleged raping and **killing**.
b) Security guard detained after raping (Ø a woman) and **killing a woman**.
c) Security guard detained *for* raping (Ø a woman) and **killing a woman**.

The 'content of saying' section of lead 5 consists of a primary clause: *POLICE detained a security guard*, and a secondary clause complex: *allegedly raping and killing a woman in a factory in Bao'an District on June 3.* Both *raping a woman* (underlined) and *killing a woman* (emboldened) are present in the secondary clause complex.

In summary 5(a) *raping* (as a noun) was unpacked to *raping a woman*; once it is, we can recognise a repetition relation with *raping (∅ a woman)* contained in the lead; Likewise *killing* (also as a noun) in this summary was unpacked to *killing a woman*, which is in a repetition relation with *killing a woman* contained in the lead.

In the lead *raping and killing a woman* is linked to the preceding proposition expressed by the primary clause by *'for'*, which once again functions as a causal connector expressing reason. In summaries 5(a) and 5(c) the propositions that repeat *raping and killing a woman* contained in the lead are also linked by this connector to the preceding propositions that summarize the proposition expressed by the primary clause in the 'content of saying' section of the lead.

The last nucleus we consider in this subsection is *rein in booming investment* in lead 6.

6) {China yesterday increased its interest rates for the first time in 18 months} *in an effort to* <**rein in booming investment** [[that has the potential to destabilise what is now the world's fourth largest economy]]>. (the lead of text 41; group 2)

   The summaries of text 41 are:

   a) China increased interest rates *to* **control booming investment**.
   b) China's central bank increased interest rates *to* **rein in booming investment**.
   c) China increased interest rates *to* **rein in booming investment**.

Lead 6 consists of a primary clause: *China yesterday increased its interest rates for the first time in 18 months*, and a secondary clause complex: *rein in booming investment that has the potential to destabilise what is now the world's fourth largest economy*. The secondary clause complex contains a ranking (i.e. non-embedded) clause *rein in booming investment*, with *booming investment* post-modified by an embedded *that* clause complex.

The proposition expressed by the primary clause in the lead is summarized by *China increased interest rates* in summaries 6(a) and 6(c), and by *China's central bank increased interest rates* in summary 6(b). The proposition *rein in booming investment* contained in the lead is repeated by *control booming investment* in the first summary, and is also repeated word for word in the last two summaries.

In the lead the proposition *rein in booming investment* is linked to the preceding proposition expressed by the primary clause by *in an effort to*, which also functions as a causal connector expressing purpose. All three summaries also link these two propositions by *'to'* (shorthand for *'in order to'*), also functioning as a causal connector expressing purpose.

To sum up, in this subsection we have seen that both the nuclei occurring in the secondary clauses and the propositions expressed by the primary clauses on which the secondary clauses are dependent on are repeated in two or more summaries, and that the causal relations in the leads that link them are also replicated in the summaries. It is worth finding out whether the relations other than causality in leads that link the initial occurrences of other type A material nuclei to some propositions are also considered as important to the meaning of the texts in which they occur. The next two sections explore this.

5.5.2. The four initial occurrences of type A material nuclei that are part of the content of saying of verbal nuclei.

There are four initial occurrences of type A material nuclei that are part of the content of saying of verbal nuclei. They are *NATO airstrikes* in lead 7, *blow up the building* in lead 8, *using a mobile phone* in lead 9, and *hold accountable those (official)* in lead 10.

We begin by looking at *NATO airstrikes* in lead 7.

7) {The United States and Russia are ratcheting up the pressure on Yugoslav President Slobodan Milosevic}, <warning that **NATO airstrikes** are inevitable unless he takes decisive measures to end the crisis in Kosovo province>. (the lead of text 38; group 2)

The summaries of text 38 are:

a) U.S. and Russia warned Milosevic of **NATO airstrikes**.
b) US and Russia warned Milosevic of **airstrikes** unless ending Kosovo crisis quickly.
c) Yugoslavia removes some, but not all, tanks and troops from Kosovo.

NATO airstrikes in the lead is repeated word for word in the first summary, and is shortened to *airstrikes* in the second summary.

Lead 7 consists of a primary clause: *The United States and Russia are ratcheting up the pressure on Yugoslav President Slobodan Milosevic*, and a secondary clause complex: *warning that NATO airstrikes are inevitable unless he takes decisive measures to end the crisis in Kosovo province*. In the lead *NATO airstrikes* is present in the underlined *that* clause complex that gives the content of the warning issued by the United States and Russia (in other words, the *that* clause complex gives the content of saying of the verbal nucleus *the United States and Russia warning*). Therefore, *NATO airstrikes* is part of the content of the warning.

In the first summary the content of the U.S. and Russia's warning (i.e. the content of saying of the verbal nucleus *U.S. and Russia warned*) is given in the *of* prepositional phrase in which *NATO airstrikes* occurs. Therefore, *NATO airstrikes* is the content of the warning. The receiver of the warning is *Milosevic*.

In the second summary the content of the warning is given in *of airstrikes unless ending Kosovo crisis quickly*. *Airstrikes* occurs in this stretch, and is therefore also part of the content of the warning. *Milosevic* is once again the receiver of the warning.

Therefore, although the syntax in the lead that links *NATO airstrikes* to *the United States and Russia warning* is different from that in the first two summaries, the semantic relation in the lead that links these two propositions is echoed in these two summaries. This is also found in three more examples discussed below.

The next nucleus we consider is *blow up the building* in lead 8.

8) (Nine months before *the attack on the American Embassy here*, U.S. intelligence officials received a detailed warning [[that Islamic radicals were plotting to **blow up the building**]]), according to Kenyan and American officials. (the lead of text 32; group 1)

The summaries of text 32 are:

a) Americans received warning beforehand to **blow up Nairobi Embassy**.
b) Americans received warning about plot to **blow up Nairobi Embassy**.
c) US received warning of **bombing U.S. embassy in Nairobi** beforehand.

The attack on the American Embassy here (italicised) and *blow up the building* (emboldened) are the two occurrences in the lead of a type A material recurrent nucleus in text 32.

The 'content of saying' section of the lead contains a ranking clause: *Nine months before the attack on the American Embassy here, U.S. intelligence officials received a detailed warning*, and an embedded clause complex (marked as [[...]]) that post-modifies *warning*.

In the lead the embedded clause complex in which *blow up the building* occurs gives the content of the warning that U.S. intelligence officials received. *Blow up the building* is

therefore part of the content of the warning. The relation in the lead that links *blow up the building* and the warning is echoed in all the three summaries. In the first summary the content of the warning that Americans received is given in the *'to'* infinitive clause that post-modifies *warning*. Therefore, *'blow up Nairobi Embassy'* is the content of the warning. In the second summary the content of the warning is given in the prepositional phrase *'about plot to blow up Nairobi Embassy'* that also post-modifies *warning*. *Blow up Nairobi Embassy* is present in this phrase and therefore is a part of the content of the warning. Likewise, in the last summary the content of the warning is given in the prepositional phrase *'of bombing U.S. embassy in Nairobi'* that once more post-modifies *warning*. Once again, therefore, *bombing U.S. embassy in Nairobi* is the content of the warning.

The next nucleus we consider is *hold accountable those (officials)* in lead 9.

9) {Seven Egyptian human rights organizations today requested Egyptian President Hosni Mubarak} <to **hold accountable those** responsible for torturing residents of a village in Egypt's countryside while investigating two capital murders last August>. (the lead of text 75; group 2)

The summaries of text 75 are:

a) Seven Egyptian Human Rights Organizations demand formal investigation of torture event.
b) Egyptian human rights organizations demand **holding torturers accountable**.
c) Seven Egyptian human rights organizations demand **holding Copts' torturers accountable.**
d) Human rights organizations urge to **hold Copt torturer accountable.**

Lead 9 consists of a primary clause and a secondary clause complex. The primary clause is *Seven Egyptian human rights organizations today requested Egyptian President Hosni Mubarak*. The secondary clause complex is *to hold accountable those responsible for*

torturing residents of a village in Egypt's countryside while investigating two capital murders last August.

In the lead what was requested by the seven Egyptian human rights organizations (i.e. the content of saying of the verbal nucleus *seven Egyptian human rights organizations requested*) is given in the secondary clause complex. *Hold accountable those* is part of the secondary clause complex, and is therefore part of the content of saying of the verbal nucleus *seven Egyptian human rights organizations requested.*

In summary 9(b) what was demanded by Egyptian human rights organizations is given in the non-finite clause projected by *Egyptian human rights organizations demand.* Therefore, *holding torturers accountable* is the content of saying of the verbal nucleus *Egyptian human rights organizations demand.*

In summary 9(c) what was demanded by seven Egyptian human rights organizations is likewise given in the non-finite clause projected by *Seven Egyptian human rights organizations demand.* Therefore, *holding Copts' torturers accountable* is also the content of saying of the verbal nucleus *Seven Egyptian human rights organizations demand.*

In summary 9(d) what was urged by Human rights organizations is given in the 'to' infinitive clause projected by *Human rights organizations urge.* Therefore, *hold Copt torturer accountable* is also the content of saying of the verbal nucleus *Human rights organizations urge.*

Once again, the semantic relation in the lead that links *hold accountable those* to *seven Egyptian human rights organizations requested* is echoed in the last three summaries.

The last nucleus we consider in this subsection is *using a mobile phone* in lead 10.

10) ([[**Using a mobile phone** during a thunderstorm]] could kill people), doctors warned today. (the lead of text 25; group 1)

The summaries of text 25 are:

a) Doctors warned that **using mobile phones** in thunderstorm could kill people.
b) Doctors warned that **using mobile phones** in thunderstorm could be life-threatening.
c) **Using a mobile phone** in thunderstorm could kill people.

In lead 10 the content of the warning by doctors (i.e. the content of saying of the verbal nucleus *doctors warned*) is given in the stretch projected by *doctors warned*, namely *using a mobile phone during a thunderstorm could kill people*. Therefore, *using a mobile phone* is part of the content of saying of the verbal nucleus *doctors warned.*

Once more, we can see that the relation in the lead that links *using a mobile phone* to *doctors warned* is echoed in the first two summaries. In these two summaries what was warned by doctors is given in the *'that'* clause projected by *Doctors warned.* Therefore, in these two summaries *'using mobile phones'* is also part of the content of saying of the verbal nucleus *doctors warned.*

To conclude the present subsection, the syntax in the lead that links the four initial occurrences of type A material nuclei to the verbal nuclei may be different from that in the summaries of the texts in which they occur, but the semantic relation in the lead that link them is echoed in the summaries.

In subsections 5.5.1 and 5.5.2, we have seen that when two or more nuclei contained in the lead are repeated in two or more summaries of the text in which they occur, the semantic relation in the lead between these nuclei is also repeated in the summaries. In subsection 5.5.3(a) below we will further test this provisional conclusion.

