

The Informational Role of Nominal Expressions in English Texts

Von der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades einer Doktorin der Philosophie genehmigte Dissertation

vorgelegt von

Hongying Xu

Berichter: Univ.-Prof. Dr. phil. Stella Neumann
Dr. Lise Fontaine

Tag der mündlichen Prüfung: 03.02.2023

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

To my father

Abstract

This thesis explores the informational role of nominal expressions in English texts by taking a multi-dimensional approach. While the field of nominal expressions has been studied from a variety of perspectives, their informational role in text however remains a relatively underdeveloped area. The main goal of this thesis is to examine how nominal expressions convey information values in different texts.

The research presented in this thesis investigates four main areas: (i) information distributions of nominal expressions in English texts; (ii) the relationship between linguistic forms of nominal expressions and type of information status they represent in English texts; (iii) the interplay between information status of nominal expressions and thematic structures of English texts; and (iv) the interaction between information status of postmodifiers and integration in meaning in English texts.

This thesis combines text analysis from a functional-discourse perspective such as Systemic Functional Linguistics with corpus linguistics and cognitive linguistics methodologies. It first reconsiders Prince's (1981) theoretical classifications of information status and then conducts a corpus analysis of 3095 nominal expressions. The corpus contains four comparable genres that are selected from the American National Corpus, namely Travel guide, News report, Government document and Essay.

This research has the following findings: (i) information distributions of nominal expressions reveal differences between the texts due to varied cognitive and situational characteristics; (ii) linguistic forms of nominals show multiple correspondences to types of information values they convey in the texts; (iii) information distributions of nominals in Theme and Rheme positions are different from those of the complete texts; and (iv) postmodifiers of complex nominals show varied tendencies to convey information values when they are involved in different types of integration in meaning.

This thesis shows how nominal expressions structure information values at the levels of group/phrase, clause and text and provides empirical evidence that the

informational role of nominal expressions is an important index of English text types.

Acknowledgements

Pursuing PhD was completely Brand-new information to me, when I first set out on this journey several years ago. Having nearly finished the last stage of this long, challenging and fascinating journey, I finally can express my sincere gratitude to the people who are so supportive, insightful, and important throughout.

First and foremost, I want to express my heartfelt thanks to my co-supervisors, Dr. Lise Fontaine at Cardiff University and Prof. Dr. Stella Neumann at RWTH Aachen University (alphabetical order), for your constant support, encouragement and suggestions. Any mistakes in this thesis of course are my responsibility alone, but the success of this thesis is the result of countless times of discussions I have had with both Lise and Stella whose contributions are extremely valuable. Dear Lise and dear Stella, thank you so much! Both of you have taught me how to do good science and have made me a better person. I could not have wished better supervisors!

I would like also express my gratitude to Professor Peng Xuanwei. He has shown me how fascinating it is in doing research. Many thanks for his patience, guidance and supervision during my master studies. I have learned so much from him.

This thesis would not have been possible without the help from the proofreaders and also my dear friends. They are Dr. Zhu Xiaotong, Dr. Liu Zhiping, Prof. Dr. Zhang Mingjin, Ms Celina Brost and Ms Alessa Busch. Thank you so much for your helpful comments on early versions of the chapters of this thesis. Your support is very important for my doctoral studies. I have been really lucky to be friends with you great people.

Special thanks also go to Ms Ina Thiemann, Dr. Ekaterina Lapshinova-Koltunski and Dr. Mark-Christoph Müller for helping me use the annotation tool of this thesis; and Dr. Stephan Baumann for answering my long list of questions concerning classifying information.

I am also deeply grateful to Jürgen Osthoff, the librarian of English and American studies library. Jürgen, thank you for being my family and my dear friend in Germany. All cups of coffee you've made and all conversations we've had everyday become part

of the best memories I have ever had during my doctoral studies. Jürgen, thank you for making our library my favorite place, especially the linguistic room where I have studied nearly every day in the past few years.

I would like to thank my close friends without naming them individually. It is important for me to have them throughout my life. They are always patient, supportive and helpful to relieve my anxiety. Thank you for keeping me motivated, focused, and confident.

In addition, I want to express my gratitude to my colleagues for their valuable advices and feedbacks in our colloquiums; and the Erasmus+ Student Exchange program for offering me a scholarship to study one semester at Cardiff University.

Most of all, I owe my family a great debt of gratitude. Both my parents and my sister support and encourage me throughout my life. They make me believe that I have the potential to do anything I want to. That's why someone like me who was born and raised in a very small town of Northeast China managed to do doctoral studies here. I have great parents who have showed me how to be a person with determination, strength, and kindness. Thank you mom and dad for always being my rock and my shield. You are the pillar of my life. I am also very grateful that you have given me a sister. I am so lucky to always have her by my side. Thank you for being my precious sister and my best friend. Thank you for taking care of our parents and me. Thank you for your constant love and support! If my father were alive today, he would be very proud. I am so sorry that I have not got to this point soon enough. I would like to dedicate this thesis to my father.

Contents

Abstract	i
Acknowledgements	iii
List of Tables	ix
List of Figures	xii
List of Abbreviations	xiii
1 Introduction.....	1
1.1 Goal of this thesis.....	3
1.2 Overview of the thesis.....	5
2 State of the Art.....	7
2.1 The notion of information	7
2.2 Main frameworks of classifying information status.....	11
2.2.1 Classifying information status from the perspective of speakers	11
2.2.2 Assumed Familiarity by Prince (1981).....	13
2.2.3 Classifying information status from multiple perspectives	25
2.3 Linguistic forms of nominals expressing types of information values	31
2.4 Information status of nominal expressions and thematic structure	35
2.5 NP-complexity and integration in meaning.....	37
2.6 Conclusion: Motivating a multi-dimensional approach to the informational role of nominal expressions in English texts	40
3 Methodology: A multi-dimensional framework	42
3.1 Annotated data of four genres from MASC	42
3.1.1 The manually annotated sub-Corpus from the American National Corpus (MASC)	42
3.1.2 Annotated data of the four genres selected from MASC.....	44

3.2 A multi-dimensional framework	46
3.2.1 Annotation attributes for nominal expressions	46
3.2.2 Annotation attributes for information status and thematic structure	62
3.3 The annotation tool.....	73
3.4 Summary	75
4 Information distributions of nominal expressions in texts.....	76
4.1 Information distributions of nominal expressions of individual categories.....	76
4.1.1 Brand-new unanchored	79
4.1.2 Brand-new anchored	81
4.1.3 Non-containing Inferrable.....	84
4.1.4 Containing Inferrable.....	89
4.1.5 Unused	91
4.1.6 Situationally evoked	94
4.1.7 Textually evoked	96
4.2 Information distributions of nominal expressions of all categories	102
4.3 Summary	107
5 Linguistic forms of nominal expressions and types of information status	110
5.1 Outline of various nominal expressions in the dataset.....	111
5.2 Linguistic forms of nominals and information status they convey in the texts	117
5.2.1 Linguistic forms of nominals conveying Brand-new unanchored information	117
5.2.2 Linguistic forms of nominals conveying Brand-new anchored information	131
5.2.3 Linguistic forms of nominals conveying non-containing Inferrable information	134

5.2.4 Linguistic forms of nominals conveying containing Inferrable information	143
5.2.5 Linguistic forms of nominals conveying Unused information	145
5.2.6 Linguistic forms of nominals conveying Situationally evoked information	147
5.2.7 Linguistic forms of nominals conveying Textually evoked information	149
5.3 Summary	159
6 The interplay between information status of nominal expressions and thematic structures	162
6.1 Relationship between texts and thematic structures in information distributions of nominal expressions	163
6.2 Information distributions of nominal expressions in Theme and Rheme positions	166
6.2.1 Information distributions of Theme	167
6.2.1.1 Participant	174
6.2.1.2 Circumstance	178
6.2.2 Information distributions of Rheme	189
6.2.3 Contrastive analysis of Theme and Rheme in information distributions of nominal expressions	199
6.3 Summary	205
7 The interaction between information status of complex nominal expressions and integration in meaning	207
7.1 Outline of postmodifiers of the corpus dataset	208
7.2 Information status of postmodifiers in types of expansion	211
7.3 Summary	222
8 Conclusion	225

8.1 Main contributions of this thesis	225
8.2 Limitations of the research	226
8.3 Directions for further research	227
References.....	229
Appendix.....	248
Appendix A: Raw frequencies of varied linguistic forms of nominal expressions representing types of information in English texts.....	248
Appendix B: Information distributions of nominal expressions and thematic structures	254
Appendix C: Raw frequencies of postmodifiers representing types of information in different types of expansion	259
Appendix D: Source texts	261

List of Tables

Table 2–1: Annotation labels of the <i>r</i> -level (Riester and Baumann 2017: 5)	28
Table 2–2: Annotation labels of the <i>l</i> -level (Riester and Baumann 2017: 22)	30
Table 2–3: Logico-semantic relations of complex nominals adapted from Halliday and Matthiessen (2014)	38
Table 3–1: Basic information of the data for final annotation	45
Table 3–2: Annotation attributes for nominal expressions	47
Table 3–3: Annotation attributes for information status and thematic structure	62
Table 4–1: Information distributions of nominal expressions in the texts	78
Table 4–2: The distributions of Brand–new unanchored information per text	79
Table 4–3: The distributions of Brand–new anchored information per text	81
Table 4–4: The distributions of non–containing Inferrables per text	84
Table 4–5: The distributions of containing Inferrables per text	89
Table 4–6: The distributions of Unused information per text	91
Table 4–7: The distributions of Situationally evoked information per text	94
Table 4–8: The distributions of Textually evoked information per text	96
Table 4–9: Key factors of the given–new chains per text	98
Table 4–10: Ratios of the shortest given–new chains to all the chains per text	100
Table 4–11: Ratios of non–nominal antecedents to Textually evoked items per text	101
Table 4–12: Information distributions of the four genres	105
Table 5–1: Types of nominal expressions per text	113
Table 5–2: Types of pronouns per text	114
Table 5–3: Types of nominal expressions with determiners per text	115
Table 5–4: Types of nominals expressing Brand–new unanchored information per text	118
Table 5–5: Types of determiners of expressing Brand–new unanchored information per text	128
Table 5–6: Types of nominals expressing Brand–new anchored information per text	

.....	132
Table 5–7: Types of nominals expressing non–containing Inferrable information per text.....	135
Table 5–8: Types of determiners of expressing non–containing Inferrable information per text.....	139
Table 5–9: Types of nominals expressing containing Inferrable information per text.....	143
Table 5–10: Types of nominals expressing Unused information per text.....	145
Table 5–11: Types of nominals expressing Situationally evoked information per text.....	148
Table 5–12: Types of nominals expressing Textually evoked information per text...	149
Table 5–13: Types of pronouns expressing Textually evoked information	154
Table 5–14: Types of determiners expressing Textually evoked information per text.....	156
Table 6–1: The number of texts verifying the hypothesis of correlation	165
Table 6–2: The number of genres verifying the hypothesis of correlation	165
Table 6–3: The distributions of nominal expressions in Theme and Rheme positions per text.....	167
Table 6–4: Information distributions of nominals in Theme positions per text.....	168
Table 6–5: Average information distributions of Themes in the Travel guide and News report texts.....	169
Table 6–6: Information distributions of nominals as Participants in Theme positions per text.....	175
Table 6–7: Average information distributions of Participants in the Travel guide and News report texts.....	176
Table 6–8: Information distributions of nominals as Circumstances in Theme positions per text.....	179
Table 6–9: Average information distributions of Circumstances in the Travel guide and News report texts.....	180
Table 6–10: Information distributions of nominal expressions in Rheme positions per	

text	189
Table 6–11: Average information distributions of Themes in the Travel guide and News report texts	191
Table 6–12: Information distributions of nominal expressions in Theme and Rheme positions per text.....	198
Table 7–1: Distributions of postmodifiers in types of expansion per text	208
Table 7–2: Information status of postmodifiers in elaboration and the remaining types of expansion per text	213
Table 7–3: Information status of postmodifiers in enhancement per text.....	221

List of Figures

Figure 2–1: Assumed Familiarity (Prince 1981: 237).....	14
Figure 2–2: The system of Identifiability (Lambrecht 1994: 109)	26
Figure 2–3: The Givenness Hierarchy (Gundel <i>et al.</i> 1993).....	32
Figure 3–1: The MMAX2 Attribute Window and Text Window.....	73
Figure 3–2: The MMAXQL Query Console of Markable Tuples	74
Figure 3–3: MMAXQL Query Console of Statistics	74
Figure 4–1: The given–new chains in Np2 and Tg1	99
Figure 4–2: Information distributions of the ten texts	103
Figure 5–1: Proper nouns of Np3 Np4 and Np5 expressing Brand–new unanchored information	120
Figure 5–2: Proper nouns of Tg1 and Tg1 expressing Brand–new unanchored information	122
Figure 5–3: Proper nouns of Np6 Essay and Government document expressing Brand– new unanchored information.....	123
Figure 5–4: The remaining types of nominals expressing Brand–new unanchored information in Tg2 and Np6	130
Figure 5–5: Given–new chains in Np5 and Np6.....	151
Figure 5–6: Multiple correspondence between linguistic forms of nominal expressions and information status in English texts	160
Figure 6–1: Brand-new unanchored frequencies of the texts and their thematic structures.....	164

List of Abbreviations

ANC	American National Corpus
BID	Balanced Information Distribution
BNC	British National Corpus
CD	Communicative Dynamism
COCA	Corpus of Contemporary American English
MASC	Manually Annotated Sub–Corpus of American National Corpus
OANC	Open American National Corpus
<i>of</i> -NP	Noun Phrase with relator <i>of</i>
SFL	Systemic Functional Linguistics
USA/US	United States of America

1 Introduction

When we read texts, we usually have an intuition that some are more reader-friendly and require less mental effort to process, while some are not. Example (1–1) illustrates two instances of different texts selected from the Manually Annotated Sub-Corpus of American National Corpus:

- (1–1) a. The tragedy was that they were ignored or discounted. There was an explosion in risky subprime lending and securitization, an unsustainable rise in housing prices, widespread reports of egregious and predatory lending practices, dramatic increases in household mortgage debt, and exponential growth in financial firms’ trading activities, unregulated derivatives, and short-term “repo” lending markets, among many other red flags. Yet there was pervasive permissiveness; little meaningful action was taken to quell the threats in a timely manner. (*Conclusions of the Financial Crisis Inquiry Commission*)
- b. As capital of Europe’s most explosive economy, Dublin seems to be changing before your very eyes. New construction is everywhere, the streets buzz, traffic is increasingly congested, and in the frenetic pace of rush hour everyone in Dublin seems intent on changing places with everyone else. At night the streets are crowded with people bent on having a good time. Prosperity is in the air; the roar of the “Celtic Tiger” can clearly be heard. But this is not the whole picture. The proverbial hospitality and warm welcome are still here. (*Dublin and the Dubliners*)

The instance in (1–1a) is selected from the genre of Government document and that in (1–1b) occurs in the genre of Travel guide. The two instances are clearly different. We as readers have little doubt that the instance in (1–1b) seems more reader-friendly

and requires less effort to process compared with the text in (1–1a).

From a linguistic point of view, this difference can be attributed to certain features of English. According to Biber *et al.* (1998), nominal expressions carry the majority of lexical content of a text. Differences between texts to a large extent are revealed by the differences between their nominal expressions¹. Many grammatical and semantic features of English nominal expressions have been established as sensitive indexes of registers in a number of studies, such as determiner, modifier and structural complexity etc. However, the role of nominal expressions structuring information in different texts still remains a relatively understudied area.

Taking a close look at example (1–1), there are several differences between the two instances in terms of the informational role of nominal expressions. First, the instance in (1–1a) has a different topic and communicative purpose from that in (1–1b). (1–1a) is concerned with a financial and economic issue and aims to explain what caused a tragedy, while (1–1b) provides a description of a well-known city. The difference might be reflected by the information distributions of nominals throughout the texts. Second, linguistic forms of nominals in the two instances are also different and this to a large extent could influence the type of information they convey in the texts. For example, the instance in (1–1a) has a cataphoric expression “*the tragedy*” that expresses information in the following clause. However, no such expressions were found in (1–1b). Third, the text in (1–1a) has fewer nominals in the departure points of clauses compared with that in (1–1b), which means that the information distributions of nominals are also different at the level of the clause. Furthermore, the example in (1–1a) contains more complex nominals than that in (1–1b), such as “*an explosion in risky subprime lending and securitization*”, “*an unsustainable rise in housing prices*”, and “*widespread reports of egregious and predatory lending practices*”. These nominals integrate more chunks of information and as a consequence information is integrated differently through nominals in the two texts.

Previous studies have typically focused on one of the above-mentioned aspects

¹ The term ‘nominal expression’ or simply ‘nominal’ in the current study is used as a general term that includes nouns, noun phrases or nominal groups.

drawing on limited empirical evidence from complete texts. The aim of this study is to address this gap in the literature by providing a detailed account of the informational role of nominal expressions and by developing an extensive empirical approach.

1.1 Goal of this thesis

This thesis aims to explore four main aspects of the informational role of nominal expressions in English texts.

The first aspect of interest is the information distributions of nominal expressions in different genres or text types². Most previous studies (e.g. Mathesius 1975; Firbas 1992; Halliday 1967; Chafe 1994) have restricted their analysis of nominal information distribution to text excerpts. However, given that nominals express the majority of lexical content of texts and many of their characteristics have been established as sensitive indexes of text types (Biber *et al.* 1998; Biber and Conrad 2009; Neumann 2014), it is reasonable to assume that the information values of nominal expressions might also serve to indicate differences between texts.

The second aspect under study is the linguistic forms of nominal expressions which express different types of information status. The correspondence between nominal forms and information status has been discussed in great detail in the literature from a variety of perspectives. For example, some studies have considered which forms express a given type of information status (e.g. Ariel 1996 2001), while others have

² Concerning genre or text type, we will draw on two related but different perspectives from the Systemic Functional Linguistics (henceforth SFL) framework. The first stems from Halliday's original view on text type. He regards text type as register that describes the context of situation with three parameters comprised of field, tenor and mode. Field specifies the topic of the utterance; tenor characterizes the relationship between the participants; and mode is concerned with the way in which language is transmitted (Neumann 2014: 15–16). Register is also referred to as 'genre', which comes from French, meaning 'type' (e.g. Halliday 1978: 145; Halliday & Hasan 1985; Matthiessen *et al.* 2010: 106–107). The other is what Martin and his colleagues have developed: it is a cultural notion that manipulates the choice and configuration of field, tenor and mode; it is "the purposeful, goal oriented aspects of text", where purpose is "treated not as part of register, but as an underlying semiotic in its own right" and "taken as determining the distinctive beginning–middle–end structures used to distinguish genres and sub–genres" (Martin 1984: 25). This thesis will follow the Hallidayan sense of genre or text type.

discussed a one-to-one correspondence between linguistic forms of nominal expressions and information status (e.g. Gundel *et al.* 1993; Lambrecht 1994). However, the relationship between the two has not been explored exhaustively in a large dataset.

The third aspect concerns the relationship between information status of nominal expressions and thematic structures in English texts. After Halliday (1967) separated information from thematization, many scholars investigated thematic development in English texts (e.g. Daneš 1974; Kopple 1991; Ghadessy 1995; Hasan and Fries 1995; Mauranen 1996; Liu and Tucker 2015). However, less research has been conducted into the relationship between the information status of nominal expressions and their thematic structures. Although the unmarked patterns Given ^ New and Theme ^ Rheme proposed by Halliday (1994) are generally accepted, no research has explored this relationship in different text types with a more specified information taxonomy.

The last aspect is the complexity of nominal expressions and types of information status they encode in text. Previous literature has explored the complexity of nominals more from a structural perspective (e.g. Berlage 2014), type of information complex nominals encode remains a relatively understudied area. As shown in example (1–1), complex nominal expressions in theory encode more information. Their postmodifiers expand the head nouns both grammatically and semantically. The expansion naturally leads to an integration in meaning. However, the influence of integration in meaning on information status of nominals still remains an understudied field.

This thesis aims to gain insights into the informational role of nominal expressions in English texts. Based on the above, it will specifically address the following questions:

- How are the information values of nominal expressions distributed in different English texts?
- What type of information values is expressed by the various linguistic forms of nominal expressions?
- What is the relationship between the information values of nominal expressions and their occurrences in Theme and Rheme positions?
- What type of information does postmodifiers convey in terms of integration in

meaning?

1.2 Overview of the thesis

The remainder of the dissertation is organized as follows. Chapter 2 provides an overview of the state of the art. This includes a review of previous literature related to the notion of information, different frameworks of classifying information values, linguistic forms of nominals expressing certain type of information, the relation between information and thematic structures and expansion types of complex nominal expressions. It summarizes the main contributions, identifies the intersection of linguistic research on nominal expressions, information status and text types and motivates a multi-dimensional approach to explore the informational role of nominals in English texts.

Chapter 3 provides a detailed description of the methodology. This mainly introduces the corpus design, the annotation tool and the principles of the analysis. The corpus contains 3095 nominal expressions gathered from ten texts of four comparable English written genres selected from the Manually Annotated Sub-Corpus of American National Corpus (MASC). This chapter describes all annotation tags of nominal expressions in considerable detail and further adapts Prince's (1981) information taxonomy in order to classify information values more precisely.

Chapter 4 to 7 present and analyze the data results. Chapter 4 focuses on the information distributions of nominal expressions in English texts, thereby filling the first research gap identified in Section 1.1. It presents frequency distributions of nominals in each information category and different patterns of information throughout the texts. This chapter will also consider differences among the texts in relation to information values.

In Chapter 5, I present the analysis of the relationship between the linguistic form of nominal expressions and the types of information status they represent, which addresses the second research gap. This chapter will provide a detailed account of the

frequencies of types of nominals when representing the same information status and show their similarities and differences between the texts.

Chapter 6 presents the information distributions of nominal expressions in Theme and Rheme positions of the texts, thereby filling the third research gap. This chapter first explores the relation between texts and their thematic structures in the information distributions of nominal expressions by examining the nominals of complete texts and then illustrates the information distributions in Theme and Rheme positions respectively by using nominals in each clausal position. This chapter will show how nominals distribute types of information at the level of clause and reveal their discourse functions in different clausal positions of texts.

Chapter 7 examines the influence of types of integration in meaning on information distributions of complex nominal expressions in English texts, which fills the last research gap identified in Section 1.1. This chapter will present how nominals as postmodifiers distribute information status when they expand the head nouns through logico-semantic relations in different texts.

The study will end with a conclusion in Chapter 8, which summarizes the main findings, evaluates the methodology and provides directions for further research.

2 State of the Art

The various key aspects of information and nominal expressions have been investigated from a variety of perspectives in the linguistic field. Many scholars have proposed different notions of information, established classifying frameworks of information status, and explored the relations between various linguistic features of nominal expressions and information. The main goal of this chapter is to offer an overview of the above areas. These areas serve as theoretical foundations to explore the informational role of nominal expressions presented in this thesis. This is important as it helps motivate the methodology presented in this thesis towards establishing a multi-dimensional approach to understanding the informational role of nominal expressions in English texts.

This chapter is organized as follows. Section 2.1 introduces the notion of information from multiple perspectives. Section 2.2 reviews the strengths and weaknesses of selected frameworks for classifying information status. Section 2.3 presents previous studies on the relations between linguistic forms of nominals and types of information status. Section 2.4 gives a description of the studies on interaction between information status and clausal positions. Section 2.5 presents previous accounts on the structural complexity of nominal expressions and ways of expanding complex nominals. Finally, Section 2.6 summarizes the whole chapter and motivates the multi-dimensional approach to this thesis.

2.1 The notion of information

There are mainly two types of information, ritual and objective. Ritual information is of pure interpersonal nature (Prince 1981). For example, when two acquaintances say *hello* on the street, they indicate that one speaker is aware of the presence of another

and they greet in a friendly way. This is in contrast to objective information that provides knowledge about things, events and so on. Given that the present study aims to explore the informational role of nominals, objective information in English text is the main concern. This section mainly introduces multiple features of information from different theoretical perspectives and the relationships between context, text and background knowledge, and types of givenness.

The earliest studies on information were conducted by the Prague School. Mathesius (1975) characterizes information in the Old-vs.-New sense to define thematic structure. Firbas (1964, 1992) studies information in terms of Communicative Dynamism (CD). CD is “a phenomenon constantly displayed by linguistic elements in the act of communication” (Firbas 1992:7). Information is defined as linguistic elements that display degrees of CD in accordance with the position of linguistic elements in sentences or subclauses. The degrees of CD are determined by several factors, which are linear modification, the immediate relevant context, the semantic relations and content of linguistic elements, and prosodic prominence (for spoken language only). The degree of dynamism becomes higher with the development of word order.

The notion of information proposed by the Prague School is further developed by Systemic Functional Linguistics (henceforth SFL). Halliday (1994: 296) defines information as “the tension between what is already known or predictable and what is new or unpredictable... it is the interplay of new and not new that generates information in the linguistic sense”. The “not new” and “new” are “information that is presented by the speaker as recoverable (Given) and not recoverable (New) to the listener” (Halliday 1994: 298). Recoverable information here includes two types. One is the preceding text inherited from Firbas (1992), and the other is “something that is in the situation, like I and you; or in the air, so to speak; or something that is not around at all but that the speaker wants to presents as Given for rhetorical purpose” (Halliday 1994: 298), which has extended the scope of context by adding extratextual features.

Different from Firbas (1992), Halliday (1967) regards a clause as the basic unit of information, namely “information unit”. One information unit is realized by tone group.

In spoken text, Given information is deaccented as the background, while New information receives a pitch accent as it carries particular importance (cf. Allerton 1978; Baumann 2006). New information is defined as the “information focus” (Halliday 1967: 203), which is also referred as “emphasis” or “importance” by marking the informative part of a message to hearers.

Information is also a phenomenon that has been explored in psycholinguistics. Chafe (1970, 1973, 1974, 1976, 1987 and 1994) defines information in terms of “consciousness”, stating that consciousness has a focus which constantly moves from one item to another (Chafe 1994: 29). When consciousness is activated, it is only “a small part of the experiencer’s model of surrounding world” and is surrounded by the so-called “peripheral consciousness” that provides context for the focus, i.e. basic knowledge of the ongoing interaction, such as space, time, location and participants (Chafe 1994: 29). The activation of consciousness determines the changes of information status, which is divided into three types, namely active, semi-active or fully active (Chafe 1974: 112). Besides, Chafe shares the same idea with Halliday on the central role of speakers, claiming that “what a speaker shares with his addressee must be part of what is in the speaker’s consciousness at the time” (Chafe 1994: 29). Chafe has mainly emphasized the relation between information and mind.

In addition to the theoretical approaches to study information, insight is also provided into distinguishing information from meaning in terms of the cognitive representation system of human information processing. Givón (2001: 43) indicates information is conveyed by proposition that is made out of words and is typically represented by grammar-coded clause, while meaning is a concept that is expressed by sound-coded word. This implies that information is different from conceptual meaning. It is stored as proposition at the level of clause, which is also closely related to the mental world, external world, culture or a combination of them (Givón 2001: 8).

The difference between meaning and information is further discussed by Lambrecht (1994: 46), indicating that “meaning is expressed either in individual words or via relations established between words, information can strictly speaking only be conveyed relationally, via propositions”. Individual words are “the elements of

information”, which are called lexical information or referential information. Compared with lexical / referential information, propositional information changes “the hearer’s mental representation of the world” by establishing relations (Lambrecht 1994: 43).

Propositional and lexical/referential information are further developed into “relational Newness/Givenness” and “referential Newness/Givenness” by Gundel (2003). Referential givenness concerns the assumed mental state of a discourse entity that is stored in the speaker’s and hearer’s mind and it is explored by most of the frameworks of information classification. Relational givenness is established by “two complementary parts, X and Y, of a linguistic or conceptual representation, where X is given in relation to Y, and Y is new in relation to X” (Gundel 2003: 4). Relational givenness focuses on the pragmatic functions of referring expressions at the syntactic level. Given lexical items can turn to be structurally new for pragmatic reasons (cf. Halliday 1967).

Unlike the above accounts mainly focusing on the central role of speakers played in conveying information, Clark and Haviland (1974, 1977; also see Clark and Wilkes–Gibb 1986; Schober and Clark 1989; Clark 1992) take the role of hearers into consideration. They regard information as the knowledge shared by both speakers and hearers and propose the Given–New Contract. The contract is based on the Cooperative Principle (Grice 1975), indicating that the speaker agrees to convey given information that he/she assumes the hearer already knows and to convey new information that he/she assumes the hearer does not yet know (Clark and Haviland 1977: 4). Both speakers and hearers need to refer to the contextual and textual clues and some pre-existing knowledge that one is assumed to have at the time of utterance. In addition, the speaker is supposed to provide only one direct antecedent for given information. If not, the hearer is assumed to use one of the three strategies, namely bridging, addition and reconstructing. The difference between information with direct and indirect antecedents has been discussed by Clark and Haviland (1977). Their psycholinguistic experiments show that information with direct antecedents requires less mental effort to comprehend. However, the experiments only tested self-created conversations, which “leave basic

questions unresolved so long as they are isolated from observations of natural language, and from crucial introspective evidence as well” (Chafe 1994: 173).

This section has introduced previous accounts of the notion of information, with particular reference to multiple theoretical approaches to study information, the relationship between information and context and text, the difference between referential and relational information and the central role of speakers in conveying information. The previous accounts indicate that information is a complex phenomenon that is situated contextually at the intersection of functional linguistics, cognitive linguistics and psycholinguistics. This results in varied frameworks for classifying information status, which will be presented in Section 2.2.

2.2 Main frameworks of classifying information status

As introduced in Section 2.1, information studies have mainly focused on the central role of speakers. Most frameworks developed from the studies classify information status from the perspective of speakers, with few from multiple perspectives. This section will present a selection of main frameworks of classifying information status. These frameworks involve the most relevant aspects in exploring the informational role of nominals in English texts. By comparing their properties and weaknesses, this study points out the great influence of Prince’s (1981) information taxonomy and identifies its inconsistencies and ambiguities in classification. In doing so, I will motivate the annotation scheme and I provide further definitions for classifying information which are needed for the current study.

2.2.1 Classifying information status from the perspective of speakers

This section mainly presents two frameworks for classifying information status from the perspective of speakers, namely Recoverability/Predictability and Saliency.

Recoverability/Predictability is summarized by Prince (1981) based on types of

information status the terms classify rather than the notions of information. Recoverability is proposed by SFL and it has two categories, namely recoverable (Given) and irrecoverable (New) (see Section 2.1). Recoverable information is presented as something that has been mentioned before or exists in the situation, while irrecoverable information is something that is not mentioned or expected (Halliday and Matthiessen 2014: 118). Predictability is developed by Kuno (1972, 1974, 1978), which classifies information into “old” and “new”. Old information is the sentence element that is predictable from the preceding text, while the new is unpredictable (Kuno 1978: 282–283). Recoverability and Predictability cannot be regarded as the same. Based on their definitions, recoverable information also includes shared background knowledge that both writers and readers take for granted, such as *the sun*. However, *the sun* is unpredictable to readers if it occurs for the first time in a text. Besides, according to Halliday (1967), a referent that has occurred in the preceding text can still be classified as irrecoverable if it is presented as the information focus of a clause. However, under the notion of predictability, the referent is always predictable in terms of the occurrence in the preceding text.

The second representative framework is based on the sense of Saliency proposed by Chafe (1970, 1974, 1976, 1987 and 1994). Saliency refers to the particularity of a thing, entity or event the speaker assumes in the hearer’s consciousness. Different from Predictability / Recoverability, Chafe recognizes a third category of information status between Given and New information, namely accessible. Chafe (1994: 165) points out that the approach proposed by Halliday (1994) “does not recognize a degree of activation cost (or recoverability) that is intermediate between given and new”. However, classifying information status only into three types by Chafe is considered as being less comprehensive. Lambrecht (1994: 100) indicates that theoretically speaking, mental representations may have no upper limit to the number and types of cognitive status in the course of conversation. Chafe’s framework does not contain all cognitive statuses to classify information. This weakness is acknowledged by Chafe (1994: 69), indicating that “information in the mind may be in any one of at least three activation states”.

According to Prince (1981: 231), Recoverability/Predictability and Saliency are mutually dependent, indicating that “if a speaker assumes that the hearer can predict that some particular item or items will occur in some particular position within a sentence, then the speaker must assume that it is appropriate that the hearer have some particular thing in his/her consciousness”.

Drawing insights from the literature, Prince (1981) proposes an information taxonomy. This framework mainly takes the speaker’s perspective to describe the assumption of a listener’s mental state of discourse entities, namely Assumed Familiarity. It will be introduced in Section 2.2.2.

2.2.2 Assumed Familiarity by Prince (1981)

“Assumed Familiarity” proposed by Prince (1981) is regarded as the most influential framework for classifying information status (Riester and Baumann 2017). This section presents an overview of “Assumed Familiarity” and discusses the strengths and weaknesses by comparing it with other relevant approaches. It aims to motivate a further defined framework based on Prince (1981), which can be more transparent for classifying information values in text for this study.

Figure 2–1 displays the “Assumed Familiarity” proposed by Prince (1981). This framework classifies information values into three main categories, namely New, Inferrable and Evoked. Each of the three categories is further divided into subtypes. A detailed description of each type is as follows.