5.5.3(a) The initial occurrences of two type A material nuclei that contextualize

When one proposition contextualizes another, these two propositions are linked by temporal relations. The initial occurrences of two contextualizing type A material nuclei are *kidnapped a 5-year-old girl* in lead 11 and *hit the Indonesian island of Java* in lead 12.

We first consider *kidnapped a 5-year-old girl* in lead 11.

11) {A man was shot dead by police on Monday morning} *after* <he **kidnapped a 5-year-old girl** in downtown Beijing and refused police persuasion>, authorities said. (the lead of text 40; group 1)

The summaries of text 40 are:

a) A man was shot *after* **kidnapping a girl**.
b) Man shot *after* **kidnapping a girl** and refusing police persuasion.
c) Man shot dead *after* **kidnapping** and refusing police persuasion.

The content of saying section of the lead consists of a clause and a clause complex in a hypotactic relation. The primary clause is *A man was shot dead by police on Monday morning*. The secondary clause complex is *he kidnapped a 5-year-old girl in downtown Beijing and refused police persuasion. 'Kidnapped a 5-year-old girl'* is present in the secondary clause complex.

'Kidnapping a girl' in the first two summaries is in a repetition relation with *'kidnapped a 5-year-old girl'* contained in the lead. In the last summary *'kidnapping'* (probably best treated as a noun) was unpacked to *'kidnapping a girl'*; once it is, we can also recognize a repetition relation with *'kidnapped a 5-year-old girl'* contained in the lead.

The proposition expressed by the primary clause in the 'content of saying' section of the lead is summarized by *A man was shot* in the first summary, by *Man shot* in the second summary, and also by *Man shot dead* in the last summary. The proposition *he kidnapped a 5-year-old girl in downtown Beijing* contained in the secondary clause complex in the lead is summarized by *kidnapping a girl* in the first two summaries, and by *kidnapping* (as a noun) in the last summary.

In the lead the proposition *he kidnapped a 5-year-old girl in downtown Beijing* is linked to the preceding proposition expressed by the primary clause by *after*, which functions as a temporal connector. In the three summaries the propositions that summarize these two propositions are also linked by this connector.

The other contextualizing type A material nucleus is *hit the Indonesian island of Java* in lead 12.

12) {A desperate search for survivors is continuing today} *after* <a massive earthquake **hit the Indonesian island of Java**, killing more than 4,300 people>. (the lead of text 68; group 2)

The summaries of text 68 are:

a) Search for survivors continued *after* an earthquake **struck Java**.
b) Search for survivors continued *after* an earthquake **hit Java**.
c) Search for survivors continued in earthquake-hit Java.

Lead 12 consists of a primary clause: *A desperate search for survivors is continuing today*, and a secondary clause complex: *a massive earthquake hit the Indonesian island of Java, killing more than 4,300 people*. *'Hit the Indonesian island of Java'* is present in the secondary clause complex.

Both *'struck Java'* in summary 12(a) and *'hit Java'* in summary 12(b) are in a repetition relation with *'hit the Indonesian island of Java'* contained in the lead. In summary 12(c) the noun group *'earthquake-hit Java'* was not suitable for unpacking because criterion 2 in Figure 3.15 was not satisfied in that *'hit'* does not function as Thing (i.e. the head noun) in this noun group.

The proposition expressed by the primary clause in the lead is summarized by *Search for survivors continued* in all the three summaries. The proposition *a massive earthquake hit the Indonesian island of Java* contained in the secondary clause complex in the lead is summarized by *an earthquake struck Java* in the first summary, and by *an earthquake hit Java* in the second summary, but is reduced to the nominal group *earthquake-hit Java* in the last summary, a phenomenon we will discuss in subsection 5.5.7.

In the lead the proposition *a massive earthquake hit the Indonesian island of Java* is linked to the preceding proposition by *after*. In the first two summaries the propositions that summarize these two propositions are also linked by this connector.

The two instances we have discussed in this subsection confirm the conclusion above that when two or more nuclei contained in the lead are repeated in two or more summaries of the text in which they occur, the relationship in the lead between these nuclei are also often echoed in the summaries.

5.5.3(b) The six initial occurrences of type B material nuclei that also contextualize

There are six material nuclei that also contextualize the propositions occurring in the primary clause (or clause complex) in (the 'content of saying' section of) the leads, but they

are not repeated in two or more summaries of the texts in which they occur. These nuclei are *killing more than 4,300 people* in lead 12, *killing at least 156 people* in lead 13, *gather information* in lead 14, *kicked the door* in lead 15, *retrieved three more bodies* in lead 16, and *a mine in northern China was flooded* in lead 17.

Let us first look at lead 12 again, reproduced below for convenience of discussion. But this time the nucleus we are considering is not *hit the Indonesian island of Java* but *killing more than 4,300 people*.

12) {A desperate search for survivors is continuing today} *after* <a massive earthquake hit the Indonesian island of Java, **killing more than 4,300 people**>. (the lead of text 68; group 2)

The summaries of text 68 are:

a) Search for survivors continued after an earthquake struck Java.
b) Search for survivors continued after an earthquake hit Java.
c) Search for survivors continued in earthquake-hit Java.

Killing more than 4,300 people is present in the secondary clause complex (marked as <...>), which is linked to the preceding primary clause (marked as {...}) by *after* (a temporal connector). This nucleus is not repeated in any summary of text 68. On the other hand, the proposition expressed by the primary clause in the lead is summarized by *Search for survivors continued* in all the three summaries.

The next type B material nucleus we consider is *killing at least 156 people* in lead 13.

13) {Typhoon Babs brought torrential rains and landslides to Taiwan and lashed Hong Kong with strong winds Sunday} *after* <**killing at least 156 people** in the Philippines and leaving hundreds of thousands homeless>. (the lead of text 60; group 2)

The summaries of text 60 are:

a) Typhoon Babs floods Taiwan, lashes Hong Kong, and heads toward China.
b) Babs brought Taiwan torrential rains, Hong Kong winds.
c) Babs brings strong winds and torrential rains to Taiwan and Hong Kong.

In the lead *killing at least 156 people* is present in the secondary clause complex (marked as <...>) that is linked to the preceding primary clause complex (marked as {...}) by *after* (a temporal connector). This nucleus is also not repeated in any of the summaries of the text. On the other hand, the proposition *Typhoon Babs brought torrential rains to Taiwan* contained in the primary clause complex of the lead is summarized by *Typhoon Babs floods Taiwan* in the first summary, by *Babs brought Taiwan torrential rains* in the second summary, and by *Babs brings torrential rains to Taiwan* in the last summary. The proposition *Typhoon Babs lashed Hong Kong with strong winds* also contained in the primary clause complex is summarized by *Typhoon Babs lashes Hong Kong* in the first summary, by *Babs brought Hong Kong winds* in the second summary, and by *Babs brings strong winds to Hong Kong* in the last summary.

We have seen that the propositions occurring in the primary clause (or clause complex) of the two leads discussed above are summarized in two or more summaries of texts 60 and 68 respectively. Let us consider whether this phenomenon is also found in the other four examples discussed below.

We next consider *gather information* in lead 14.

14) WELLINGTON – {New Zealand children have suffered convulsions} *as* <officials scramble to **gather information** following reports of similar reactions to the flu vaccination in Australia>. (the lead of text 18; group 2)

The summaries of text 18 are:

a) New Zealand kids react to flu vaccine.
b) New Zealand children have suffered convulsions after flu vaccinations.
c) New Zealand kids suffered convulsions following vaccinations.

Gather information is present in the secondary clause (marked as <...>) that is linked to the preceding primary clause (marked as {...}) by *as*, which functions as a temporal connector.

Once again, *gather information* is not repeated in any of the summaries of text 18, but the proposition expressed by the primary clause in the lead is summarized by *New Zealand children have suffered convulsions* in the second summary, and by *New Zealand kids suffered convulsions* in the last summary.

We then consider *kicked the door* in lead 15.

15) {A man died in a fall into an elevator shaft} *after* <he allegedly **kicked the door** open Tuesday>. (the lead of text 57; group 2)

The summaries of text 57 are:

a) Drunken man died in elevator fall.
b) A drunken man died in a fall into an elevator shaft.
c) A drunken man died after falling into an elevator shaft.

In lead 15 *kicked the door* is present in the secondary clause that is linked to the preceding primary clause by *after* (again a temporal connector)

Again *kicked the door* is also not repeated in any of the summaries of text 57, but the proposition expressed by the primary clause of the lead is summarized in all the three summaries.

Therefore, the phenomenon we observed earlier in leads 12 and 13 is also found in leads 14 and 15. We may draw the following provisional generalization: when the nucleus that occurs in the secondary clause of a lead is not repeated in two or more summaries of the text in which it occurs, one or more propositions that appear in the primary clause on which it is dependent are instead often repeated in the summaries. We will further test this

generalization with leads 16 and 17 given below.

We next consider *retrieved three more bodies* in lead 16.

16) DHAKA, June 16 (Xinhua) – {The death toll from early Tuesday's devastating rain-triggered landslides in Bangladesh's two southeastern districts rose to 54}, *as* <rescuers **retrieved three more bodies** including that of an army personnel Wednesday>, officials said. (the lead of text 50; group 1)

The summaries of text 50 are:

a) Death toll from Bangladesh landslides rises to 54.
b) Death toll from rain-triggered landslides in Bangladesh rose to 54.
c) **Three more bodies retrieved**, adding to 54 killed in Bangladesh landslides.

In lead 16 *retrieved three more bodies* is present in the secondary clause that is linked to the preceding primary clause by *as* (again as a temporal connector).

Retrieved three more bodies in the lead is repeated by *three more bodies retrieved* in the last summary only. On the other hand, once again the proposition expressed by the primary clause in (the 'content of saying section' of) the lead is summarized in the first two summaries, and is paraphrased by the part after the comma in the last summary.

The next type B material nucleus we consider is *a mine in northern China was flooded* in lead 17.

17) {RESCUE workers set up equipment and prepared yesterday to find survivors} *after* <**a mine in northern China was flooded**>, Xinhua news agency reported. (the lead of text 51; group 1)

The summaries of text 51 are:

a) Rescuers set up equipment and prepare to find trapped miners.
b) Rescue workers set up equipment to search for survived miners.
c) Rescuers prepare to find survivors trapped in a flooded mine.

The 'content of saying' section of lead 17 consists of a primary clause complex and a secondary clause. *A mine in northern China was flooded* is present in the secondary clause, which is linked to the preceding primary clause complex by *after* (once again a temporal connector).

A mine in northern China was flooded in the lead is not repeated in the first two summaries, but is reduced to the nominal group *a flooded mine* in the last summary. The way in which this nominal group is derived is similar to the way in which *earthquake-hit Java* contained in summary 12(c) is derived. We will discuss the nominal groups that are derived in such a way in subsection 5.5.7.

Once again, *RESCUE workers set up equipment* contained in the first half of the primary clause complex in the lead is summarized by *Rescuers set up equipment* in summary 17(a), and is repeated word for word in the first half of summary 17(b). *(RESCUE workers) prepared to find survivors* contained in the second half of the primary clause complex of the lead is summarized by *(Rescuers) prepare to find trapped miners* in the first summary, by *(Rescue workers) search for survived miners* in the second summary, and by *(Rescuers) prepare to find survivors* in the final summary.

To sum up this subsection, we have seen that when the nuclei occurring in the secondary clauses of the leads are not repeated in two or more summaries of the texts in which they occur, one or more propositions that appear in the primary clauses on which they are dependent are instead often repeated in the summaries. This finding supports the conclusion we drew earlier from tables 5.2 and 5.3 that the nuclei that are conceived of by my informants as important to the meanings of the texts in which they occur tend to appear in

major types of clause in (the 'content of saying' section of) the leads. As noted earlier, presenting the propositions that are important to the texts' meanings in major types of clause in (the 'content of saying' section of) the leads may be a means that news writers often employ to foreground these propositions. These propositions could have been backgrounded in a secondary clause in (the 'content of saying' section) of the leads. For instance, lead 12 could have been written in the following way:

{A massive earthquake hit the Indonesian island of Java}, <killing more than 4,300 people and leading to a desperate search for survivors that continued today>. (rewritten lead 12)

I would argue that had the lead been written as such, *a desperate search for survivors that continued today*, which, as noted earlier, was conceived of by my informants as important to the meaning of the text, would have been backgrounded because it had been presented in a secondary clause.

The constraint on the length of summary that the summary writers were asked to produce (i.e. approximately 10 words) seems to have an impact on the summary writers' decisions as to whether to include contextualizing information. It seems that the contextualizing information will be included in the summaries provided that sufficient space is left in the summaries after the propositions that are contained in the primary clauses in (the 'content of saying' section of) the leads are summarized. For instance, after summarizing the proposition expressed by the primary clause in the content of saying section of lead 11, all three summaries have sufficient space left to include the proposition (the man) *kidnapped a 5-year-old girl* contained in the secondary clause complex of the lead. On the other hand, after summarizing the propositions contained in the primary clause complex of lead 13, summaries 13(a) and 13(c) already contain 11 and 12 words respectively, leaving no space to include the proposition *killing at least 156 people* contained in the secondary clause

complex of the lead; summary 13(b) contains 8 words, likewise leaving no sufficient space to include this proposition because had it been included, this summary would have exceeded 10 words.

In the examples we have discussed in subsections 5.5.1, 5.5.2 and 5.5.3(a), the way in which two or more nuclei in a lead are linked is preserved by my summary writers because the linkages in the leads are often replicated in the summaries. In the subsections that follow, we will be discussing proposition splitting and inference, with which two propositions in a lead that are not linked in any way are represented by my summary writers in such a way that they become linked in some way in the summaries. Also we will be discussing proposition integration, with which two propositions in a lead that are linked in some way are represented by my summary writers in such a way that one qualifies an element contained in the other proposition.

5.5.4 Proposition splitting

We first consider proposition splitting. In this subsection we will discuss four initial occurrences of type A material nuclei, namely *stabbed eight elementary school students* in lead 18, *the assassination of President Pak Chong-hui* and *the slaying* in lead 19, and *ravaging southern China* in lead 20. Each of these nuclei occurs in a clause (or clause complex) that qualifies an element of another clause. Let us first consider *stabbed eight elementary school students* in lead 18.

18) {A PUBLIC prosecution has been brought against a man in Nanping, East China's Fujian Province}, <who allegedly **stabbed eight elementary school students** to death and wounded another five>, the People's Procuratorate of Nanping announced Saturday. (the lead of text 13; group 1)

The summaries of text 13 are:

a) School stabbing suspect was prosecuted in Fujian Province.
b) A man was prosecuted for **stabbing elementary school students**.
c) A man in Fujian was prosecuted for **attacking children**.

The 'content of saying' section of lead 18 consists of a primary clause and a secondary clause complex. The primary clause is *A PUBLIC prosecution has been brought against a man in Nanping, East China's Fujian Province*, and the secondary clause complex is *who allegedly* ***stabbed eight elementary school students*** *to death and wounded another five*. The secondary clause complex qualifies *a man*, an element in the primary clause.

Halliday and Matthiessen (2004: 426) say that there is no direct relationship between the clause that qualifies an element and the 'outer' clause containing that element, and that the relationship between these two clauses is an indirect one, with the qualified element as intermediary. Therefore, *stabbed eight elementary school students* contained in the secondary clause complex has no direct relationship with the proposition expressed by the primary clause.

Stabbed eight elementary school students in the lead is repeated by *stabbing elementary school students* contained in the second summary and by *attacking children* contained in the last summary. In the first summary the nominal group *School stabbing suspect* was not suitable for unpacking because once again criterion 2 in Figure 3.15 is not satisfied in that *stabbing* does not function as Thing in this nominal group.

Unlike *stabbed eight elementary school students* in the lead, neither *stabbing elementary school students* in summary 18(b) nor *attacking children* in summary 18(c) is represented in a qualifying relative clause. Summary 18(b) consists of two clauses in a hypotactic relation.

The primary clause is *'A man was prosecuted'*, and the secondary clause is *'stabbing elementary school students'*. These two clauses are linked by *'for'*, which functions as a causal connector. Therefore, *stabbing elementary school students* is expressed as the reason why the man was prosecuted. Summary 18(c) also consists of two clauses in a hypotactic relation. The primary clause is *'A man in Fujian was prosecuted'*, and the secondary clause is *'attacking children'*. These two clauses are likewise linked by *'for'*, which also functions as a causal connector. Therefore, *attacking children* is also expressed as the reason why the man was prosecuted.

This example is in line with Van Dijk and Kintsch's (1983) finding that the hearer/reader (information receiver) sometimes constructs a propositional textbase which need not be congruent with the surface structure indicated in the texts produced by the speaker/writer (information giver). 'Proposition splitting' is one of the ways in which the information receivers rewrite the representations of the propositions in the discourse produced by the information givers (Van Dijk and Kintsch, 1983: 133). It occurs when the speaker/writer represents proposition A as specifying or qualifying an element contained in proposition B, but the hearer/reader represents these two propositions in a way that they are directly linked in some relation. 'Proposition splitting' is found in summaries 18(b) and18(c).

It would be worth investigating the reason(s) why some summary writers used a 'proposition splitting' strategy, but this is beyond the scope of my present study. Van Dijk and Kintsch suggest that the use of the proposition splitting strategy is related to pragmatics: if the hearer/reader believes, according to his or her knowledge, opinions, goals, or interests, that the modifying information may be needed for later use, either in understanding the same discourse or in order to form expectations about the course of events, he or she will tend to

upgrade this modifying information to the status of separate fact (Van Dijk and Kintsch. 1983: 132-133).

The next two type A material nuclei we consider are *the assassination of President Pak Chong-hui* and *the slaying* in lead 19 given below.

19) The joint military-prosecution investigation team, <which probed **the assassination of President Pak Chong-hui** by then Central Intelligence Agency Director Kim Chae-kyu in October 1979>, had found no evidence [[that then Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in **the slaying**]], a former team member testified recently. (the lead of text 1; group 1)

The summaries of text 1 are:

a) No evidence showed Chong was an accomplice in **the slaying of Pak**.
b) No evidence showed that Chong involved in **Pak's assassination**.
c) Chong was not found to involve in **the assassination of Pak**.

The 'content of saying' section of the lead consists of a primary clause complex and a secondary clause. The primary clause complex is *The joint military-prosecution investigation team had found no evidence that then Army Chief of Staff and martial law commander, Gen. Chong Sung-hwa, was an accomplice in the slaying*. It contains a ranking (i.e. non-embedded) clause *The joint military-prosecution investigation team had found no evidence*, with *evidence* post-modified by an embedded *that* clause (marked as [[...]]). *The slaying*, which occurs in the embedded clause, has no direct relationship with the proposition expressed by the ranking clause. The secondary clause (marked as <...>) qualifies *The joint military-prosecution investigation team*, the subject of the ranking clause. *The assassination of President Pak Chong-hui*, which occurs in the secondary clause, also does not have a direct relationship with the proposition expressed by the ranking clause.

Proposition splitting is also found in all of the three summaries. In the first summary *Chong was an accomplice in the slaying of Pak*, which summarizes the proposition expressed by

the embedded *that* clause in the lead, does not modify any element but is directly linked to the preceding proposition *No evidence showed* which seems to paraphrase the ranking clause contained in the primary clause complex of the lead. In the second summary *Chong involved in Pak's assassination*, which also summarizes the proposition expressed by the embedded *that* clause in the lead, is also linked to the preceding clause *No evidence showed that*, which, as just noted, seems to paraphrase the ranking clause contained in the primary clause complex of the lead. In the last summary *(Chong) involve in the assassination of Pak*, which once again summarizes the proposition expressed by the embedded *that* clause in the lead, is also linked to the preceding clause *Chong was not found.*

Proposition splitting is also found in one of the summaries of the text in which lead 20 occurs.

20) {The death toll from fierce rainstorms [[**ravaging southern China** this week]] has climbed to 70}, <with five people previously listed as missing in Jiangxi Province being found dead on Sunday>. (the lead of text 15; group 2)

The summaries of text 15 are:

a) 70 died as rainstorms **battered southern China**.
b) 70 killed in rainstorms **battering southern China**.
c) 70 killed in ravaging storms in southern China.

Lead 20 consists of a primary clause complex (marked as {...}) and a secondary clause complex (marked as <...>). The primary clause complex contains a ranking clause *The death toll from fierce rainstorms has climbed to 70*, with the nominal group *fierce rainstorms* post-modified by an embedded clause (marked as [[...]]). *Ravaging southern China* contained in the embedded clause has no direct relationship with the proposition expressed by the ranking clause. The secondary clause complex also consists of a ranking clause *five people being found dead on Sunday*, with the subject of this clause, namely *five*

people, being post-modified by the embedded clause *'previously listed as missing in Jiangxi Province'*.

Ravaging southern China in the lead is repeated by *battered southern China* contained in the first summary, and by *battering southern China* contained in the second summary. Once more, the nominal group *ravaging storms in southern China* in the last summary was not suitable for unpacking because criterion 2 in Figure 3.15 was not satisfied in that *ravaging* does not function as Thing in this nominal group.

The proposition *(fierce rainstorms are) ravaging southern China* in the lead is reduced to the nominal group *ravaging storms in southern China* in the last summary, a phenomenon that will be discussed in subsection 5.5.7.

Proposition splitting is found in summary 20(a). This summary consists of two clauses in a hypotactic relation. The primary clause is *70 died*, and the secondary clause is *rainstorms battered southern China.* These two clauses are linked by *as*, which functions as a temporal connector. Therefore, the summary writer of 20(a) expresses *rainstorms battered southern China* as the temporal circumstance that indicates when 70 (people) died.

The representation of *battering southern China* in summary 20(b) is, on the other hand, identical to the representation of *ravaging southern China* in the lead. In this summary *battering southern China* is the embedded clause that also post-modifies the noun *rainstorms*.