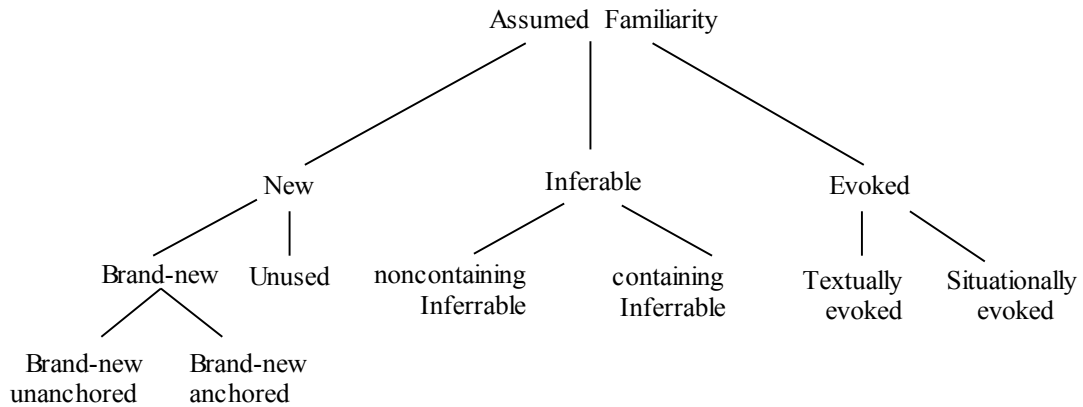


Figure 2–1: Assumed Familiarity (Prince 1981: 237)

New

New information, as the name suggests, is newsworthy and occurs in a text for the first time. It has two subtypes: Brand–new and Unused, which are illustrated by example (2–1) selected from Prince (1981). Brand–new information is composed of Unanchored and Anchored. Brand–new unanchored information is mentioned the first time and unknown to hearers/readers, such as “*a beautiful dress*” in example (2–1a). Brand–new anchored information, though newly created by speakers/writers, is linked to or contained within other discourse entities. “*A rich guy I know*” in (2–1b) is a case in point. Unused information is also defined as “permanent registry” by Kuno (1972), “culturally copresent” by Clark and Marshall (1981) and “unique referents” by Quirk *et al.* (1985), indicating that it has unique identities and references and is taken for granted by hearers/readers without further context and text, such as “*Rotten Rizzo*” in (2–1c).

- (2–1) a. I bought *a beautiful dress*. (Brand–new + attribute)
 b. *A rich guy I know* bought a Cadillac. (Brand–new Anchored + attribute)
 c. *Rotten Rizzo* can’t have a third term. (Unused + Attribute)

According to Baumann and Riester (2012), the distinction between Brand–new and Unused information is problematic. First, it is peculiar to put proper names,

indefinite and definite nominals into the same category from a semantic perspective. Second, nominal expressions referring to places, names are difficult to be regarded as known or unknown objectively due to different intended recipients of text (Riester 2008a, 2008b). Normally “*George Clooney*” conveys Unused information, while “*Harry Smith*” is Brand–new unanchored (Baumann and Riester 2012). However, “*Harry Smith*” could also convey Unused information if the intended recipients are assumed to know him. Like “*Rizzo*” in example (2–1c), some people may have no idea of the reference without the same knowledge shared with the writer/speaker.

The above indicates that Baumann and Riester’s (2012) review of Prince’s (1981) framework is also problematic. Their first point implies that indefinite expressions cannot convey the same type of information as proper names. This actually acknowledges a one–to–one correspondence between information status and linguistic forms of nominal expressions. However, Lambrecht (1994: 79) points out that “there is no one–to–one correlation between identifiability or non–identifiability of a referent and grammatical definiteness or indefiniteness of the noun phrase designating the referent”. Although it may seem odd to think of proper names and indefinite nominals as both being in the same category from a semantic perspective, it is possible that both linguistic forms of nominal expressions convey the same information value. The second point, the controversy surrounding the objective categorization of places and names, reveals different shared background knowledge possessed by hearers/readers. This has been discussed by Prince (1981, 1992), who makes it clear that information status must be classified on the basis of familiarities the hearers/speakers can reasonably be assumed to have.

Besides, the term “anchored” in Prince (1981) also seems confusing (cf. Baumann and Riester 2012). Brand–new anchored information is linked to or contained within other noun phrases (Prince 1981: 236). Prince (1981) does not indicate clearly how to identify an “anchor”, i.e. whether it is expressed through morphological features, semantic meanings or a combination of both. Furthermore, there is a certain ambiguity between Brand–new anchored and Inferrable information. This will be discussed in detail under the category of Inferrable.

Inferrable

Inferrable information is assumed to be inferred by hearers/readers through reasoning from Evoked or other Inferrable discourse entities (Prince 1981: 236). Types of inference are required for the Inferrables. Culture-based and logical inference, such as member to set or set to member are common Inferences in categorizing Inferrables.

(2–2) I got on *a bus* yesterday and *the driver* was drunk.

In example (2–2) given by (Prince 1981: 233), “*the driver*” is Inferrable from “*a bus*”, since the hearers/readers are expected to infer the relationship between bus and driver via logical reasoning and assumed knowledge.

Prince’s (1981) description here is problematic to some extent. It does not specify how readers use inference to categorize Inferrables and how Inferrables are different from other information. Assumed knowledge is not only required to categorize Inferrables, but is also needed for Unused information. Prince’s (1981) description here brings a certain ambiguity between Inferrable and Unused information. She classified *salt* in a recipe as Unused information, claiming that the writer assumes the readers take salt for granted in the text. However, *salt* here also carries some property of Inferrable by establishing a part–whole relationship with other ingredients.

Obviously, inference is crucial to classifying Inferrables, but is not fully specified by Prince (1981). In comparison, it is discussed in great detail in terms of Bridging by Clark and Haviland (1974, 1977). Bridging information, though not co–referential, is contextually dependent and linked to Given information that is expressed by definite nominal expressions. Definiteness establishes types of semantic associations with other discourse entities, such as possession, attribution or different kinds of part–whole relations (Clark 1977; Winston *et al.* 1987). Bridging is also called bridge anaphora, associative anaphora (Poesio and Vieira 1998; Asher and Lascarides 1998; Vieira and Poesio 2000; Poesio 2004) or indirect anaphoric reference (cf. Riester and Baumann 2017).

A general scale of inference is required for establishing abstract implicit and flexible mental relations in Bridging. The inference is triggered by a linguistic stimulus that activates knowledge in long-term memory, which involves a variety of prototypical mental representations. Thus inference has also been explored in terms of a series of cognitive notions, such as activation, mental model and prototype.

Inference is mainly studied by two approaches. The first focuses on human categorization, which is represented by family resemblance (Wittgenstein 1953) and the studies of prototype (e.g. Weinreich *et al.* 1968; Berlin and Kay 1969; Kay and McDaniel 1978; Labov 2004; Lakoff 1987; Berlin *et al.* 1974; Berlin 1976, 1978; Rosch 1973, 1978, 1983; McGloskey and Glucksburg 1978; Barsalou 1983). The second concerns stereotypic mental representations of world knowledge and information retrieval. They are mainly discussed in terms of Frames (Minsky 1975; Fillmore 1975, 1977, 1985), Scripts (Schank and Abelson 1977), Scenarios (Sanford and Garrod 1981), Schemata (Bartlett 1932; van Dijk 1977; Anderson 1978, 2019; Rumelhart and Ortony 1977; Rumelhart 1975, 1980; Anderson and Pearson 1988; Kecskes 2012), Cultural Model (Holland and Quinn 1987; D'Andrade 1992), Prototype and Mental Model (Johnson-Laird 1983; Fauconnier 1994).

These theories, though with different terms, display many commonalities (cf. Tannen 1979). They are summarized as follows:

- (i) Representations of background knowledge
They all contain mental representations of background knowledge that we use in inference and discourse interpretation.
- (ii) Existence of obligatory/default elements
Stereotypic representations of world knowledge that are stored in the long-term memory will be partially activated under specific situations. They are obligatory or default elements in texts, such as drivers of buses and waiters in restaurants.
- (iii) Culture-based and individually different
Inference not only depends on different cultural backgrounds, but is also

influenced by different interpretations of intended text recipients (cf. Brown and Yule 1983: 248).

- (iv) Definiteness and uniqueness of the discourse referents in stereotypic situations

Discourse referents identified through inference are usually expressed by definite nominal expressions. Definiteness is an important feature to indicate familiarity to hearers/readers (Brown and Yule 1983: 252; Haviland and Clark 1974; Clark 1978: 313)

- (v) Dynamic and flexible

Inference is dynamic and flexible. Discourse entities identified from inference “can be expected, but not guaranteed” (Brown and Yule 1983: 240) in a discourse.

- (vi) Interaction between pre-existing knowledge, text and context

Inference is established on world knowledge, the development of text and situational context.

In addition, Inferrable has a special subclass, namely containing Inferrable. What is inferred off is properly contained within the Inferrable itself (Prince 1981: 236). Each containing Inferrable is expressed by at least two nominal expressions, one of which serves as trigger and usually conveys Given information. Example (2–3) illustrates two containing Inferrables.

- (2–3) a. Hey, *one of these eggs* is broken. (Prince 1981: 233)
b. *The purpose of this chapter* is to generalize Weinreich’s statement...
(Prince 1981: 248)

“*One of these eggs*” in (2–3a) is a set-member containing Inferrable. “*These eggs*” as the trigger conveys Situationally evoked information, which is known to readers as features of extratextual context and text itself (details of Situationally evoked will be introduced later). “*The purpose of this chapter*” in (2–3b) is a cultural-based containing

Inferrable, triggered with “*this chapter*”. Similarly, both triggers are embedded anaphora of the containing Inferrables (cf. Baumann and Riestler 2012). The examples also show that nominals within containing Inferrables usually form semantic relations, with particular reference to the nominal head of a complex phrase and its possessor or nominal argument (cf. Riestler and Baumann 2017: 10).

Information status of trigger entities is not only fixed to the category of Situationally evoked. Prince’s (1981) analysis shows the trigger entities can also convey non-containing Inferrable, Textually evoked and Unused information. Instances of Textually evoked and Unused triggers are given in example (2–4):

- (2–4) a. In speaking of “functions”, I do not intend to raise here *the many issues that attach to the notion of “functionalism” in the social sciences, and more generally, in the philosophy of the sciences and humanistic disciplines*.
- b. In their methodological reflections on worlds of human knowledge, *scholars such as Ernst Cassirer and Kenneth Burke* have found that question of function, and in human action, the question of function known as purpose, indispensable.

Prince (1981) classified “*the many issues that attach to the notion of “functionalism” in the social sciences, and more generally, in the philosophy of the sciences and humanistic disciplines*” in (2–4a) as containing Inferrable, claiming that “*the notion of functionalism*” as a trigger conveys Textually evoked information in terms of the occurrence of “*functions*” in the preceding text. “*Scholars such as Ernst Cassirer and Kenneth Burke*” in (2–4b) was also categorized as containing Inferrable. Prince (1981) indicated that “*Ernst Cassirer and Kenneth Burke*” conveys Unused information.

Prince’s (1981) criteria for the triggers for containing Inferrables are problematic in several aspects. First, containing Inferrables as the subtype of Inferrables are expected to be contextually and textually dependent discourse entities. However, when triggers convey Unused information, the containing Inferrables are not dependent on

any specific context and text and they are regarded as context-free expressions (Riester and Baumann 2017). Second, containing Inferrables with Unused triggers contradict the Familiarity Scale, namely Evoked > Unused > Inferrable > containing Inferrable > Brand-new anchored > Brand-new unanchored (Prince 1981: 245). According to the scale, containing Inferrables are assumed to be more familiar to the readers/hearers than Brand-new anchored referents. However, Prince (1981) classifies Unused information as a subtype of New information and containing Inferrables as accessible. Unused information with newsworthiness is expected to be less familiar to readers and requires more mental effort to process in theory. In this case, the familiarity displayed by containing Inferrables with Unused triggers is not necessarily higher than Brand-new anchored referents. Third, containing Inferrables with non-containing Inferrable triggers causes the same problem as those with Unused triggers. Readers are required to make inference twice to process the containing Inferrables. The first inference is for the non-containing Inferrables as triggers, and the second is to infer the containing ones based on the triggers. In theory, the readers put more mental effort to process containing Inferrables, which makes them less familiar than the non-containing ones. This also contradicts the Familiarity Scale.

In addition, it is also problematic that Prince (1981) classified the embedded clauses of the containing Inferrables displayed by example (2–4) as known information. Taking a closer look at the example in (2–4a), it is controversial to categorize “*the many issues that attach to the notion of “functionalism” in the social sciences, and more generally, in the philosophy of the sciences and humanistic disciplines*” as containing Inferrable. According to Schmid (2000), both the shell noun “*the many issues*” and the subordinate clause basically convey new information, since the noun phrase is presented as the focus of attention of the sentence and it is the only cohesive tie to the preceding text is the word “*function*”. Schmid (2000: 330) points out even the shell noun and the subordinate clause occur as topics, they are very rare to convey given information. This is illustrated by his example in (2–5):

(2–5) Villaverde, southern Madrid, finds the big Socialist hope has a lot to prove

By John Hooper

For Spain's ruling Socialists, "Superjudge" Baltasar Garzón is an almost priceless asset. *His decision to stand for them at the general election on June 6* seems to guarantee something that has been called into question too much of late – their honesty.

In example (2–5), "*his decision to stand for them at the general election on June 6*" would be categorized as containing Inferrable based on the classification of Prince (1981), claiming that the content of the decision is contained within the nominal expression itself. However, Schmid (2000: 331) states that the nominal expression actually conveys new information. Although it is linked to the preceding text through the possessive determiner "*his*" and the pronoun "*them*", it is not accessible to the readers. The headline in the example marks the beginning of the article, it is less likely for readers to be familiar with the content of "*his decision*". The strategy of presenting new information as given (Peng 2014) is "a rhetorical gambit by the writer", which is not rare (Schmid 2000: 331).

As indicated above, Prince's (1981) classification shows a certain ambiguity between Inferrable and information of other categories. The first fuzzy boundary is between Inferrable and Brand–new anchored information. This is caused by Prince's (1981) descriptions. Nominals of both categories are defined to be linked to the preceding text, but Prince (1981) does not specify types of links. The second fuzzy boundary is between containing Inferrable and Unused information. Prince (1981, 1992) indicates that this is caused by different intended text recipients. Different readers may classify some discourse entities differently if they have different general and specialist background knowledge. In Prince's (1981) analysis, the discourse entity "*scholars such as Ernst Cassirer and Kenneth Burke*" illustrated in example (2–4b) could also convey containing Inferrable information, if the readers are not familiar with scholars of the Prague School. However, Prince (1981) has not fully explored the potential for Inferrables to be somewhat problematic and in need of further constraints for their identification.

Evoked

Evoked information is expressed by nominal expressions that the readers/hearers can retrieve from contextual and textual grounds (Prince 1981). It has two subtypes, namely Situationally Evoked and Textually Evoked.

Situationally evoked information refers to discourse participants and salient features of the context shared by the speakers/writers and hearers/readers. The discourse participants share joint attention on certain discourse entities. Classifying Situationally evoked information may also involve pointing gestures (Baumann and Riester 2012) or visual identification (Diessel 2006: 465). The information is more common to occur in spoken text such as face-to-face dialogue. It is typically expressed by indexicals or deictic expressions, such as pure indexicals *I, you, here* and demonstratives *this, that, these* (cf. Kaplan 1989).

Textually evoked information refers to discourse entities that have occurred in the preceding text. It could be repeated in varied linguistic forms. Some examples are given by Baumann and Riester (2012) in (2–6):

- (2–6) On my way home *a dog* barked at me.
- a. *The dog* belongs to my neighbor's oldest son.
 - b. *The animal* belongs to my neighbor's oldest son.
 - c. *It* belongs to my neighbor's oldest son.

In example (2–6), “*a dog*” is repeated by three ways: full repetition, hypernym and pronoun. Although the repeated forms are different, they denote the same referent without adding more information. Unlike example (2–6), the Textually evoked entity “*the fierce German Shepherd*” in example (2–7) is more specific and informative than the antecedent “*a dog*” without changing its experiential identity.

- (2–7) On my way home, a dog barked at me.
- The fierce German Shepherd* seemed to be quite aggressive.

- (2–8) The later Uriel Weinreich (1966: 399) observed: “whether there is any point to semantic theories which are accountable only for special cases of speech – namely humorless, prosaic, banal prose – is highly doubtful.” “The purpose of this chapter is to generalize *Weinreich’s statement* and to remove *the qualification*: linguistic theories accountable only for such cases of speech cannot be consistently justified...”

Although classifying Textually evoked information is explicit in general, there are some controversial cases with Prince’s (1981) analysis. In example (2–8), Prince (1981: 250) classifies the nominal expression “*Weinreich’s statement*” as Inferrable based on “a high degree of metalinguistic inference”.

This type of nominal is semantically general, which is defined as general noun by Halliday and Hasan (1976: 274–275) and shell noun by Schmid (2000). It serves as “a metalanguage for the clause” by concerning “the nature of the clause or sentence as a message in the text itself” (Winter 1992: 133; cf. Ivanič 1991: 94; Francis 1994: 83). In theory, “*Weinreich’s statement*” requires more mental effort to process, since its antecedent is a proposition that contains more chunks of information than those with nominals as antecedents. It has a lower degree of givenness than pronouns like “*it*” displayed by (2–6c).

However, “*Weinreich’s statement*” in example (2–8) could also be classified as Textually evoked, as it refers to the proposition that has occurred in the preceding text. Since its antecedent is a proposition containing many information chunks, “*Weinreich’s statement*” serves as an economic linguistic unit for the writer to mention it again in the following text (Schmid 2000: 370). There is supportive evidence provided by Schmid (2000) to classify the same type of nominals like “*Weinreich’s statement*” as Textually evoked information. He (ibid: 346–347) points out that if the speakers cannot assume that the propositional antecedent is still known to the readers, they will apply one of two strategies to repeat the information, by reformulating it either in a non–restrictive clause or a restrictive appositive clause.

Analyzing antecedents involves two considerations: (i) textual distance or anaphoric distance between the Textually evoked discourse entities and their antecedents (Schmid 2000: 346; cf. Givón 1983, 1984, 2001; Sun and Givón 1985; Lambrecht 1994; Kunz 2010; Baumann and Riester 2012); and (ii) certain intervening information, such as the competition between possible antecedents (Schmid 2000: 346; cf. Clancy 1980; O'Brien and Albrecht 1991; Strube and Hahn 1999; Ariel 2001). Example (2–9) is an illustration given by Schmid (2000: 347):

(2–9) Mohamed al Fayed remained at odds with Princess Diana's representatives yesterday after claiming that the crash in which she and his son, Dodi, died was the result of a conspiracy. In an unprecedented statement, her office attacked speculations about the Princess's death and said it was upsetting her sons, princes William and Harry.

Crash investigators in Paris also refused to be drawn on Mr. Fayed's claim, saying they would wait until the inquiry was over. "Mr. Fayed has his own opinions but we are not prepared to comment on them. The investigation is already before a French judge and only when he concludes his investigations will we have an opinion," said one of the investigating team. *The conspiracy theorists' view that the death of the Princess was no accident* remained, at first, confined to sections of the Arab media...

"*The conspiracy theorists' view that the death of the Princess was no accident*" in (2–9) presents the same idea with its antecedent that is expressed by the first paragraph of this example. However, with a long textual distance and too much intervening information in between, the writer obviously assumed that the propositional antecedent would not be known, therefore reformulated as a restrictive appositive clause.

In comparison, "*Weinreich's statement*" illustrated by example (2–8) does not apply the strategies to reformulate its antecedent into restrictive or non-restrictive clauses. It has a shorter textual distance to its antecedent and there is no intervening information in between. The writer assumed that it was known information to the readers. It should

be classified into the category of Textually evoked.

However, it is hard to measure the textual distances and intervening information between Evoked nominals and their antecedents in a text. Baumann and Riester (2012) apply a different strategy from Prince (1981) and Schmid (2000) in classifying nominal expressions with propositional antecedents. They regard nominals expressing facts, properties, issues, events or states as valid in the full text after they have been introduced to the readers. These nominals will be classified as given information in any case.

2.2.3 Classifying information status from multiple perspectives

Frameworks presented in the previous sections only focus on the perspective of speakers. This section is mainly concerned with three selected frameworks for classifying information status from multiple perspectives.

The first selected framework introduced here is proposed by Prince (1992). It includes two perspectives in classifying information status, namely knowledge and consciousness. Knowledge is concerned with Hearer–status, which refers to “the speaker’s beliefs about the hearer’s beliefs” (Prince 1992: 301). Hearer–status can be further divided into Hearer–old and Hearer–new. Consciousness is about Discourse–status, which assesses the information status “from the point of view of the discourse model being constructed during discourse processing” (Prince 1992: 303).³ Discourse–status contains two subtypes, Discourse–old or Discourse–new.

Discourse–status and Hearer–status are different, but they are interdependent. Hearer–new information is necessarily Discourse–new based on the assumption that the hearers are assumed to remember what has been mentioned before. Discourse–old information is also Hearer–old. Discourse–new information can be either Hearer–old

³ Prince (1981: 235) indicates that “a text is a set of instructions from a speaker to a hearer on how to construct a particular Discourse Model”. The model contains discourse entities, attributes and links between entities. Discourse entities as discourse-model objects are also known as discourse referents, which represent individual(s), substances, concepts and etc. All these entities are expressed by nominals expressions in a text.

or Hearer–new.

Basic correspondence could be established between Discourse/Hearer–status and Assumed Familiarity proposed by Prince (1981). When nominal expressions convey both Discourse–old and Hearer–old information, they are Evoked; when nominals are Discourse–new and Hearer–old, they convey Unused information; and when nominals convey both Discourse–new and Hearer–new information, they are presented as Brand–new discourse entities. The relation between Discourse/Hearer–status and Inferrable is more complex. According to Prince (1992: 309), Inferrables “are technically Hearer–new and Discourse–new but depend upon beliefs assumed to be Hearer–old”, and they also contain Discourse–old triggers. Inferrables have indeterminate values in Discourse/Hearer–status (also see Loock 2013).

The second selected framework is proposed on the basis of Identifiability by Lambrecht (1994). It is displayed by Figure 2–2:

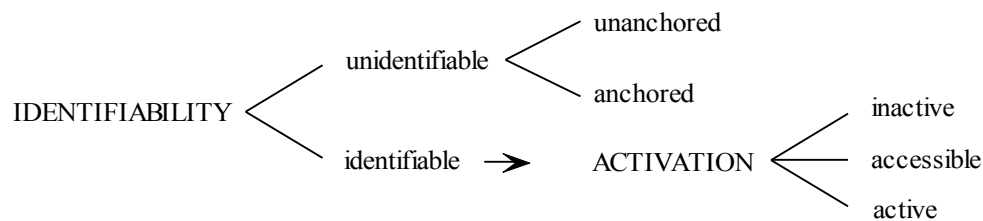


Figure 2–2: The system of Identifiability (Lambrecht 1994: 109)

This framework contains two main categories. The first is Identifiability, which “has to do with a speaker’s assessment of whether a discourse referent is already stored in the hearer’s mind or not”; and the second is Activation inherited from Chafe (1994), which “has to do with the speaker’s assessment of the status of the representation of an identifiable referent as already ‘activated’, as merely ‘accessible’, or as ‘inactive’ in the mind of the hearer at the time of the speech act” (Lambrecht 1994: 76). Identifiable information can be inactive, accessible and active. Unidentifiable information remains inactivated in the hearer’s consciousness, which is further classified into anchored and unanchored by referring to Prince’s (1981) terminology. Interpretations of each category are exactly the same with Chafe’s (1994) and Prince’s (1981) frameworks.

The third framework the *RefLex* scheme is proposed by Riester and Baumann (2017) (also see Baumann and Riester 2012). It includes two levels: (i) the referential level (henceforth *r*-level); and (ii) the lexical level (henceforth *l*-level). Table 2–1 displays annotation labels of the *r*-level. The labels are mainly adapted from Prince’s (1981) Assumed Familiarity, but they are different in several aspects. First, Unused information is further divided into two subtypes, namely *r*-unused-unknown and *r*-unused-known. Second, Situationally evoked information is further specified with pointing gestures or gaze. The *r*-environment class is added to capture more features of face-to-face communications. Third, anaphoric distance between referents and their antecedents is taken into account. If there are five clauses between a nominal and its co-referential antecedent in written text, the nominal will be labelled as *r*-given-displaced⁴. In addition, the *r*-bridging and *r*-bridging-contained labels, though similar in terminology, are not subtypes of the same category as the non-containing and containing Inferrable defined in Prince (1981). The *r*-bridging-contained label is applied to context-free expressions as Unused information. They are described as globally unique discourse entities that could be processed without specific contextual grounds. Furthermore, the *r*-level also takes into account other features, such as the optional features idioms, “+predicative” and “+generic”.

⁴ According to Riester and Baumann (2017: 8), the choice of a distance of five clauses (in written texts) or intonation phrases is arbitrary to some extent.

Tag	Contextual class
<i>r-given-sit</i>	Referents contained in text-external context (communicative situation)
<i>r-environment</i>	
<i>r-given</i>	Referents mentioned in previous discourse context
<i>r-given-displaced</i>	
<i>r-cataphor</i>	Discourse-new entities that depend on other expressions in the discourse context
<i>r-bridging</i>	
<i>r-bridging-contained</i>	Globally unique entities that are discourse-new and independent of the discourse context
<i>r-unused-unknown</i>	
<i>r-unused-known</i>	
<i>r-new</i>	Non-unique, discourse-new entities
<i>r-expletive</i>	Non-referring expressions
<i>r-idiom</i>	Optional features
+ <i>generic</i>	
+ <i>predicative</i>	

Table 2–1: Annotation labels of the *r*-level (Riester and Baumann 2017: 5)

The distinction between *r*-bridging-contained and *r*-unused information is not always as clear as the labels define. Riester and Baumann (2017: 10–11) thus propose the Permutation test. The test is specified with examples illustrated by (2–10).

Permutation test: Try to dislocate the embedded argument of a complex definite description to the left. If the remaining “anaphor” is still interpretable in relation to the dislocated “antecedent”, assign the label *r*-bridging-contained. If not, assign one of the *r*-unused labels.

(2–10) a. *The construction of the new townhall* will start next year.

Permutation: *A new townhall* will be built, and *the construction* will start next year.

Result: *r*-bridging-contained

b. *The swimming pool of the new townhall* created discontent among the

voters.

Permutation: They built *a new townhall*, and ⁵**the swimming pool*
created discontent among the voters.

Result: *r–unused–unknown*

c. John says that we should ask *his hairdresser*.

Permutation: John/He says that we should ask **the hairdresser*.

Result: *r–unused–unknown*

Although example (2–10) shows that the Permutation test helps us distinguish the ambiguity between *r–bridging–contained* and *r–unused* to some extent, there are several problems. First, the test cannot apply to all ambiguous cases. It is based on the syntactic structure of complex nominal expressions, especially on those with the relator “*of*”. For example, “*scholars such as Ernst Cassirer and Kenneth Burke*” illustrated by example (2–4b) does not contain an anaphoric embedded argument, but it was still categorized either as *r–bridging–contained* or as *r–unused*. Second, the test result cannot guarantee an actual description of information status of nominal expressions without considering the actual contextual and textual environments in written discourses. Third, the classification of *his hairdresser* (2–10c) seems controversial here. In fact, it is more like a Brand–new anchored entity, since it was presented in Rheme position as the unmarked focus of attention of the clause. Besides, it is also difficult to set up clear criteria for classifying semantic relations.

⁵ The symbol “*” here means the anaphor is not interpretable.

Tag	Saliency class
<i>l-given-same</i>	active, i.e. salient concepts
<i>l-given-syn</i>	
<i>l-given-super</i>	
<i>l-given-whole</i>	
<i>l-accessible-stem</i>	semi-active, i.e. derivable concepts
<i>l-accessible-sub</i>	
<i>l-accessible-part</i>	
<i>l-new</i>	inactive concepts

Table 2–2: Annotation labels of the *l*-level (Riester and Baumann 2017: 22)

Table 2–2 displays the annotation labels of the *l*-level, which is mainly based on Chafe (1994). They mainly explore the givenness of content words such as nouns, adjectives, verbs and (content) adverbs. These content words establish different semantic relations (or sense relations) through lexical meanings that are stored in readers’/hearers’ long-term memory. The semantic relations at the *l*-level are thus independent of context and text (Lyons 1977; Hasan 1999; cf. Kunz 2010).

The *l*-level annotation is purely based on lexical meanings of content words. It does not have any interaction with the *r*-level.

- (2–11) a. Look at the funny dog over there! I like *that dog*.
b. Look at the funny dog over there! It makes me think of *Anna’s dog*.

In example (2–11) given by Riester and Baumann (2017: 23), both “*that dog*” and “*Anna’s dog*” were annotated at the *r*-level. However, at the *l*-level, only the word “*dog*” was annotated without considering the determiner. In fact, the two levels annotate different discourse entities.

To sum up, Section 2.2 has provided an overview of the selected frameworks for classifying information values. Drawing insights from Halliday (1967), Kuno (1978) and Chafe (1976), Prince (1981) taxonomy has proposed a more fine-grained framework that classifies information values into seven types. It has inspired the recent proposals of information classification, namely Prince (1992), Lambrecht (1994) and

Riester and Baumann (2017). However, as indicated above, Prince (1992) leaves Inferrable unidentified; Lambrecht (1994) only intermingles terminologies proposed by Chafe (1994) and Prince (1981) without classifying information values into more subtypes; and the *l*-level of the *RefLex* scheme established by Riester and Baumann (2017) only annotates words without their determiners, which is not related to the classification of information. In comparison, Prince's (1981) taxonomy is fine-grained and it is also feasible for it to be applied to analyze information distributions of nominal expressions in written texts. Therefore, this study will classify information values based on the adapted model from of Prince (1981). Clear definitions will be provided in the methodology of this study (Chapter 3) to avoid the ambiguities between some of the information categories identified in Section 2.2.2.

2.3 Linguistic forms of nominals expressing types of information values

Previous sections of this chapter have presented some salient notions of information and frameworks for classifying information status of nominal expressions. This section presents previous accounts of the relation between linguistic forms of nominals and information status they express.

The initial account on the relation between linguistic forms of nominal expressions and their information status can be dated back to Harris (1771: 215–216), claiming that the use of indefinite articles denotes unknown referents, while the use of definite articles denotes known referents. His research establishes a grammatical correlation between unknown/known and indefinite/definite articles in English. This finding indicates that definiteness of nominal expressions is closely related to type of information status they encode.

The relation between definiteness of nominals and information status also attracts other scholars' attention. Brown and Yule (1983: 189) point out that speakers usually introduce new information with indefinite referring expressions and intonational prominence, while given with attenuated syntactic and phonological forms. Their

observation also shows that it is impossible to establish a classifying framework of information status without considering the linguistic forms of nominal expressions presented by speakers.

Gundel *et al.* (1993) propose a one-to-one correlation between cognitive status and the linguistic forms of referring expressions, namely the Givenness Hierarchy. It is shown in Figure 2-3:

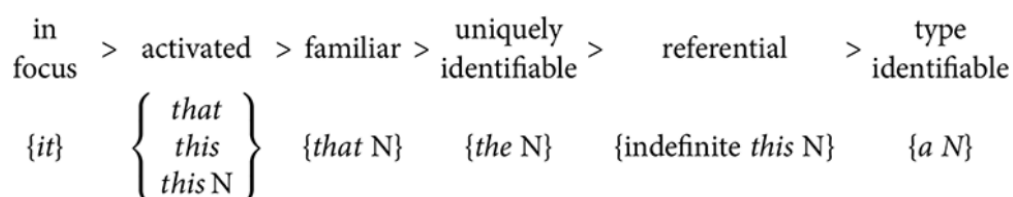


Figure 2-3: The Givenness Hierarchy (Gundel *et al.* 1993)

The Givenness Hierarchy assumes that the readers use different linguistic forms as signals to process the cognitive status of referents (cf. Ariel 1988). The cognitive status decreases in givenness from left to right within the hierarchy. One status could entail the other lower ones. For example, if a referent is in focus, it is also activated, familiar, uniquely identifiable, referential and type identifiable to the readers; if a referent is only type identifiable, it cannot be classified into other types of cognitive status.

The Givenness hierarchy is problematic in several aspects. First, the one-to-one relation between linguistic forms and cognitive status of referents is considered as a major drawback (Kunz 2010: 44). Lambrecht (1994: 79) indicates that “there is no one-to-one relation between identifiability or non-identifiability of a referent and grammatical definiteness or indefiniteness of the noun phrase designating that referent” (cf. Fontaine *et al.* 2023). Second, it does not include all types of linguistic forms that represent cognitive status. Plural nominals, bare nouns and proper expressions cannot be mapped onto the hierarchy (also see Kunz 2010: 44). Third, although Gundel *et al.* (1993) applied this hierarchy to Chinese, Japanese and Russian, they ignored the fact that some languages do not have exactly the same definiteness system as English (cf. Kunz 2010) and they mark definiteness by word order, numeral or case marking (see Lambrecht 1994). Some languages use more types of linguistic forms to represent the

same cognitive status. Furthermore, linguistic forms of referents should not be identified as the only factor to classify their cognitive status. As shown in Section 2.3, the classification of information status could be influenced by many other factors, such as word order, textual distance to antecedents and semantic relations etc.

Unlike the approaches presented above, Prince (1992) explores the relation between definiteness and information status of nominal expressions by taking into consideration subjecthood. Her research shows four points: (i) Hearer–old discourse entities are typically represented by definite nominals and Hearer–new by indefinite; (ii) Discourse–status has no analogous marking of definiteness of nominals; (iii) activated pronominals are more likely to be subject than full nominal expressions; and (iv) linguistic forms of nominals are not helpful in classifying certain information status with particular reference to Inferreds (A review of Prince’s (1981, 1992) frameworks can be seen in Section 2.2).

The relation between definiteness and information status of nominals is difficult to define. Prince (1992) points out that the concept of definiteness to some extent is problematic within English itself. As a formal category, one nominal is identified as definite if it is represented by definite or demonstrative articles, personal pronouns, proper expressions and universal quantifiers. The identification becomes more complex in natural language as some nominals may not carry obvious (in)definite features, for example “*how population–linked pressures affect U.S. cities*” (Prince 1992: 311). Lambrecht (1994: 81) states that “a ‘specific indefinite NP’ is one whose referent is identifiable to the speaker but not to the addressee, while a ‘non–specific indefinite NP’ is one whose referent neither the speaker nor the addressee can identify at the time of utterance”. An indefinite nominal could still be identifiable to the readers if it refers to a specific referent (also see Jones 2014; Fontaine *et al.* 2023: 47). There is no clear-cut boundary between the grammatical category definiteness and the semantic term specificity (also see Hawkins 1978; Heim 1982; von Heusinger 2002a, 2002b; Rijkhoff 1989, 2002).

Previous literature also explores the relation between choice of referring expressions and grammatical roles. Grosz *et al.* (1995) propose the Centering Theory,

ranking the accessibility of referring expressions based on the syntactic positions and linguistic forms. The central assumption is that “pronominalization and subject position are possible linguistic mechanism for establishing and continuing some entity...” (Grosz *et al.* 1995: 13). The scale SUBJECT > OBJECT(S) > OTHER is thus suggested for referring expressions to be the focus of our attention. However, some scholars (e.g. Rambow 1993; Strube and Hahn 1999) point out that it may not well apply for a free-order-language, for example German, as the object can precede the subject and the syntactic function and position are not merged as English.

All the above investigations indicate givenness is closely related to certain linguistic forms of nominal expressions. Chafe (1976, 1994, 1996) suggests a direct correlation between pronouns and their cognitive status in spoken texts. Unstressed pronouns are chosen for retrieving activated information, while stressed nominal expressions represent semi-active and inactive discourse entities. Ariel (1985a, 1985b, 1988, 2001; 2014) proposes the Accessibility Theory exploring all linguistic forms of nominals expressing Given information. The theory regards nominal expressions as linguistic markers to speakers/readers “to retrieve a certain piece of Given information from his memory indicating to him how accessible this piece of information is to him at the current stages of the discourse” (Ariel 2001: 29). The linguistic markers form a continuous scale within Given information with multiple degrees of accessibility, as illustrated as follows:

Full name + modifier > full name > long definite description > short definite description > last name > first name > distal demonstrative + modifier > proximate demonstrative + modifier > distal demonstrative + NP > proximate demonstrative + NP > distal demonstrative (-NP) > proximate demonstrative (-NP) > stressed pronoun + gesture > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > zero

(Ariel 1990: 73)

The Accessibility Theory aims to cover all linguistic features of nominal

expressions, including definiteness, determiner, modifier and accent. It is not only designed for English but also for other natural languages. Not all features are supposed to exist in one language. However, the theory only focuses on nominal expressions expressing Given information and neglects those conveying New information.

This section has introduced previous studies on the relation between information status and the linguistic forms of nominal expressions. It mainly demonstrates that: (i) the one-to-one correlation cannot be established in natural languages; (ii) definite nominals especially pronouns, demonstratives, and possessives, are more likely to express Given information; and (iii) information status is not only influenced by definiteness of nominals but also their syntactic positions. The last point will be further specified in the next section.

2.4 Information status of nominal expressions and thematic structure

Section 2.3 shows that information status is not only related to linguistic forms of nominal expressions but also syntactic positions. This section presents selected previous literature on the relation between information status of nominals and their syntactic positions mainly from a functional perspective, with particular reference to SFL.

Original studies on the relation stem from the Functional Sentence Perspective, which defines information based on word order. According to Mathesius (1975: 156; also see Garvin 1964; Daneš 1974), “the usual position of the theme of an utterance is the beginning of the sentence, whereas the rheme occupies a later position, i.e. we proceed from what is already known to what is being made known”. Firbas (1992) defines thematic elements as those with the lowest communicative dynamism and rhematic elements as those with the highest. Theme basically conveys known information, while Rheme conveys the unknown.

Different from the Functional Sentence Perspective, Halliday (1994: 299) indicates that a speaker will choose the Theme from within what is Given and locate the focus,

the climax of the New, somewhere within the Rheme. This forms the unmarked pattern between the two systems, namely Given ^ Theme and New ^ Rheme. However, Given + New and Theme + Rheme are not the same thing. Halliday (ibid) argues that at any point of the discourse process, the speaker could make choices that are against the background of what has been mentioned and what has happened before, which will override the unmarked pattern of Theme within Given, New within Rheme. Invented examples in (2–12) display both unmarked and marked patterns between thematic structure and information structure in declarative sentences. The Theme-Rheme boundary is marked by “*”.

- (2–12) a. *I* (Given) * *had a little lemon tree* (New).
b. *A lemon tree* (New) * *I* (Given) *had*.
c. *On Sunday night* (New) * *I* (Given) *left*.

In (2–12a), the relationship of thematic structure and information structure is typically unmarked, with “*I*” conveying Given information in Theme and “*a little lemon tree*” conveying New in Rheme. In (2–12b), the case contrasts with (2–12a). Although “*I*” and “*a little lemon tree*” are participants in both instances, “*a little lemon tree*” in (2–12b) as Complement of the clause occurs in the Theme position and turns the relationship of thematic structure and information structure into a marked case. In (2–12c), “*Sunday night*” is embedded within a prepositional phrase. The phrase serves as a circumstance in the Theme position of the clause and conveys New information to the readers/hearers. The examples show that complements and circumstances in Theme position indicate a marked pattern of Theme within New, Given within Rheme.

According to Halliday and Matthiessen (2014), the interaction of thematic structure and information structure carries the rhetorical gist of a clause. Fries (1995a) finds that there is a relation between the purposes of texts and information status in the Theme and Rheme positions of written advertisements. Ghadessy (1995) points out that the placement of information in Theme and Rheme positions of clauses can be regarded as one of the ways of distinguishing written sports commentary and other registers.

Downing (1995) presents the complementarity of Theme and New information by analyzing Chaucer's General Prologue. Peng (2014) proposes the strategy of "presenting New as Given" in Theme position when analyzing a fictional text. Xu and Peng (2020 2021) investigate information distributions in Theme and Rheme positions of special constructions in Mandarin, namely *Duiyu* and *on the stage were sitting X / Y died his father*. The interaction of thematic structure and information structure "gives a composite texture to the discourse" (Halliday 1994: 299) and provides a perspective on distinguishing text types.

2.5 NP-complexity and integration in meaning

As shown in the previous sections, some nominal expressions of certain information category have a higher degree of syntactic complexity, such as containing Inferrables expressed by nominals with embedded phrases and clauses illustrated by Prince (1981) (see Section 2.2). Although these complex nominals have the same information status as the simple ones, they contain more chunks of information. This section aims to present selected literature on the syntactic complexity of nominal expressions and its influence on information, namely how complex nominal expressions are expanded into larger units.

The phenomenon of complex nominal expressions has been widely explored from multiple perspectives, such as the comprehensive grammar of English (Huddleston 1984; Quirk *et al.* 1985; Biber *et al.* 1999; Huddleston *et al.* 2002), Systemic Functional Grammar (Halliday and Matthiessen 2014); Functional Grammar (Dik 1997a, 1997b), Functional Discourse Grammar (Hengeveld and Mackenzie 2006), Lexical-Functional Grammar (Bresnan 1982, 2001) and Cognitive Grammar (Langacker 1991, 2004). These approaches have one thing in common: nominal expressions are basically divided into two types in terms of syntactic complexity, simple and complex.

Previous literature indicates that nominals become more complex with an increase of pre- and post-modifiers, though without a clear-cut boundary between simple and

complex nominals (e.g. Wasow 1997; Hawkins 2004). According to Berlage (2014: 255), postmodification is the key factor that determines complexity of nominal expressions in English and the complexity could be measured by three parameters: (i) the length of nominal expressions, (ii) the structural complexity measured by phrasal nodes and (iii) the degree to which nominal expressions are sentential, which is closely related to the first two parameters.

With more words and embedded structures, nominal expressions become more complex and are expanded both grammatically and semantically. Halliday (1994) proposes logico-semantic relations that expand and link nominals within complex nominals from a functional perspective. Table 2–3 displays them with instances adapted from Halliday and Matthiessen (2014). There are two dimensions of logico-semantic relations related to complex nominal expressions and their postmodifiers, namely taxis and expansion. The first dimension, taxis or interdependency, contains parataxis and hypotaxis. Parataxis is the relation between two elements of equal status while hypotaxis is the relation between elements of different status, one is dominant while the other is dependent.

Taxis Expansion	Paratactic	Hypotactic
Elaboration	depositors – the people who provide the money;	the house that Jack built;
(=)	his latest book, ‘ <i>The Jaws of Life</i> ’;	the house being built by Jack;
		the house by the bridge;
		the person to take pictures;
		his latest book ‘ <i>The Jaws of Life</i> ’
Extension	All the King’s horses and all the King’s men;	the people whose house we rented;
(+)		the incoming government unlike its predecessor;
		his teacup instead of the bread and butter
Enhancement (x)	All those on board, and hence all the crew;	the one I’ve always done the most for;

Table 2–3: Logico-semantic relations of complex nominals adapted from Halliday and Matthiessen (2014: 559)

The second dimension of logico-semantic relationships is the system of expansion, which is composed of three main types, namely elaboration, extension and enhancement.⁶ Their definitions are specified as follows:

- (i) Elaboration: one expression expands another by further specifying, expositing, exemplifying, clarifying, refining or adding descriptive attributes. It can be simplified as the notation “=”, which means ‘equals’;
- (ii) Extension: one expression expands another by adding something new, providing alternatives or replacements. It can be simplified as the notation “+”, which means ‘is added to’;
- (iii) Enhancement: one expression expands another by qualifying it with circumstantial feature of time, place, cause or condition. It can be simplified as the notation “x”, which means ‘is multiplied by’.

Halliday and Matthiessen (2014: 461–487)

The logico-semantic relations of complex nominals influence how meaning is integrated within the unit. According to Halliday and Matthiessen (2014: 430), meaning integration becomes tighter by combining clauses into a clause complex both semantically and grammatically. This principle could also be applied to nominal complex, with the meaning of nominals combined in one unit. Degrees of integration tightness could be displayed by different logico-semantic relations. For example, *his book The Jaws of Life* and *his book, the Jaws of Life* have no significant difference in

⁶ Besides the three types of expansion, locutions and ideas can also be embedded and function as qualifiers within complex nominal expressions. This is considered as Projection, as in *the thought that she might one day be a queen* (Halliday 1994: 264). However, “*she might one day be a queen*” is different from “*Jack built*” in “*that house that Jack built*” illustrated in Table 2–3. Since it can stand on its own independently, it is difficult to argue for it to be classified as a postmodifier in a complex nominal. This actually leads to a discussion about the fuzzy boundary between complex nominal expressions and clauses (cf. Schmid 2000: 25; Keizer 2007: 262), which is beyond the aim of this research. Therefore, projection is not included as a parameter that determines the complexity of nominals in the present study.

semantic meaning, however, they contain different pieces of information. Nominals of *his book The Jaws of Life* indicate a tighter integration in meaning, since they are more tightly related within one nominal complex by hypotaxis, both semantically and grammatically. Since meaning integration is compatible with information integration from the linguistic sense, logico-semantic relations could be used to show how information is integrated within nominal complexes.

2.6 Conclusion: Motivating a multi-dimensional approach to the informational role of nominal expressions in English texts

This chapter has provided an overview of previous literature on information and nominal expressions from multiple perspectives, including the notion of information, classifying frameworks of information status, the relationships between information status of nominals and the three factors, namely linguistic forms, thematic structures and complexities in grammar and meaning.

This chapter has also identified several understudied fields in relation to the informational role of nominal expressions in English texts. First, Prince's (1981) information taxonomy as the most influential and fine-grained framework has distinct advantages over the others. However, the criteria for classifying Brand-new anchored, Inferrable and Unused need to be more specified (Section 2.2). Her framework is thus selected as the basic framework for the present thesis to classify information values of nominal expressions. Second, previous literature mostly focus on proposing various classifying frameworks of information status without applying them to analyzing different texts. Information distributions of nominal expressions in complete texts still remain understudied. Third, the relationship between linguistic forms of nominals and information status were explored with either information status of particular nominals or linguistic forms of particular information status (Section 2.3). Previous studies haven't provided detailed account of this relationship by including all types of linguistic forms and information status. Another understudied field is the relationship between

information structure and clausal positions. The unmarked relationship Given ^ New and Theme ^ Rheme is generally acknowledged, but most previous accounts only focus on Theme without exploring information distributions of nominal expressions in both clausal positions of comparable text types (Section 2.4). Furthermore, as discussed in Section 2.5, in comparison to structural complexity, the complexity of nominal expressions in information has still remain understudied.

This chapter has presented the complex and vast nature of the information of nominal expressions from the most important and relevant literature and laid theoretical foundations for the current thesis. The next chapter will present a multi-dimensional approach by combining text analysis with relevant cognitive linguistics concepts , which is necessary to shed light on the informational role of nominals in English texts.

3 Methodology: A multi-dimensional framework

Having laid the theoretical foundations in Chapter 2, this chapter aims to present a multi-dimensional approach to achieve the main aim of the present study, namely the informational role of nominal expressions in English texts. Relevant features of nominal expressions will be analyzed in texts, including information values, linguistic forms, thematic structures and logico-semantic relations. This chapter provides a detailed account of the data collection and methodological framework used in the analysis.

This chapter is organized as follows. Section 3.1 outlines the source data selected from the American National Corpus. Section 3.2 presents a multi-dimensional framework for the informational role of nominal expressions in English texts. Section 3.3 gives a brief introduction to the annotation tool for the corpus analysis. Finally Section 3.4 offers a summary of the whole chapter.

3.1 Annotated data of four genres from MASC

In the following, Section 3.1.1 first presents the motivations for choosing the MASC as the data source and then Section 3.1.2 introduces the reason for choosing four genres and outline the final dataset annotated for the present study.

3.1.1 The manually annotated sub-Corpus from the American National Corpus (MASC)

The present study selected the Manually Annotated Sub-Corpus of American National Corpus (MASC) as data source. The MASC contains about 500,000 words of written texts in contemporary American English and transcribed speeches. The corpus is

primarily extracted from the Open American National Corpus⁷ (OANC) with free access (Ide *et al.* 2002; Ide and Suderman 2004; Ide 2008). The OANC is a growing subset text corpus of the American National Corpus (ANC) produced by native English speakers from 1990 onward (Fillmore *et al.* 1998).

Based on Ide *et al.* (2010), the MASC has the following advantages. First, it selects texts with a balanced distribution of a broad range of genres, which is different from most corpora with free access. Each genre accounts for 5% or 6% of the whole MASC. Second, compared with other corpora, such as the British National Corpus (BNC), it includes traditional genres as well as updated ones, such as emails, tweets and blogs. Although the present analysis does not contain the genre of web data, they can be annotated in later comparative studies. Third, it has a clear distinction between spoken and written data. As the present study only focuses on written texts, it is important to ensure that all the data analyzed had originated from written text. In addition, it enables full-text linguistic annotation. Some online research corpora, such as the Corpus of Contemporary American English (COCA), are only available selectively through web browsers, without an access to full data and all annotations on account of copyright.⁸ Data in the MASC is not only accessible through a web browser but can also be downloaded without charge. It is also important to note that the whole OANC and its sub-corpora are available in multiple formats, which is suitable for a wide variety of annotating softwares. More importantly, the sentence boundaries of the MASC have been validated with manual annotations. This is helpful in identifying the departure points of the sentences, which is closely related to the annotation of thematic structures in this study (see Section 3.2.2).

Therefore, the MASC has been selected as an appropriate data source to explore the informational role of nominal expressions in English texts, given the clear distinction between spoken and written text, the free availability of full-text annotation as well as its wide adaptability to annotation software.

⁷ <http://www.anc.org/>

⁸ https://en.wikipedia.org/wiki/American_National_Corpus

3.1.2 Annotated data of the four genres selected from MASC

As indicated before, texts of four written genres were selected from the MASC for the present study, namely Travel guide, Essay, Government documents and News report. The first commonality the genres have is a fundamental purpose on conveying information rather than developing interpersonal relationships (also see Biber and Conrad 2014: 109). As informational written genres, they focus on adding something new to our existing knowledge. The second similarity among the four genres is that no direct personal connection is established between the readers and writers. Unlike emails or letters, personal details do not play a role in the genre analysis. The third important aspect concerns is the way they are written. Writers have more time for planning, revising and editing the texts. This is closely related to the linguistic features of nominal expressions (also see Biber and Conrad 2014: 114).

Although they have similarities as written genres, they are published for different purposes. Specifically, News report describes current events with newsworthiness; Travel guide provides all kinds of information about a place; Essay as a subgenre of academic prose, is a piece of writing with arguments about a particular topic and Government document is any information that is published by a government agency in a position of authority. Besides, we also have the intuition that Travel guide and News report are often easier to read than Essay and Government documents. Travel guide and News report are usually regarded as general interests with a larger scale of intended audience. They focus on simply presenting information without developing arguments. In contrast, Essay and Government documents are written for readers with specialist background knowledge of particular topics. They are expected to offer further explanations and interpretations. Both similarities and differences between the texts of four genres motivate us to explore the informational differences of nominal expressions.

Table 3–1 provides basic information of the data annotated for the present thesis, which includes the document name and its abbreviation, genre, word account and the frequencies of nominal expressions in each text. The data contains 13,243 words and 3095 nominal expressions in total. Two genres, Travel guides and Newspaper, contain

multiple texts, with 1960 and 1992 words, 459 and 514 nominal expressions respectively. Only one text is included in the genres Essay and Government document. This is caused by a number of reasons. Since text length is assumed to be one of the factors having an impact on the number and linguistic features of nominal expressions, multiple texts were selected for the genres of Travel guide and News report to ensure a sufficient corpus size to allow generalizable findings. For example, Np2 only has 44 nominals. If it is the only text representing the genre of News report, it would be difficult to find more varied features of nominals or draw general conclusions. The word count and number of nominals of Travel guide, News report and Essay do not differ greatly. Furthermore, the corpus annotation was carried out manually. The classification of information values is based on the working memory of the annotator. Texts with a large number of nominals make corpus annotation unmanageable. Although the text of Government documents selected here clearly contains more nominals than texts of the other three genres, it is the shortest under the genre of Government documents in the MASC.

Document name with abbreviation	Genre type	NP frequency	word count	NP percent (%)
IntroHongKong (Tg1)	Travel guides	191	824	23.18
IntroDublin (Tg2)	Travel guides	268	1136	23.59
wsj_2465 (Np1)	Newspaper	187	752	24.87
A1.E1.-NEW (Np2)	Newspaper	44	188	23.4
20000415-APW_ENG_NEW (Np3)	Newspaper	75	275	27.27
wsj_0026 (Np4)	Newspaper	52	215	24.19
wsj_0158 (Np5)	Newspaper	92	325	28.31
wsj_0027 (Np6)	Newspaper	64	237	27
Conclusions of the Financial Crisis Inquiry Commission (Gd)	Government documents	1540	6788	22.69
Homosexuality (Ey)	Essay	582	2503	23.25

Table 3–1: Basic information of the data for final annotation⁹

⁹ Note that some of the word counts do not correspond with those given by MASC because irrelevant information such as references, text numbers and authors' affiliations were omitted in the present study.

3.2 A multi-dimensional framework

This section presents the linguistic features of nominal expressions, with particular reference to varied attributes and embedded values analyzed in this study. Each attribute has a name and a set of possible values, which can be selected for nominals. When editing the annotation scheme, attributes and values are shown in a sequence of tags. For example, the tag `attribute_modification` indicates that the name of the attribute is `modification` of nominal expressions; under the attribute, the tag value `modification_pre+Head` means one of the possible values is nominal expressions with premodifiers and head nouns. In the following, Section 3.2.1 and Section 3.2.2 describe the annotation attributes for nominal expressions and information status respectively.

3.2.1 Annotation attributes for nominal expressions

Given that the current study cannot cover all possible information aspects in English texts, it is restricted to the informational contributions of nominal expressions. Non-nominal elements, such as the preposition *in* in *sightseeing in Hong Kong*, are excluded from the corpus analysis. The nominal expressions analyzed in this study mainly contain two kinds: those that independently function as clause elements and those embedded as constituents within complex nominal expressions.

The attributes for nominal expressions in the manual annotation were defined based on the different properties discussed in Chapter 2. They are listed and described in Table 3-2. Their embedded values are elaborated as follows:

Main Tag	Description
attribute_ne	Basic forms of nominal expressions
attribute_modification	Basic modifying types of nominal expressions
attribute_continuity	Types of complex nominal expressions in terms of extraposition
attribute_expansion	Basic types of expansion
attribute_interdependency	Basic types of interdependency
attribute_lexical	Types of semantic relations formed by nominal expressions
attribute_lexical_set	Sets formed by nominal expressions that establish a semantic relation
Non-annotated nominal expressions	Idioms and expletives

Table 3–2: Annotation attributes for nominal expressions

Attribute_ne: this label is only assigned to the segmented nominal elements in terms of linguistic forms. A detailed description of the set of values are given as follows:

- (i) value_ne_zero: bare expressions structured only with nominal Heads. E.g. *fate, luck, hotels*
- (ii) value_ne_proper: proper nominals that indicate personal names, geographical names, objects and commercial products, holidays, months, years, days of the week, religions and relational concepts, persons with unique public functions, public buildings, institutions, laws etc., political parties and languages and nationalities. E.g. *Shakespeare, USA, Hong Kong, 1 July, 1997, 2008*

Note: the above two values are different from the value_modification_Head of the attribute_modification. This is shown by example (3–1):

- (3–1) a. *Shopping* never ends. (Tg1)
- b. *Sightseeing in Hong Kong* starts at sea level with the enthralling water traffic — a mix of freighters, ferries, tugs, junks, and yachts. (Tg1)

In the corpus annotation of this thesis, both “*shopping*” and “*sightseeing*” are labelled as `value_ne_zero` under the attribute of noun forms. However, under the attribute of modification, “*shopping*” in (3–1a) is labelled as `value_modification_Head`, “*sightseeing*” in (3–1b) is annotated as `value_modification_Head+post`. The annotation of `attribute_ne` is only based on the linguistic form of a nominal element, while the annotation of `attribute_modification` is based on the structure of a complete nominal.

(iii) `value_ne_pronoun`: the label indicates various kinds of pronouns, which contains possessive, first/second/third personal, demonstrative, reflexive, reciprocal and other pronouns.

(a) `value_ne_pronoun_possessive`: possessive pronouns, e.g. *hers, ours*;

(b) `value_ne_pronoun_personal`

(b1) `value_ne_pronoun_personal_first`: first personal noun: *I*;

(b2) `value_ne_pronoun_personal_second`: second personal pronoun, e.g. *you*;

(b3) `value_ne_pronoun_personal_third`: third personal pronoun, e.g. *she, he, they*;

(c) `value_ne_pronoun_demonstrative`: demonstrative pronoun, e.g. *this, that*;

(d) `value_ne_pronoun_reflexive`: reflexive pronoun, e.g. *myself, himself*;

(e) `value_ne_pronoun_reciprocal`: reciprocal pronoun, e.g. *each other, one another*;

(f) `value_ne_pronoun_other`: pronouns that do not belong to the above categories, e.g. *others, nothing*;

(iv) `value_ne_determiner`: nominal expressions with determiners are mainly divided into three subgroups: expressions with articles, possessive or demonstrative determiners (`determiner_central`), expressions with determiners that occur before the three subcategories (`determiner_pre`) and expressions with other determiners (`determiner_other`).

(a) `value_ne_determiner_pre`: nominal expressions that are further premodified

by other determiners in front of articles, possessives and demonstratives (cf. Biber *et al.* 1999). E.g. *such a crowded place, both these cakes*.

(b) value_ne_determiner_central: nominal expressions premodified by articles

(b1) value_ne_determiner_central_article:

b1.1) value_ne_determiner_article_indefinite: expressions premodified by indefinite articles, e.g. *an intoxicating place, a steady business*;

b1.2) value_ne_determiner_article_definite: expressions premodified by definite articles, e.g. *the street, the people*.

(b2) value_ne_determiner_central_possessive: expressions premodified by possessive determiners, e.g. *its people, their usual topics of conversation*;

(b3) value_ne_determiner_central_demonstrative: expressions premodified by demonstrative determiners, e.g. *these words, this beautiful city*

(c) value_ne_determiner_other: expressions premodified by other determiners.

E.g. *another inviting spot, two major racetracks*

(v) value_ne_other: nominal expressions that do not belong to the above categories.

There are mainly two subtypes. First is those with premodifiers that cannot be categorized into the above labels, e.g. *left-wing politics*. Second is the *of*-NP construction, within which there are more than two nominal elements and the first one does not have any determiners, e.g. *one of the world's greatest population densities*.

Attribute_modification

(i) value_modification_of: nominal expressions that describe fractions and proportions or contain the relator *of* in between, like *nearly one in 10 mortgage borrowers* from Government document and *one of the world's greatest population densities* from Tg1.

Note 1: *Of*-NP constructions were annotated as a whole unit without further segmentation. It also applies to those with more than one *of*relators or with other kinds of embedded phrases in between, see example (3-2).

- (3–2) a. Panic fanned by ***a lack of transparency of the balance sheets of major financial institutions***, coupled with a tangle of interconnections among institutions perceived to be “too big to fail,” caused the credit markets to seize up. (Gd)
- b. The CRA was enacted in 1977 to combat “redlining” by banks — ***the practice of denying credit to individuals and businesses in certain neighborhoods*** without regard to their creditworthiness. (Gd)

Although *of*-NP constructions can be further categorized into different types based on various semantic, syntactic and pragmatic features, the present thesis is only concerned with information values in text. Secondly, it is difficult to classify the head nouns of *of*-NPs and the elements introduced by the relator “*of*” as modifiers or complements (e.g. Francis *et al.* 1998: 176–199; Hawkins 1981; Keizer 2007; Traugott 2008a, 2008b; Schönthal 2016), see example (3–3). While annotating *of*-NP constructions with varied linguistic features is not directly relevant for the current research, it does cause some issues for the corpus analysis.

- (3–3) a. On Tuesday the conference got word of another atrocity, ***the assassination in Medellin of two employees of El Espectador***, Colombia’s second-largest newspaper. (Np1)
- b. ***The enactment of legislation in 2000 to ban the regulation by both the federal and state governments of over-the-counter (OTC) derivatives*** was a key turning point in the march toward the financial crisis. (Gd)

Thirdly, many *of*-NP constructions contain at least one embedded nominal expression and establish semantic relationships within themselves. Based on Prince’s (1981) information taxonomy, whole constructions are more likely to express containing Inferrable information.

Exception: In written English, an *of*-NP construction can be separated by some punctuation, usually commas, such as the instance in (3–4). In this case, nominals on either side of the relator “*of*” are separately annotated without regard to Note 1.

- (3–4) While there was *some awareness of*, or at least a debate about, *the housing bubble*, the record reflects that senior public officials did not recognize that a bursting of the bubble could threaten the entire financial system. (Tg1)

In (3–4), “*some awareness*” and “*the housing bubble*” are not presented as a joined unit and represent two pieces of information. Therefore, the two expressions were annotated separately in the corpus analysis by leaving out the relator “*of*”.

Note 2: Additional postmodifiers of the *of*-NP constructions are annotated under the tag of `value_postmodifier`. “*Risky mortgages*” in example (3–5) as the postmodifier of “*trillions of dollars*” is labelled as `value_postmodifier`.

- (3–5) *Trillions of dollars in risky mortgages* had become embedded throughout the financial system, as mortgage-related securities were packaged, repackaged, and sold to investors around the world. (Gd)

- (ii) `value_modification_Head`: nouns without any modifiers, such as *shopping* in *shopping is everywhere*;
- (iii) `value_modification_pre+Head`: nominal elements only with premodifiers, for example *an intoxicating place* in *it is an intoxicating place*;
- (iv) `value_modification_Head+post`: nominal elements only with postmodifiers, for example *voices* in *you don't hear voices raised in anger*;
- (v) `value_modification_pre+Head+post`: nominal elements with both premodifiers and postmodifiers, for instance *other factors* in *other factors that influence cities all*

over the world;

(vi) value_postmodifier: nominal elements as postmodifiers, such as *cities* and *the world* in *other factors that influence cities all over the world*. Some nominal expressions have several postmodifiers. They were further annotated in terms of postmodifying order, which is represented by a set of values in letter notation, namely a, b, c, d and d+:

- (a) value_postmodifier_a
- (b) value_postmodifier_b
- (c) value_postmodifier_c
- (d) value_postmodifier_d

Since the postmodification system of nominal expressions is progressive (Halliday 1979, 1994; Halliday and Matthiessen 2014), it is not possible to list the sequence of all postmodifiers. Those in a more than four-word modifying position were thus all labeled as “d+”. Example (3–6) is an illustration:

(3–6) This report catalogues the corrosion of mortgage–lending standards and *the securitization pipeline* that transported *toxic mortgages* (postmodifier_a) from *neighborhoods* (postmodifier_b) across *America* (postmodifier_c) to *investors* (postmodifier_d) around *the globe* (postmodifier_d+). (Gd)

Note 2: Certain nominal expressions can modify more than one head nouns. In such cases, all the head nouns were assigned to the feature of “+post”. Example (3–7) illustrates the point:

(3–7) This report describes *the events and the system that propelled our nation toward crisis*. (Gd)

“*The events*” and “*the system*” in (3–7) happen to share the same postmodifying

clause “*that propelled our nation toward crisis*”. In addition, both “*the events*” and “*the system*” are premodified by the definite article. Thus both were labeled under the attribute of modification as “pre+Head+post”.

Note 3: Expressions of non-defining relative clause (also called non-restrictive clause) were not annotated as postmodifier.

Non-defining relative clauses are different from defining relative clauses in two aspects: information integration and level of language. As far as information is concerned, non-defining relative clauses add more information to something that is regarded to be fully specific. Strictly speaking, the addition indicates a relation of extension rather than elaboration (Halliday 1994: 227–228). Furthermore, non-defining clauses are typically marked off with commas or dashes in written texts (e.g. Biber *et al.* 1999: 603; Halliday 1994: 228). They form a separate tone group that signals the apposition relationship in English (Halliday 1994: 228). For this reason, non-defining clauses are not semantically integrated as a whole with the nominals occurring before the comma or dashes.

As far as the level of language is concerned, non-defining relative clauses are not restricting to defining nominal expressions. Their domain of antecedents can also be a clause, such as the instance in (3–8) given by Halliday (1994: 227). They are not postmodifiers within nominal expressions, which is different from *whose house we rented* in *the people whose house we rented* (see Chapter 2, Section 2.5). Therefore, non-defining relative clauses were not analyzed as postmodifiers of nominal expressions in the present study.

(3–8) From then on we started winning prizes, *which turned out to be very easy*.

Note 4: There is a grey area between postmodifier and complement.

The syntactic relations between the clauses and the head nouns are not defined clearly (e.g. Quirk *et al.* 1985: 1231–1260; Herbst 1988). If the instance is workable for either test as follows, then it is not labeled as postmodifier in the corpus analysis.

Test 1: For fact–clauses, try to leave out the noun heads of a complex unit. If the remaining clause can stand on its own independently, do not annotate them as postmodifier. If not, assign the postmodifier label.

Example 1: *the fact that so many other funds were exposed to the same risks as those hedge funds* (Gd)

Test: So many other funds were exposed to the same risks as those hedge funds. ☑

Result: Do not annotate *that*–clause as postmodifier.

Analysis: “*That*” here serves as a binder instead of a relative (cf. Halliday and Matthiessen 2014: 494). There is no meaning gap between the head noun (cf. Biber *et al.* 1999: 645) and the clause. The whole unit can also be transformed to two similar sentences: *the fact is so many other funds were exposed to the same risks as those hedge funds* and “*so many other funds were exposed to the same risks as those hedge funds*” is a fact and furthermore the use is often motivated by pragmatic, stylistic or rhetorical purpose (cf. Keizer 2007: 273; Schmid 2000: 24).

Test 2: For *to*–clauses, try to dislocate the postnominal part to the left and add the copula *be* between it and the noun head. If the transformation is interpretable, do not assign them as postmodifier. If not, assign the postmodifier label.

Example 2: But it is a good idea ***to think and stop it***. (Schmid 2000: 24)

Test: To think and stop it is a good idea. ☑

Result: Do not annotate the *to*–clause as postmodifier.

Analysis: The transformation, though less common than the original one, is acceptable. Obviously, it contributes none representation to the noun head.

The two tests, however, cannot work well for all the possible sequences of nominal

heads followed by clauses. Some are highly marginal instances, such as the example given by Schmid (2000: 25), *a great place to live and work in*. In addition, there is no clear-cut boundary between complex nominal expression and clause. In some cases, it is impossible to make a strict separation between complementing/appositive, relative and adverbial clauses as postmodifiers (cf. Keizer 2004; Keizer 2007: 262), with particular reference to *wh*-clauses, such as the example given by Schmid (2000: 26) that *a time when Washington believed that there was a real threat that southeast Asia would fall under communist rule*. Cases of the two controversial types were analyzed as postmodifiers in this study. Further syntactic explorations of the boundary between nominal and clause is beyond the scope of this study.

Attribute_continuity: This label applies to the nominal Heads of postmodified expressions. For a variety of syntactic or pragmatic reasons, a nominal expression may not be presented continuously and where this is not the case, it is necessary to add an attribute of continuity. Such an attribute has three values:

- (i) value_continuity_null: head nouns without postmodifiers;
- (ii) value_continuity_continuous: head nouns with postmodifiers that occur right after, for instance *trillions of dollars* in *trillions of dollars risky mortgage*;
- (iii) value_continuity_discontinuous: head nouns with postmodifiers that do not occur right after, see example (3–9):

- (3–9) a. To alleviate the problem, the government has become the city’s major landlord with *the construction of massive apartment blocks that, though they have every modern facility, average only 9m² (100 square ft.) in size*.¹⁰ (Tg1)

¹⁰ In example (3–9a), “*though they have every modern facility*” was not analyzed as the postmodifier of “*the construction of massive apartment blocks*”. Like the non-defining clauses, it is assumed to have a separate tone group from the *of*-NP construction and they do not form an integrated information unit or semantically function as a whole.

- b. Dr. Evelyn Hooker, a heterosexual psychologist, conducted *a ground-breaking study* in the mid-1950s *that went along similar reasoning as Freud*. (Ey)

The two instances in (3-9) are discontinuous nominal expressions, with interruptions between the head nouns and postmodifiers. The instance in (3-9a) is interrupted by comma, with “*though they have every modern facility*” inserted. The head noun “*the construction of massive apartment blocks*” was assigned with “value_continuity_discontinuous”. In (3-9b), the head noun “*a ground-breaking study*” is also separated from its postmodifier “*that went along similar reasoning as Freud*” by the prepositional phrase “*in the mid-1950s*”.

Attribute_expansion: The label is restricted to the postmodifiers and appositions of nominal expressions.

- (i) value_expansion_null: nominal expressions that are not involved in types of expansion;
- (ii) value_expansion_elaboration: nominal expressions that are involved in elaboration;
- (iii) value_expansion_extension: nominal expressions that are involved in extension;
- (iv) value_expansion_enhancement: nominal expressions that are involved in enhancement;
- (v) value_expansion_other: nominal expressions that are involved in expansion more than once. As *voices, anger, motorists, their horns* in (3-11), they are involved in both elaboration and extension, parataxis and hypotaxis.