5.5.5. Inference

Inference occurs when two propositions are not linked in any way in a lead but they are represented by summary writers in such a way that they are linked in some way. In other words, there is no direct relationship in a lead between two propositions, but summary writers infer a relation between them. There are three examples in which summary writers infer a relation not explicitly stated in a lead that holds between two propositions. Let us first consider *providing paid services* in lead 21.

21) Police in Beijing's Chaoyang district on Tuesday discovered 557 hostesses **providing paid services**, which is against the law, in four luxury night clubs, and suspended the clubs' business for six months, the Beijing Times reported Thursday. (the lead of text 31; group 1)

The summaries of text 31 are:

a) Four night clubs suspended for **providing paid** hostesses **services**.
b) Four night clubs suspended due to **provision of paid services**.
c) Four night clubs suspended for paid services.

Providing paid services in the lead is repeated by *providing paid (...) services* contained in the first summary. Once the nominalization *provision of paid services* in the second summary was unpacked, we can also recognize a repetition relation with *providing paid services* contained in the lead.

The lead can be analyzed in two ways. The first way considers that the clause complex *providing paid services, which is against the law, in four luxury night clubs* post-modifies the preceding nominal group *557 hostesses*. When it is analyzed in such a way, the 'content of saying' section of the lead is made up of two parts in a paratactic relation. The first part contains a ranking clause: *Police in Beijing's Chaoyang district on Tuesday discovered 557 hostesses*, with the nominal group *557 hostesses* post-modified by the embedded clause

complex noted above. *Providing paid services* is present in the embedded clause complex. The second part is a single clause: *suspended the clubs' business for six months.*

The second way of analyzing the lead considers that the clause complex *providing paid services, which is against the law, in four luxury night clubs* is projected by the preceding mental clause *Police in Beijing's Chaoyang district on Tuesday discovered.* The clause complex and the mental clause make up the first part of the 'content of saying' section of the lead. The second part is *suspended the clubs' business for six months.* The two parts stand in a paratactic relation.

No matter which way the lead is analyzed, an inference strategy is used by summary writers. In the lead there is no direct relationship between *providing paid services* and *suspended the club's business*, but in the first two summaries a causal relation holds between the two propositions.

'Which is against the law' that follows, and comments on, *providing paid services* may be the cue that led the summary writers of the first two summaries to infer a causal relation between the provision of paid services and the suspension of the night clubs because presumably it is the behaviour that is against the law that resulted in the suspension of the night clubs.

The next type A material nucleus we consider is *murdering his neighbour Zhao Zuohai* in lead 22 given below.

22) {Zhao Zhenshang in central China's Henan province has been jailed for ten years}. <accused of **"murdering" his neighbor Zhao Zuohai**>. However, the "victim" reappears recently, u-turning an old-time case that lays bare custody torture by police officers and sends public confidence on local authorities into a tailspin. (the lead of text 29; group 2)

The summaries of text 29 are:

a) Zhao was jailed for 10 years for **killing** before the victim reappeared.
b) Zhao was imprisoned because of **killing**. The victim returned 10 years later.
c) Zhao was wrongly imprisoned for 10 years, casting doubt on justice.

Lead 22 contains two sentences. The first sentence consists of a primary clause (marked as {...}) and a secondary clause complex (marked as <...>).

Once the nominalization *killing* that occurs in the first two summaries is unpacked, we recognize a repetition relation with *"murdering" his neighbor Zhao Zuohai* contained in the lead.

Inference is also found in the first two summaries. In the first summary *killing* is linked to the preceding proposition by the conjunctive preposition *for*, which expresses reason. In the second summary *killing* is also linked to the preceding proposition by the conjunctive prepositional phrase *because of*, which also expresses reason. Therefore, in the first two summaries a causal relation holds between the killing of his neighbour and the imprisonment of Zhao Zhenshang, but such a relation between the two propositions is not explicitly stated in the lead. *'Accused of'* in the lead may be the cue that led the summary writers of the first two summaries to infer a causal relation between the two propositions because presumably the imprisonment of Zhao Zhenshang was due to the criminal behaviour that he was accused of.

The two examples we have discussed concur with the finding reported in Caron *et al* (1988: 315). They found that the reader's ability to recall a target sentence where a causal relation between it and the cue sentence had to be inferred was almost as accurate as when the causal relation between the sentences was explicitly stated in the text. Their finding suggests two points. The first is that when the causal relation between two events is not explicitly stated in a text but is inferable by readers, the recall of these events is not affected. This is shown by the two examples we have discussed. The second point is that the causal relationships evoked by *against the law* in lead 21 and *accused of* in lead 22 do not seem to be weaker than the versions that make explicit use of conjunctions (e.g. *because, for*) or conjunctive prepositions/prepositional phrases (e.g. *due to*) to indicate the relation. The first sentence of lead 22 could have been written as *Zhao Zhenshang in central China's Henan province has been jailed for ten years for "murdering" his neighbor Zhao Zuohai.* Replacing *'accused of'* and the preceding comma with the conjunction *'for'* would make the causal relationship between the imprisonment of Zhao Zhenshang and the murder of his neighbour more explicit, but perhaps not stronger.

The third and final type A material nucleus we consider in this subsection is *129 people were killed* in lead 23.

23) {Typhoon Babs raced toward southern China on Saturday}, <leaving behind a trail of mess on the Philippines' main island [[where at least **129 people were killed** and hundreds of thousands were forced to flee their homes]]>. (the lead of text 30: group 2)

The summaries of text 30 are:

a) Typhoon Babs, **killing 129** in Philippines, heads toward China.
b) Babs heads for southern China after **killing 129** in Philippines.
c) Babs leaves parts of the Philippines in a "state of calamity."

Lead 23 consists of a primary clause (marked as {...}) and a secondary clause complex (marked as <...>). The secondary clause complex contains an embedded clause complex (marked as [[...]]) that post-modifies the nominal group *the Philippines' main island*.

129 people were killed in the lead is in a repetition relation with *killing 129* contained in the first two summaries.

Inference is once again found in summary 23(b). This summary consists of two clauses in a hypotactic relation. The primary clause is *Babs heads for southern China*, and the secondary clause is *killing 129 in Philippines*. They are linked by a temporal connector *after*, which indicates that Babs' movement towards south China happened after it had killed 129 people in the Philippines. But such a relation between the two propositions is not explicitly stated in the lead.

Leaving behind contained in the lead may lead summary writer of 23(b) to infer a temporal relation that holds between Babs' killing of people in the Philippines and its movement towards southern China after that.

Inference is not, on the other hand, found in summary 23(a). This summary consists of two clauses in a hypotactic relation. The primary clause is *Typhoon Babs heads toward China*. which summarizes the proposition expressed by the primary clause in the lead. The secondary clause *killing 129 in Philippines* qualifies *Typhoon Babs*, an element in the primary clause of the summary. Therefore, in this summary there is also no direct relationship between Babs' killings of people in the Philippines and its movement towards southern China.

5.5.6 Proposition integration

'Proposition integration' is the opposite of 'proposition splitting'. 'Proposition integration' occurs when proposition A is linked to proposition B in some way in the text produced by the speaker/writer, but the hearer/reader represents one of these two propositions as specifying a constituent contained in the other proposition. This phenomenon was found in only one example, given below.

24) {The leader of a 14-year-old Kurdish insurgency, considered a terrorist by Turkey and its most-wanted criminal, has been arrested in Rome}, <setting up a major battle over **his extradition** with Italy>. (the lead of text 28; group 2)

    a) Turkey seeks **extradition of Kurdish rebel leader** arrested in Rome.

Lead 24 consists of a primary clause complex (marked as {...}) and a secondary clause (marked as <...>). *His extradition,* which is the initial occurrence of a type A material nucleus in text 28, is present in the secondary clause. The phrase *setting up* in the initial position of the secondary clause indicates that Ocalan's extradition is an issue that arose after the arrest of Ocalan. But such a relation is not shown in summary 24(a). This summary consists of a ranking clause *Turkey seeks extradition of Kurdish rebel leader*, with *Kurdish rebel leader* post-modified by the embedded clause *arrested in Rome*. Consequently, there is no relation between the extradition of Kurdish rebel leader and the arrest of the leader.

5.5.7 Reduction

In the examples discussed in the previous subsections, we have seen that a clause contained in a lead may be reduced to a nominal group in summary. By doing so, the process of the clause is reconstrued by summary writers as the pre-modifier of a head noun. Van Dijk and

Kinstch (1983: 123) refer to this kind of phenomenon as reduction. There are four instances in which reduction occurs. Let us first consider lead 12 and summary 12(c).

12) {A desperate search for survivors is continuing today} after <a massive earthquake hit the Indonesian island of Java, killing more than 4,300 people>. (the lead of text 68; the second group)

    c) Search for survivors continued in earthquake-hit Java.

The clause *a massive earthquake hit the Indonesian island of Java* (underlined) in the lead is reduced to the nominal group *earthquake-hit Java* (also underlined) in summary 12(c). The process of the clause, namely *hit*, was reconstrued as a pre-modifier of the head noun (i.e. *Java*) of the underlined nominal group in the summary.

We next consider lead 17 and summary 17(c) given below.

17) {RESCUE workers set up equipment and prepared yesterday to find survivors} after <a mine in northern China was flooded>, Xinhua news agency reported. (the lead of text 51; the first group)

    c) Rescuers prepare to find survivors trapped in a flooded mine.

Similarly, the clause *a mine in northern China was flooded* (underlined) in lead 17 is reduced to the nominal group *a flooded mine* (also underlined) in summary 17(c). The process of the clause, namely *flooded,* was also reconstrued as a pre-modifier of the head noun (i.e. *mine*) of the underlined nominal group in the summary.

Therefore, we can see that in both examples a clause in the lead is reduced by summary writer to a single nominal group in which the process of the clause is reconstrued in the summary as the pre-modifier of a head noun. Reduction is also found in the two instances discussed below.

18) {A PUBLIC prosecution has been brought against a man in Nanping, East China's Fujian Province}, <who allegedly stabbed eight elementary school students to death and wounded another five>, the People's Procuratorate of Nanping announced Saturday. (the lead of text 13; the first group)

a) School stabbing suspect was prosecuted in Fujian Province.

Once again, the clause *who* (standing for *'a man'*) *allegedly stabbed eight elementary school students to death* contained in lead 18 is reduced to the nominal group *school stabbing suspect* in summary 18(a). The process of the clause, namely *stabbed,* is again reconstrued as a pre-modifier of the head noun (i.e. *suspect*) of the underlined nominal group in the summary.

20) {The death toll from fierce rainstorms [[ravaging southern China this week]] has climbed to 70}, <with five people previously listed as missing in Jiangxi Province being found dead on Sunday>. (the lead of text 15; the second group)

c) 70 killed in ravaging storms in southern China.