- (3-11) But it is also efficient, with one of the best transportation systems anywhere, and for such a crowded place, quiet — you don’t hear *voices raised in anger, motorists sitting on their horns, or loud boomboxes* (Tg1)

Note 1: examples like *the examiner's assessment, a brilliant work* given by Halliday and Matthiessen (2014: 559), which crosses the borderline between elaboration and projection, are annotated as elaborating parataxis in the current analysis.

From the perspective of traditional grammar, the two elements are considered to be in apposition to each other if they refer to the same entity. They form a typical relationship of elaboration. At the same time, it can be categorized as a case of projection as the second element is projected by the first with the single quotes. However, the projection is not included in the current annotation scheme. The ultimate annotation lies on elaboration.

Note 2: If the Head nouns inherently carry enhancing features and are elaborated by embedded clauses, all elements of the nominal complexes were annotated as `value_expansion_elaboration`.

(3–12) *The time I like best* is the hour before dawn. (Halliday and Matthiessen 2014: 501)

Some Head nouns can construe circumstantial relations within themselves, such as time, place and reason etc. (Halliday 1994: 247). “*The time*” in example (3–12) is a case in point. It is enhancing inherently and is also elaborated by the embedded clause “*I like best*”. Cases like (3–12) were labeled as “`value_expansion_elaboration`”.

Note 3: In some cases, the category `value_expansion_elaboration` is difficult to distinguish from the category `value_expansion_enhancement`.

The relationship of expansion can be interpreted in two ways and we can illustrate it with the example taken from Halliday (1994: 245), *the problem with asking directions*. *Asking directions* not only elaborates the primary element by specifying *the problem*, but also adds a circumstantial feature to *the problem* by indicating the cause. There is no obvious advantage of categorizing one over the other. Such instances were analyzed

as value_expansion_other due to the combined meanings.

Attribute_interdependency

(i) value_interdependency_null: nominal expressions that indicate no interdependency;

(ii) value_interdependency_parataxis: nominal expressions involved in parataxis;

(a) value_interdependency_parataxis_1;

(b) value_interdependency_parataxis_2;

(c) value_interdependency_parataxis_3;

(d) value_interdependency_parataxis_4;

(e) value_interdependency_parataxis_4+;

Paratactic structures were represented by numeral notation 1 2 3..., which follows Halliday (1994: 218). As the structure is continuing (Halliday 1979, 1994; Halliday and Matthiessen 2014), it's not possible to list all the sequences of paratactic elements. Nominals were thus all labeled as "4+" if they occur after the fourth nominal in parataxis. For example:

(3–13) You will also read about read about the forces at work behind the breakdowns at Moody's, including *the flawed computer models* (parataxis_1), *the pressure from financial firms that paid for the ratings* (parataxis_2), *the relentless drive for market share* (parataxis_3), *the lack of resources to do the job despite record profits* (parataxis_4), and *the absence of meaningful public oversight* (parataxis_4+). (Gd)

(iii) value_interdependency_hypotaxis: nominal expressions involved in hypotaxis.

(a) value_interdependency_hypotaxis_a;

(b) value_interdependency_hypotaxis_b;

(c) value_interdependency_hypotaxis_c;

(d) value_interdependency_hypotaxis_d;

(e) value_interdependency_hypotaxis_d+.

Hypotactic structures were represented by the same letter notation used for

postmodification. The dominant element is labeled as “hypotaxis_a” and then the dependents are *b, c, d...* Likewise, as the hypotactic structure is progressive, nominals were all labeled as “d+” if they occur after the third postmodifier within the nominal complex. See example (3–14):

- (3–14) This report catalogues *the corrosion of mortgage–lending standards and the securitization pipeline* [hypotaxis_a] that transported *toxic mortgages* [hypotaxis_b] from *neighborhoods* [hypotaxis_c] across *America* [hypotaxis_d] to *investors* [hypotaxis_d+] around *the globe* [hypotaxis_d+]. (Gd)

Attribute_lexical

The Attribute_lexical label applies to nominal expressions that establish synonymy, antonymy, meronymy/holonymy, hypernymy/hyponymy and other semantic relations. It is worth noting that semantic relations here are relations of lexical meanings and are independent of specific context and text.

- (i) value_lexical_null: nominal expressions that do not form any semantic relations;
- (ii) value_lexical_synonymy: expressions that indicate identical experiential meanings;

Note that abbreviations, such as *the European Union* and *the EU*, are annotated as partial repetitions of given entities, which is different from Riester and Baumann’s (2017) approach that annotates *the European Union* and *the EU* as synonyms (more details can be seen in the label Evoked_Textually).

- (iii) value_lexical_antonymy: expressions that indicate opposite experiential meanings;
- (iv) value_lexical_meronymy/holonymy: expressions that indicate part–whole relations;
- (v) value_lexical_hypernymy/hyponymy: expressions that indicate general–specific relations;
- (vi) value_lexical_other: expressions that form other types of semantic relations.

Attribute_lexical_set

Nominal expressions within a semantic relation are marked as a set. Each set contains at least two expressions. As illustrated in (3–15), “*gambling*”, “*cards*” “*mahjong*” “*the lottery*” and “*the horses*” are in one meronymy/holonymy set.

- (3–15) You’ll also notice that *gambling* is a passion, whether it be *cards*, *mahjong*, *the lottery*, or *the horses*. (Tg1)

Non-annotated nominal expressions

Not all nominal expressions have values in corpus annotation. There are mainly two classes of non-annotated expressions: idiom and expletive.

- (i) Idiom: an idiom is a fixed expression with a meaning that is not derived from the sum of the meanings of its components, such as (3–16a), and it can be rephrased by other syntactic categories that do not contain nominals, as illustrated in (3–16b). In addition, nominal expressions within idiomatic phrases, such as “*terms*” in “*in terms of*” and “*light*” in “*in the light of*”, have no value as content words. Nominals in idioms cannot introduce discourse entities or have references and they are not included in the corpus analysis.

- (3–16) a. To paraphrase Shakespeare, *the fault lies not in the stars*, but in us.
(Gd)
b. *to go back to the drawing board* = to start all over (Riester and Baumann, 2017: 12)

- (ii) Expletive: an expletive has no referential meanings but occupies a syntactic position that usually functions as dummy Subject. Example (3–17) illustrates the point:

- (3–17) a. *It* won’t rain on you in Dublin all the time. (Tg2)

- b. As this report goes to print, *there* are more than 26 million Americans who are out of work, cannot find full-time work, or have given up looking for work. (Gd)

Other annotation conventions

(i) Coordination

Coordinated nominal expressions, which do not fall into *of*-NP constructions as *2007* and *2008* in *the events of 2007 and 2008*, are labeled separately as independent units. In this way, their coordination or any of the conjuncts can form co-referentiality. It is illustrated by example (3–18):

(3–18) *The Conservatives* and *the Social Democrats* have found an agreement.

a. *They* decided not to raise taxes.

b. It was *the Conservative Party* who had promised this to their voters.

(Riester and Baumann 2017: 21)

(ii) The annotation of numerals

Numerals mainly occur as determiners and their forms can be influenced by the syntactic structure of nominal expressions that begin with the multiplier “1”. For example, *one million* can be freely replaced by *a million*. In the present thesis, the article was analyzed as one part of the numeral instead of an independent determiner and therefore the nominal expression *one million dollars* was labeled as “determiner_other” and “pre+Head” in the annotation scheme as other numerals with measuring units.

(iii) The annotation of nominals that do not refer to unique entities

As indicated before, the annotation scheme aims to classify all nominal expressions that convey information in English text. Although some expressions do not refer to unique entities, such as *every Wednesday*, they are annotated in the present research as informational elements.

3.2.2 Annotation attributes for information status and thematic structure

Attributes for the information status in text for the manual annotation were defined on the various properties discussed in Chapter 2. They are listed and described in Table 3–3. The embedded values are elaborated following this table. The given examples are mainly selected from the data analysis for the present study.

Main Tag	Description
attribute_thematic	Thematic status of nominal expressions
attribute_infovalue	Information status of nominal expressions in text
attribute_givennew_set	Sets formed by nominal expressions referring to the same discourse entities

Table 3–3: Annotation attributes for information status and thematic structure

Attribute_thematic

- (i) value_thematic_null: this label is used for nominal expressions that realize no thematic functions, such as those representing titles, authors’ affiliations or appendixes attached at the end of a text;
- (ii) value_thematic_Theme: this label refers to the expression that locates at the point of departure of a clause. It is further annotated as circumstance and participant in terms of the role in transitivity. Unmarked Theme is typically mapped on to Subject as participant, as “*Timex*” in (3–19):

(3–19) *Timex* is a major U.S. producer and seller of watches. (Np4)

- (a) value_thematic_TopicalTheme_circumstance: this label is used for the expression functioning as circumstance in the experiential structure of a clause;
- (b) value_thematic_TopicalTheme_participant: this label is used for the expression functioning as participant in the experiential structure of a clause.

Regardless of register-specific considerations, a Theme is considered to be marked in the forms illustrated in example (3–20), which are all closely related to nominal expressions. It can be seen that the marked Theme is either an adjunct serving as circumstance or even more infrequently, a complement serving as participant in a clause (see Halliday and Matthiessen 2014: 100). However, as far as information status is concerned, there is no one-to-one correlation to syntactic functions. Marked Themes can convey both Brand-new and Evoked information based on Prince's (1981) framework. Therefore, the annotation of Theme is only bounded by basic transitivity roles.

- (3–20) a. *Over **the past 12 years***, at least 40 journalists have died there. (Np1)
- b. ***Two things*** we need to comment on. (Halliday and Matthiessen 2014: 101)
- c. We are aware of our responsibility to our critics. We are also aware of our responsibility to the author, who probably would not have authorized the publication of these pages. ***This responsibility*** we accept wholly, and we would willingly bear it alone. (Halliday and Matthiessen 2014: 99)
- d. *After all, except for **music***, what did they have in common? (Halliday and Matthiessen 2014: 103)
- e. *From **this crossroads town*** follow the main road south through increasingly arid landscapes towards Rembitan, a pretty little village claiming a 17th-century mosque, then Sade. (Halliday and Matthiessen 2014: 103)

Note 1: WH-elements in interrogatives are not annotated in corpus analysis

In a WH-interrogative, the WH-element functions as Theme and represents the piece of missing information that is required by the addressor, which is obligatorily new (also see Halliday and Matthiessen 2014: 101). The annotation of the WH-element does not contribute to this study which aims to explore

variations in the information distribution of nominal expressions.

Note 2: The annotation of Themes in complex clauses is located at the level of clause.

For example, both *some* and *it* are labelled as Theme in *some would say it is a gene, passed on from parents to child*.

(iii) value_thematic_Rheme: this label is used for nominal expressions that do not occur in the departure point of a clause.

Attribute_infovalue

As indicated above, the annotation of information status in this study is mainly refined from Prince's (1981) framework. Although Section 2.2 provided a detailed overview of the properties and weaknesses, additional criteria are needed to cope with ambiguous cases to achieve a fine-grained and precise classification. For convenience, the definition of each category is briefly repeated here.

(i) value_infovalue_New: this label refers to an entity that has not occurred in the previous text;

(a) value_infovalue_New_Brand-new: this label is applied to a new entity introduced by speaker/writer;

(a1) value_infovalue_Brand-new_anchored: the expression is linked to other discourse entities via another expression that was mentioned previously; To avoid ambiguity, several points concerning the anchor need to be specified. First, the anchor should only convey evoked information. This has not been specified in Prince's (1981) framework. Second, it obviously repeats the semantic and morphological features of an evoked referent. Third, it cannot establish a semantic relation with the anchored referent. For example "*a rich guy*" and "*I know*" in "*a rich guy I know*" do not establish a semantic relation as "*the topic*" and "*my thesis*" in "*the topic of my thesis*" do.

(a2) value_infovalue_Brand–new_unanchored: this label is assigned to a complete new discourse entity, which cannot be derived from background knowledge or the preceding text.

(b) value_New_Unused: the label is used for an entity assumed to be known by hearer/reader without specific context and text;

This label is not further categorized into known and unknown types as Riester and Baumann (2017) suggest. There are two main reasons. Apart from the ambiguous distinction between known and unknown, further categorization is more functional for spoken corpus analysis. It can be illustrated with another example from Baumann and Riester’s (2012).

(3–21) We are sitting in the lobby.

a. *Harry Smith* came up to us.

b. *George Clooney* came up to us.

In example (3–21a), “*Harry Smith*” was classified as an unknown Unused item. The name is part of the knowledge shared by the interlocutors at the same setting and time of the utterance, which is more likely to be seen in conversations between friends. Unlike the case in example (3–21), the writers and readers show a rather distant relationship in the texts of four genres analyzed in this study. Therefore, it is not worth using the known/unknown distinction here.

(ii) infovalue_Inferrable: this label is used for discourse entities that are classified via inference;

(a) value_Inferrable_non–containing: this label is applied to entities inferred via logical reasoning and contextual and textual environments;

Note 1: Non–containing Inferrables are different from the category Brand–new anchored in four main ways

First, they do not necessarily carry the same semantic or morphological features with their antecedents/triggers; second, they form semantic relationships with their antecedents/triggers; third, the entities tend to be definite, usually with linguistic features signified; and furthermore the triggers of inference are not restricted to specific evoked entities, and they can also be propositions.

Note 2: Discourse entities within semantic relations are not necessarily counted as non-containing Inferrables.

(3–22) *A bird* is sitting in the tree. It has just lost *a feather*. (Riester and Baumann 2017: 9)

As indicated in Section 2.2, expressions within semantic relations, especially whole–part relations, are inclined to be Inferrables. However, the inclination becomes slight or even untenable in some cases. In example (3–22), the two indefinite expressions make a typical contribution to the whole–part relation. Some accounts (e.g. Riester and Baumann 2017: 9) treat “*a feather*” as a bridging anaphor that corresponds to the non-containing Inferrable in Prince’s (1981) terminology. This approach causes two problems in corpus annotation.

First, their categorization is not based on specific contextual and textual grounds. Second, this categorization violates the strategy of a balanced information distribution (BID) (Doherty 2006). In general, information is distributed in a balanced way with given and new progressing in order. In English, this strategy will only be violated when certain elements need to be highlighted for a specific pragmatic purpose. The violation can be seen from marked constructions in written text, such as Topicalization and Raising.

However, the instance in (3–22) does not meet the conditions for violating the BID. Based on the syntactic structure and the ordinary context scenario, new things are expected to facilitate the information unfolding after the given entity “*it*”. The indefinite expression “*a feather*” cannot be activated from the readers’/hearers’ long-term

memory. Therefore, discourse entities within semantic relations like “*a feather*” in (3–22) are not necessarily annotated as non–containing Inferrables in this study.

- (b) `value_Inferrable_containing`: this label is used for discourse entities that are inferred from themselves.

Containing Inferrables are considered as the most complex type of information. They can be difficult to distinguish from Brand–new anchored, Unused and non–containing Inferrable information (see Chapter 2, Section 2.2).

Note 1: Containing Inferrables are further defined to meet the following conditions in order to distinguish them from other types of information.

First, they are considered as contextually dependent entities in the present thesis, which is different from previous approaches (cf. Riester and Baumann 2017). Second, they should be represented by complex nominals, with at least two entities inside, and the entities within the nominal should form a certain semantic relation. Third, the triggers of containing Inferrables can only convey Evoked information, either situationally or textually. Nominal expressions with triggers of non–containing Inferrable and Unused are labeled as `non–containing_Inferrable` in the present annotation.

Note 2: There is invariably a degree of uncertainty in annotating Inferrables.

As indicated before (see Section 2.2.2), there is no clear boundary in inferencing. Containing Inferrables are typical for different interpretations on account of the diversity of intended text recipients in written texts.

- (iii) `infovalue_Evoked`: given entities by context and text.

- (a) `value_Evoked_Situationally`: entities that are evoked from the text–external context. The following are typical cases in written genres:

- (a1) Expressions refer to participants in a text, i.e. first and second personal pronouns;

- (a2) Expressions refer to the time of utterance, such as *the present*;
- (a3) Expressions refer to the setting and location of an utterance, which is more frequent in topographic reports and verbal maps of some territories (also see Halliday and Matthiessen 2014: 100; Riestler and Baumann 2017: 5), such as *this crossroads town* in (3–20e);
- (a4) Expressions refer to the intervals relative to the time of utterance, such as *last month, 100 years ago*.

(b) value_Evoked_Textually: entities that are evoked from the previous text. Although some cases have been reviewed in Section 2.2.2, more examples are listed as follows:

Full repetition without change in Head and modifier, see example (3–23):

- (3–23) The debate of *homosexuality* has been one of the most long lasting and controversial ones ever. What, exactly, causes *homosexuality*? (Ey)

Partial repetition with reduction in modifiers or abbreviations, which is illustrated in example (3–24):

- (3–24) a. She disagreed with the popular belief at the time that homosexuality was *a mental illness*... Fortunately, many prominent psychiatrists also believed that homosexuality was not *an illness*. (Ey)
- b. OTC derivatives contributed to the crisis in three significant ways. First, one type of derivative — *credit default swaps (CDS)* — fueled the mortgage securitization pipeline. (Gd)

Repetition with addition. See example (3–25):

- (3–25) *The Wicklow Mountains*... are visible from everywhere... *the beautiful Wicklow Mountains* and the Wicklow Mountains National Park provide a

more rugged countryside... (Tg2)

Pronominal repetition, which is displayed by example (3–26):

- (3–26) And while *multinational chains* have made inroads, **they** seem as less blatant here than elsewhere. (Tg2)

General word/abstract anaphors referring to entities, events, facts, propositions etc., whose antecedents are not restricted to nominal expressions. See example (3–27):

- (3–27) a. But it was not until 1869 that the term “*homosexual*” was first used...
At this point in time, Karl Maria Kertbeny used **the word** in a pamphlet which fought to repeal the current anti-homosexual laws of Prussia.
(Ey)
- b. The greatest tragedy would be to accept the refrain that *no one could have seen this coming and thus nothing could have been done*. If we accept **this notion**, it will happen again. (Gd)
- c. Exciting, mysterious, glamorous — **these words** have described Hong Kong for at least a century. (Tg1)

Epithets with postponed newsworthiness. See example (3–28):

- (3–28) On my way home, a dog barked at me. *The fierce German Shepherd* seemed to be quite aggressive. (Baumann and Riester 2012: 130)

Semantic relations, for example synonymy, hyperonymy/hyponymy, meronymy / holonymy and paraphrases as the most frequent ones (cf. Kunz 2010). See example (3–29):

(3–29) For instance, *heterosexual males* have an average of five times as many children as homosexual ones... Because consequently *heterosexual men* contribute five times as much genetic information to the next the gene pool. (Ey)

(b1) value_ Evoked_ Textually_ distance: the label is used to investigate and distinguish different textual distances between the new and given entities within a coreferential tie (given–new set) and it has four measures listed as follows:

(b1.1) value_ distance_ same: the coreferential antecedent of an expression occurs in the same sentence;

(b1.2) value_ distance_ adjacent: the coreferential antecedent of an expression occurs earlier in the previous sentence;

(b1.3) value_ distance_ further: the coreferential antecedent of an expression occurs earlier in the previous two or three sentences;

(b1.4) value_ distance_ textual: the coreferential antecedent of an expression occurs earlier in the previous four or more than four sentences.

The present thesis assumes a referent to be valid through the whole text. Once it is mentioned again, it will be classified as Given information regardless of the textual distance to its antecedent in the text (also see Yule 1981; Kunz 2010; Baumann and Riester 2012). Note that calculating the textual distance between the Evoked referents and their antecedents is arbitrary to some extent, as more findings are required from psycholinguistic studies that cannot be conducted in this study.¹¹

¹¹ It is a matter of debate whether the information status of a given referent is preserved throughout a text and whether the textual distance between the referent and its antecedent can be regarded as a crucial factor. Chafe (1970: 40) indicates that the Given status remains activated, by giving the example that an indefinite expression “*a letter*” that occurred on page 13 is still activated when it is mentioned again on page 118 in the form of “*the note*”. According to previous accounts (Chafe 1987; Arnold 1998: 22; Lambrecht 1994: 90), the deactivation of a given referent is mainly caused by three factors: (i) a linear decay mechanism of cognitive activation, (ii) interference from competing

Attribute_givennew_set

Nominals referring to the same discourse entity within a coreferential chain are linked as a set. Each set contains at least two entities. If the antecedent of a Given referent is embedded within an *of*-NP construction that contains more than one discourse entity, the *givennew_set* is drawn to the referent and the complete *of*-NP construction.

In addition, some given expressions may have more than one antecedent, the *givennew_set* consists of all the involved expressions, such as the link between *the Conservatives and the Social Democrats* and *they* in example (3–18) *The Conservatives and the Social Democrats have found an agreement. They decided not to raise taxes.* Besides, the *givennew_set* label only applies to the Given referents with nominals as their antecedents. No *givennew_set* is drawn between Given referents and their antecedents that are not expressed by nominals.

Other annotation conventions for the informational aspect

(i) Quoted/direct speech

Nominals within the quoted/direct speech are regarded as an inseparable part of the complete text and are analyzed in the same way with the non-quoted contents in this study.

- (3–30) Once upon a time there was a dear little girl who was loved by everyone who looked at her, but most of all by *her grandmother* (...) One day her mother said to her: “Come, Little Red Riding Hood, here is a piece of cake and a bottle of wine; take them to *your grandmother*, (...)” (Riester and Baumann 2017: 22)

The annotation of quoted speech in this study is different from Riester and

antecedents and (iii) the heterogeneity of a referent. Clark and Sengul (1979) examined the first factor and their experiment proves that Given referents show a significant difference between the one-clause textual distance and the two-or-three clause distance.

Baumann's (2017) approach since their approach is only concerned with the quoted speech in fictional dialogues, such as example (3–30). It is reasonable to classify “*your grandmother*” as Unused information rather than Evoked, since it is presented by “*her mother*” as something taken for granted. The example is selected from the beginning of a text, in which there are not many fictional characters and dialogues. However, in other cases, such as the well-known novel *Pride and Prejudice* written by Jane Austin, the expressions of one quoted speech are connected with the narration of the complete story. If we still adhere to Riester and Baumann's (2017) approach, we would need to ask whether they should be annotated separately without considering the connections.

Besides, referents introduced by the quoted speech can be mentioned again in the text. Since information values of these referents usually remain activated, it would be unjustified to analyze each quoted speech as an isolated unit without considering the connections with the preceding text. For example, “*the cartel*” in (3–31) is obviously presented as given information by partially repeating “*the Medellin drug cartel*”. If the quoted speech is separately annotated, then a coreferential tie cannot be established.

- (3–31) Robert Merkel, a former U.S. attorney handling drug indictments in Florida, doesn't think for a minute that Castro's much publicized trials of high officials engaged in the drug trade mean he has broken off with *the Medellin drug cartel*. “If *the cartel* succeeds in blackmailing the Colombian authorities into negotiations, the cartel will be in control and Fidel can exploit his past relationships with them,” he told the Journal's David Asman recently. (Np1)

(ii) Annotation of headings

Headings of each text were annotated in the current thesis.

3.3 The annotation tool

The corpus annotation was accomplished by means of MMAX2 (Müller and Strube 2006). MMAX2 helps us manually annotate various categories and their subcategories with further defined attributes. It can also mark discourse relations by linking different elements in text. MMAX2 contains elements that specify attributes and relations, their names, their types and a sequence of embedded value tags, which has been described in detail in Section 3.2.

For each text, an independent annotation project was established. Within each project, nominal expressions are defined as Markables. The markables were then assigned to the labels indicating varied linguistic features in the annotation scheme. Those features can be seen via the MMAX2 Attribute Window, with labels on the left and markables of each text on the right. In addition, MMAX2 can draw links between markables based on different semantic relations they establish in text, which can be visualized through the Markable Set Browser. The above is displayed by Figure 3–1:

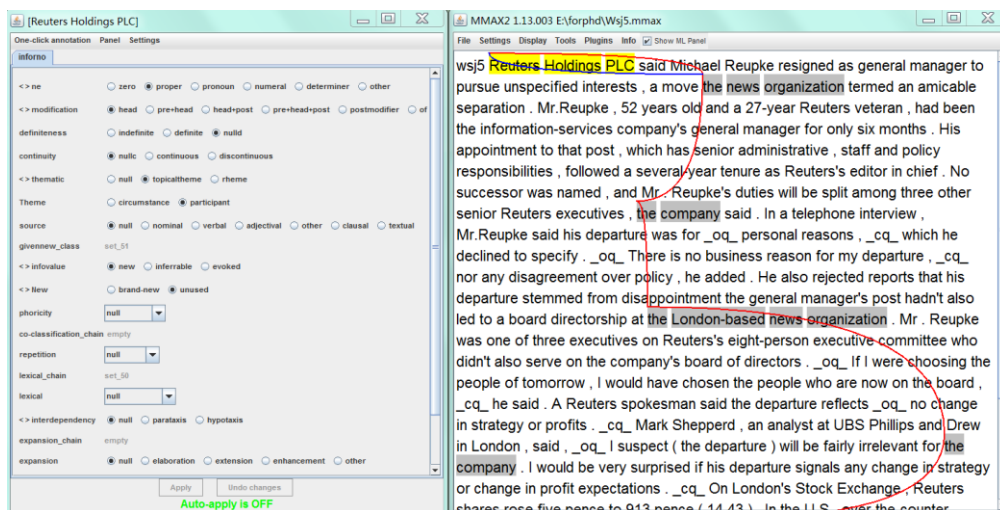


Figure 3–1: The MMAX2 Attribute Window and Text Window

The results of annotating the assigned labels of each project were collected through MMAXQL. MMAXQL is a Multi–Level Query Language of MMAX2 implemented to provide queries on annotated markables. In particular, MMAXQL can be used for two purposes: (i) detect and browse the coded elements with specific features that can be

combined with more than one assigned labels from different attributes. For example, Figure 3–2 illustrates that nominals with determiners convey evoked information and there is one clause between them and their antecedents in the text; and (ii) quantify the proportions of the coded elements. Figure 3–3 shows a statistical query of the markables illustrated in Figure 3–2.

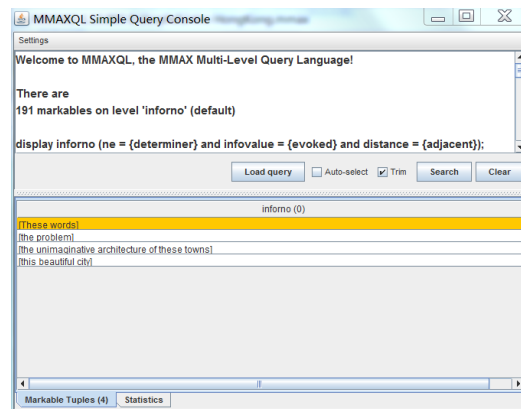


Figure 3–2: The MMAXQL Query Console of Markable Tuples

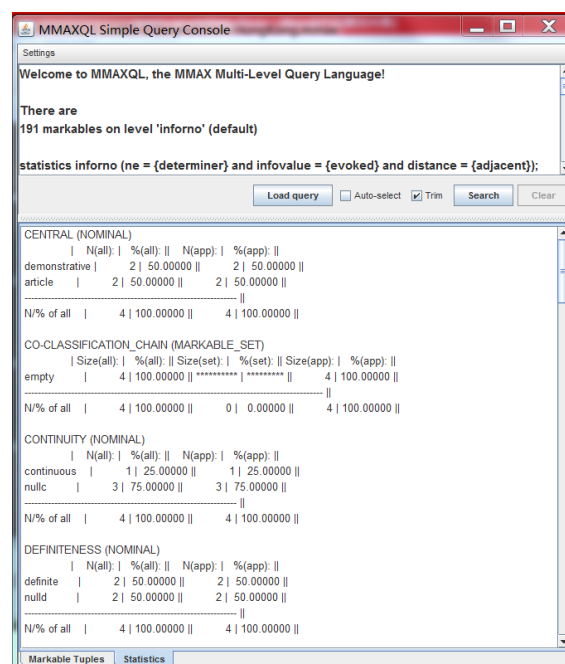


Figure 3–3: MMAXQL Query Console of Statistics

Although MMAX2 has not been updated for years, it is an appropriate annotation tool for the present study. It allows for a fine–grained analysis of the informational role of nominal expressions in English texts. The annotation enables us to explore the

information distributions of nominal expressions in different text types, to examine the multi-faceted features of varied nominal expressions in conveying types of information and to investigate the informational properties of nominal expressions in structuring texts by detecting the variation of the semantic and co-referential ties.

3.4 Summary

This chapter has provided a detailed description of the multi-dimensional methodology undertaken in the present study. It has elaborated the selection of the corpus, the annotation attributes and values both in nominal and informational aspects with particular reference to the adapted model of Prince's (1981) information taxonomy and the removal of ambiguities between certain linguistic categories and the use of the annotation tool MMAX2. Results of the corpus analysis will be presented in the following chapters to shed light on the informational role of nominal expressions in English texts.

4 Information distributions of nominal expressions in texts

As stated in Chapter 1, the main goal of the present study is to explore the informational role of nominal expressions in English texts. Chapter 2 and 3 have already provided theoretical foundations and the multi-dimensional approach to this study. The current chapter presents the empirical results of the analysis of information distributions of nominal expressions in English texts.

As stated in Chapter 2, previous literature so far mostly focused on proposing frameworks for classifying information status and exploring relationships of linguistic features of nominal expressions and information status. Information distributions of nominal expressions in complete texts have been a relatively understudied field. This chapter aims to provide a detailed account of this field. As previously noted in Chapter 3, nominals were manually annotated with varied information status based on the classifying model adapted from Prince (1981). The aim of this chapter is mainly achieved by looking at the frequency distributions of nominal expressions representing types of information status in English texts.

This chapter is organized as follows. Section 4.1 presents information distributions of nominal expressions in individual categories. Section 4.2 describes the distributions in all categories. Both sections will discuss the similarities and differences between the texts in the informational role of nominals. Finally, Section 4.3 offers a summary of this chapter.

4.1 Information distributions of nominal expressions of individual categories

Information distributions of nominal expressions in individual categories can be helpful in describing contextual and textual characteristics of a text. This section presents the analysis results based on the frequencies of each complete text.¹² Before analyzing the

¹² To increase readability, the frequencies presented in tables and figures were rounded to two decimal places, except for divisibility or zero distribution.

results of individual categories, Table 4–1 illustrates the raw frequencies and percentages of all information categories in the texts, which provides an overview of information distributions of nominal expressions in the texts. The abbreviations of Brand-new unanchored, Brand-new anchored, non-containing Inferrable, containing Inferrable, Unused, Situationally evoked and Textually evoked are: BN, BN.A, I, I.C, U, E.S, and E.T respectively.

Infor Text	BN		BN.A		I		I.C		U		E.S		E.T		SUM	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tg1	59	30.89	11	5.76	53	27.75	3	1.57	19	9.95	1	0.52	45	23.56	191	100
Tg2	69	25.75	7	2.61	62	23.13	6	2.24	40	14.93	2	0.75	82	30.6	268	100
Np1	49	26.2	12	6.42	34	18.18	1	0.53	21	11.23	4	2.14	66	35.29	187	100
Np2	14	31.82	1	2.27	8	18.18	1	2.27	5	11.36	1	2.27	14	31.82	44	100
Np3	18	24	3	4	14	18.67	0	0	8	10.67	0	0	32	42.67	75	100
Np4	18	34.62	2	3.84	4	7.69	3	5.77	8	15.38	0	0	17	32.69	52	100
Np5	36	39.13	3	3.26	8	8.7	1	1.09	7	7.61	0	0	37	40.22	92	100
Np6	28	43.75	1	1.56	7	10.94	1	1.56	4	6.25	0	0	23	35.94	64	100
Gd	271	17.6	26	1.69	400	25.97	80	5.19	57	3.7	4	0.26	702	45.58	1540	100
Ey	113	19.42	8	1.37	128	21.99	17	2.92	25	4.3	2	0.34	289	49.66	582	100

Table 4–1: Information distributions of nominal expressions in the texts

4.1.1 Brand–new unanchored

Brand–new unanchored (BN) information is closely related to the degree of newsworthiness of a text. It indicates the amount of new information that cannot be retrieved from our shared knowledge.¹³ “*A telephone interview*” in example (4–1) is a case in point:

(4–1) In *a telephone interview*, Mr. Reupke said his departure was for “personal reasons,” which he declined to specify. (Np5)

Text	Gd	Ey	Np3	Tg2	Np1	Tg1	Np2	Np4	Np5	Np6
Freq. (%)	17.6	19.42	24	25.75	26.2	30.89	31.82	34.62	39.13	43.75

Table 4–2: The distributions of Brand–new unanchored information per text

Table 4–2 illustrates the distributions of Brand–new unanchored information per text ranked in ascending order.¹⁴ As a whole, Brand–new unanchored information is important in structuring the texts. The highest frequency (Np6 43.75%) is nearly equal to half of all the information conveyed by a text. This reflects the main characteristic of written texts: focusing on conveying new information (Biber and Conrad 2014: 109) so that the addressee’s knowledge in a certain field can be increased (cf. Lambrecht 1994: 120). Second, the texts vary considerably in the distribution, with the maximal margin of 26.15% between Government document (17.6%) and Np6 (43.75%). The variation reflects different tendencies of the texts. Texts of News report and Travel guide contain more Brand–new unanchored information and thus indicate higher degrees of newness than Essay and Government document. It precisely demonstrates the difference in communicative purpose. News report and Travel guide texts aim to introduce and

¹³ In Prince (1981: 230), the term “Shared Knowledge” is interpreted as a kind of givenness that the speaker assumes the hearer knows, assumes or can infer a particular thing (but is not necessarily thinking about it). The thesis explicitly uses it or other terms like “background”, “shared background knowledge” and “background knowledge” in a general sense, which simply means the information that is assumed to know, either specialized or common–sense.