The clause *(rainstorms were) ravaging southern China this week* contained in lead 20 is likewise reduced to the nominal group *ravaging storms in southern China* in summary 20(c). The process of the clause, namely *ravaging,* is likewise reconstrued as a pre-modifier of the head noun (i.e. *storms*) of the underlined nominal group in the summary.

My summary writers used both the 'reduction' strategy and 'proposition integration' strategy to represent a proposition expressed by a clause contained in a lead in such a way that the proposition is packaged into a nominal group in the summary. The nominal groups derived from 'reduction' are different from those derived from 'proposition integration' in two ways. Firstly, the nominal groups derived from 'reduction' are more concise than those derived from 'proposition integration'. For instance, if summary 12(c) had been rewritten as *Search for survivors continued in Java hit by earthquake* (replacing *earthquake-hit*[18] with *hit*

[18] *Earthquake-hit* was considered as a compound item.

by earthquake), the rewritten summary would have contained two more words than the original summary.

Secondly, 'proposition integration' involves rankshifting, whereas 'reduction' involves grammatical metaphor. With 'proposition integration', the summary writers represent two propositions in a lead in such a way that one is written as an embedded clause post-modifying a constituent of the other proposition. On the other hand, 'reduction' involves reconstruing the process of a clause as the epithet/classifier of the head noun in a nominal group. The reconstrual of a process as epithet/classifier is a type of grammatical metaphor (Halliday and Matthiessen, 1999: 247; Martin and Rose, 2003: 105).

Another strategy, nominalization, is more often used by summary writers than 'reduction' and 'proposition integration' to rewrite propositions expressed by clauses in the original texts in such a way that they are represented in nominal groups contained in summaries. With nominalization, the processes of clauses in the original texts are reconstrued as the Things (i.e. the head nouns) of nominal groups in summaries, and incidentally the participants that are involved in the processes are either left out or reconstrued by summary writers as the modifiers of the Things in nominal groups. Therefore, both nominalization and reduction strategies involve grammatical metaphors, but of different types (see Halliday and Matthiessen, 1999: 246-248).

5.5.8 The remaining and final example

The final nucleus we consider is *a knife attack on children* in lead 25.

25) (The man [[who was detained after **a knife attack on children** at an east China kindergarten Thursday]] carried out *the attack* due to personal humiliations), police said on Friday. (the lead of text 26; group 1)

The summaries of text 26 are:

a) Xu **attacked kindergarten children** to revenge on personal humiliation.
b) Xu **attacked kindergarten children** for revenge on personal humiliation.
c) Xu **attacked kindergarten children** due to personal humiliations.

A knife attack on children and *the attack* are the two occurrences in the lead of a type A material nucleus in text 26. The 'content of saying' section of this lead contains a ranking clause *The man carried out the attack due to personal humiliations*, with the subject of this clause post-modified by an embedded clause (marked as [[...]]). *A knife attack on children*, emboldened, is present in the embedded clause, while *the attack*, italicised, is present in the ranking clause noted above.

Once *a knife attack on children* and *the attack* contained in the lead are unpacked, we recognize a repetition relation with *attacked kindergarten children* (also emboldened) contained in all of the three summaries.

The first two summaries contain *revenge on*, but this meaning is not present in the lead. It is, however, present in sentence 2 of the text, given below.

Jiang Wenxiang, chief of the Public Security Bureau in Taixing City, Jiangsu Province. said Xu Yuyuan, 46, told police the attack was "his revenge on society."

This instance once again shows that the information that is conceived of by readers as important to the meanings of a hard news text may not be always present in the lead.

5.6 Conclusions

The first part of this chapter (sections 5.2, 5.3 and 5.4) has shown that the initial occurrences of type A and type B material nuclei tend to appear in different types of clause. The initial

occurrences of type A material nuclei tend to occur in major types of clause in (the 'content of saying' section of) the leads, while the initial occurrences of type B material nuclei tend to occur in minor types of clause in (the 'content of saying' section of) the leads. The implication is that news writers tend to foreground the information that is important to the meaning of a text in primary or independent clauses in (the 'content of saying' section of) the leads, and background the information that is less important to the text's meaning in secondary or embedded clauses in (the 'content of saying section' of) the leads.

The second part of the chapter (section 5.5) has examined the initial occurrences of type A material nuclei that are, however, found in secondary or embedded clauses in (the 'content of saying' section of) the leads. In subsections 5.5.1, 5.5.2 and 5.5.3(a), we found that when two or more propositions in a lead are repeated in two or more summaries of the text in which they occur, the relationship in the lead that links the propositions is also often echoed in the summaries. The implication is that when two or more propositions in a lead are considered by readers to be important to the meaning of the text in which they occur, the relationship between the propositions is also thought of as important to the text's meaning. After that, in subsections 5.5.4 and 5.5.5 we looked at 'proposition splitting' and 'inference' respectively. With these two strategies, the two propositions that are not linked in any way in a lead are represented by my summary writers in such a way that they are, however, linked in some way in the summaries. On the other hand, with proposition integration discussed in subsection 5.5.6, we found that the two propositions that are linked in some way in a lead are represented by my summary writers in such a way that they are, however, not so linked in the summaries. Section 5.5.7 discussed 'reduction', with which the proposition that is expressed by a clause is packaged by my summary writer in a single nominal group.

If we consider the conclusions given in the two previous paragraphs together, we can identify the circumstances in which the nuclei occurring in minor types of clause contained in (the 'content of saying' section of) the leads are thought of by readers to be important to the meaning of the texts in which they occur, and also the strategies that readers use to write summaries for hard news texts.

In subsections 5.5.1 and 5.5.3(a), we have seen that when the nuclei that occur in the secondary clauses in the ('content of saying' section of) leads are, however, thought of by readers as important to the meaning of the texts in which they appear, it is often the case that the nuclei occurring in the primary clauses on which they are dependent are also thought of as important to the texts' meaning. What is more, the relations in leads that link the secondary clause nuclei to the primary clause nuclei are often replicated in the summaries of the texts in which they occur. Therefore, replication in summaries of the linkages that occur in the leads is a strategy that is often used by readers to write summaries for hard news texts.

The nuclei discussed in subsection 5.5.4 occur in qualifying relative clauses or embedded clauses contained in the leads to post-modify elements of other clauses. Therefore, no direct relationships hold between them and the propositions occurring in major types of clause in (the 'content of saying' section of) the leads. These nuclei are, however, thought of as important to the meaning of the texts in which they occur because they are represented by my summary writers in such a way that they become linked in some way to the propositions occurring in major types of clause in (the 'content of saying' section of) the leads. Likewise, the nuclei discussed in subsection 5.5.5 also occur in secondary or embedded clauses contained in (the 'content of saying' section of) the leads, but they are thought of as

important to the meaning of the texts in which they occur. This is because summary writers infer a relation holding between them and the propositions occurring in major types of clause in (the 'content of saying' section of) the leads, even though such a relation is not explicitly stated in the text. Therefore, the instances discussed in the two subsections show that readers may create in summaries linkages that do not occur in the leads.

In addition to the two strategies noted above, namely replication in summaries of linkages that occur in the leads, and creation in summaries of linkages that do not, however, occur in the leads, readers also use the nominalization and reduction strategies to write summaries. Readers use these two strategies to rewrite the propositions expressed by clauses in the original texts in such a way that they are represented by nominal groups in summaries. With the nominalization strategy, the processes of clauses in the original texts are reconstrued as the Things (i.e. the head nouns) of nominal groups in summaries, and incidentally the participants that are involved in the processes are either left out or reconstrued by summary writers as the modifiers of the Things in nominal groups. The analysis of summaries shows that this strategy is very often employed by readers, perhaps because it enables summary writers to represent propositions in nominal groups, which may function as elements of other propositions so that a summary consists of two or more propositions. With the reduction strategy, on the other hand, the processes of clauses in the original texts are reconstrued as the pre-modifiers of head nouns in nominal groups.

Chapter 6 Conclusions

6.1 Introduction

Having discussed the findings of my study in the previous two chapters, this final chapter notes the areas of innovation and contribution of my study, and brings my findings and those noted in the works of other authors together with a view to identifying better the strategies that are often used by readers to write summaries and the relationships between cohesion in texts and the texts' meanings.

6.2 Areas of innovation and contribution of my present study

My present study has explored two areas noted below that have not been dealt with by earlier studies.

First of all, while most earlier studies on cohesion have analyzed texts in terms of the repetitions of single elements, my present study has analyzed texts and their summaries in terms of repetitions of the combinations of Processes and their Mediums (i.e. the nuclei). As noted in the Introduction chapter, the nucleus of a clause can be seen as a miniature representation of the proposition expressed by the clause. Therefore, analyzing texts in terms of repetitions of nuclei is in effect analyzing the propositions that recur in the text. My present study has analyzed hard news texts and their summaries in terms of recurrent nuclei, with a view to exploring the effects of cohesion in texts on readers' interpretation of the texts' meanings. Chapters 4 and 5 of my thesis show that analyzing texts and their summaries in such a way enhances our understandings of the characteristics of hard news texts, of the relationships between cohesion in texts and the texts' meanings, and of the strategies that readers use to write summaries for hard news texts. We will discuss findings in these regards in detail in the sections that follow.

Secondly, earlier studies have explored nominalizations in terms of a variety of aspects, e.g. their occurrences in children's development of language use (see Painter *et al*, 2007), their functions in the unfolding of texts (see Hoey (1983)), the different roles they play in different text types (see Martin (1993)), the ways of unpacking them (see Halliday and Matthiessen (1999)). However, the circumstances which are not suitable for unpacking nominalizations have not been explored. Some work reported in my thesis has filled in the void. I have used flowcharts to spell out the criteria that need to be satisfied to ensure that it is appropriate to unpack nominalizations. Figure 3.9 gives the criteria for unpacking nominalizations in texts, while figure 3.11 gives the criteria for unpacking nominalizations in summaries.

6.3 Patterns of lexis in a text and the meaning of the text

It has been suggested by such earlier works noted in chapter 2 as Hoey (1991), Scott (1997, 2000, 2004, 2010), Benbrahim and Ahmad (1994), and Barzilay and Elhadad (1997) that the patterns of the lexis in a text are often indicative of what the text is about. This point is once again evidenced by my finding that more often than not one or more recurrent nuclei in a hard news text are thought of by readers as important to the meaning of the text.

Having said that the information that is important to the meaning of a text is often repeated in the text, two caveats need pointing out. The first is that not all the reiterated information in a text is thought of by readers as important to the meaning of the text. This is not only evidenced by the observation that in each text contained in categories 1, 2 and 3 of Figure 4.1 given in chapter 4, some recurrent nuclei are thought of as important to the meaning of the text in which they occur, while others are not, but is also shown by the observation of the three texts contained in category 4 that none of the recurrent nuclei contained in each of these

texts are thought of as important to the meaning of the text in which they occur. An issue arises: what is the characteristic of the reiterated information that is thought of as important to the meaning of the text in which they occur? Table 4.1 given in chapter 4 shows that the recurrent nuclei that are thought of as important to the meaning of the texts in which they appear have a strong tendency to occur in the leads. Moreover they tend to occur in either primary or independent clauses contained in the leads (see section 5.4 of chapter 5). On the other hand, observations of the recurrent nuclei that are not thought of as important to the meaning of the texts in which they appear show that they tend not to occur in the leads. When they do occur in the leads, they tend to occur in either secondary or embedded clauses.

The second caveat is that the information that is important to the meaning of a text is not always repeated in the text. There are cases in which the nucleus that is thought of as important to the meaning of the text in which it appears occur only once in the text. For example, in text 80 the nucleus *sixty-one police were injured* is thought of as important to the meaning of this text but it occurs only in the lead and is not repeated in the discourse after the lead. Observations of this kind of nuclei show that their sole occurrences also have a strong tendency to be found in the lead of the texts in which they occur.

The reiteration in a text of information that is important to the meaning of the text is a prominent feature of hard news texts (White, 1997). Another prominent feature of this text type is suggested in the discussion of the first caveat noted above, that is, the leads very often summarize the gist of the texts in which they occur (Bell, 1991; White, 1997; Van Dijk, 1988). The summarization function of the leads is noted by many authors, and is once again evidenced by my finding that the nuclei that are thought of as important to the meaning of the texts in which they appear have a strong tendency to occur in the leads. White (1997: 412)

characterizes hard news texts in terms of these two features, noting that this text type is characterized by the repetitions in the text of the point of newsworthy impact as originally set out in the headline and lead. His statement indicates that elevation to the headline and lead and reiteration are often employed by news producers to foreground the propositions that are important to the meaning of the texts in which they occur. However, his statement should not be taken as suggesting that the propositions that occur in the leads are all important to the meaning of the texts in which they occur. Tables 5.2 and 5.3 given in chapter 5 show that some nuclei that occur in the leads are thought of as important, whereas other nuclei that also occur in the leads are, on the other hand, not thought of as important to the meaning of the texts in which they occur. We will discuss information distribution in the leads in the section that follows.

6.4 Information distribution in the leads

My present study has found that the nuclei in the leads that are thought of by readers as important to the meaning of the texts in which they occur tend to occur in primary or independent clauses, whereas those that are not thought of as important to the meaning of the texts in which they occur tend to occur in secondary or embedded clauses. The finding has two implications. First of all, it indicates that information distribution in the leads is not random. The finding concurs with Matthiessen and Thompson's (1988: 289) statement that the ways in which information is distributed are determined by the writer's goals, that is, the propositions that realize the writer's central goals tend to occur in main clauses, whereas those that are supplementary or ancillary to the central goals tend to occur in subordinate clauses, and is also in line with Thompson's (1987) finding that the events that are central to the theme of a narrative tend not to be construed in hypotactically dependent and embedded clauses. Secondly, the finding indicates a third prominent feature of hard news texts, that is,

the propositions in a text that are important to the meaning of the text are not only often elevated to the lead, but also foregrounded in major types of clause.

However, it is important to note that the propositions in the leads that occur in secondary or embedded clauses are not always thought of by readers as peripheral to the meaning of the texts in which they occur. My present study found that some nuclei in the leads that are thought of as important to the meaning of the texts in which they appear occur in these two types of clauses. When the nuclei in the leads that occur in secondary clauses are thought of as important to the meaning of the texts in which they appear, it is often the case that the nuclei occurring in the primary clauses on which they are dependent are also thought of as important. The summaries written for the texts in which they occur show that the ways in which the secondary clause nuclei relate to the primary clause nuclei are also thought of as important to the meaning of the texts in which they occur (for the replications in the summaries of the linkages that occur in the leads, see section 4.3.2 of chapter 4 and sections 5.5.1, 5.5.2 and 5.5.3(a) of chapter 5). When the nuclei in the leads that occur in embedded clauses are thought of by readers as important to the meaning of the texts in which they appear, we often find that readers infer relations holding between them and the nuclei occurring in primary or independent clauses. In this situation, the relations inferred by readers that hold between them and the nuclei occurring in primary or independent clauses do not occur in the leads (For the creation by readers of relations that hold between the nuclei occurring in embedded clauses and those occurring in primary or independent clauses, see sections 5.5.4 and 5.5.5 of chapter 5). The creation by readers of linkages that do not occur in the leads shows that readers make inferences when they read. There are often some signals in a text that lead readers to make particular inferences (Hoey, 1979, 1983, 2001; McCarthy, 2002).

6.5 The strategies that summary writers often use to write summaries for hard news texts

Kintsch and Van Dijk (1983) and Van Dijk (1988) have noted three strategies that are often used by readers to write summaries, namely *deletion*, *generalization* and *construction*. Van Dijk (1988: 116) notes that *deletion* occurs when the information that is not further used as a presupposition for the interpretation of the rest of the text is deleted, *generalization* occurs when similar properties are relevant for different members of a set, and *construction* occurs when several partial acts or events are subsumed by an overall macroact or macroevent. My present study has found four more strategies, namely replication in summaries of linkages that occur in leads, creation in summaries of linkages that do not, however, occur in leads, nominalization and reduction. These four strategies are explained below.

In the last paragraph of the previous section, we have seen that the way in which two or more propositions in a lead are linked is often replicated in summaries. The replication in the summaries of the linkages that occur in the leads shows that the conjunctive relation holding between two or more propositions in a lead is important to the meaning of the text in which they occur. We will discuss in the section that follows in more detail the contribution of conjunctive relations holding between the propositions of a text to the meaning of the text. We have also seen that summary writers may create linkages that do not occur in leads. This occurs when two or more propositions do not stand in a relation in the original text, but summary writers infer a relation holding between them.

The two strategies noted above, namely replication in summaries of linkages that occur in leads and creation in summaries of linkages that do not, however, occur in leads, concern the ways in which two or more propositions that occur in the original text are represented in

summaries. The nominalization and reduction strategies, on the other hand, are often employed by summary writers to rewrite a proposition expressed by a clause contained in the original text in such a way that it is expressed by a nominal group in a summary. With the nominalization strategy, the processes of clauses in the original texts are reconstrued as the Things (i.e. the head nouns) of nominal groups in summaries, and incidentally the participants that are involved in the processes are either left out or reconstrued by summary writers as the modifiers of the Things in nominal groups. The analysis of summaries shows that this strategy is very often employed by readers, perhaps because it enables summary writers to represent propositions in nominal groups, which may function as elements of other propositions so that a summary consists of two or more propositions. With the reduction strategy, on the other hand, the processes of clauses in the original texts are reconstrued as the pre-modifiers of head nouns in nominal groups.

6.6 Conjunctive relations and the meaning of a text

It is suggested in the works of many linguists, such as Eggins (1994, 2004), Halliday and Hasan (1976), Halliday (1994), Halliday and Matthiessen (2004), Hoey (1983, 2001), Martin (1992), Martin and Rose (2003), Martin *et al* (2010), Mann and Thompson (1988a, 1988b), McCarthy (2002), Matthiessen and Thompson (1988), Quirk *et al* (1972, 1985), Thompson (1996, 2004, 2005, 2010) and Winter (1977, 1994), that conjunctive relations holding between the propositions contained in a text are important to the meaning of the text. But these works provided relatively little empirical evidence in support of this view. The view is, however, supported by my finding that when two or more nuclei in a text are thought of by readers as important to the meaning of the text, the relationship in the lead that links them is also thought of as important to the meaning of the text.

The works of psychologists concerning the effects of conjunctive relations on the recall of the events they connect also contribute to our understanding of the relationships between the conjunctive relations that hold between the propositions contained in a passage and the meaning of the passage. Causal relations have been explored more thoroughly than other kinds of relation. It has been found that the recall probabilities of events linked in causal relations are better than that of events linked in other kinds of relation or not linked in any way (Black and Bern, 1981; Bradshaw and Anderson, 1982; Caron *et al*, 1988; Van den Broek and Lorch, 1993). Black and Bern (1981) is widely considered as the first work that drew attention to the prominent effects of causal relations on the recalls of events. They found that 'causally related events in narratives were remembered better than events that were not causally related' (1981: 267). The prominent effects of causal relations on the recalls of events were also confirmed by other researchers. For instance, Caron *et al* (1988) found that the second sentence, cued by the first, was much better recalled when the two sentences were connected by 'because' or when subjects thought of a causal relationship between them than when they are connected by 'and' or 'but'. They (1988: 320) also noted that causal relation was more effective than other kinds of relation in triggering in readers the construction of a coherent framework. The findings of these works suggest that the association of events linked in a causal relation is stronger than that of events linked in other kinds of relation. This view is supported by the finding reported in Van den Broek *et al* (2000) and Simner and Pickering (2005) that when the informants were asked to write a sentence to follow the one they have just read, they tended to compose a sentence that depicted an event that was causally related to the one depicted in the sentence they had read.

The prominent effects of causal relations on the recalls of events have attracted many other psychologists to explore the characteristics of causal relations. Keenan *et al* (1984) and

Myers *et al* (1987) examined the relationships between the degrees of causal relatedness between events and the recall probabilities of the events. Both studies found that the recall of an event that is linked to a cue sentence in an intermediate degree of causality was much better than those linked to the cue sentence in high or low degrees of causality. The reason for the superior recall of the event linked to the cue sentence in an intermediate degree of causality may be that readers need to generate a proposition that enables them to perceive a causal relation between the recalled sentence and the cue sentence, and elaborative processing results in better memory (see Anderson and Reder, 1979). On the other hand, the recall of the event linked to the cue sentence in a high degree of causality is less effective because the recalled event is implied by the cue sentence and therefore minimal processing was necessary for readers to think of a causal linkage between them. The recall probability of the event linked to the cue sentence in a low degree of causality is lowest because the relation between it and the cue sentence is too tenuous to facilitate memory (Keenan *et al*, 1984: 125).

Trabasso and his colleagues have examined the relationships between the number of causal relations that an event holds with other events in a narrative and the reader's evaluation of the importance of the event. They firstly divide a story into statements. Each statement is expressed by a clause, and depicts an event. After that, they consider all pairs of statements and identify the pairs in which a causal relation holds between the two statements. They then construct a causal network for each story that represents the causal linkages between all pairs of events contained in the story, and identify the causal chain in the network. A causal chain begins with the statement that introduces the protagonist, sets the time and locale and initiates the story's action, and ends with the statement that indicates the protagonist' accomplishment of his/her goal when the goal is accomplished, or the consequence of the protagonist' failure to achieve his/her goal when the goal is not accomplished. Thus, the causal chain is made up

of the opening and closing statements of the chain and the statements after the opening statement that are successively linked in such a way that they eventually lead to the closing statement. The statements that do not occur in the causal chain are referred to by Trabasso and his colleagues as dead ends. Dead end statements are either not linked to any other statement in the story (i.e. they do not occur in the causal network), or although they occur in the causal network, they do not eventually lead to the closing statement of the causal chain. It is found that both the recall probability and readers' rating of degree of importance of an event to the meaning of the text in which it occurs are associated with whether it occurs in the causal chain of the story, and the number of causal relations that it has with other statements in the chain. Statements that are in the causal chain are recalled much more often than those that are not in the causal chain; and the more causal linkage a statement has with other statements, the more often it is recalled (Trabasso and Sperry, 1985; Trabasso and Van den Broek, 1985; Van den Broek, 1988, 1990). The statements that are in the causal chain of a story are thought of by readers as more important to the meaning of the story than those that are not in the chain, and causal chain statements that have many causal linkages to other statements are thought of by readers as more important to the meaning of the text in which they occur than those that have few causal linkages to other statements (Trabasso and Sperry, 1985; Trabasso and Van den Broek, 1985, 1986).