¹⁴ The abbreviations of the texts were presented in Section 3.1.2 of Chapter 3.

describe unknown things and they are expected to be more informational, while the Essay and Government texts in the dataset focus on explaining certain phenomena or stating opinions about an issue, which are expected to be persuasive. Taking the average values of the texts of News report and Travel guide (33.25% and 28.32% respectively) into consideration, the News report text has a more prominent position in conveying newsworthiness. This further highlights its major characteristic by “focusing more on current newsworthy events” (Biber and Conrad 2014: 118). Finally, Essay (19.42%) and Government document (17.6%) do not demonstrate a big gap in the frequency of nominals expressing Brand-new unanchored information, which is only 1.82%. Taking a look at the contents, the Essay text discusses the cause of homosexuality, and the Government document text endeavors to explain the causes of the U.S. financial and economic crisis in 2008. The margin then comes as no surprise, since both texts do not aim to describe newsworthy events, but to find out the truth about some controversial topics. In this sense, Essay and Government document are similar to some extent.

Texts of the same genre also show marked differences in conveying Brand-new unanchored information. For Travel guide, the difference value between Tg1 and Tg2 is 5.14%, which is even wider than that between Government document and Essay. Taking a closer look at the content, Tg1 is about Hong Kong and Tg2 is an introduction of Dublin. Both describe popular aspects of the well-known cities, such as people, accommodation, economy and sightseeing both in the cities and suburbs. However, besides the economic and cultural domains, Tg1 adds a political aspect by introducing the policy of Hong Kong’s peaceful handover to China and the changes during and after the transition. As “topic is the most important situational factor influencing vocabulary choice” (Biber and Conrad 2014: 46), the political addition in Tg1 directly provides an opportunity to use nominal expressions in that specific topic domain. In this case, more new information is introduced to the text. It is not surprising that Tg1 contains a larger amount of Brand-new unanchored information than Tg2.

For News report, though all the texts (Np1–6) focus on describing newsworthy events, they vary considerably. The maximum differential value is 19.75% (24% of Np3 and 43.75% of Np6), which even exceeds the complete distributions of Essay and

Government document. When looking at the specific purposes of the texts, it can be found that both Np5 and Np6 are concerned with resignations of employees, Np2 and Np4 focus on introducing the new measures taken by governments to boost economy, and Np1 and Np3 mainly describe the prominent social problems that attract the attention of international communities.

Texts as Np1 and Np3 contain more background information compared with the other two sets, as they have reported and discussed the same issues for many times and put an emphasis on offering evaluations of the current situation or suggestions for the future. On the one hand, the difference in specific topics provides an explanation for the variation of the information distribution, but on the other hand, it indicates certain similarity between the texts within the same specific topics. In Table 4–2, three pairs of texts, Np5 and Np6, Np2 and Np4, Np1 and Np3, contain similar frequencies of nominals expressing Brand–new unanchored information.

4.1.2 Brand–new anchored

Brand–new anchored (BN.A) information is of secondary importance in evaluating the newness of a text after Brand–new unanchored information. As a subtype of Brand–new information, it refers to new information that has both semantic and grammatical links to Evoked information in text (see Chapter 3, Section 3.2.2). In (4–2), “*whose brother*” with “*a man*” as the anchor is a case in point:

(4–2) This comes from a man *whose brother*, Guillermo, was murdered in 1986.
(Np1)

Text	Ey	Np6	Gd	Np2	Tg2	Np5	Np4	Np3	Tg1	Np1
Freq. (%)	1.37	1.56	1.69	2.27	2.61	3.26	3.84	4	5.76	6.42

Table 4–3: The distributions of Brand–new anchored information per text

Table 4–3 shows the distributions of Brand–new anchored information in all the

texts in ascending order. In general, Brand–new anchored information plays a much less important role than the unanchored in structuring text, with the highest frequency lower than 7%. That can be partially explained by its categorical criterion. As indicated in Chapter 3, anchored items are both semantically and grammatically linked to expressions in the preceding text. In the linguistic sense, the link is usually realized by the premodifiers of nominal expressions, with particular reference to personal determiners. The personal determiners are inherently anaphoric and refer to discourse entities that occur before, which function as anchors for nominals expressing anchored information. Nominal expressions of other types might not have potential for being categorized into the information category. Nominals with participial modifiers, for example *flashing lights*, *growing problem* and *exhausting task* (Quirk *et al.* 1985: 588), do not contain anchors. Some concrete expressions as head nouns, such as *rain*, cannot be modified by anaphoric determiners. In this case, Brand–new anchored items have fairly restricted linguistic representations both in anchors and head nouns, which to some extent explains their relatively low frequencies in the texts. Further discussion will be provided in Chapter 5 when presenting the relationship between linguistic forms of nominal expressions and types of information status they express in the texts.

In addition to the linguistic restrictions, the low frequency might also be influenced by another factor: relevance. Brand–new unanchored expressions are a violation of relevance. According to the principles of cognitive and communicative relevance (Sperber and Wilson 1986/1995; Wilson 2010; Levinson 1989; Bach 1999; cf. Huang 2012: 27), human cognition tends to be geared to the maximization of relevance, and every utterance indicates a presumption of its own optimal relevance. When one anaphoric nominal is introduced to the text, readers tend to relate it with discourse entities and propositions occurring in the preceding text and try to establish semantic relations between them. Furthermore, the writer, an expert in predicting readers' activation states (Sanders and Canestrelli 2012: 203), is supposed to introduce related items by following the given–new contract in order to convey information successfully (e.g. Clark and Haviland 1977). Given the anaphoric implications and the principles of cognitive and communicative relevance, it would be possible that readers tend to

consider the nominals as Inferrable information with semantic relations with some discourse entities and propositions in the preceding text and their background knowledge. However, a few nominals, though with anaphoric links, still do not form semantic relations with Given information. New experiential identities will be created after readers fail to establish the relevance. This might be an explanation for the lower frequency of nominals conveying Brand–new anchored information.

Returning back to the data analysis, another finding from Table 4–3 is that there are differences across texts of four genres. The News report and Travel guide texts generally contain more Brand–new anchored information compared with the Essay and Government document texts. This can be attributed to the communicative purposes of the texts. As indicated before, Brand–new anchored information, though less important than the unanchored, is a secondary way of increasing newsworthiness to text. The data in Table 4–2 and Table 4–3 suggests that News report and Travel guide contain more frequencies of nominals conveying both anchored and unanchored information to display higher degree of newsworthiness.

For the individual texts under the same genre, the situation is complex. In Travel guide, Tg1 (5.76%) contains more anchored information than Tg2 (2.61%), with a gap of 3.15%. The tendency is the same as that in the unanchored category. Therefore, compared with Tg2, Tg1 has a higher degree of newsworthiness. The individual texts of News report do not display regular patterns. The gaps between some of the texts are even larger than texts of different genres. For example, the difference between Np1 and Np3 is 2.42%, while the gap between Essay and Government document is only 0.32%. There are three possible explanations for the above. The first is the frequencies of Brand–new anchored information do not display many differences between the texts within the same genre. Second, the frequencies of all the texts are very low, which makes it hard to find general patterns. The third explanation is that not many individual texts were analyzed under the genres of Travel guide and News report, the small scale of corpus thus cannot provide reliable conclusions for the differences.

4.1.3 Non-containing Inferrable

Non-containing Inferrable (I) items are a subtype of Inferrables. They are considered to be intermediate between New and Evoked entities (also see Chafe 1994; Baumann and Riester 2012). As the most complex case in text (Prince 1981: 236), it is categorized by a combination of shared background knowledge, semantic relations, logical reasoning and contextual and textual information. Table 4–4 illustrates the frequencies of non-containing Inferrables in the texts in ascending order.

Text	Np4	Np5	Np6	Np2	Np1	Np3	Ey	Tg2	Gd	Tg1
Freq. (%)	7.69	8.7	10.94	18.18	18.18	18.67	21.99	23.13	25.97	27.75

Table 4–4: The distributions of non-containing Inferrables per text

First of all, News report texts have fewer non-containing Inferrables compared with texts of the other three genres. The difference can be explained by two points from the perspective of situational characteristics of text: (i) communicative purpose and (ii) shared background knowledge. First, as indicated before, the general purpose of News report is to bring sufficient newness to readers, it thus has less potential for introducing Inferrables that require expressions in the preceding text as triggers. Second, although News report bears some similarity with Travel guide in describing newsworthy events, Travel guide has a rather standardized way of describing travelling. Travel guide texts are more familiar to readers as part of their shared experience. Some situation-bound utterances (see Kecskes 2000, 2002, 2010, 2012) or idealized entities (Lakoff 1987; cf. Kecskes 2012), for example visitors, accommodation, transportation and the like, are activated under the TRAVEL schema, which are expected to occur in texts.¹⁵ Taking a closer look at the Travel guide texts, the activation of conceptual knowledge is confirmed by many instantiated non-containing Inferrables, such as *the impression of the visitor, food, and hotels* in Tg1 and *the visitor, traffic and Dublin’s food* in Tg2.

¹⁵ Schemata theory mainly concerns the stereotypic mental representations of world knowledge and information retrieval (see Section 2.2). The term “schema” as the central concept in the theory represents a high level of knowledge that guides readers to understand a situation (Kecskes 2012: 179).

Essay and Government document also have corresponding systems of the conceptual knowledge that underlie non-containing Inferrables guided by context and text. Taking Government document as example, many of the non-containing Inferrables are categorized through a FINANCIAL schema, which is the manifestation of financial knowledge, such as *investors*, *a free market economic system* and *risky subprime lending and securitization*. Unlike Travel guide, the knowledge is only possessed by a small group of people, i.e. specialist background knowledge (see Biber and Conrad 2014: 43).

It is also important to note that both general and specialist knowledge occur in texts. Besides the financial items, entities categorized with the help of general knowledge like *its causes*, *the losses* and *the next step* form an integral part of the Government document text. Some non-containing Inferrables can carry both characteristics and demonstrate various semantic relations, for example, the antonymy formed by *borrowers* and *lenders* and the meronymy established by *the credit markets*, *the stock market*, *the subprime mortgage markets*, *the financial markets* and *markets*. Basically, non-containing Inferrables can be divided into three types in terms of the shared background knowledge: general, specialist and a combination of both.

Besides the simple semantic relations formed only between entities, non-containing Inferrables are also categorized through other relations that can bind more than one entity or proposition in text. This point can be illustrated by example (4–3):

- (4–3) On 1 July, 1997 the British Crown Colony of Hong Kong reverted to Chinese sovereignty as a Special Administrative Region of the People’s Republic of China. Today Hong Kong remains a capitalist enclave with its laws and rights intact, and China has promised that Hong Kong will continue in this fashion for at least 50 years... Around ***the time of the transition*** there was much speculation about how ***things*** would change.
(Tg1)

The item “*the time of the transition*” in (4–3) is a non-containing Inferrable that is

generated from the previous sentence. The relation between entity and proposition is established in a very natural way as both the verb “*continue*” and the prepositional phrase “*at least 50 years*” in the example indicate a changing process that needs to take a period of time. Similarly, the second non-containing Inferrable is understood by linking it to the previous proposition, in which the remained capitalist enclave and the continuation of the fashion actually are part of the “*things*”. In terms of cohesion (Halliday and Hasan 1976: 274), “*things*” belongs to the class of general noun, which is a lexical item with general reference.

(4–4) Our task was first to determine what happened and how it happened so that we could understand why it happened. Here we present our conclusions. We encourage the American people to join us in making their own assessments based on the evidence gathered in our inquiry. If we do not learn from history, we are unlikely to fully recover from it. Some on Wall Street and in Washington with a stake in the status quo may be tempted to wipe from memory the events of this crisis, or to suggest that no one could have foreseen or prevented them. This report endeavors to expose *the facts*, identify *responsibility*, unravel *myths*, and help us understand how the crisis could have been avoided. (Gd)

With the development of the text, non-containing Inferrables can be generated by complex relations as more and more information are accumulated as background knowledge. The items in (4–4) represent Inferrables that are categorized through a string of propositions. When processing the first sentence, it becomes clear that three points are going to be elaborated in their task: WHAT, HOW and WHY and they have a direct bearing on the items *the facts*, *responsibility* and *myths*. From a semantic perspective, *the facts* concludes *what happened*, *responsibility* is related to *how it happened* and *myths* explains *why it happened*. The second and third sentences then provide information about the comments of their task. Given that, however, it is not appropriate to simply equate the three points with the three items, since their references

are no longer the same. “*The fact*” implicates there are more findings based on their evidence besides “*what happened*”. “*Responsibility*” means someone should be blamed for both “*what happened*” and “*how it happened*”. “*Myths*” includes things unknown to the public in addition to the cause of the happening. They all contain more newsworthiness than the first sentence. The abstract relations established for comprehending the three items involve several propositions. Such non-containing Inferrables have an impact on the discourse coherence.

Studies on the role of coherence in text comprehension have been mainly developed from the perspective of psycholinguistics (e.g. Kintsch 1970 1978 1998; Kintsch and van Dijk 1978; van Dijk and Kintsch 1983; Keenan *et al.* 1984; Sanders *et al.* 1992; Albrecht and O’Brien 1993; Myers, *et al.* 1994; Graesser *et al.* 1994; Sanders 1997; Graesser *et al.* 2001; Das and Taboada 2018; Becker *et al.* 2020; Saux *et al.* 2021; Wannagat *et al.* 2021). Many of them have discussed the relation between inference and coherence, taxonomies of coherence and inference, and how they facilitate information processing between sentences. More importantly, they provide insight into how readers comprehend different text types. Functional studies, for instance SFL, mainly focus on the role of writer (Hasan 1999). However, “utterances are socially mediated to anticipate readers’ possible objections and engage them in appropriate ways” (Hyland 2001). Taking both writer and reader roles into account, the text analysis can show a deeper understanding of language production and comprehension. Non-containing Inferrables can be regarded as a parameter to explore this aspect, with particular reference to the complex ones involving several propositions in text.

Besides, the complexity of non-containing Inferrables is also worthwhile to mention. As one complex Inferrable is derived from more than one proposition, it requires more mental effort to build the inference. Although Table 4–4 shows Tg1 contains more Inferrable frequencies than Government document, it does not mean that it is more difficult for readers to comprehend. This can be supported by the instances in (4–4) selected from Government document. The complexity of non-containing Inferrables to some extent can be used to explain why we feel the Essay and Government document texts are more difficult to comprehend than those of News report

and Travel guide. Apart from type of shared background knowledge, the complexity of Inferrables also has an impact on text comprehension. Due to the limited space, it is impossible to further categorize all Inferrables in terms of complexity. More discussion will be provided in Chapter 5 when analyzing the linguistic forms of non-containing Inferrables in the texts of four genres.

Moving to the analysis of the individual texts of the same genre, News report shows significant differences, with a clear dividing line between Np1, Np2, Np3 and Np4, Np5, Np6. Each of the first three texts contains nearly 20% non-containing Inferrables, while for the latter, it is less than 10%. Another interesting finding is that Np1 has exactly the same distributions with Np2 (18.18%). When looking at the specific purposes, Np1 and Np2 are clearly different. Their same distributions of non-containing Inferrables might be a coincidence of calculation. For those having the same specific purpose, only one set demonstrates some similarity, which contains the highest Inferrable frequencies in News report texts: Np1 and Np3, with a slight variation of 0.49%. For the other two sets, Np2 and Np4, Np5 and Np6, the differences are 10.49% and 2.24% respectively. For Travel guide, Tg1 and Tg2 has a gap of 4.62%, which is smaller than those of News report.

The differences among the texts within the same genre are even bigger compared with some texts of different genres. For example, the difference between Essay and Tg2 is only 1.14%. This is not closely related to the specific purpose or shared knowledge of a text. Taking a closer look at the specific non-containing Inferrables in the texts of News report and Travel guide, most of them were categorized by simple semantic relations established between discourse entities rather than propositions. Less distinctive properties of non-containing Inferrables were found to explain the differences. One possible explanation for the differences among texts of the same genre might be the linguistic forms of nominal expressions of text. If a text contains more *of*-NP constructions that form part-whole relations, it is more likely to have a higher frequency of non-containing Inferrables. This will be further discussed in Chapter 5.

4.1.4 Containing Inferrable

Containing Inferrables are the other subtype of Inferrables. Different from the non-containing items, they are inferred from themselves as “containing the trigger within the description” (Prince 1992: 13). The categorical criteria in the present annotation are different from Prince’s (1981, 1992). She allows various types of triggers and treats the containing Inferrables of written texts as “multi–receiver discourse... where the sender either is not sure of the receivers’ knowledge/beliefs, or where s/he believes that there are relevant differences among the receivers” (Prince 1992: 308). As indicated in Chapter 3, the triggers of containing Inferrables in the current thesis should only be nominals expressing Evoked information, either situationally or textually, through which prototypical relations can be established between the triggers and the Inferrables. Table 4–5 shows the distributions of containing Inferrables of the texts in ascending order.

Text	Np3	Np1	Np5	Np6	Tg1	Tg2	Np2	Ey	Gd	Np4
Freq. (%)	0	0.53	1.09	1.56	1.57	2.24	2.27	2.92	5.19	5.77

Table 4–5: The distributions of containing Inferrables per text

First, it is a surprise that not all the texts have containing Inferrables. Second, the frequencies do not feature prominently in text, with the highest lower than 6%. Like Brand–new anchored, the frequencies can also be attributed to the restriction of linguistic forms of nominal expressions. Based on the categorical criteria, a prototypical relation with triggers can only be established when there are at least two discourse entities within a nominal expression. In this case, simple nouns, for instance pronouns and bare nouns, have no potential for expressing containing Inferrables.

The containing Inferrables share the same characteristic with non–containing ones. As text–motivated entities, they are inferred from establishing meaning relations and thus contribute coherence and cohesion to text construction. The relations are all simple, which are only between entities expressed by nominal expressions. A typical example is given in (4–5):

(4–5) According to Darwin’s theory of natural selection, the advantageous traits are passed on, while *the disadvantageous ones* eventually die out. For instance, heterosexual males have an average of five times as many children as homosexual ones, as a female is required to reproduce children, yet homosexuals are, by definition, not sexually attracted to females. Therefore, from the evolutionary standpoint, homosexuality becomes *one of the disadvantageous traits*. (Ey)

The containing Inferrable “*one of the disadvantageous traits*” in example (4–5) contains a part–whole relationship established between “*one*” and the trigger “*the disadvantageous traits*”. The trigger is a Textually evoked expression, since it repeats “*the disadvantageous ones*” that occurs in the previous sentence. Besides, the example also indicates another feature of containing Inferrables: the triggers can serve as a reminder of Evoked entities, since they are repeated within the containing Inferrables. In this way, the represented information is reactivated, which is easy for readers to process.

Table 4–5 also displays that both Tg1 (1.57%) and Tg2 (2.24%) have lower frequencies compared with Essay (2.92%) and Government document (5.19%). Texts of News report do not generate any clear patterns. Four of them, Np3, Np1, Np5 and Np6, contain very few containing Inferrables, and the frequencies are all lower than 1.6%. In addition, it is hard to see general patterns from texts of the same genre. In Travel guide, the difference between Tg1 and Tg2 is 0.67%, which is wider than texts of different genres. For example the difference between Tg1 and Np6 is only 0.01%. Taking a closer look at the specific items of the two texts, there are three items in Tg1: *the flag of China*, *many of them* and *the unimaginative architecture of these towns*; and six in Tg2: *the roar of the Celtic Tiger*, *the center of the city*, *the north and south sections of the city*, *parts of the city’s magic*, *the garden oases of the city* and *the ancient sites of Ireland*. These containing Inferrables simply indicate part–whole and property–thing relations, which are not closely to communicative purposes or mental frames of varied

text types. In News report, the examination is the same with that of Travel guide. The only one item in Np1, *public enemy No.1 of press freedom*, and three items in Np4, *many types of watches*, *other types of watches* and *the main beneficiaries of the president's action*, do not indicate significant difference in types of both semantic relations and containing Inferrables.

The distributions of containing Inferrables of the texts do not reveal clear differences closely related to situational characteristics, but rather show similarity in semantic relations established within the Inferrables themselves.

4.1.5 Unused

Unused information is something that can be taken for granted by writers and readers, though with the first occurrence in text. Based on the definition, another prominent feature of the Unused item is that their unique experiential identities contribute to the ease of readers' comprehension. For example, *the sun*, *China* and *Winston Churchill* as Unused items in any text would be understood by readers without any difficulty. Table 4–6 shows the distributions of Unused information in the texts in ascending order.

Text	Gd	Ey	Np6	Np5	Tg1	Np3	Np1	Np2	Tg2	Np4
Freq. (%)	3.7	4.3	6.25	7.61	9.95	10.67	11.23	11.36	14.93	15.38

Table 4–6: The distributions of Unused information per text

First, all the texts have Unused information and the differences in frequencies are clear, with a range of 11.68% between Np4 and Government document. The second main finding here is that, compared with the Travel guide and News report texts, the Government document and Essay texts contain fewer Unused items, each of which is nearly 4%. To explain the above variations, it is necessary to examine specific expressions. Example (4–6) illustrates all the Unused expressions of Essay and Np4:

(4–6) a. *ancient Greece, his book The Symposium, Plato, Greek mythology,*

Hercules, Zeus, Homer, Achilles, Patroclus, Karl Maria Kertbeny, Germany, Berlin, the Nazis, the Urning theory of Karl Ulrichs, King William I, the German Kingdoms, the time of their unification, Sigmund Freud, The Rosa Parks of the gay movement, the American Psychiatric Association Board of Trustees, the Diagnostic and Statistical Manual, the American Psychological Association, Northwestern University, Boston University School of Medicine, Darwin's theory of natural selection (Ey)

- b. *The White House, President Bush, the U.S., the Virgin Islands, the U.S. Generalized System of Preferences, the Philippines, Thailand, U.S. Trade Representative Carla Hills* (Np4)

The first finding from example (4–6) is the difference in the types of shared background knowledge. In Essay, many of the expressions are associated with ancient Greece, Germany and psychiatry. While in Np4, they mainly concern the common aspects of the U.S. and two developing countries. Essay involves more items that are activated from the specialist knowledge of psychiatry.

As Prince (1981: 235) points out, “the hearer may be assumed to have a corresponding entity in his/her own model”. The assumption in fact is based on the speaker’s evaluation of the hearer’s background knowledge. In the Essay text, terms like *the American Psychiatric Association Board of Trustees, the Diagnostic and Statistical Manual, the American Psychological Association* obviously are familiar to a specific group of audience that possesses knowledge in psychology. Compared with the Unused expressions of Np4, the writer of Essay clearly assumes that readers know more about homosexuality as background knowledge.

The second main finding is the similarity Essay and Np4 in the Unused discourse entities in example (4–6). Some of the Unused items form semantic relations, for example the holonymy established by *Greek mythology, Hercules, Zeus, Homer, Achilles* and *Patroclus* in Essay, and in Np4, the two co-hyponyms, *the Philippines* and *Thailand*. In addition, it is important to note that the Unused items in both texts are

motivated by specific contexts. Although their unique experiential identities are taken for granted by readers, different contexts contribute to the activation of particular aspects of Unused information in a text. Just like salt in a recipe, the name *Sigmund Freud* is taken for granted in an article that is related to psychiatry. If *Sigmund Freud* occurs in a menu representing a kind of sandwich, its information value is not Unused to the customers who only know the cause of homosexuality.

Returning to compare the Unused frequencies in the texts, the lower frequencies in the Government and Essay texts demonstrate that both tend to involve less shared background knowledge compared with the News report and Travel guide texts, regardless of either the general or specialist knowledge. On average, texts of Travel guide and News report contain almost the same Unused frequencies, with a slight variation of 2.02%. Based on the similarity, it is necessary to take a look at the specific nominals conveying Unused information of the Travel guide texts. Those of Tg1 are presented in example (4–7):

(4–7) *Hong Kong, English, Beijing's announced policy of maintaining Hong Kong's prosperity and stability, the West, Queen Elizabeth, the Union Jack, the New Territories, Chinese, Cantonese, South China, Macau, Victoria Peak, Hong Kong Island, Victoria Harbor, the MTR rail line, the Kowloon peninsula, Lantau, Lamma, Cheung Chau* (Tg1)

Instances in example (4–7) shows that Tg1 is also similar to Np4 in the type of shared background knowledge and semantic relations indicated by nominals conveying Unused information.

For the individual texts of the same genre, News report shows clear differences in the Unused frequencies, with the biggest gap of 9.13% between Np4 and Np6. For Np5 and Np6, they contain the lowest frequency of nominals conveying Unused information, with a small gap of 1.36%. The difference between Np1 and Np3 is only 0.56%. Np2 and Np4 have a gap of 4.02%, which is even larger than the Unused frequencies of Government document. It should also be noted that the differences

between Np1, Np2 and Np3 are not clear, especially between the first two texts with the slightest variation of 0.13%. For Travel guide, Tg2 contains more frequencies than Tg1, with a difference of 4.98%. Taking a look at the Unused expressions of Tg1 and Tg2, they do not indicate any prominent features that can distinguish them from other texts within the same genre. Although Unused items can demonstrate the difference in types of shared background knowledge between different genres, they may not be an appropriate indicator for the individual texts with different specific purposes within the same genre. The above findings can also be partially attributed to the data analyzed for the current thesis. As not many individual texts were annotated, it is difficult to predict general tendencies or explain differences between the texts within the same genre.

4.1.6 Situationally evoked

Situationally evoked information, as one subtype of the Evoked category, refers to the extratextual entities in the physical context of the communication, i.e. the setting. It basically contains the places and time that are shared by both writers and readers (also see Biber and Conrad 2014: 44). Table 4–7 shows the distributions of Situationally evoked information of the texts in ascending order.

Text	Np6	Np5	Np4	Np3	Gd	Ey	Tg1	Tg2	Np1	Np2
Freq. (%)	0	0	0	0	0.26	0.34	0.52	0.75	2.14	2.27

Table 4–7: The distributions of Situationally evoked information per text

The first main finding here is that just like the containing Inferrables, not every text has items of the category. Four out of the six texts in News report contain zero Situationally evoked information. Second, the frequencies of the other texts are very low, with the maximum of 2.27% by Np2. Third, comparing the four different genres, Travel guide has more Situationally evoked items than Government and Essay. The average of Situationally evoked frequencies of the News report texts is higher than texts of the other genres.

The above aspects reveal a prominent feature of written texts that distinguishes them from the spoken ones: writers and readers do not share the same setting. To explore the characteristics of Situationally evoked information in the written texts, it is important to take a look at the specific items. All Situationally evoked items of the texts are presented in (4–8)¹⁶:

- (4–8) *our country, the status quo, the world, you, this report* (Gd)
you, the world (Ey)
you (Tg1)
us, the world (Tg2)
we, this week, the hemisphere, this week (Np1)
the world (Np2)

The above expressions reflect three main characteristics of the written texts in describing the communicative situations. First, most of them refer to places and readers. The linguistic expressions and the experiential identities of the Situational evoked entities are exactly the same, either *the world you*, or *we/us*. Np1 is the only text that involves the reference of time by “*this week*”. Second, the Situationally evoked items are not an indicator of types of written texts. Their frequencies are more influenced by the number of nominal expressions of a text rather than the different communicative purposes. For example, both Essay and Tg2 contain two Situationally evoked expressions, their frequencies are different. That is mainly caused by the difference in text length, since the Essay text has a larger number of nominals than Tg2. In addition,

¹⁶ “*The world*” and “*the country*” in example (4–8) refer to the physical context of the communication, which are the places like *this room*, *this house* and *this school* shared in many spoken texts. Some may argue that “*the world*” and “*the country*” convey Unused information in terms of the uniqueness of their experiential identities. However, they are different from typical Unused entities like *the sun* and *the moon*. Their references are not unique. For example, science fictions contain imaginary worlds and countries in the year of 2083 that are assumed to be shared by the writers and readers. “*The world*” and “*the country*” could be used to refer to places of some particular settings in a specific time period, while the referents of *the sun* and *the moon* remain the same in any text types.

the Situationally evoked entities here are all concerned with general aspects of setting, without referring to specific information of readers or location. It manifests a low degree of interactiveness between readers and writers in written texts (cf. Biber and Conrad 2014: 42). Writers have a rather distant relationship with readers, since they are not in the same setting.

Based on the above, Situationally evoked items can be regarded as an observable indicator for distinguishing written and spoken texts. However, they play a less important role in distinguishing written texts with various communicative purposes. The linguistic expressions have demonstrated a rather fixed pattern, with particular reference to the second person pronouns. More discussions about the linguistic forms of nominals and information status they express will be provided in Chapter 5.

4.1.7 Textually evoked

Textually evoked items, as the other subtype of the Evoked category, denote information that can be referred back in the preceding text. The categorization is based on the sameness of experiential identity. Compared with the other categories, it is the only one that directly deals with coreferential chains (cf. Kunz 2010: 72; Kunz and Lapshinova–Koltunski 2018).

Text	Tg1	Tg2	Np2	Np4	Np1	Np6	Np5	Np3	Gd	Ey
Freq. (%)	23.56	30.6	31.82	32.69	35.29	35.94	40.22	42.67	45.58	49.66

Table 4–8: The distributions of Textually evoked information per text

Table 4–8 shows the distributions of Textually evoked information of all the texts in ascending order. On the whole, Textually evoked information plays a central role in structuring text, with the minimum above 20% and the maximum close to 50%. Second, there is a marked difference between the texts. The biggest gap is 26.1% between Tg1 and Essay. Furthermore, Government document and Essay contain higher frequencies than Travel guide and News report, which can be explained by the difference in

communicative purposes. As indicated before, Travel guide and News report focus on conveying newness by describing newsworthy things rather than developing opinions and providing explanations for certain issues. It is reasonable that both genres contain less Given information than Government document and Essay.

One prominent feature of Textually evoked items is the direct connection with co-referential chains. Within the chain, the experiential identities can be repeated in varied ways (also see Halliday and Hasan 1976; cf. Kunz 2010). Example (4–9) is a case in point:

(4–9) The debate over *homosexuality* has been one of the most long–lasting and controversial ones ever. What, exactly, causes *homosexuality*...? Therefore, neither homosexuals nor heterosexuals really have a “choice” in *the matter*... In an interview in 1903, he professed *his beliefs*: “I am... of the firm conviction that homosexuals must not be treated as sick people...” ... In 1935, he furthered *his claims* when he wrote a now famous “Letter to an American Mother” of a homosexual... (Ey)

The first “*homosexuality*” in (4–9) is fully repeated in the next sentence, and it also forms hyponymy/hyperonymy with the general noun “*the matter*”. This phenomenon is defined as *chain intersection* by Kunz and Lapshinova–Koltunski (2018). Similarly, “*his beliefs*” and “*his claims*” also indicates chain intersection by being both co-referential and synonymous.

In addition, previous studies on Textually evoked items especially those within coreferential chains have displayed differences in the focus of a text, text continuity and degree of accessibility between the types of registers across languages (for example Kunz 2010; Neumann 2014). However, those studies have not taken the total nominal expressions of a text into consideration. The current thesis also includes the chains in the annotation, namely the Attribute_givennew_set (see Section 3.2.2).

The annotation mainly concerns two aspects. The first is the relation between the number of chains (henceforth **N-chain**) and the total number of nominal expressions

(henceforth **N-total**) of a text. The second aspect is the dominance of the most frequent expression per text. It is defined as **Dominance** in this study, indicating the frequency of the most dominant referent (N-domi) to the number of all nominals in the text (N-total), i.e. Freq.-domi / N-total. Table 4–9 displays the above aspects of the texts.

Text	Tg1	Tg2	Np1	Np2	Np3	Np4	Np5	Np6	Ey	Gd
N-chain	20	32	29	4	12	11	12	5	72	177
N-total	191	268	187	44	75	52	92	64	582	1540
Freq.-domi	18	30	9	6	12	5	10	12	40	80
Dominance (%)	9.42	11.19	4.81	13.64	16.00	9.62	10.87	18.75	6.87	5.19

Table 4–9: Key factors of co-referential chains per text

First, the number of co-referential chains does not suggest an absolute association with the number of nominal expressions. For example, Np3 has the same numbers of sets with Np5, but Np3 contains more nominal expressions in total; Np1 contains more co-referential chains but fewer nominals than Tg1. There is no definite positive correlation between the total number of nominals and the number of chains. Second, the value of Dominance reveals different degrees of focalization of the texts. In general, Travel guide and News report have a higher degree of Dominance than Government document and Essay.

In addition, most previous research focuses on the longest co-referential chains of a text (e.g. Neumann 2014; Kunz and Lapshinova-Koltunski 2018). However, less insight has been developed in exploring the other ones, with particular reference to the shortest chains. The shortest co-referential chain only contain two nominals, within which one referent is only repeated once in the text. Although it has no direct relation to the degree of Dominance of a text, it might influence it to some extent. When one text has a larger proportion of shortest co-referential chains, will the degree of Dominance also be higher? A reliable answer for that needs a further examination of specific Textually evoked items. Figure 4–1 displays all the given–new chains and the numbers of expressions involved in the chains in Np2 and Tg1. For example, the figure illustrates that in Np2, the given–new chain “DUBAI” contains four expressions having the same experiential identity.

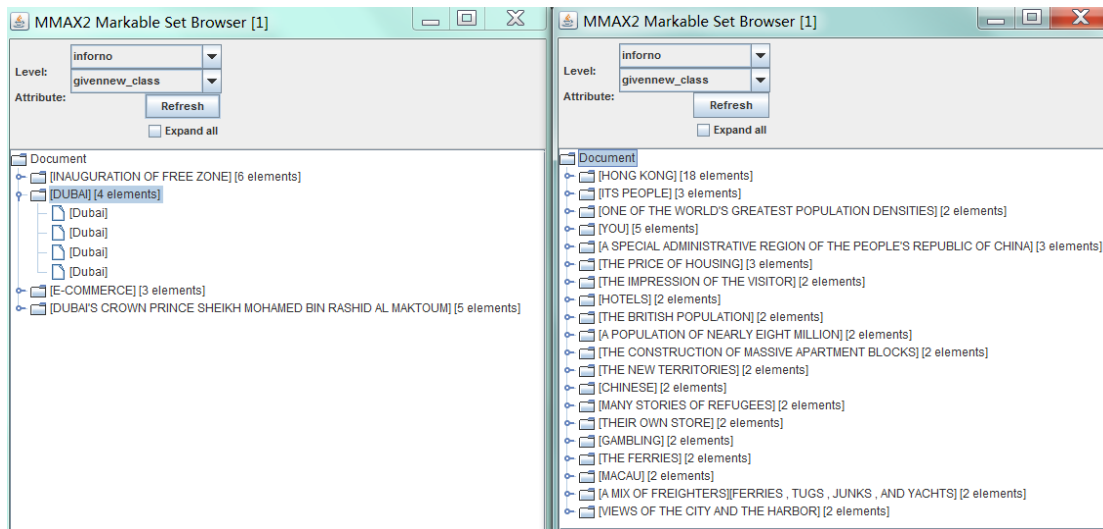


Figure 4–1: The given–new chains in Np2 and Tg1

As Figure 4–1 indicates, Np2 has no co–referential chains that only involve two textual elements. The most dominant item *inauguration of free zone*, is repeated six times. The other three items *E–commerce*, *Dubai* and *Dubai’s Crown Prince Sheikh Mohamed Bin Rashid Al Maktoum*, occurred three, four and five times respectively in Np2. Although they are less frequent, there is not a big difference in the degrees of Dominance with *Inauguration of free zone*. In Tg1, *Hong Kong* as the most dominant item has been mentioned 18 times. The other items only occurred two or three times. Fifteen out of twenty chains are the shortest in total. With a large proportion of the shortest chains, the most frequent item even becomes more dominant in a text, since it receives less competition from the other Textually evoked items. In this case, *Hong Kong* is more focalized than *inauguration of free zone* and Tg1 has a higher degree of dominance than Np2.