In the light of Trabasso *et al*'s finding that causal chain statements are important to the meaning of a text, we may predict that when a nucleus occurring in the secondary clause of a lead are causally linked to the primary clause nucleus, the secondary clause nucleus is more important to the meaning of the text than any other nuclei in the text that are not causally linked to the primary clause nuclei. This prediction is confirmed by the examples discussed in section 5.5.1 of chapter 5, where we saw that in each example, the secondary clause nucleus

that is causally linked to the primary clause nucleus occur in the summaries, while any other nucleus in the text that is not causally linked to the primary clause nucleus do not occur in the summaries.

Trabasso *et al*'s works indicate that in a narrative the causal relations holding between the statements are important to the coherence and meaning of the text. A similar finding has been noted by other authors for non-narrative texts, that is, the conjunctive relations holding between the parts of a non-narrative text are important to the coherence and meaning of the text. The implication is that there is no stark boundary line between narrative and non-narrative texts. There are two more pieces of evidence that show the boundary is vague. One is given by Hoey's (2001: 187) finding that the culturally popular patterns of text organization apply promiscuously to narrative and non-narrative text alike. The other reason for believing the boundary is vague is that both narrative and non-narrative texts can be characterized in terms of the relations that hold between adjacent and non-adjacent parts contained in a text. Trabasso *et al* found that a causal relation holds between adjacent and also non-adjacent statements of a narrative. Similarly, Thompson *et al*'s analysis of text in terms of rhetorical structure show that a rhetorical relation holds between adjacent and also non-adjacent text spans in a non-narrative text.

Their works also suggest that readers often make of the relations that hold between the propositions of a text so as to understand the coherence of the text. This view is supported by my findings that the way in which two or more propositions in a lead are linked is often replicated in summaries, and that summary writers may create linkages that do not, however, occur in leads. One more piece of evidence in support of the view is given by Mann and Thompson's (1988a, 1988b) analysis of the rhetorical relations holding between the parts of

a text. They show that a coherent text can be characterized by the rhetorical relations holding between the parts of the text, and that a text in which the relations holding between its parts are conceivable is more coherent than a text in which the relations holding between its parts are difficult to conceive.

6.7 Future research avenues

Future research may embark on four avenues. My present study has examined the characteristics of the recurrent nuclei in a text that occur in a majority of summaries. Future research may examine the language phenomena that occur in a majority of summaries of a text but do not occur in the text itself. One may wish to identify the nuclei that occur in a majority of summaries of a text but do not occur in the text, and then examine the content in the text that is summarized by the nuclei contained in the summaries. Text 78 is an example in which there is a nucleus, *punish misbehaving/errant teachers*, which occurs in the summaries but does not occur in the text itself. In section 4.4 of chapter 4, I have shown that the examination of the content in the text that is summarized by *punish misbehaving/errant teachers* reveals a way in which readers summarize the text.

We however need to be aware of a defect of using summaries to identify important information. There may be cases in which readers remember a convincing example given in the text they read that makes them believe in the point the example illustrates, but the example is too specific to be included in the summaries. If this happens, the summaries would not contain the example that is nonetheless important to the meaning of the text.

One may also wish to explore whether my finding that when two or more nuclei in a lead are repeated in the summaries of the text, the relationship in the lead that holds between them is

also replicated in the summaries also holds if one removes the lead from each of the hard news texts that one collects for his/her study, and asks informants to write summaries to give what they think of as the gist of the remaining text after the lead. The replication in the summaries of the linkage that occurs in the lead as noted in my finding may be caused by the possibility that summarizers believe the lead summarizes the text that they read and replicate what is given in the lead. It is interesting to explore whether the finding also holds if informants read and write summaries for the discourse after the lead.

It is also worth exploring the strength of association produced by various kinds of conjunctive relations in linking events. Black and Bern (1981), Bradshaw and Anderson (1982), and Caron *et al* (1988) found that the association of events linked in causal relations is much stronger than that of events linked in other kinds of relations, suggesting that different kinds of conjunctive relations may produce varying degrees of association between events. Previous studies, however, have not compared the strengths of association produced by other kinds of relations than causality in linking events. In order to do so, one may collect for the different kinds of conjunctive relations under consideration an equal number of hard news texts, with the lead of each text containing a linkage that holds between two or more propositions. After that one analyzes the leads and the summaries written by informants for the texts in terms of nuclei, and examines the probability of replication in the summaries of linkages that occur in the leads.

Another path that future research may take is analyzing texts in terms of both Hoey's (1991) method and the method proposed by Trabasso and his colleagues with a view to evaluating the two approaches. The two approaches are complimentary because Winter (1977, 1994) classifies clause relations into two major types, matching and sequence, and Hoey's approach

analyzes the matching relations between sentences in a text, while Trabasso *et al*'s approach analyzes the sequence relations holding between the propositions of a text. Both approaches identify propositions that are important to the meaning of a text. It is interesting to evaluate both approaches in terms of the questions given below.

1. How often the propositions that are identified by an approach as important are indeed also judged by readers as important to the meaning of a text.
2. Are there any propositions that are judged by readers as important to the meaning of a text but fail to be identified by an approach as important? If so, why does the approach fail to identify them as important?

Exploring the first question identifies the probabilities of both approaches in successfully identifying propositions that are important to a text's meaning. Exploring the second question identifies their weaknesses.

There is still a great deal of work to be done on the relationship between text production and text comprehension. I have made only a start on this work, and I invite you to join me.

References

Anderson, J. R. and Reder, L. M. 1979. 'An elaborative processing explanation of depth of processing', in L. Cermak and F. Craik (eds.), *Levels of Processing in Human Memory*. Hillsdale, NJ: Erlbaum, pp385-403.

Barzilay, R. and Elhadad, M. 1997. 'Using lexical chains for text summarization'. in Mani, I. and Maybury, M (eds.) (1999), *Advances in Automatic Text Summarization*. Massachusetts: The MIT Press, pp111-121.

Bell, A. 1991. *The Language of News Media*. Oxford: Blackwell.

Benbrahim, M. and Ahmad, K. 1994. *Computer-Aided Lexical Cohesion Analysis and Text Abridgement*. Computing Sciences Report CS-94-11. Guildford (UK): University of Surrey.

Berber Sardinha, A. P. 1997. *Automatic Identification of Segments in Written Text*. Unpublished PhD thesis, University of Liverpool.

Berber Sardinha, A. P. 2001. 'Lexical segments in text', in Scott, M. and Thompson, G (eds.) (2001), *Patterns of Text: in Honour of Michael Hoey*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp213-237.

Black, J. B. and Bern, H. 1981. 'Causal coherence and memory for events in narratives'. *Journal of Verbal Learning and Verbal Behavior*, 20, pp267-275.

Bloor, T. and Bloor, M. 1995. *The Functional Analysis of English: A Hallidayan Approach*. Beijing: Foreign Language Teaching and Research Press.

Bradshaw, G. L. and Anderson, J. R. 1982. 'Elaborative encoding as an explanation of levels of processing'. *Journal of Verbal Learning and Verbal Behavior*, 21, pp165-174.

Brown, G. and Yule, G. 1983. *Discourse Analysis*, Cambridge: Cambridge University Press.

Caron, J., Micko, C., and Thüring, M. 1988. 'Conjunctions and the recall of composite sentences'. *Journal of Memory and Language*, 27, pp309-323.

Carter, R. and McCarthy, M. 1988. 'Lexis and discourse: vocabulary in use', in Carter, R. and McCarthy, M. (eds.) (1988), *Vocabulary and Language Teaching*. London and New York: Longman, pp201-220.

Eggins, S. 1994. *An Introduction to Systemic Functional Linguistics* (1st edition). London: Pinter Publishers.

Eggins, S. 2004. *An Introduction to Systemic Functional Linguistics* (2nd edition). London: Continuum.

Francis, G. 1994. 'Labelling discourse: an aspect of nominal-group lexical cohesion'. in M. Coulthard (ed.), *Advances in Written Text Analysis*, London and New York: Routledge. pp83-101.

Francis, G., Hunston, S., and Manning, E. 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar* (2nd edition). Beijing: Foreign Language Teaching and Research Press.

Halliday, M.A.K. 1998. ''Things and relations: Regrammaticising experience as technical knowledge'', in Martin, J. R. and Veel, R. (eds.), *Reading Science: Critical and Functional Perspectives on Discourse of Science*. London and New York: Routledge, pp185-235.

Halliday, M.A.K. 1999. "The grammatical construction of scientific knowledge: the framing of the English clause", in J. Webster (ed.), *The Language of Science*. London and New York: Continuum, pp102-134.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. Beijing: Foreign Language Teaching and Research Press.

Halliday, M.A.K. and Hasan, R. 1985. *Language, Context, and Text: Aspects of Language in a Social Semiotic Perspective* (2nd edition). Oxford: Oxford University Press.

Halliday, M.A.K. and Matthiessen, C.M.I.M. 1999. *Construing Experience through Meaning: a Language-Based Approach to Cognition*. London and New York: Continuum.

Halliday, M.A.K. and Matthiessen, C.M.I.M. 2004. *An Introduction to Functional Grammar* (3rd edition). London: Hodder Arnold.

Hasan, R. 1984. 'Coherence and cohesive harmony', in J. Flood (ed.), *Understanding Reading Comprehension*. Delaware: International Reading Association, pp181-219.

Heyvaert, L. 2003. 'Nominalization as grammatical metaphor: on the need for a radically systemic and metafunctional approach', in Anne-Marie Simon-Vandenbergen, Miriam Taverniers and Louise J. Ravelli (eds.) (2003), *Grammatical Metaphor: Views from systemic functional linguistics*. Amsterdam/Philadelphia: John Benjamins, pp65-99.

Hoey, M. 1979. *Signalling in Discourse*. Discourse Analysis Monographs, Birmingham: ELR, University of Birmingham.

Hoey, M. 1983. *On the Surface of Discourse*. London: George Allen and Unwin. republished (1991) by English Studies Unit, University of Nottingham.

Hoey, M. 1986. 'Overlapping patterns of discourse organization and their implications for clause relational analysis of Problem-Solution texts', in C. R. Cooper and S. Greenbaum (eds.), *Studying Writing: Linguistic Approaches*. London: Sage, pp187-214.

Hoey, M. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. 1994. 'Signalling in discourse: a functional analysis of a common discourse pattern in written and spoken English', in M. Coulthard (ed.), *Advances in Written Text Analysis*. London and New York: Routledge, pp26-45.

Hoey, M. 2001. *Textual Interaction: an Introduction to Written Discourse Analysis*. London and New York: Routledge.

Hoey, M. and O'Donnell, M. B. 2007. 'The beginning of something important...'. presentation given in the Aston Corpus Symposium (4 May 2007).

Hoey, M. and O'Donnell, M. B. 2008. 'Lexicography, grammar, and textual position'. *International Journal of Lexicography*, Vol. 21, No.3, pp293-309.

Keenan, J, M., Baillet, S, D., and Brow, P. 1984. 'The effects of causal cohesion on comprehension and memory'. *Journal of Verbal Learning and Verbal Behavior*, 23, pp115-126.

Kintsch, W. and Keenan, J. M. 1974. 'Recall of propositions as a function of their position in the hierarchical structure', in Kintsch, W (1974) *The Representation of Meaning in Memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, pp137-140.

Kintsch, W. and Van Dijk, T.A. 1978. 'Toward a model of text comprehension and production'. *Psychology Review* 85, pp363-394.

McCarthy, M. 1988. 'Some vocabulary patterns in conversation', in Carter, R. and McCarthy, M. (eds.) (1988), *Vocabulary and Language Teaching*. London and New York: Longman. pp181-200.

McCarthy, M. 2002. *Discourse Analysis for Language Teachers*. Shanghai: Shanghai Foreign Language Education Press.

Mani, I. and Maybury, M. (eds.). 1999. *Advances in Automatic Text Summarization*. Massachusetts: The MIT Press.

Mann, B. 1999. 'An introduction to Rhetorical Structure Theory (RST)'. Assessed online at http://www.di.uniba.it/intint/people/fior_file/INTINT05/RST.pdf.

Mann, W. and Thompson, S. 1988a, 'Rhetorical structure theory: a theory of text organization', in Livia Polanyi (ed.), *The Structure of Discourse*. Norwood, N.J.: Ablex. Also available as ISI/RR-87-100.

Mann, W. and Thompson, S. 1988b. 'Rhetorical structure theory: Toward a functional theory of text organization'. *Text* 8(3), pp243-281.

Marcu, D. 1999. 'Discourse trees are good indicators of importance in text', in Mani, I. and Maybury, M. (eds.) (1999), *Advances in Automatic Text Summarization*. Massachusetts: The MIT Press, pp123-136.

Martin, J. R. 1989. *Factual writing: exploring and challenging social reality* (2nd edition). Oxford: Oxford University Press.

Martin, J. R. 1992. *English Text: System and Structure*. Beijing: Peking University Press.

Martin, J. R. 1993. "Life as a noun: arresting the universe in science and humanities", in Halliday, M.A.K. and Martin, J. R (eds.), *Writing Science: Literacy and Discursive Power*. London: Falmer Press, pp221-267.

Martin, J. R. 1997. 'Analyzing genre: functional parameters', in F. Christie and J. R. Martin (eds.), *Genre and Institutions: Social Processes in the Workplace and the School*. London: Cassell, pp3-39.

Martin, J. R., Matthiessen, C.M.I.M., and Painter, C., 2010. *Deploying Functional Grammar*. Beijing: The Commercial Press.

Martin, J. R. and Rose, D. 2003. *Working with Discourse: Meaning beyond the Clause*. Beijing: Peking University Press.

Matthiessen, C.M.I.M. 2002. 'Combining clauses into clause complexes: a multi-faceted view', in J. Bybee and M. Noonan (eds.), *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*. Amsterdam: John Benjamins.

Matthiessen, C.M.I.M. and Thompson, S. A. 1988. 'The structure of discourse and 'subordination'', in J. Haiman and S. A. Thompson (eds.), *Clause Combining in Grammar and Discourse*. Amsterdam: Benjamins, pp275-329.

Maury, P. and Teisserenc, A. 2005. 'The role of connectives in science text comprehension and memory'. *Language and Cognitive Processes*, 20/3, pp489-512.

Myers, J., Shinjo, M., and Duffy, S. 1987. 'Degree of causal relatedness and memory'. *Journal of Memory and Language*, 26/4. pp453-465.

Painter, C., Derewianka, B., and Torr, J. 2007. 'From microfunction to metaphor: learning language and learning through language', in R. Hasan., C.M.I.M. Matthiessen., and J. Webster (eds.) (2007), *Continuing Discourse on Language*, Vol.2, pp563-588.

Parsons, G. 1991a. *Cohesions and Coherence: Scientific Texts*. Nottingham: University of Nottingham, Department of English Studies.

Parsons, G. 1991b. 'Cohesion and coherence: scientific texts', in Ventola, E. (ed.), *Functional and Systemic Linguistics: Approaches and Uses*. Berlin: Mouton de Gruyter, pp415-429.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1972. *A Grammar of Contemporary English*. London and New York: Longman.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Reder, L. M. and Anderson, J. R. 1980. 'A comparison of texts and their summaries: memorial consequences'. *Journal of Verbal Learning and Verbal Behaviour* . 19. pp121-134.

Scott, M. 1996. *WordSmith Tools Version 1*, Oxford: Oxford University Press.

Scott, M. 1997. "PC Analysis of Key Words -- and Key Key Words". *System*. Vol. 25. No. 2 pp233-245.

Scott, M. 2000. 'Focusing on the Text and Its Key Words', in L. Burnard and T. McEnery (eds.), *Rethinking Language Pedagogy from a Corpus Perspective*, Vol. 2. Frankfurt: Peter Lang, pp103-122.

Scott, M. 2001, "Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs", in M. Ghadessy, A. Henry, and R.L. Roseberry (eds.) *Small corpus studies and ELT: theory and practice*. Amsterdam: Benjamins, pp47-67.

Scott, M. 2004. *WordSmith Tools Version 4 Manual.* Oxford: Oxford University Press.

Scott, M. 2008. *WordSmith Tools Version 5*. Liverpool: Lexical Analysis Software.

Scott, M. 2010. *WordSmith Tools Help*. Liverpool: Lexical Analysis Software.

Scott, M. and Thompson, G (eds). 2001. *Patterns of Text: in Honour of Michael Hoey*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Scott, M. and Tribble, C. 2006. *Textual Patterns: Keyword and Corpus Analysis in Language Education*. Amsterdam: Benjamins.

Simner, J. and Pickering, M. 2005. 'Planning causes and consequences in discourse'. *Journal of Memory and Language*, 52, pp226-239.

Sinclair, J. McH. 1993. 'Written discourse structure', in Sinclair (2005), *Trust the Text: Language, Corpus and Discourse.* London and New York: Routledge, pp82-101.

Sinclair, J. McH *et al.* 2001. *Collins COBUILD English Dictionary for Advanced Learners*. London: HarperCollins Publisher.

Swales, J. 1990. *Genre Analysis: English in Academic and Research settings*. Cambridge: Cambridge University Press.

Tang, Q. Y. 2006. *Packet Nouns and Information Packaging in Discourse*. Shanghai: Shanghai University Press.

Thompson, G. 1994. *Collins COBUILD English Guides 5: Reporting*. London: HarperCollins Publisher.

Thompson, G. 1996. *Introducing Functional Grammar* (1st edition). Beijing: Foreign Language Teaching and Research Press.

Thompson, G. 2003. 'The elided participant: presenting an uncommonsense view of the researcher's role', in Anne-Marie Simon-Vandenbergen, Miriam Taverniers and Louise J. Ravelli (eds.) (2003), *Grammatical Metaphor: Views from systemic functional linguistics*. Amsterdam/Philadelphia: John Benjamins, pp257-278.

Thompson, G. 2004. *Introducing Functional Grammar* (2nd edition). London: Arnold.

Thompson, G. 2005. 'But me some buts: A multi-dimensional view of conjunction'. *Text* 25(6), pp763-791.

Thompson, G. 2010. 'Conjunctive relations', presentation given in the LinC Summer School in Systemic Functional Linguistics, The University of Cardiff.

Thompson, S. A. 1987. ''Subordination' and narrative event structure', in Tomlin. R. S (ed.) (1987), *Coherence and Grounding in Discourse*. Amsterdam: Benjamins.

Trabasso, T. and Sperry, L. 1985. 'Causal relatedness and importance of story events'. *Journal of Memory and Language*, 24, pp 595-611.

Trabasso, T. and Van den Broek, P. 1985. 'Causal thinking and the representation of narrative events'. *Journal of Memory and Language*, 24, pp 612-630.

Trabasso, T. and Van den Broek, P. 1986. 'Causal networks versus goal hierarchies in summarizing text'. *Discourse Processes*, 9, pp1-15.

Van den Broek, P. 1988. 'The effects of causal relations and hierarchical position on the importance of story statements'. *Journal of Memory and Language*, 27, pp1-22.

Van den Broek, P. 1990. 'Causal inferences and the comprehension of narrative texts', in A. C. Graesser and G. H. Bower (eds.) *Inferences and Text Comprehension*. San Diego: Academic Press, pp175-196.

Van den Broek, P. and Lorch, Robert F. Jr. 1993. 'Network representations of causal relations in memory for narrative texts: evidence from primed recognition'. *Discourse Processes*, Vol. 16, Issue 1&2, pp75-98.

Van den Broek, P., Linzie, B., Fletcher, C., and Marsolek, C. 2000. 'The role of causal discourse structure in narrative writing'. *Memory and Cognition*, 28(5). pp711-721.

Van Dijk, T. A. 1986. 'News schemata', in C. Cooper and S. Greenbaum (eds.). *Studying Writing: Linguistic Approaches*. Beverly Hills, CA: Sage, pp155-186.

Van Dijk, T. A. 1988. *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum.

Van Dijk, T. A. and Kintsch, W. 1983. *Strategies of Discourse Comprehension.* New York: Academic Press.

Ventola, E. 1987. *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*. London: Frances Pinter.

White, P. 1997. 'Death, disruption and the moral order: the narrative impulse in mass-media "hard news" reporting', in Toolan, M (ed.) (2007), *Critical Discourse Analysis: Critical Concepts in Linguistics*, Vol. 3, London and New York: Routledge, pp390-424.

Widdowson, H. G. 1978. *Teaching Language as Communication.* Oxford: Oxford University Press.

Widdowson, H. G. 2004. *Text, Context, and Pre-text: Critical issues in Discourse Analysis.* Oxford: Blackwell.

Winter, E. O. 1977. 'A clause relational approach to English texts: a study of some predictive lexical items in written discourse'. *Instructional Science* 6/1, pp1-92.

Winter, E. O. 1994. 'Clause relations as information structure: two basic text structures in English', in M. Coulthard (ed.), *Advances in Written Text Analysis.* London and New York: Routledge, pp46-68.