Then returning back to the numbers presented in Table 4–9, Tg1 has a Dominance value of 9.42%, while Np2 has a larger value of 13.64%. If the analysis is only based on the value of Dominance, it is natural to summarize that Np2 has a more dominant focus than Tg1. However, this summary may be not convincing enough considering the impacts from other coreferential chains. To have a more accurate observation, it is necessary to add another factor as reference: the frequency of the shortest coreferential

chains (henceforth **N-short-chain**) to all the chains of a text, which is defined as the minimum degree of dominance (henceforth **D-min**) in this study. Table 4–10 illustrates all the D-min values of the texts:

Text	Tg1	Tg2	Np1	Np2	Np3	Np4	Np5	Np6	Ey	Gd
N-short-chain	15	23	15	0	8	9	3	2	36	82
N-chain	20	32	29	4	12	11	12	5	72	177
D-min (%)	75	71.88	51.72	0	66.67	81.82	25	40	50	46.33

Table 4–10: Ratios of the shortest coreferential chains to all the chains per text

Both Table 4–9 and Table 4–10 suggest that there is no obvious association between the Dominance and D-min values and the number of nominal expressions of a text. Second, as the focus of a text tends to be more dominant if it has high values on both aspects, Travel guide is more focalized than Government document and Essay and Essay is more focalized than Government document. For the individual texts of Travel guide, Tg1 and Tg2 are close to each other in general. Although Table 4–9 shows that the Dominance value of Tg1 (9.42%) is a little bit lower than that of Tg2 (11.19%), the D-min value of Tg1 (75%) is higher than that of Tg2 (71.88%) in Table 4–10. Taking both Dominance and D-min values into account, Tg1 and Tg2 do not differ from each other significantly in the focalization.

The News report texts do not indicate a general pattern of topic continuity and focalization. Although most of the texts have higher values of Dominance, the D-min values vary considerably. Taking a closer look at the specific texts, the value of Dominance of Np2 is higher than most of the other texts, but the minimum degree of Dominance (D-min) value is zero. Similarly, Np1, Np4, Np5 and Np6 display the same tendency, with high values of one aspect but rather low values of another. Np3 is the only exception, which shows similar tendencies of Dominance and D-min, with both values ranking the second highest place within the News report texts.

Combining the findings of coreferential chains with Textually evoked distributions in the four genres, there are not many clear associations between the two. The data of Table 4–8, Table 4–9 and Table 4–10 only suggest that the values of Dominance in

Travel guide, Government document and Essay tend to be negatively related to the Textually evoked distribution. News report texts do not indicate any obvious tendencies with considerable differences in Dominance and D–min values.

As indicated in Chapter 3, Section 3.2.2, not all items within coreferential chains are categorized as Textually evoked information. The antecedents of Textually evoked items can convey other types of information (also see Kunz 2010). Therefore, the ratio of the coreferential elements to the number of nominal expressions is higher than the Textually evoked information of a text. More importantly, coreferential chains do not include all antecedents of Textually evoked items. Some antecedents are not expressed by nominal expressions, they can be clauses, adjectives or verbs. Table 4–11 illustrates the ratios of the Textually evoked items with non–nominal antecedents to all evoked nominals, which is defined as **R–nonN** in this thesis.

Text	Tg1	Tg2	Np1	Np2	Np3	Np4	Np5	Np6	Ey	Gd
R-nonN (%)	6.67	7.32	4.55	0	3.13	5.88	5.41	4.35	12.46	11.26

Table 4–11: Ratios of non–nominal antecedents to Textually evoked items per text

Table 4–11 suggests clear differences in the non–nominal antecedents of Textually evoked items between the four genres. First, News report texts have the lowest R–nonN value, the average of which is only 3.89%. In contrast, the R–nonN value of Essay is the largest. It does not differ clearly from that of Government document, with a gap of 1.2%. Third, the average R–nonN value of Travel guide is 7%, which is higher than that of News report, but is lower in comparison to Essay and Government document. The individual texts, Tg1 and Tg2, do not vary from each other considerably, with a narrow gap of 0.65%. In addition, for News report, the R–nonN values of the individual texts show clear differences, with a range from 0 to 5.88%.

The above can be regarded as an indicator of different types of Textually evoked items in conceptualizing complex information across texts. The non–nominal antecedents represent ideas, properties, relations or statements that are rather abstract and complex with large information chunks. Textually evoked items with such antecedents allow writers to conceptualize and integrate the large piece of information

into entities, which is also one type of shell nouns defined by Schmid (2000). The integrating function makes it easier to mention those entities again, since they are a “cognitively more economical linguistic unit than a clause” (Schmid 2000: 370). However, comparing those with nominal antecedents, they are cognitively heavier by requiring more mental effort to integrate the abstractness and complexity. In theory, a text with a larger R–nonN value is assumed to be more abstract and requires more mental effort to comprehend.

In addition, the R–nonN values presented in Table 4–11 provide evidence for the general expectations of readers that texts of Government document and Essay are more difficult to understand than those of Travel guide and News report. Compared with News report and Travel guide, Essay and Government document are expected to contain more Textually evoked items with non–nominal antecedents, since the higher frequencies can demonstrate more textual abstractness and complexity. This result also partially confirms Schmid’s (2000: 379) finding that “shell nouns are more frequent in texts on abstract topics that are written in a neutral or formal style and serve an expository or argumentative function”.

4.2 Information distributions of nominal expressions of all categories

As stated above, the aim of the current chapter is to capture one of the features of the informational role of nominal expressions in English text, with particular reference to information distribution. Section 4.1 has explored the distributional properties of each information category. Section 4.2 will further explore the distributions of all information categories. Figure 4–2 displays the distributions of the texts:

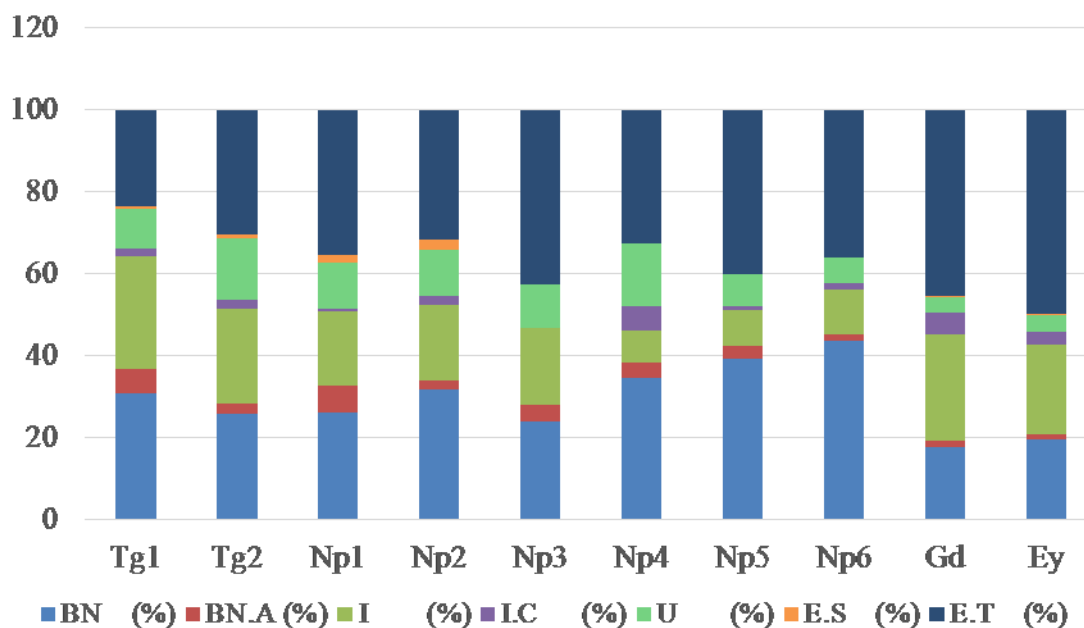


Figure 4-2: Information distributions of the ten texts

As the figure shows, information of the three categories, Brand–new unanchored, Textually evoked and non–containing Inferrable, contribute most to the texts, except for Np4 with more Unused information than non–containing Inferrables. Second, with a further comparison of the three types of information, Brand–new unanchored and Textually evoked items play a more important role in most of the texts, except for the Government document and Essay texts that have more non–containing Inferrables than Brand–new unanchored information. The result is consistent with Peng’s (2014) research, which explored the informational feature of a literary text by applying Prince’s (1981) taxonomy. Although there are some differences in classifying the sub–types of Inferrables, the criteria for the other categories are exactly the same. It can be concluded that the major distributions of information in written discourses is realized by Brand–new and Textually evoked items.

Third, for the other four categories, a preferred scale of information distributions of English text is $U > BN.A > I.C > E.S$. At first glance, more containing Inferrables occur in half of the texts. However, it is still less important than Brand–new anchored items. Np3 can function without containing Inferrables, which indicates that some texts can function well without such information. Similarly, Situationally evoked information

also does not occur in all the texts and it is regarded as the least prominent feature of written texts (also see Section 4.1.6; cf. Peng 2014). Based on the above, a preferred hierarchy or scale for what type of information expressed by nominal expressions is distributed in English written texts, tends to develop as follows:

The Scale of Information Distributions of nominals in English written Text

$$\begin{array}{l} \text{E.T} \\ \text{BN} \end{array} > \text{I} > \text{U} > \text{BN.A} > \text{I.C} > \text{E.S}$$

Note that the scale can change more or less according to the type of English written genres. Besides containing Inferrable and Situationally evoked, information of the other categories, especially those ranked at the bottom of the scale might not occur in some text. Second, Brand–new unanchored and Textually evoked information occur in all types of written texts, since each text needs Brand–new unanchored information to convey newsworthiness and the topic needs to be continued through Textually evoked items. Information of the two categories are the most prominent in information distributions of a text. Since the scale is proposed on the basis of information distributions of ten texts, it still needs to be further confirmed with a larger corpus.

In addition to the scale, it is also important to explore informational patterns of the complete texts and the relations among all the information categories in terms of the frequencies of nominal expressions. If some of the categories are related to others, their information distributions could be predictable to some extent. For example, if there is a negative correlation between Brand-new anchored and containing Inferrable, we only need to analyze the information distribution of one category. That of the other decreases. Since the exploration is more concerned with general tendencies, Travel guide and News report are analyzed by the average frequencies of the individual texts. The results are illustrated with the other two genres in Table 4–12:

Infor Genre	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg-avg.	28.32	4.19	25.44	1.91	12.44	0.64	27.08	100
Np-avg.	33.25	3.56	13.73	1.87	10.42	0.74	36.44	100
Gd	17.60	1.69	25.97	5.19	3.70	0.26	45.58	100
Ey	19.42	1.37	21.99	2.92	4.30	0.34	49.66	100

Table 4–12: Information distributions of the four genres

As the table suggests, in information distributions of nominal expressions, the Travel guide text here is more similar to the News report except for the frequencies of non-containing Inferrables, while the Essay and Government document texts resemble each other in all types of information values. Compared with the Essay and Government document texts, the Travel guide and News report texts tend to contain more information of Brand–new, Unused and Situationally evoked categories, but less of the other categories. This difference reveals two general patterns of information distributions across texts of the four genres. Travel guide and News report texts tend to develop the texts in a more straightforward way. They introduce newsworthiness by directly presenting new discourse entities, activating more entities that are already in the mental models of readers, and involving more things shared in the extratextual environment. However, the Government document and Essay texts are more likely to establish both simple and complex relations between discourse entities and propositions throughout texts, with more Textually evoked and non–containing Inferrable information as the empirical evidence.

The different patterns of information distributions of nominals in complete texts could be mainly explained by the situational characteristics of genres. As indicated before, Travel guide and News report texts aim to describe newsworthy items to general audience. It is not surprising to see that texts of both genres contain more Brand–new information and require less background information shared between writers and readers. In comparison, both Essay and Government document texts analyzed in this study are concerned with explaining the causes of some phenomena, which contain facts by simple and complex relations established between items and propositions

(Schmid 2000: 66). Furthermore, both are written for some particular audiences with specialist knowledge about a certain topic.

More surprisingly, Table 4–12 suggests that Travel guide and News report do not highlight newsworthiness by containing considerably more Brand–new unanchored frequencies than the Textually evoked. For example, in the News report texts, the frequencies of Brand–new unanchored information is even lower than those of Textually evoked information. The finding goes against the intuition that texts focusing on describing newsworthy events might have more Brand–new information than the other categories of information. The newsworthiness of the Travel guide and News report texts turns to be more noticeable in comparison to the Essay and Government document texts.

One possible explanation for the above might be the role of Textually evoked information in constructing text. Compared with nominals expressing the other categories of information, Textually evoked nominals play a dominant role in developing text. Written texts aim to communicate information (cf. Biber and Conrad 2014: 109) and their newsworthiness is carried by Brand–new information. With the development of text, discourse entities conveying Brand–new information at the beginning will turn to be known to readers by reoccurring in the text, i.e. Textually evoked. This could be exemplified by co-referential chains with nominals conveying Brand–new information as antecedents. Without Textually evoked information, it is hard to continue the topic of the texts and to provide background information for newly introduced items. Furthermore, Textually evoked information is also important for readers to comprehend texts, since it can function as a trigger for Brand–new anchored and Inferrable items, which is connected with several types of information in the texts.

For the relations between the frequencies of individual information categories, the Brand–new unanchored and Textually evoked items are evenly distributed in Travel guide and News report respectively. Both genres have a similar ratio of Brand–new unanchored to Textually evoked items, each of which is close to 0.39. Second, the frequencies of Brand–new anchored and Textually evoked items suggest a negative relationship. The two types of information tend to develop in the opposite direction

across the four genres. Given the low frequencies of Situationally evoked information, Textually evoked items are the main source of functioning anchors for Brand–new anchored items. However, the negative relationship seems to indicate that there are other factors that influence the Brand–new anchored distribution. In addition, Brand–new unanchored information displays a positive relation with the Situationally evoked across the four genres. If one genre contains more Brand–new unanchored information than the other, then it is more likely to have more nominals expressing Situationally information. For the other categories, no obvious relations have been found across the four genres. As the frequencies of some information categories are really low, it is difficult to provide general claims for the differences.

4.3 Summary

The goal of this chapter was to present and discuss the results of information distributions of nominal expressions in English texts. Unlike the previous literature, this chapter has provided a detailed description of information status of nominals in four comparable written genres based on the reinterpreted framework of Prince (1981), thereby addressing the first research question identified in Chapter 1. Specifically, Section 4.1 has analyzed and discussed the frequencies of each information category; and Section 4.2 has examined the general features of all categories in the texts.

Several important findings have been presented in the chapter. First, information distributions of nominal expressions reveal different characteristics of the texts. The frequencies of individual information categories in the four genres are different in the degree of newsworthiness by the Brand–new and Textually evoked information, varied types of shared background knowledge and semantic relations by the Inferrable and Unused information, and low level of interactiveness of written text by the Situationally evoked information. Brand–new unanchored, non–containing Inferrable and Textually evoked categories play an important role in structuring the texts than Situationally evoked, Brand–new anchored and containing Inferrable information.

The frequencies of the main information categories have indicated that texts have different patterns of distributing information. Both the Travel guide and News report texts focus on describing newsworthy discourse entities with more general shared knowledge. Their newsworthiness is more highlighted when comparing with Government document and Essay rather than the distributions of Textually evoked information within the texts themselves. The Travel guide texts are less newsworthy than the News report, since they have lower frequencies of nominals expressing Brand–new information but higher frequencies of non–containing Inferrable and Unused. Compared with Travel guide and News report, Government document and Essay contain more non–containing Inferrables that are categorized from specialist knowledge and complex relations established between discourse entities and propositions.

Second, three new terms, namely Dominance, D–min and R–nonN were proposed in the analysis of Textually evoked information. Dominance has explored the distributions of the longest given–new chains (or co–referential chains), D–min has indicated the frequencies of the shortest chains, and R–nonN has explored the influence of non–nominal antecedents of the texts. Unlike previous studies that only explored the longest chains, the other two terms have provided new insights into observing focalization and topic continuity of a text.

Third, nominals conveying the same information value are still different in terms of their antecedents in the texts. Some of them integrate more chunks of information and contain a higher degree of abstractness and complexity when their antecedents are represented by a string of words or propositions. Such instances are more likely to be seen in the Government document text.

Furthermore, this chapter has also examined the relations between the frequencies of the information categories. Not many regularities were found from the analysis results. Brand–new unanchored information is positively associated with the Situationally evoked, while Brand–new anchored information is negatively associated with the Textually evoked.

In conclusion, this chapter has presented information distributions of nominal expressions in English texts. The next chapter explores the relationship between

linguistic forms of nominals and information status they represent in the texts.

5 Linguistic forms of nominal expressions and types of information

status

As stated in Chapter 1, the main goal of the present study is to explore the informational role of nominal expressions in English texts. Chapter 4 has presented the information distributions of nominal expressions in English texts. The goal of this chapter is to discuss the relationship between linguistic forms of nominal expressions and types of information they represent in the texts.

As we saw in Chapter 2, previous literature has so far mostly focused on information status of particular nominals, linguistic forms of particular information status, or a one-to-one correlation between the two. The actual relation between types of linguistic forms and information status in varied English texts has remained under-investigated. With this in mind, this chapter provides a detailed account of this relationship.

Specifically, the aim of the chapter is achieved by exploring the linguistic forms of nominals in representing categories of information based on the reinterpreted framework of Prince (1981). As previously noted in Chapter 3, the forms of nominal expressions were manually annotated into five main types: bare noun, proper noun, pronoun, nominal expression with determiner and nominals that do not belong to the above categories. Among them, pronouns and nominal expression with determiners were further annotated (see Section 3.2.1).

This chapter is organized as follows. Section 5.1 presents the distributions of various types of nominal expressions of the corpus dataset. The following Section 5.2 discusses these results and explains what type of information status is represented by each type of nominal expression. Finally, Section 5.3 brings the chapter to an end with a summary of the relation between linguistic forms of nominal expressions and the represented information status in different texts.

5.1 Outline of various nominal expressions in the dataset

This section focuses on the distributions of various nominal expressions in the corpus dataset, with particular reference to the similarities and differences between texts in the frequencies of linguistic forms. Table 5–1 illustrates both the raw frequencies and proportions of all types of nominal expressions in the texts.

The statistics show some similarities between the texts of Travel guide. Except for the frequency of proper nouns, Tg1 and Tg2 are similar in the other types of nominals. In comparison, the patterns of the individual texts of News report are not consistent. In the distributions of bare nouns, two groups of texts indicate similarity. The first is Np1 (7.49%), and Np4 (5.77%), with a gap of 1.72%, and the second group is Np3 (9.33%), Np5 (10.87%) and Np6 (10.94%). In the distributions of proper nouns, similarities could be seen from two sets of texts, which are Np4 (32.69%) and Np6 (31.25%), and Np1 (27.81%) and Np2 (25%). In the frequency of pronouns, it is surprising to find that Np4 does not contain any, which is the only instance of the corpus dataset. The remaining texts then can be divided into two groups. The first group includes Np1 (11.23%), Np3 (14.67%) and Np5 (9.78%), in which the latter two both indicate similarity to Np1, with the respective differences of 3.44% and 1.45%. The second group is constituted by Np2 (6.82%) and Np6 (4.69%). In the distributions of nominal expressions with determiners, two groups of texts show clear resemblance. The first is composed of Np1 (44.92%), Np2 (40.92%) and Np5 (44.57%), and the second contains Np4 (32.69%) and Np6 (34.38%). In the distributions of the remaining types of nominal expressions that were labelled as *other-ne*, there are also two groups of texts demonstrating similarity. The first group has three texts, Np1 (8.56%) Np2 (11.36%) Np3 (13.33%), and the second contains two, Np5 (18.48%) and Np6 (18.75%).

Government document contains the highest frequency of nominals with determiners but the lowest of proper nouns across texts. Essay has the largest proportion of proper nouns but the least of all other types of nominal expressions. In general, the News report texts contains the lowest frequency of bare nouns, with an average value of 10.05%, which is in clear contrast with those of the Travel guide texts, with the

highest value of 17.74%. In contrary to the frequency of pronouns, nominals with determiners are frequent in texts of four genres.

As indicated above, pronouns and nominals with determiners were annotated in further detail. Table 5–2 and Table 5–3 display the distributions of subtypes respectively. The statistics in Table 5–2 indicate that there are clear differences in the frequency of pronouns. Except for Np4 not having pronouns, the other texts use personal pronouns frequently. Over 70% pronouns of each text are personal. In contrast, pronouns of other types are very rare. No texts contain reciprocal pronouns. Less than three texts contain possessive and reflexive pronouns and only four texts have demonstrative pronouns. The remaining pronouns that do not belong to the previous categories, for example *something* and *no one*, are present in six texts.

N-type Text	bare		proper		pron		det.		other-ne		Sum	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tg1	35	18.32	34	17.8	15	7.85	82	42.93	25	13.09	191	100
Tg2	46	17.16	71	26.49	21	7.84	102	38.06	28	10.45	268	100
Np1	14	7.49	52	27.81	21	11.23	84	44.92	16	8.56	187	100
Np2	7	15.91	11	25	3	6.82	18	40.91	5	11.36	44	100
Np3	7	9.33	28	37.33	11	14.67	19	25.33	10	13.33	75	100
Np4	3	5.77	17	32.69	0	0	17	32.69	15	28.85	52	100
Np5	10	10.87	15	16.3	9	9.78	41	44.57	17	18.48	92	100
Np6	7	10.94	20	31.25	3	4.69	23	35.94	11	17.19	64	100
Gd	193	12.53	132	8.57	194	12.6	699	45.39	322	20.91	1540	100
Ey	95	16.32	67	11.51	140	24.05	214	36.77	66	11.34	582	100

Table 5–1: Types of nominal expressions per text

N-type Text	possessive		personal		demonstrative		reflexive		reciprocal		other-pron.		Sum	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tg1	0	0	13	86.67	0	0	0	0	0	0	2	13.33	15	100
Tg2	1	4.76	15	71.43	1	4.76	1	4.76	0	0	3	14.29	21	100
Np1	0	0	17	80.95	3	14.29	0	0	0	0	1	4.76	21	100
Np2	0	0	3	100	0	0	0	0	0	0	0	0	3	100
Np3	0	0	11	100	0	0	0	0	0	0	0	0	11	100
Np4	0	0	0	0	0	0	0	0	0	0	0	0	0	100
Np5	0	0	8	88.89	0	0	0	0	0	0	1	11.11	9	100
Np6	0	0	3	100	0	0	0	0	0	0	0	0	3	100
Gd	2	1.03	160	82.47	15	7.73	7	3.61	0	0	10	5.15	194	100
Ey	0	0	102	72.86	11	7.86	4	2.86	0	0	23	16.43	140	100

Table 5–2: Types of pronouns per text

N-type Text	pre-de.		possessive-de.		demonstrative-de.		article		other-det.		Sum	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tg1	1	1.22	13	15.85	3	3.66	54	65.85	11	13.41	82	100
Tg2	0	0	4	3.92	6	5.88	87	85.29	5	4.9	102	100
Np1	0	0	12	14.29	3	3.57	56	66.67	13	15.48	84	100
Np2	0	0	2	11.11	0	0	15	83.33	1	5.56	18	100
Np3	0	0	2	10.53	0	0	16	84.21	1	5.26	19	100
Np4	0	0	0	0	1	5.88	8	47.06	8	47.06	17	100
Np5	0	0	5	12.2	1	2.44	23	56.1	12	29.27	41	100
Np6	0	0	3	13.04	0	0	17	73.91	3	13.04	23	100
Gd	3	0.43	84	12.02	66	9.44	455	65.09	91	13.02	699	100
Ey	1	0.47	28	13.08	11	5.14	131	61.21	43	20.09	214	100

Table 5-3: Types of nominal expressions with determiners per text

Table 5–3 illustrates the distributions of nominal expressions with types of determiners. First it can be seen that those with articles play the most important role in all the texts, with a minimum frequency of 47.06% (Np4) and maximum of 85.29% (Tg2). Second, the remaining types of determiners that were labelled as *other-det.* in Table 5–3, for instance *other developing nations*, *many shops* and *three Nobel Prize winners*, are of secondary importance, with the lowest frequency of 4.9% (Tg2) and the highest of 47.06% (Np4). Third, Np4 does not have any possessive determiners, and Np2, Np3 and Np6 do not contain demonstrative determiners. Except for Tg2 and Np4, the remaining texts have more frequencies of possessive determiners than the demonstrative ones. In comparison, nominals with predeterminers, such as *such a crowded place*, indicate a very low frequency and they only occur in three texts. In addition, the individual texts of Travel guide and News report do not show regular patterns within the same genre, though with some similarities in type of determiners.

The description of types of nominal expressions reveals the potential for expressing information status. For example, nominals with determiners express more types of information status, since their frequencies are higher than other types of nominals in all the texts. In contrast, the absence of certain types of nominal expression in some texts as compared to others is also revealing. Since reciprocal pronouns do not occur in any texts, they will not be discussed in this chapter. In addition, the distributions of linguistic forms can also reflect different situational characteristics of texts and genres. For example, the frequency of proper nouns in Table 5–1 indicates that compared with Government document, the Travel guide texts contain more nominals with unique experiential identities. It is not surprising, as the Travel guide texts aim to introduce places of interest to readers, in which case geographical names are more likely to occur. More characteristics of texts can be revealed under types of information status.

Having presented the distributions of various nominal expressions of the corpus dataset, the next section will explore the correspondence between the linguistic forms of nominal expressions and types of information status they convey in the texts.

5.2 Linguistic forms of nominals and information status they convey in the texts

This section investigates the linguistic forms of nominals in expressing types of information status in terms of the adapted taxonomy of Prince (1981). Based on the scale of information distributions in English written texts (see Chapter 4, Section 4.2), different information values do not play the same role in structuring text. Brand–new unanchored, non–containing Inferrable and Textually evoked expressions occur most frequently in all texts. Therefore, those expressions will be considered with a more detailed analysis by further exploring the subtypes of nominals with determiners and pronouns.

5.2.1 Linguistic forms of nominals conveying Brand–new unanchored information

This section mainly analyzes and discusses the linguistic forms of nominal expressions in expressing Brand–new unanchored status. Table 5–4 displays the distributions of five major forms by percentages.¹⁷ The raw frequencies are presented as Table A1 in the Appendix of this thesis.

The statistics in Table 5–4 show interesting similarities and differences. As would be expected, pronouns have the least potential for expressing Brand–new unanchored information in each text. Although some pronouns were used to express Brand–new information in the Travel guide texts and Essay, the frequency is very low, merely with the highest of 3.54%. There is a well–established tendency for pronouns to function as referential coherence (Halliday and Hasan 1976) and to signal a high degree of accessibility of the referent (Ariel 2001: 9). However, indefinite pronouns are an exception. They do not identify a referent specifically and their indefinite status can

¹⁷ The blue shaded areas in the tables of the present chapter are used to denote that certain type of nominals were not found in the complete text. For example, Table 5–4 indicates that Np4 does not contain any pronouns expressing types of information (also see Table 5–1). Texts that are not marked with the blue color, namely Np1, Np2, Np3, Np5 and Np6, do not contain pronouns expressing certain Brand–new unanchored information.

convey new information in text. Example (5–1) is a case in point:

(5–1) There are *many stories of refugees* who arrived with ***nothing*** in their pockets, set up a small sidewalk stall, worked diligently until they had their own store, and then expanded it into a modest chain. (Tg1)

“*Nothing*” in (5–1) is considered newsworthy, as “*many stories of refugees*” in the preceding text are presented as unknown information and readers do not have any clue to knowing details of the stories.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	22.03	1.69	1.69	54.24	20.34	100
Tg2	18.84	13.04	1.45	52.17	14.49	100
Np1	20.41	22.45	0	53.06	4.08	100
Np2	7.14	14.29	0	42.86	35.71	100
Np3	22.22	27.78	0	11.11	38.89	100
Np4	16.67	11.11	0	27.78	44.44	100
Np5	22.22	11.11	0	38.89	27.78	100
Np6	21.43	14.29	0	35.71	28.57	100
Gd	17.71	6.64	0	47.6	28.04	100
Ey	13.27	18.58	3.54	48.67	15.93	100

Table 5–4: Types of nominals expressing Brand–new unanchored information per text

The uses of the remaining linguistic forms for expressing Brand–new unanchored information are more frequent in text. First, it is not surprising to see that bare nouns play an important role in expressing Brand–new unanchored information, since they do not contain modifiers that indicate phoric relations to context and text. This characteristic corresponds to the newsworthiness of Brand–new unanchored entities. It is also noticeable that variations among some texts is surprising, with the biggest gap of 15.08% existing between Np2 (7.14%) and Np3/Np5 (22.22%). Example (5–2) illustrates the only instance of Np2 showing that the bare noun “*e-commerce*” occurs

in the heading rather than the body of the text.

(5–2) Inauguration of free zone in Dubai for *e-commerce* (Np2)

In contrast, Np3 and Np5 have exactly the same frequencies of bare nouns, which are similar to those of Np6 (21.43%) and Np1 (21.41%). These frequencies suggest that the News report texts seem not completely arbitrary in selecting bare nouns referring to Brand–new unanchored entities. Similarity was also found from the Travel guide texts. The difference between Tg1 (22.03%) and Tg2 (18.84%) in the frequencies of bare nouns expressing Brand–new unanchored information is only 3.19%. As for the other two genres, Government document (17.71%) indicates similar frequencies with Tg2 (18.84%) and Np4 (16.67%), with a gap of 1.13% and 1.04% respectively; Essay contains the second lowest frequency of Brand–new unanchored bare nouns, which is 6.13% higher than Np2.

Regardless of individual differences within the Travel guide and News report texts, their average frequencies of bare nouns are 20.44% and 18.35% respectively. Neither shows a striking difference from the frequency of Government document. In contrast, the frequencies of Essay are very low, especially when compared with those of the Travel guide texts, with a gap of 7.17%. One possible explanation is that Essay requires more precise descriptions and prefers other types of linguistic forms to express Brand–new unanchored information. That will be discussed with further evidence from the frequencies of proper nouns and determiners in the Essay text.

As indicated in Table 5–4, it is also common for proper nouns to express Brand–new unanchored entities. Proper nouns are used to denote referents with unique identities. According to Biber *et al.* (1999: 245), proper nouns can be typically divided into the following categories in terms of referring content: personal names, geographical names, objects and commercial products, holidays, months, days of the week, religions and relational concepts, persons with unique public functions, public buildings, institutions, laws etc., political parties and languages and nationalities.

Proper nouns from all categories can be used to convey Brand–new unanchored information.

- (5–3) To supplement my findings from research, I have conducted personal interviews with *two adolescent homosexual males*, **James Dobbens** and **Daniel Woods**. (Ey)

In example (5–3), “*James Dobbens*” and “*Daniel Woods*” are the “*two adolescent homosexual males*”, which have not occurred in the preceding text and were not shared by readers as background knowledge. They are newsworthy to readers.

Some texts have shown clear differences in the distributions of proper nouns expressing Brand-new unanchored information, with the biggest gap of 26.09% found between Tg1 (1.69%) and Np3 (27.78%). Of all the texts, Government document (6.64%) contains the second lowest frequency of proper nouns, while Essay (18.58%) has the third largest proportion. Interestingly, under the genre of News report, Np2 and Np6 (14.29%), Np4 and Np5 (11.11%), were found to have the same frequencies. However, the difference between some texts is easy to notice. For example, Np3 and Np4/Np5 have the biggest gap of 16.67%.

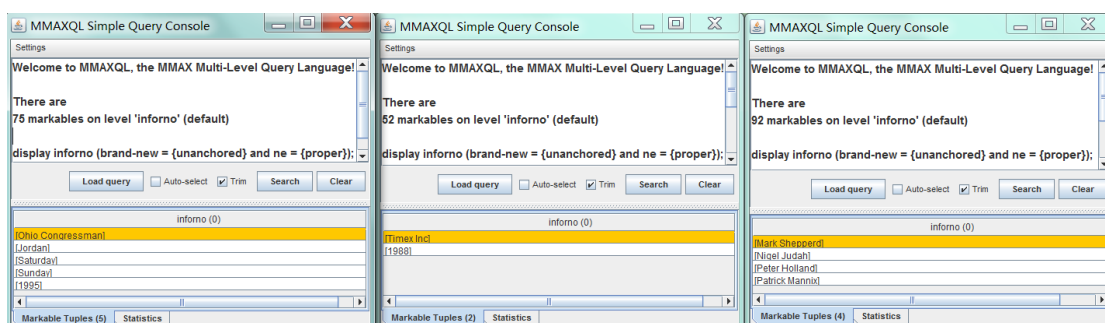


Figure 5–1: Proper nouns of Np3 Np4 and Np5 expressing Brand–new unanchored information

Figure 5–1 displays all the instances of proper nouns representing Brand–new unanchored information in Np3, Np4 and Np5 respectively via the MMAX2 Query

Console. As the figure suggests, proper nouns in Np4 and Np5 do not indicate many different semantic features. The Markable Tuples of the MMAX2 Query Console show that Np4 contains a name of company and a specific year, i.e. *Timex Inc* and *1988*, while Np5 only has personal names, namely *Mark Shepperd*, *Nigel Judah*, *Perter Holland*, and *Patrick Mannix*. In comparison, there are three categories of proper nouns in Np3, namely political party, geographical name and specific day or year. When there are more categories involved in a text, the distributions of proper nouns with Brand–new unanchored status seem to increase.

As with the News report texts, the individual texts of Travel guide also differ from each other in the distributions of proper nouns that express Brand–new unanchored information, with a clear gap of 11.35%. Figure 5–2 displays all the instances of Tg1 and Tg2 respectively. Taking into account the characteristics of the Travel guide texts, it is not surprising to see that the two texts contain three types of proper nouns, which indicate historical date, geographical names and public buildings. Tg1 has only one proper noun, which provides temporal information. However, in Tg2, the proper nouns involve two types of references: geographical names and public buildings. Besides, Tg1 and Tg2 are different in selecting categories of proper nouns. One possible explanation could be that compared with Tg1, Tg2 includes more sub–topics, which is indicated by categories of proper nouns. The observation is the same with that of the News report texts.

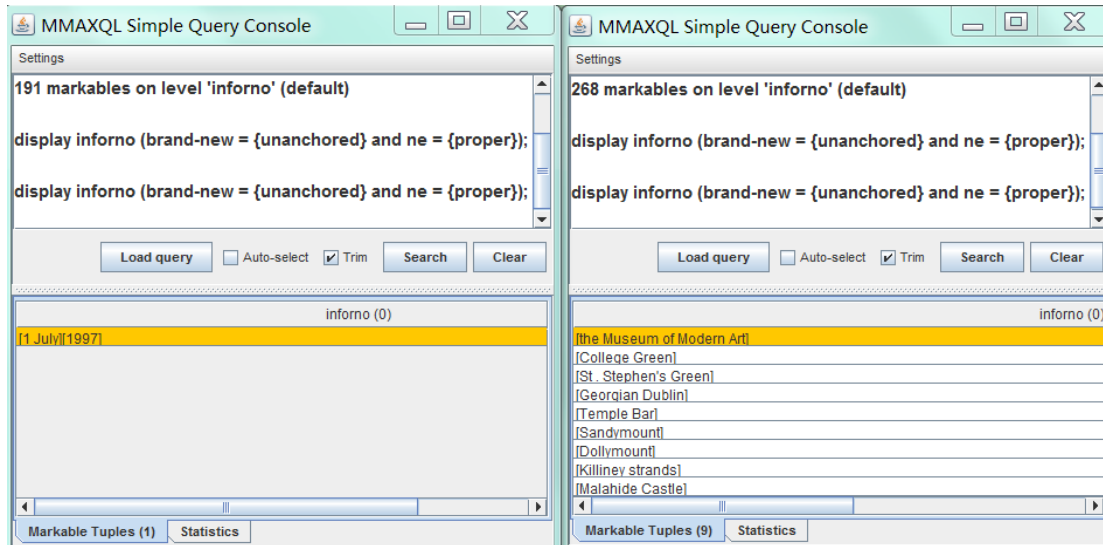


Figure 5–2: Proper nouns of Tg1 and Tg1 expressing Brand–new unanchored information

Regardless of the individual differences within the same genre, the average values of the Travel guide and News report texts are 7.37% and 13.1% respectively. The Travel guide texts do not differ from Government document clearly in the frequency of proper nouns that express Brand–new unanchored information, with a small gap of 0.73%. In contrast, the difference becomes more noticeable when the other genres are compared, with particular reference to the differences between Essay and Government document texts which contain the highest and lowest frequencies respectively.

To compare the texts of four genres, the current analysis selects Np6 as an example of the News report texts, since the distributions of proper nouns are close to the average value, with a slight variation of 1.19%. In addition to the instances of the Travel guide texts presented in Figure 5–2, the proper nouns of Np6, Essay and Government document are illustrated respectively by Figure 5–3. It aims to show that the three genres are different in selecting the category of proper nouns that express Brand–new unanchored information.

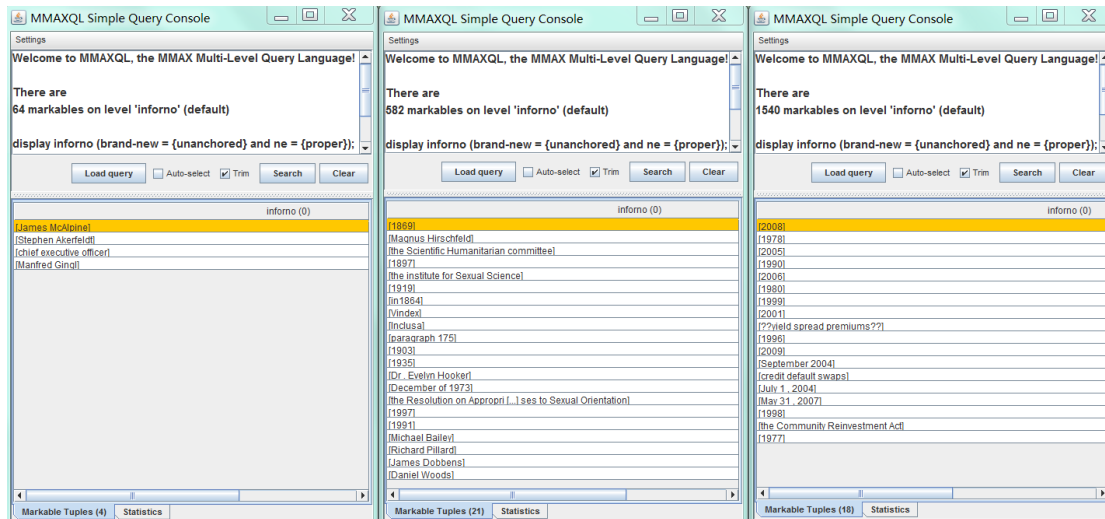


Figure 5–3: Proper nouns of Np6, Essay and Government document expressing Brand–new unanchored information

As indicated by the Markable Tuples of the MMAXQL Query Console in Figure 5–3, Np6 contains proper nouns of personal name and company position; most of the instances of Government document refer to specific day or year; and Essay involves more than two categories of proper nouns that indicate specific year, institute, personal name and articles of law. In comparison to the other three genres, Essay shows more variety in the category of proper nouns expressing Brand–new unanchored status. One possible explanation for the finding can be attributed to the characteristics of different genres, with particular reference to the general and specific topical domains. As indicated by Biber and Conrad (2014: 46), topic is the most important situational factor influencing lexical choice.

The above analysis shows that the use of proper nouns reveals the characteristics of texts. In the Travel guide texts, geographical names are expected to occur in the texts, since the topical domain involves information about places of interest, restaurants and accommodation for visitors. The News report texts focus on describing unknown people, locations and temporal information of certain events, in which proper nouns indicating personal and institutional names, geographical names and specific days, occur frequently. Government document, regarding the content analyzed for the current thesis, aims to offer a series of events that caused the 2007–2008 Financial Crisis in

chronological order and it comes as no surprise that most of the proper nouns refer to years. In comparison, the Essay text entitled *Homosexuality* in the dataset involves several subtopics: the origin of documented homosexuality, famous debates, history of curing it as mental illness and personal interviews, which can lead to both wide usage and variety of proper nouns.

Now the analysis moves to the distributions of nominals with determiners. In general, Table 5–4 indicates nominals with determiners play an important role in structuring Brand–new unanchored information. Five texts have a frequency distribution of around 50% and they are Tg1 (54.24%), Tg2 (52.17%), Np1 (53.06%), Essay (48.67%) and Government document (47.6%). Essay is similar to Government document in the distribution of determiners, with a slight variation of 1.07%. The Travel guide texts, Tg1 and Tg2, also indicate similar frequencies, with a small gap of 2.07% between them. The average frequency of the Travel guide texts is the highest, which is up to 53.21%. The above shows that texts of the three genres, Travel guide, Essay and Government document demonstrate similarity in the Brand–new unanchored information expressed by nominal expressions with determiners.

In comparison, except for Np1, the frequencies of the other News report texts are clearly low. The average value is only 26.06%, which is the minimum of the texts selected from four genres. Furthermore, except for the frequencies in Np5 (38.89%) and Np6 (35.71%), the other News report texts show striking differences, with the highest frequency of 41.95% between Np1 (53.06%) and Np3 (11.11%).

To provide reliable explanations for the similarities and differences between texts of different genres, it is necessary to take the characteristics of types of determiners into account. Example (5–4) illustrates two types in terms of the definiteness of central determiner. In (5–4a), “*an intoxicating place*” is a typical Brand–new unanchored entity that is denoted by the indefinite article “*an*”. While in (5–4b), though with the definite article, “*the example of birds*” is still unknown to readers.

(5–4) a. With its vibrant atmosphere and night–and–day activity it is ***an***

intoxicating place. (Tg1)

- b. He gave *the example of birds*, which are born with instincts such as knowing how to build a nest, or catch their prey. (Ey)

Example (5–5) illustrates another representative type of nominal expressions with determiners that indicate quantity of the referent, either indefinite or definite. In (5–5a), the indefinite quantifier of “*some hormonal theories*” does not contain any phoric relations with other discourse entities in the preceding text and apparently this expression is irrecoverable from the context and text. In comparison, the specific quantities indicated by “*450 editors and publishers*” in (5–5b) and the negative reference of “*no one*” in (5–5c) are brand–new to readers.

- (5–5) a. he believed in *some hormonal theories* as a cause of homosexuality, but this only led to unsuccessful attempts to “cure” homosexuals with the use of hormone injections. (Ey)
- b. The slaughter in Colombia was very much on the minds of *450 editors and publishers* from Latin America, the United States, the Caribbean and Canada attending the 45th general assembly of the Inter–American Press Association in Monterrey, Mexico, this week. (Np1)
- c. Some on Wall Street and in Washington with a stake in the status quo may be tempted to wipe from memory the events of this crisis, or to suggest that *no one* could have foreseen or prevented them. (Gd)

Example (5–6) illustrates another type of nominal expressions with determiner. “*Each and every night*” is embedded within some prepositional phrase and the determiner is inclusive, which refers to all nights. This expression does not indicate any anaphoric relation and conveys Brand–new unanchored information.

- (5–6) Firms depended on tens of billions of dollars of borrowing that had to be

renewed *each and every night*, secured by subprime mortgage securities...
(Gd)

In example (5–7), the nominal expression “*more than double the level held in 1990*” is different from those illustrated by previous examples. Based on Biber *et al.*’s (1999: 258) categorization, there are two determiners in front of the head noun, the predeterminer “*more than double*” and the central determiner “*the*” the central determiner. With more than one determiner, more information of the referent is provided via the premodifier of the head noun. Such an instance was only found in Government document.

(5–7) By 2005, the 10 largest U.S. commercial banks held 55% of the industry’s assets, *more than double the level held in 1990*. (Gd)

In example (5–8), “*this respect*” is a nominal expression with a demonstrative determiner and “*this*” indicates a cataphoric relation to the following content, which is unknown to readers (cf. Biber *et al.* 1999: 274). Demonstrative determiners expressing Brand–new unanchored information were only found in Tg2 and Government document.

(5–8) Nonetheless, we make the following observation about government housing policies — they failed in *this respect*. As a nation, we set aggressive homeownership goals with the desire to extend credit to families previously denied access to the financial markets. (Gd)

Possessive determiners, such as “*their*” in “*their own theories and beliefs*” in the Essay text possessive determiners, cannot express Brand–new unanchored information, since they inherently indicate anaphoric relations to discourse entities in the preceding text. The above also shows that definiteness is not the only factor that influences

information status, some definite nominals can still express new information. Furthermore, expressions with some determiners are not found in all the genres and the difference in type of determiners might be related to genres. Hence, the following focuses on analyzing different types of determiners in text.

As indicated in Section 3.2.1, determiners are further annotated into three subtypes: Predeterminers, Central determiners and the remaining ones. Central determiners are then categorized into article, possessive and demonstrative determiners and those with articles are further labelled in terms of definiteness. The remaining types of determiners basically are the place for nominals that cannot be put in Predeterminers and Central determiners, like the entities in example (5–5) and (5–6). Among them, only possessive determiners cannot be used to express Brand–new unanchored information. Table 5–5 displays the distributions of the remaining determiners per text.¹⁸

As expected, the proportions of predeterminers and demonstrative determiners are small. Although they are found in Government document and Tg2, the frequencies are very low, with all less than 3%. As indicated in example (5–7), predeterminers function to specify the identities of nominals and they are low in frequency among types of determiners and are rare in conveying Brand–new unanchored information. Demonstrative determiners here are used to build cataphora, which is also a relatively rare phenomenon in English text in comparison to the anaphora (e.g. Trnavac and Taboada 2016). Furthermore, not all texts have the remaining types of determiners. For example, Np6 does not contain definite articles and Np3 has no determiners of the remaining types expressing Brand–new unanchored information.

Third, although some texts indicate clear similarities, they are not closely related to genres. The frequency of indefinite articles is similar in two groups of texts. The first is Tg2 (52.78%), Essay (52.73%) and Np3 (50%) and the second is Np5 (35.71%), Tg1 (40.63%) and Np1 (42.31%). In the distributions of nominals with definite articles, three groups, Tg1 (25%) and Government document (24.81%), Np1 (19.23%) Np4 (20%) and Essay (21.82%), Np2 (71.43%) and Np6 (70%) all show similar distributions.

¹⁸ Raw frequencies of Table 5–5 are presented in Appendix A, Table A2.

In the proportions of the remaining types of determiners, Tg1 (34.38%) has a frequency which is similar to Np6 (30%) and Np1 (38.46%) respectively. The frequency of Np4 (20%) has a small gap between that of Essay (25.45%). While for the other texts, there are no clear similarities.

N-type Text	pre-de. (%)	dem. (%)	article- <i>a</i> (%)	article- <i>the</i> (%)	other-det. (%)	Sum (%)
Tg1	0	0	40.63	25	34.38	100
Tg2	0	2.78	52.78	38.89	5.56	100
Np1	0	0	42.31	19.23	38.46	100
Np2	0	0	14.29	71.43	14.29	100
Np3	0	0	50	50	0	100
Np4	0	0	60	20	20	100
Np5	0	0	35.71	0	64.29	100
Np6	0	0	0	70	30	100
Gd	1.55	1.55	28.68	24.81	43.41	100
Ey	0	0	52.73	21.82	25.45	100

Table 5–5: Types of determiners of expressing Brand–new unanchored information per text

Unsurprisingly, Table 5–5 also illustrates that nominals with indefinite articles are more likely to express Brand–new unanchored information. Except for Np2 Np5 and Government document, the other texts contain the highest frequency of occurrence. In contrast, most texts do not contain many definite articles, with Tg2, Np2, and Np3 as exceptions. The finding is not surprising based on the characteristics of type of articles. As for the remaining types of determiners, the distributions do not show regular patterns.

Although no striking similarities are illustrated by the texts of the same genre, Table 5–5 still shows some general tendencies. Compared with Government document, the Travel guide texts tend to use more nominal expressions with both indefinite and definite articles rather than those with the remaining types of determiners to express Brand–new unanchored information. Interestingly, the Essay text indicate a tendency to introduce unknown discourse entities through nominals with indefinite articles. This

might be related to the semantic features of referents and needs to be explored by further research.

Finally, the analysis moves to the remaining types of nominals that were labelled as *other-ne* in Table 5–4. Two typical examples are given in (5–9):

- (5–9) a. During his four days in Iraq, Hall said he wanted to investigate reports from relief agencies that a quarter of Iraqi children may be suffering from ***chronic malnutrition***. (Np3)
- b. Individuals and political action committees in the sector made more than \$1 billion in ***campaign contributions***. (Gd)

In (5–9a), the head “*malnutrition*” is premodified by the adjective “*chronic*” and the “*contributions*” in (5–9b) has a noun “*campaign*” as premodifier. Both nominals indicate no phoric relations to the other discourse entities and they cannot be taken for granted as shared knowledge by the readers.

Table 5–4 indicates that the remaining types of nominals in most texts play an important role in expressing Brand–new unanchored information except for Np1 (4.08%). There are both similarities and differences between texts across different genres. Three groups are similar in the frequencies: (i) Np5 (27.78%), Np6 (28.57%) and Government document (28.04%), with 0.79% as the largest difference among them; (ii) Tg2 (14.49%) and Essay (15.93%), with a gap of 1.44%, and (iii) Np2 (35.71%) and Np3 (38.89%), with a difference of 3.18%. In addition, texts within the same genre also illustrate differences and similarities. Although some News report texts display similarities, the biggest difference in the frequency is up to 39.64% and resides between Np4 (44.44%) and Np1 (4.08%). The Travel guide texts, Tg1 (20.34%) and Tg2 (14.49%), are similar in the frequency of the remaining types of nominals.

Besides the differences between individual texts within the same genre identified above, the average values of the Travel guide and News report texts are 17.42% and 29.91% respectively. The Travel guide texts are similar to Essay in the frequency of the

remaining types of nominals, with a small difference of 1.49% and the News report texts show a similar distribution to Government document, with a gap of 1.87% in between. The two groups are however clearly different. In seeking to provide explanations for where those differences are, Figure 5–4 illustrates all the instances of Tg2 and Np6 via the MMAX2 Query Console, for which the respective frequencies are close to the average values of the Travel guide and News report texts.

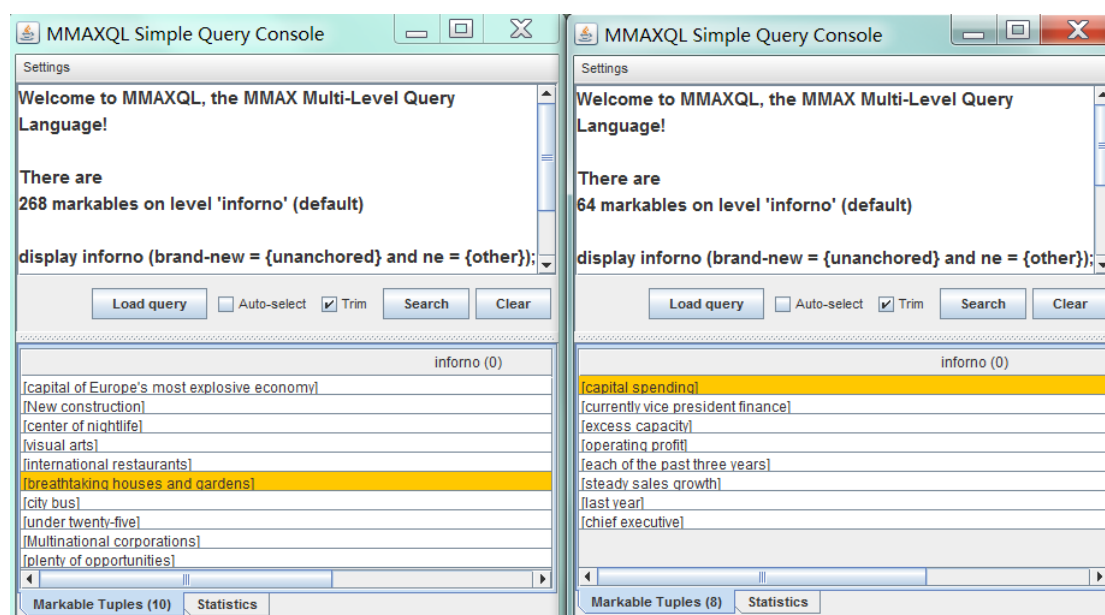


Figure 5–4: The remaining types of nominals expressing Brand–new unanchored information in Tg2 and Np6

In Tg1, the adjective “*breathtaking*” as the premodifier of “*houses and gardens*” indicates the writer’s subjective attitude towards the view. It carries interpersonal meaning, and is an example of an attitudinal Epithet (Halliday and Matthiessen 2014: 376), which was only found in the Travel guide texts. In comparison, the instances of Np6 tend to describe objective properties, for example “*steady sales growth*” that can be confirmed by specific numbers, or denote a particular subtype of the entity, such as “*operating profit*” that means the profit from a company’s business operations. The premodifiers of these expressions are defined as Classifier by Halliday and Matthiessen (2014: 377), which also were found in Essay and Government document texts.

The identified differences are closely related to situational characteristics of the

texts. Although the Travel guide and News report texts focus on describing newsworthy things for general audiences, the News report texts aim to report facts with limited subjective comments, in which case the writer's position is not important. Similarly Essay and Government document texts are expected to analyze and explain certain phenomenon in a neutral position. They focus on factual information with objective properties and verifiable evaluations, and writers' personal opinions are not normally included.

Another explanation for the differences in the remaining types of nominals might be the use of *of*-NP constructions. Figure 5-4 illustrates that Tg2 has three instances, while Np6 only contains one. The *of*-constructions also differ in functions. For example, "*plenty of opportunities*" in Tg2 is a pseudo-partitive expression (Keizer 2007: 112), with "*plenty*" as the quantifier, while "*each of the past three years*" in Np6 indicates a part-whole relationship, which is also defined as the partitive construction (Keizer 2007: 65). Reliable conclusions on the differences in the remaining types of nominals would require further in-depth analyses of types and functions of *of*-NPs in each text, which is beyond the scope of this study.

5.2.2 Linguistic forms of nominals conveying Brand-new anchored information

This section mainly presents the linguistic forms of nominal expressions conveying Brand-new anchored information. As indicated in Section 4.1.2, Brand-new anchored information does not play an important role in structuring text, with the highest frequency of 6.42% in Np1. Not many nominals were found here. Table 5-6 shows the distributions per text. The raw frequencies are presented as Table A3 in the Appendix of this thesis.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	0	0	0	81.82	18.18	100
Tg2	0	0	0	85.71	14.29	100
Np1	0	0	0	58.33	41.67	100
Np2	0	0	0	100	0	100
Np3	0	0	0	100	0	100
Np4	0	0	0	100	0	100
Np5	0	0	0	100	0	100
Np6	0	0	0	100	0	100
Gd	0	0	0	53.85	46.15	100
Ey	0	0	0	50	50	100

Table 5–6: Types of nominals expressing Brand–new anchored information per text

Table 5–6 suggests only two types of nominal expressions express the information. They are those with determiners and of the remaining types. As indicated in Chapter 3, Section 3.2.2, the categorization of Brand–new anchored information in the current thesis is further specified with two more conditions. The first is that the anchoring nominals are required to indicate clear anaphoric links to the preceding text and the second is the non–prototypical relation between anchoring and anchored expressions. Bare nouns, proper nouns, pronouns are naturally excluded from expressing Brand–new anchored information. Example (5–10) illustrates typical nominals under this information category:

- (5–10) a. There are many stories of *refugees* who arrived with nothing in *their pockets*, set up a small sidewalk stall, worked diligently until they had their own store, and then expanded it into a modest chain. (Tg1)
- b. This week, the government arrested Jose Abello Silva, said to be the fourth–ranking cartel leader. He will probably be extradited to *the U.S.* for trial under an extradition treaty President Virgilia Barco has revived. Mr. Barco has refused *U.S. troops or advisers* but has accepted U.S. military aid. (Np1)

In (5–10a), the possessive determiner “*their*” in the expression “*their pockets*” contains an anaphoric link to the “*refugees*” occurred beforehand. “*U.S. troops or advisers*” in (5–10b) belongs to the remaining types of nominals and is semantically connected with the former item “*the U.S.*”. Similarly, the anchoring and anchored expressions do not form prototypical relations.

Table 5–6 also indicates that compared with the remaining types of nominals, those with determiners are more likely to express Brand–new anchored information in all texts except Essay. Five texts of News report, i.e. Np2, Np3, Np4, Np5 and Np6, express the information only by using nominals with determiners. In addition, there are three general patterns demonstrated by the four genres. The first is expressed by the Travel guide texts, with most Brand–new anchored information expressed by determiners; the second is composed of the five texts of News report, with all Brand–new anchored information expressed by determiners; and the third pattern is expressed by Np1, Essay and Government document, since the frequencies of determiners and the remaining types of nominals are almost the same.

Given that example (5–11a) has illustrated the instance with possessive determiner, examples of nominals with possessive determiners, articles and the remaining types of determiners are now provided as follows:

- (5–11) a. With a population of nearly eight million and a total area of just over 1,095 square km (423 square miles), housing is one of *Hong Kong’s* perennial nightmares. To alleviate the problem, the government has become ***the city’s major landlord*** with the construction of massive apartment blocks that, though they have every modern facility, average only 9 square m² in size. (Tg1)
- b. On 1 July, 1997 the British Crown Colony of Hong Kong reverted to *Chinese sovereignty* as ***a Special Administrative Region of the People’s Republic of China***. (Tg1)

- c. The action came in response to a petition filed by Timex Inc. for changes in *the U.S. Generalized System of Preferences* for imports from developing nations... Timex had requested duty-free treatment for many types of watches, covered by **58 different U.S. tariff classifications**. (Np4)

“*The city’s major landlord*” in (5–11a) and “*a Special Administrative Region of the People’s Republic of China*” (5–11b) are typical nominal expressions with definite and indefinite articles that express Brand–new anchored information. Both expressions are related to the preceding items “*the city*” referring to “*Hong Kong*” and “*the People’s Republic of China*” in the form of “*Chinese sovereignty*” that occurred beforehand. They do not form semantic relations, since a city does not necessarily have a major landlord and not every country has a special administrative region as China. In (5–11c), “*58 different U.S. tariff classifications*” has a numeral determiner. Although the denoted entity is anchored to “*the U.S.*” in the previous sentence, it cannot be inferred only based on the anchor.

5.2.3 Linguistic forms of nominals conveying non-containing Inferrable information

This section mainly presents the results of the linguistic forms of nominal expressions expressing non-containing Inferrables. Table 5–7 displays the distributions per text. The raw frequencies are presented as Table A4 in the Appendix of this thesis.

As Table 5–7 indicates, not all types of nominals can express non-containing Inferrables. Proper nouns, bare nouns and pronouns are less likely to be found here. For instance, proper nouns were only found in Government document with 1% and no bare nouns exist in Np4, Np5 and Np6. Second, there are noticeable differences between the texts. Essay contains a higher frequency of nominals with determiners (60.94%), but a lower frequency of bare nouns (10.94), which is in contrast to Np2.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	33.96	0	1.89	49.06	15.09	100
Tg2	29.03	0	4.84	41.94	24.19	100
Np1	11.76	0	2.94	61.76	23.53	100
Np2	50	0	12.5	25	12.5	100
Np3	14.29	0	0	71.43	14.29	100
Np4	0	0	0	50	50	100
Np5	0	0	0	50	50	100
Np6	0	0	0	71.43	28.57	100
Gd	14	1	2.5	50.75	31.75	100
Ey	10.94	0	11.72	60.94	16.41	100

Table 5–7: Types of nominals expressing non–containing Inferrable information per text

In the distributions of bare nouns, the News report texts do not indicate a regular pattern. For example, the gap between Np1 (11.76%) and Np3 (14.29%) is only 2.53%, while the gap between Np2 (50%) and Np3 (0) is as wide as 50%. In comparison, the Travel guide texts are similar in the frequencies, with a gap of 4.93% between Tg1 (33.96%) and Tg2 (29.03%). The distributions of bare nouns play a less important role in both Essay (10.94%) and Government document (14%) texts.

There are mainly two types of non–containing Inferrables expressed by bare nouns. Both are illustrated in example (5–12):

- (5–12) a. *Shopping* never ends. (Tg1)
b. You’ll also notice that *gambling* is a passion, whether it be *cards*,
mahjong, *the lottery*, or *the horses*. (Tg1)

In (5–12a), “*shopping*” is categorized via the mental scheme of travelling that is stored in the readers’ long–term memory and it is activated by the text (cf. van Dijk 1987: 174). Unlike “*shopping*”, the categorization of “*cards*” and “*mahjong*” in (5–12b)

is based on the hyponymy established with “*gambling*” in the preceding clause.¹⁹

Besides, the three non-containing Inferrables in example (5–12) are not specified with particular references, which actually reflects the Maxim of Quantity (Grice 1975). The writer intends to provide neither less nor more information than is required to readers. Here, the bare nouns suggest that the writers do not expect the readers to construct specific references of “*shopping*” “*cards*” and “*mahjong*”, since travel guide texts are more concerned with providing general information of entertaining activities. Although bare nouns are less informative with uncertain identities (Davies and Arnold 2019: 477; Davies and Katsos 2013; Arts *et al.* 2011a, 2001b), they are still considered efficient for the readers to process (also see Piantadosi *et al.* 2011).

Government document is the only text that has proper nouns expressing non-containing Inferrable. There are only four instances and the frequency is only 1%. Example (5–13) is a case in point:

(5–13) While the vulnerabilities that created the potential for crisis were years in the making, it was the collapse of the housing bubble — fueled by low interest rates, easy and available credit, scant regulation, and toxic mortgages — that was the spark that ignited a string of events, which led to a full-blown crisis in *the fall of 2008*. The crisis reached seismic proportions in **September 2008** with the failure of Lehman Brothers and the impending collapse of the insurance giant American International Group (AIG). (Gd)

“*September 2008*” in (5–13) is an anomalous case of non-containing Inferrables compared with “*shopping*” and “*cards*” given in (5–12). The abnormality is completely

¹⁹ Some may argue that “*cards*” and “*mahjong*” in (5–12b) could be classified as Unused information, when they are considered as the games that we are familiar with and are stored as part of our shared knowledge. This classification is also reasonable, since the boundary between Unused and Inferrable is not clear-cut (see Section 2.2.2). However, in (5–12b), the occurrence of “*cards*” and “*mahjong*” is more of an elaboration of “*gambling*” rather than an emphasis of their unique identities of games. “*Cards*” and “*mahjong*” are therefore classified as Inferrables.

caused by the categorical criteria of this study. As indicated in Chapter 3, Section 3.2.2, trigger within a non-containing Inferrable conveys Textually evoked information and it forms a prototypical relationship with the non-containing Inferrable. In (5–13), the year “2008” as the trigger of “*September 2008*” occurred within the preceding nominal “*the fall of 2008*” and the “*fall*” includes “*September*”. Besides, every year has September. “*September 2008*” thus is a non-containing Inferrable. As indicated before, there are only four proper expressions conveying non-containing Inferrable information in the Government document text. The other three instances are all like “*September 2008*”.

It is important to note that the categorization is to some extent problematic with nominals conveying temporal information. Although they occur in the preceding text, they can still convey new information. On the one hand, the year “2008” occurs frequently in Government document and it introduces new items to the text every time. The readers have no access to the temporal information in advance. On the other hand, from the angle of referential givenness, the readers will store “2008” as a known discourse entity after its occurrence and the information status remains valid for the whole text (cf. Baumann and Riester 2012: 138). Although the categorization appears controversial, it does not have great influences on data analysis, with 1% non-containing Inferrables in one of the ten texts. The categorical criteria need to be further discussed. As space is limited, it is not pursued here.

The analysis now moves to pronouns. Table 5–7 indicates that pronouns do not play an important role in expressing non-containing Inferrables in the texts. Except for Np2 (12.5%) and Essay (11.72%), the other texts have a frequency that is less than 5%. Second, the News report texts are clearly different in the frequencies of pronouns expressing non-containing Inferrables. Np3, Np5 and Np6 do not have any frequencies, but Np2 contains the highest. Instances are illustrated by example (5–14):

- (5–14) a. The debate over homosexuality has been one of the most long-lasting and controversial ones ever. What, exactly, causes homosexuality?

Some would say it is a gene, passed on from parents to child. *Others* would argue that it is a result of a child’s upbringing. Still *more* would claim that it is a mental illness that can and should be cured. (Ey)

b. Sheikh Mohamed, who is also the Defense Minister of the United Arab Emirates, announced at the inauguration ceremony that “*we* want to make Dubai a new trading center.” (Np2)

The pronouns “*some*” “*others*” and “*more*” in (5–14a) as non-containing Inferrables are triggered by the expression “*the debate over homosexuality*” in the preceding text, since there are at least two parties involved in a debate. Interestingly, the pronouns do not specify accurate quantities or particular references. They are similar to the bare nouns illustrated in example (5–12), which is also a reflection of the Maxim of Quantity.

In (5–14b), “*we*” as the only non-containing Inferrable in Np2 is fairly uncommon. As a personal pronoun, it typically express Evoked information, both textually and situationally (see Chapter 4, Section 4.1.6 and Section 4.1.7). Here, “*we*” actually refers to all the government officials of the United Arab Emirates, which require readers to infer from the text. This referent includes more people besides the speaker “*Sheikh Mohamed*” himself.

Now this section moves to nominal expressions with determiners expressing non-containing Inferrables. Table 5–7 indicates that they play an important role in conveying the information and the frequencies are the highest among all types of nominals in most texts. Second, there are clear differences between texts. Government document (50.75%), Tg1 (49.06%) and Tg2 (41.94%) display lower frequencies than Essay (60.94%). The News report texts indicate some patterns. Two groups of texts have the same frequencies, namely Np3 and Np6 (71.43%), and Np4 and Np5 (50%). The average value of the News report texts is 54.94%, which is also lower than Essay.

Table 5–8 displays the distributions of types of determiners.²⁰ No demonstrative

²⁰ Raw frequencies of Table 5–8 are presented in Appendix A, Table A5.

determiner was found in the texts. Only Government document has predeterminers, with a very low frequency of 0.49%. The finding is not surprising, as both types of determiners do not occur frequently in the texts (see Section 5.1). In contrast, nominals with definite articles are more likely to express non-containing Inferrables. Except for Np4 and Np5, the texts have the highest frequencies of definite articles among all types of determiners. Indefinite articles, possessive determiners and the remaining types of determiners are of secondary importance.

N-type Text	pre-de. (%)	poss. (%)	dem (%)	article- <i>a</i> (%)	article- <i>the</i> (%)	other-det. (%)	Sum (%)
Tg1	0	26.92	0	7.69	65.38	0	100
Tg2	0	0	0	19.23	73.08	7.69	100
Np1	0	33.33	0	4.76	47.62	14.29	100
Np2	0	50	0	0	50	0	100
Np3	0	0	0	20	70	10	100
Np4	0	0	0	0	0	100	100
Np5	0	25	0	0	25	50	100
Np6	0	40	0	20	40	0	100
Gd	0.49	20.2	0	10.84	54.68	13.79	100
Ey	0	21.79	0	19.23	35.9	23.08	100

Table 5–8: Types of determiners of expressing non-containing Inferrable information per text

The Travel guide and News report texts do not show obvious patterns in the frequencies of types of determiners within the same genre. It is hard to explain the different frequencies by characteristics of text types. For example, Tg1 has a frequency of 26.92% of possessive determiners, while Tg2 has no frequency at all. In the distributions of the remaining types of determiners, the gap between Np2 (0) and Np4 (100%) is remarkably wide, reaching 100%. However, the difference here cannot be regarded as striking, given the few instances we found in the dataset.

Examples (5–15) to (5–17) display the instances of all types of determiners expressing non-containing Inferrables.

(5–15) By one measure, *their leverage ratios* were as high as *40 to 1*, meaning for *every 40 dollars in assets*, there was only 1 dollar in capital to cover losses. (Gd)

“*Every 40 dollars in assets*” in (5–15) is the only instance with a predeterminer in the dataset. It connects to the ratios “*40 to 1*” occurred in the previous clause, but provides a detailed explanation of the “*40*”. The writer aims to explain the financial measurement to enable the readers to understand “*their leverage ratios*”. The non-containing Inferrable “*every 40 dollars in assets*” is a reflection of both the Maxim of Quantity and informativeness (Grice 1975; Davies and Arnold 2019). Compared with the bare nouns illustrated by example (5–12), the nominal here is more informative with the use of determiners. This also suggests that it is possible for the writers to adjust the forms of nominals in order to provide sufficient information to the readers.

(5–16) The committee published *many books and other forms of literature*, which gave *Hirschfeld* a great amount of prestige in *his field*. (Ey)

“*His field*” in (5–16) is linked with “*Hirschfeld*” by the possessive determiner “*his*”. Besides, “*many books and other forms of literature*” also implicate that there is a field where he contributed. With both semantic and referential relations, the readers are more familiar with the non-containing Inferrable “*his field*” than the bare nouns illustrated in (5–12).

(5–17) a. Almost half of Ireland’s population is under twenty–five, and with its universities and professional schools, Dublin also has *a large student population*. (Tg2)

b. Hall said, he wanted to find out whether the United Nations or relief agencies needed to handle things differently, or whether “Iraq needs to get out of the way and let us do the job. *The Iraqi government* blames

the embargo for the malnutrition, infant mortality and other hardships.
(Np3)

The expression “*a large student population*” in (5–17a) is an instance of non-containing Inferrables with indefinite articles. It is involved in three relations. First, Dublin as the capital city of Ireland can represent the country in many aspects, such as politics, economy and education; second, Ireland is likely to host a large number of students, as it is home to plenty of educational institutions; and lastly, Dublin is expected to have a large population of students. The categorization of “*a large student population*” requires more mental effort than those inferred from simple prototypical relations, such as “*his field*” in example (5–16).

“*The Iraqi government*” in (5–17b) is an instance of non-containing Inferrables with definite articles. The categorization is the same with “*his field*” in (5–16), which has both referential and semantic relations to the trigger “*Iraq*” that occurred before.

(5–18) Timex is a major U.S. producer and seller of watches, including low-priced battery-operated watches assembled in *the Philippines* and ***other developing nations*** covered by the U.S. tariff preferences. (Np4)

“*Other developing nations*” in (5–18) is an instance of the remaining types of determiners that express non-containing Inferrable. From the perspective of cohesion (Halliday and Hasan 2012: 73–74), it forms co-classification with “*the Philippines*”, since both are members of developing nations. Non-containing Inferrables expressed by the determiner “*other*” were also found in the other texts.

Finally this section presents the analysis results for the remaining types of nominals. Table 5–7 indicates that they play an important role in expressing non-containing Inferrables in the texts, displaying the highest frequency of 50% and the lowest of 12.5%. Second, the Travel guide texts do not indicate similar frequencies. Tg1 (15.09%) and Tg2 (24.19%) show a clear gap of 9.1%. Third, the News report texts show both

similarities and differences. Three sets of texts show similar frequencies of the remaining types of nominals expressing non-containing Inferrables. Np4 and Np5 happen to contain the same frequency of 50%; the gap between Np2 (12.5%) and Np3 (14.29%) is only 1.79%; and Np1 (23.53%) and Np6 (28.57%) have a small difference of 5.04%. However, texts of different sets are clearly different. For example, the gap between Np2 and Np4 is as wide as 37.5%. Essay (16.41%) contains similar frequencies with the average value of the Travel guide texts (19.64%). Government document (31.75%) shows resemblance to the average frequencies of the News report texts (29.82%). Examples in (5–19) illustrate some instances:

- (5–19) a. The White House said Mr. Bush decided to grant duty-free status for 18 categories, but turned down such treatment for *other types of watches* “because of the potential for material injury to ***watch producers*** located in the U.S. and the Virgin Islands.” (Np4)
- b. They resulted not only in ***significant financial consequences*** but also in damage to the trust of investors, businesses, and the public in the financial system. (Gd)
- c. Perhaps ***one of the most famous studies*** on this topic was concluded in 1991 by Michael Bailey, an assistant professor of psychology at Northwestern University, and Richard Pillard, an associate professor of psychiatry at Boston University School of Medicine, and found that sexual orientation in males is largely due to genetics. (Ey)

In (5–19), the nominals shown in bold italics display three types of non-containing Inferrables. Like “*his field*” in example (5–16), “*watch producers*” in (5–19a) indicates both semantic and morphological relations with “*other types of watches*” that occurs before. In (5–19b), “*significant financial consequences*” is triggered by the preceding text that indicates specific economic damages caused by the 2007–2008 Financial Crisis. Based on the adapted framework of Prince’s (1981) information taxonomy (see Chapter

3), “*one of the most famous studies*” in (5–19c) is a non-containing Inferrable.

The remaining types of nominals are more complex than bare nouns and pronouns by containing more words. They are thus more informative according to the principle of Informativeness (e.g. Davies and Arnold 2019: 477). As indicated earlier, Essay and the Travel guide texts contain lower frequencies of the remaining types of nominals than the Government document and News report texts. This suggests that non-containing Inferrables of Essay and the Travel guide texts are less informative to some extent than those of the other two genres.

5.2.4 Linguistic forms of nominals conveying containing Inferrable information

This section focuses on the distributions of various types of nominal expressions in expressing containing Inferrables. As indicated in Section 4.1.4, containing Inferrables do not play an important role in structuring texts, with the highest frequency of 5.77% in Np4. Table 5–9 illustrates the percentages of containing Inferrables represented by varied linguistic forms per text, which actually are very few cases. The raw frequencies are presented as Table A6 in the Appendix of this thesis.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	0	0	0	66.67	33.33	100
Tg2	0	0	0	83.33	16.67	100
Np1	0	0	0	0	100	100
Np2	0	0	0	100	0	100
Np3	0	0	0	0	0	100
Np4	0	0	0	100	0	100
Np5	0	0	0	100	0	100
Np6	0	0	0	0	100	100
Gd	0	0	0	72.5	27.5	100
Ey	0	0	0	64.71	35.29	100

Table 5–9: Types of nominals expressing containing Inferrable information per text

In general, only two types of nominals are used to convey containing Inferrable in the texts, namely those with determiners and the remaining types of nominals. This result is not surprising. Containing Inferrables in the current thesis (see Chapter 3, Section 3.2.2) need to meet three conditions: (i) the containing Inferrable contains at least two items within the nominal construction; (ii) the items form prototypical relationships and (iii) at least one item as trigger of the containing Inferrable conveys Evoked information, either textually or situationally. Unlike the anomalous proper expressions such as “*September 2008*” conveying non-containing Inferrable information (see Section 5.2.3), other proper nominals, bare nouns and pronouns only represent one discourse entity. It is very rare for them to express containing Inferrable information that forms semantic relations by the constructions themselves.

Table 5–9 also indicates three distributive patterns. The first is represented by the frequencies of Tg1, Tg2, Government document and Essay, with most containing Inferrables expressed by nominals with determiners; the second is represented by the frequencies of Np2, Np4 and Np5, with all containing Inferrables expressed by nominals with determiners; and the last pattern is illustrated by the frequencies of Np1 and Np6, with all Inferrables expressed by the remaining types of nominals. Example (5–20) provides illustrations:

- (5–20) a. *The action* came in response to a petition filed by Timex Inc. for changes in the U.S. Generalized System of Preferences for imports from developing nations... U.S. trade officials said the Philippines and Thailand would be ***the main beneficiaries of the president’s action***. (Np4)
- b. According to Darwin’s theory of natural selection, the advantageous traits are passed on, while *the disadvantageous ones* eventually die out... Therefore, from the evolutionary standpoint, homosexuality becomes ***one of the disadvantageous traits***. (Ey)

In (5–20a), “*the president’s action*” has occurred before in the form of “*the action*” and forms a prototypical relation to “*the main beneficiaries*”. This makes “*the main beneficiaries of the president’s action*” a containing Inferrable. Similarly, in (5–20b), the triggering item “*the disadvantageous traits*” has occurred in the preceding text and forms a part–whole relationship with “*one of the disadvantageous traits*”. The construction “*one of + NP*” typically conveys Inferrable information, since it forms part–whole relationship within the nominal itself (cf. Prince 1981: 236).

5.2.5 Linguistic forms of nominals conveying Unused information

This section presents the results of types of nominals expressing Unused information. Table 5–10 displays the distributions per text. The raw frequencies are presented as Table A7 in the Appendix of this thesis.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	0	89.47	0	5.88	5.88	100
Tg2	2.5	80	0	15	2.5	100
Np1	0	100	0	0	0	100
Np2	0	100	0	0	0	100
Np3	0	100	0	0	0	100
Np4	0	100	0	0	0	100
Np5	0	100	0	0	0	100
Np6	0	100	0	0	0	100
Gd	1.75	54.39	0	38.6	5.26	100
Ey	0	80	0	16	4	100

Table 5–10: Types of nominals expressing Unused information per text

Table 5–10 suggests that pronoun cannot be used to express Unused information. No frequencies were found in the texts accordingly. Most frequencies were however found for proper expressions. This finding is not surprising, as proper nouns refer to discourse entities with unique identities, which is a typical characteristic of Unused information. Interestingly, all the information is conveyed by proper nominals in the

News report texts, which is different from the other texts. The remaining types of nominals can convey Unused information as well. However, their frequencies are much lower than those of proper expressions in the texts. Example (5–21) illustrates instances of bare nouns expressing Unused information.

- (5–21) a. *Coffee* has replaced the ubiquitous tea — Dublin is now almost as much a coffee city as Vienna or Seattle. (Tg2)
- b. They relaxed their underwriting standards to purchase or guarantee riskier loans and related securities in order to meet stock market analysts’ and investors’ expectations for growth, to regain market share, and to ensure generous compensation for their executives and employees — justifying *their activities on the broad and sustained public policy support for homeownership*. (Gd)

In (5–21a), “*coffee*” is taken for granted by the readers as a popular drink and naturally conveys Unused information. Unlike the categorization of “*coffee*”, “*homeownership*” in (5–21b) is classified in terms of the complex nominal “*their activities on the broad and sustained public policy support*”. When the government housing policy was introduced to the text the first time, the writer did not provide a description in greater detail. The most possible explanation is that American people as the readers have already been familiar with government activities on housing. This means the whole expression “*their activities on the broad and sustained public policy support for homeownership*” is considered as known to the readers, which conveys Unused information. “*Homeownership*” as part of the complex nominal is classified as Unused.

The instance in (5–21b) is anomalous and this is mainly caused by the categorical criteria. As stated in Chapter 3, this study annotates the information values of head nouns and their postmodifiers of nominal complexes separately without considering the non-nominals elements, such as “*on*” and “*for*” in “*their activities on the broad and*

sustained public policy support for homeownership” illustrated by (5–21b). However, it is hard to categorize information status of certain expression without referring to the information value of the whole complex nominal, like “*homeownership*” in (5–21b). Therefore, bare nouns are used to convey Unused information in the current coding scheme. As indicated in Section 5.2.3, the categorical criteria of the current thesis might be considered controversial. However, instances like “*homeownership*” are very rare in the texts and do not have great impact on statistical results.

Example (5–22) illustrates Unused information expressed by nominals with determiners and the remaining types of nominals respectively.

- (5–22) a. ***The profound events of 2007 and 2008*** were neither bumps in the road nor an accentuated dip in the financial and business cycles we have come to expect in a free market economic system. (Gd)
- b. You can view artifacts from the Bronze Age, trace the history of the Easter Rising, or revisit ***Leopold Bloom’s Odyssey*** in *Ulysses*. (Tg2)

In (5–22a), the nominal “*the profound events of 2007 and 2008*” is categorized as Unused information taken for granted by Americans, since they as readers have suffered from the Financial Crisis of 2007 to 2008 and have known the profound events before reading the report. In (5–22b), “*Leopold Bloom’s Odyssey*” is famous in *Ulysses* written by James Joyce. It is assumed to be background knowledge shared between the writer and the readers.

5.2.6 Linguistic forms of nominals conveying Situationally evoked information

This section is mainly concerned with the linguistic forms of nominals expressing Situationally evoked information. As indicated in Section 4.1.6, Situationally evoked information plays a less important role in structuring written texts. There are only 14 instances in our dataset. Table 5–11 displays the distributions per text. The raw

frequencies are presented as Table A8 in the Appendix of this thesis.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	0	0	100	0	0	100
Tg2	0	0	50	50	0	100
Np1	0	0	25	75	0	100
Np2	0	0	0	100	0	100
Np3	0	0	0	0	0	100
Np4	0	0	0	0	0	100
Np5	0	0	0	0	0	100
Np6	0	0	0	0	0	100
Gd	0	0	25	75	0	100
Ey	0	0	50	50	0	100

Table 5–11: Types of nominals expressing Situationally evoked information per text

Only two types of nominals were found here and they are pronouns and nominals with determiners. The frequencies display three patterns across all texts. The first pattern is composed of Tg1 and Np2, expressing Situationally information only by one type of nominals, either pronouns or nominals with determiners; the second is represented by Tg2 and Essay, expressing the information by two types of nominals with the same frequencies; and the third pattern is constituted by Np1 and Government document, expressing the information through a frequency of 25% of pronouns and 75% of nominals with determiners. The findings are not surprising. As shown in Section 4.1.6, the degree of interaction between the writhers and readers is very low. Few things are shared by the writers and readers in written texts, which are usually the roles of the writers and readers expressed by personal nouns such as *I*, *you* and *we*; the general environment expressed by nominals with determiners, for instance *the world*, *the country* and *the planet*; and the text itself expressed by nominals with determiners, like *this research*, *this report* and *this paper*.

(5–23) *our country, the status quo, the world, you, this report* (Gd)

you, the world (Ey)
you (Tg1)
us, the world (Tg2)
we, this week, the hemisphere, this week (Np1)
the world (Np2)

Example (5–23) illustrates all Situationally evoked nominals of the texts. All of the referents denote types of information shared by the writers and readers in written texts.

5.2.7 Linguistic forms of nominals conveying Textually evoked information

This section aims to explore linguistic forms of nominals expressing Textually evoked information. Table 5–12 shows the distributions per text. The raw frequencies are presented as Table A9 in the Appendix of this thesis.

N-type Text	bare (%)	proper (%)	pron. (%)	det. (%)	other-ne (%)	Sum (%)
Tg1	8.89	35.56	26.67	26.67	2.22	100
Tg2	17.07	36.59	19.51	26.83	0	100
Np1	0	30.3	28.79	40.91	0	100
Np2	14.29	28.57	14.29	42.86	0	100
Np3	3.13	46.88	34.38	12.5	3.13	100
Np4	0	41.18	0	29.41	29.41	100
Np5	5.41	10.81	24.32	51.35	8.11	100
Np6	4.35	52.17	13.04	30.43	0	100
Gd	12.54	11.25	26.07	38.46	11.68	100
Ey	22.84	9	41.52	21.11	5.54	100

Table 5–12: Types of nominals expressing Textually evoked information per text

The table suggests some nominals were not found in the texts to express Textually evoked information. Np1 and Np4 do not contain any frequencies of bare nouns; Tg2, Np1, Np2 and Np6 have no frequencies of the remaining types of nominals. Compared with the frequencies of proper nominals and determiners, the frequencies of pronouns

are lower in the texts, except in the Essay text. Besides, the News report texts seem more restricted in selecting nominals expressing Textually evoked information than Essay, Government document and Travel guide texts, in view of the observation that some of the texts do not contain bare nouns and the remaining types of nominals.

In the distributions of bare nouns expressing Textually evoked information, Essay (22.84%) contains the highest frequency across all texts. In contrast, most News report texts have lower frequencies of bare nouns, with an average value of only 4.53%. The frequencies of Tg1 (8.89%) and Tg2 (17.07%) differ from each other clearly, with a gap of 8.18%. The average value of the Travel guide texts (12.98%) is very close to the distribution of Government document (12.54%). Typical examples are given in (5–24):

- (5–24) a. What, exactly, causes *homosexuality*? (Ey)
b. Dubai's Crown Prince Sheikh Mohamed Bin Rashid Al Maktoum inaugurated a free zone for *e-commerce* today, called Dubai Internet City. (Np2)

The two Textually evoked nominals “*homosexuality*” and “*e-commerce*” share a common feature in the antecedents. Both repeat the headings of the texts indicating the main topics. The topicality makes “*homosexuality*” and “*e-commerce*” occur more frequently and the frequencies of bare nouns expressing Textually evoked information in Essay and Np2 are higher than the other texts.

In the distributions of proper nominals expressing Textually evoked information, Tg1 (35.56%) and Tg2 (36.59%) have similar frequencies, with a small gap of 1.03%, and the average values of the texts are clearly higher than the frequencies of Essay (9%) and Government document (11.25%). News report texts indicate both similar and different frequencies of proper nominals expressing Textually evoked information. Np1 (30.3%) and Np2 (28.57%) only have a gap of 1.73%, while the gap between Np5 (10.81%) and Np6 (52.17%) is as wide as 41.36%.

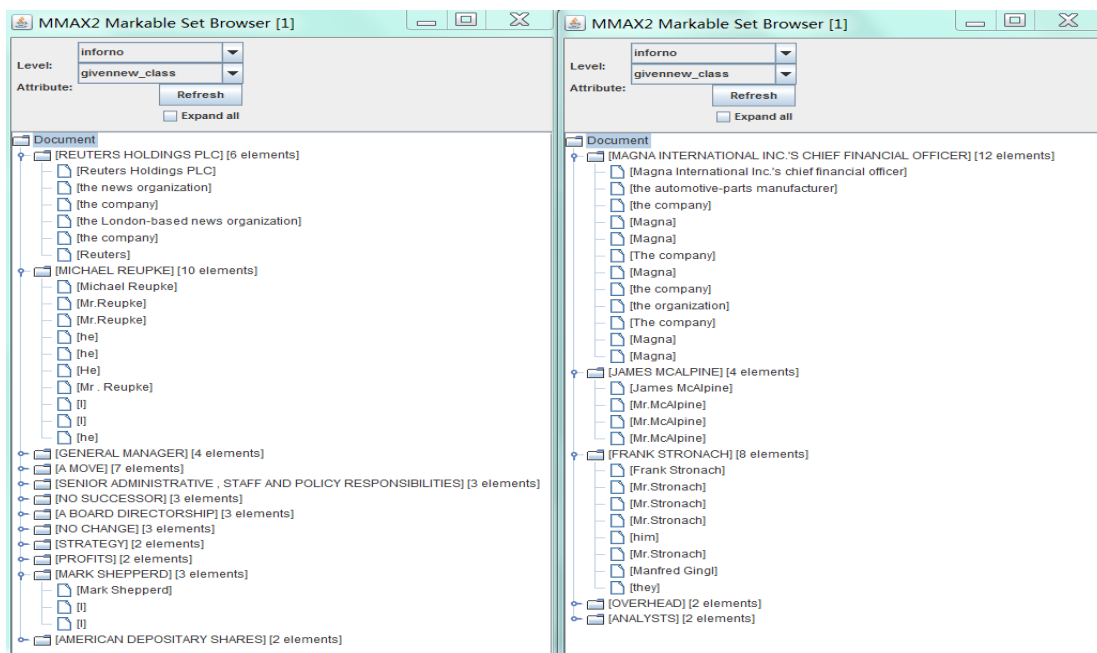


Figure 5–5: Given–new chains in Np5 and Np6

One possible explanation for the differences between the News report texts is concerned with both the repetitive forms of Textually evoked information and the topics of texts. Figure 5–5 displays the co-referential chains in Np5 and Np6. Both have three chains that start with proper expressions and the forms of repetition are illustrated by expanding the chains. Taking the topics into consideration, Np5 and Np6 are mainly about the staff resignation of some companies (also see Section 4.1.1). In comparison, the nominals in Np6 are more frequently repeated by proper expressions rather than pronouns or nominals with determiners, leading to a higher frequency of proper expressions expressing Textually evoked information.

Apparently, the topics of Np5 and Np6 are expressed by proper nominals that denote unique identities of the staff and companies, which are different from that of Essay. In the Travel guide texts, the topics of Tg1 and Tg2 focus on describing well-known cities that are typically expressed by proper nominals. The variety in topics explains the differences in frequencies of proper nominals and bare nouns expressing Textually evoked information across texts of different genres.

Now the analysis moves to the pronouns expressing Textually evoked information. The texts of Travel guide and News report do not indicate similar frequencies within

the same genre. For example, the biggest gap of the News report texts is as high as 34.38% between Np3 (34.38%) and Np4 (0). In terms of the average values of the Travel guide (23.09%) and News report (24.19%) texts, both genres do not clearly differ from Government document (26.07%) in the frequencies of pronouns expressing Textually evoked information. In comparison, Essay (41.52%) has the highest frequencies across all texts.

Pronouns are regarded as light, unspecific and non-rigid elements in the Accessibility Theory (Ariel 1996), nominals conveying Textually evoked information in the Essay text are thus more accessible to readers than the other texts. The degree of accessibility can be influenced by many factors: distance between the evoked nominals and the antecedents, intervention of potential antecedents, salience of the antecedents, and referential and relational coherence between Textually evoked nominals and the antecedents (Ariel 2001: 36; cf. Schmid 2000: 340). Examples in (5–25) illustrate above-mentioned influences:

- (5–25) a. What, exactly, causes *homosexuality*? Some would say *it* is a gene, passed on from parents to child. (Ey)
- b. *Dublin* excels in packaging its past for the visitor. You can view artifacts from the Bronze Age, trace the history of the Easter Rising, or revisit Leopold Bloom’s odyssey in Ulysses. Old buildings are being recycled; for example, the 17th-century Royal Hospital now holds the Museum of Modern Art. And ***Dublin***, a city large in expectations, is still small enough for the visitor to see most of its sights on foot. (Tg2)

The emboldened nominals in example (5–25) denote the main topics of the texts. In (5–25a), the antecedent of “*it*”, namely “*homosexuality*”, just occurred in the previous clause and the two clauses form a question-and-answer relation. In (5–25b), the second “*Dublin*” has a longer distance to the first “*Dublin*”, with two sentences in

between. Given the fact that more competing referents were introduced to the text, the writer repeated Dublin in a full form, which is more informative and specific than the pronoun “*it*” in (5–25a).

Table 5–13 presents the distributions of types of pronouns expressing Textually evoked information.²¹ Most pronouns are personal in the texts, with Tg1, Np2, Np3 and Np6 having a remarkably high frequency of 100% personal pronouns. However, this does not indicate similarities between the texts, since not many pronouns were found in the dataset, with 12 instances in Tg1, 2 instances in Np2, 11 instances in Np3 and 3 instances in Np6. In contrast, the other types of pronouns are less likely to occur in the texts. This is related to the distributions of varied pronouns in general. Personal pronouns are more frequent to be seen in texts than other types. According to Biber *et al.* (1999: 333–354), there are nearly 40,000 personal pronouns per million words in news, while the numbers of possessive pronouns, reflexive pronouns, reciprocal pronouns in news are all below 10,000 per million words. Table 5–2 in Section 5.1 also indicates that the frequencies of personal pronouns are higher than other pronouns in all texts. Only two types of pronouns are used to express Textually evoked information in the News report texts. In comparison, Tg2, Essay and Government document have at least four types of pronouns expressing the same information.

²¹ Raw frequencies of Table 5–13 are presented in Appendix A, Table A10.

N-type Text	poss. (%)	personal (%)	dem. (%)	reflex. (%)	other-pron. (%)	Sum (%)
Tg1	0	100	0	0	0	100
Tg2	6.25	81.25	6.25	6.25	0	100
Np1	0	84.21	15.79	0	0	100
Np2	0	100	0	0	0	100
Np3	0	100	0	0	0	100
Np4	0	0	0	0	0	100
Np5	0	88.89	0	0	11.11	100
Np6	0	100	0	0	0	100
Gd	1.09	86.96	7.61	3.8	0.54	100
Ey	0	84.17	6.67	3.33	5.83	100

Table 5–13: Types of pronouns expressing Textually evoked information

Although the Accessibility Theory (Ariel 1996) did not provide a detailed account of the distinctions between types of pronouns, it is reasonable to predict that some pronouns are more accessible than others. For instance, reflexive pronouns are more accessible than personal pronouns due to the short distance to the antecedent. This point is illustrated by example (5–26):

(5–26) And *the financial sector* ***itself*** has become a much more dominant force in our economy. (Gd)

In (5–26), “*itself*” occurs immediately after the antecedent “*the financial sector*”, making it easier for readers to process in relation to “*it*” illustrated by example (5–25). The degree of informativeness of type of pronouns is different even when they convey the same information status. This difference brings a new perspective in analyzing the frequencies of pronouns expressing Textually evoked information in the texts. For example, although Tg1 (26.67%) has a higher frequency of pronouns than Tg2 (19.51%) (see Table 5–12), this does not indicate that Tg1 is more accessible. As is shown by Table 5–13, the degree of accessibility of Tg1 might be lower than Tg2 in terms of its lower frequency of reflexive pronouns.

Typical examples of the other types of pronouns are given in (5–27):

- (5–27) a. *President Bush has agreed to meet within 90 days with Mr. Barco, President Alan Garcia of Peru and President Jaime Paz Zamora of Bolivia to discuss the drug problem.* It might not be a bad idea to do **that** sooner, rather than later. (Np1)
- b. To supplement my findings from research, I have conducted personal interviews with *two adolescent homosexual males, James Dobbens and Daniel Woods.* **Both** were asked how what they thought determined homosexuality (nature vs. nurture) and why, when they realized they were homosexual and how they knew, and similar questions. (Ey)

In (5–27a), “*that*” as a demonstrative pronoun refers to the discussion between the three presidents mentioned before. In (5–28b), “*both*” as an instance of the remaining types of pronouns refers to the “*two adolescent homosexual males, James Dobbens and Daniel Woods*”. Similarly, both pronouns integrate more than one piece of information into one conceptual unit, which are cognitively more complex. This integration makes it possible for the pronouns to take in new information (Schmid 2000: 370).

Now this section moves to present findings about nominals with determiners expressing Textually evoked information. Table 5–12 indicates that Tg1 (26.67%) and Tg2 (26.83%) have similar frequencies under the genre of Travel guide. The News report texts show both similarities and differences in the frequencies of determiners expressing Textually evoked information. The frequencies of two sets of texts, Np1 (40.91%) and Np2 (42.86%), Np4 (29.41%) and Np6 (30.43%), are similar. Each set indicates a gap that is lower than 2%. In contrast, Np3 (12.5%) is clearly different from Np5 (51.35%), with a big gap of 38.85%. The average values of the Travel guide and News report texts are 26.75% and 34.58% respectively. Among texts of four genres, Government document (38.46%) contains the highest frequency of determiners expressing Textually evoked information, while the lowest is contained by Essay

(21.11%).

Differences between these genres could be explained by the topics of texts and forms of repetition of Textually evoked information which have been discussed in the analysis of bare nouns and proper nouns earlier in this section. As indicated before, Government document is concerned with the Financial Crisis of 2007 to 2008. The main topic, namely “*the financial crisis*”, is mostly repeated by the form of “*the crisis*”, which is a typical nominal with determiner. Compared with the other texts, it is not surprising to see Government document has the highest frequencies of determiners expressing Textually evoked information.

N-type Text	pre-de. (%)	poss. (%)	dem. (%)	article- <i>a</i> (%)	article- <i>the</i> (%)	other-det. (%)	Sum (%)
Tg1	8.33	0	25	8.33	58.33	0	100
Tg2	0	9.09	22.73	4.55	59.09	4.55	100
Np1	0	0	3.7	0	96.3	0	100
Np2	0	0	0	16.67	83.33	0	100
Np3	0	0	0	0	100	0	100
Np4	0	0	20	0	40	40	100
Np5	0	21.05	5.26	5.26	63.16	5.26	100
Np6	0	0	0	0	100	0	100
Gd	0	13.33	23.33	5.56	56.3	1.48	100
Ey	1.64	14.75	18.03	29.51	22.95	13.11	100

Table 5–14: Types of determiners expressing Textually evoked information per text

Table 5–14 displays the distributions of types of determiners expressing Textually evoked information.²² In general, nominals with definite articles have higher frequencies than the other types of determiners. Np3 and Np6 even have a frequency of 100%. However, their raw numbers are small, with 4 and 7 instances respectively. Demonstrative determiners are of secondary importance, with the highest frequency of 25% in Tg1. The remaining types of determiners, which were labelled as *other-det.*, are relatively rare. Besides, compared with other texts, the News report texts are more

²² Raw frequencies of Table 5–14 are presented in Appendix A, Table A11.

restricted to selecting types of determiners to convey Textually evoked information, since many of them have frequencies of only one or two types of determiners. Travel guide texts have similar frequencies of demonstrative determiners and definite articles, with respective gaps of 2.27% and 0.76%. Government document and Essay are similar in the frequencies of possessive and demonstrative determiners, with small gaps of 1.42% and 5.3% respectively. Interestingly, Essay has demonstrated a more complex pattern of expressing Textually evoked information, with frequencies of all kinds of determiners. Typical examples of types of determiners are given as follows:

(5–28) *Hong Kong* is crowded — it has one of the world’s greatest population densities. But it is also efficient, with one of the best transportation systems anywhere, and for ***such a crowded place***, quiet... (Tg1)

In example (5–28), “*such a crowded place*” is a typical Textually evoked item expressed by a predetermined nominal, which obviously refers to “*Hong Kong*” in the preceding sentence. Here the predeterminer “*such*” is used to intensify the degree of population density, which is a common feature of subjectivity (cf. Biber *et al.* 1999: 282).

As indicated in Table 5–14, only Tg1 and Essay have the frequencies of predeterminers expressing Textually evoked information, with a gap of 6.69%. However, Tg1 and Essay contain only one instance. The gap is mainly caused by the various quantities of Textually evoked nominals of the complete texts rather than the differences in text types.

Example (5–29) illustrates an instance of the remaining types of determiners:

(5–29) Furthermore, there were *many occurrences of homosexual behavior* in Greek mythology; Hercules is rumored to have had 14 male lovers, and Zeus himself partook in ***such behavior***. (Ey)

In (5–29), the categorization of “*such behavior*” is the same with “*such a crowded place*” illustrated by (5–28) previously, with “*such*” linking with the antecedent “*many occurrences of homosexual behavior*” occurred in the preceding text. The “*such*” here is used to classify the specific kind of the behavior and this classifying feature is frequent in academic prose (cf. Biber *et al.* 1999: 282). It is not surprising to find the Essay text has such instances by taking into account its feature of developing arguments.

Table 5–14 shows that nominals with indefinite articles could also convey Textually evoked information. Examples are given in (5–30):

- (5–30) a. But as the report will show, *the financial industry itself* played a key role in weakening regulatory constraints on institutions, markets, and products. It did not surprise the Commission that ***an industry of such wealth and power*** would exert pressure on policy makers and regulators. (Gd)
- b. We clearly believe the crisis was a result of human mistakes, misjudgments, and misdeeds that resulted in systemic failures for which *our nation* has paid dearly... But as ***a nation***, we must also accept responsibility for what we permitted to occur. (Gd)

The categorization of “*an industry of such wealth and power*” in (5–30a) is similar to the instances in (5–28) and (5–29), with “*such*” indicating an anaphoric relation to the antecedent “*the financial industry itself*” occurred in the preceding text. Although the nominal starts with an indefinite article, it is actually definite. Its special construction with the relator “*of*” denotes that “*an industry*” refers to a particular experiential identity that which has “*such wealth and power*”. “*A nation*” in (5–30b) is indefinite based on the linguistic form. However, the nominal here actually has the same reference with “*our nation*” in the preceding sentence, which conveys Textually evoked information.

The final part of this section presents findings of the remaining types of nominals.

Table 5–12 indicates that they do not play an important role in expressing Textually evoked information. The instances were only found in six texts with four of them having a frequency that is lower than 10%. Government document and Essay have higher frequencies of the remaining types of nominals than the Travel guide texts. The News report texts are clearly different. Np4 has the highest frequency of 29.41% across all texts, while Np1 Np2 and Np6 do not contain any frequencies.

(5–31) The White House said President Bush has approved *duty-free treatment* for imports of certain types of watches... Imports of the types of watches that now will be eligible for *duty-free treatment* totaled about \$37.3 million in 1988, a relatively small share of the \$1.5 billion in U.S. watch imports that year, according to an aide to U.S. Trade Representative Carla Hills. (Np4)

Example (5–31) illustrates an instance of the remaining types of nominals expressing Textually evoked information. The first “*duty-free treatment*” here already occurred at the very beginning of the text. As one of the topical elements, it was fully and more frequently repeated than the other Textually evoked nominals. Therefore, it is not surprising that Np4 has the highest frequency of the remaining types of nominals.

5.3 Summary

The goal of this chapter was to explore the relationship between linguistic forms of nominal expressions and information status they represent in English texts. Unlike the previous literature, this chapter provides a detailed description of this relationship by taking into account of linguistic forms of nominals representing all categories of information in ten texts selected from four comparable genres, thereby addressing the second research question identified in Chapter 1. Specifically, Section 5.1 has outlined the frequency distributions of various nominals in the texts. Section 5.2 has presented

findings about the relationship between linguistic forms of nominals and information status they represent.

This chapter has presented several important findings. First, every linguistic form of nominal expressions can be used to represent more than one type of information. This provides some empirical evidence for Lambrecht’s (1994: 79) claim that “there is no one-to-one correlation between identifiability or non-identifiability of a referent and grammatical definiteness or indefiniteness of the noun phrase designating that referent”. In fact, this chapter suggests multiple correspondences between the linguistic forms of nominal expressions and information values they represent in the texts. See Figure 5–6:

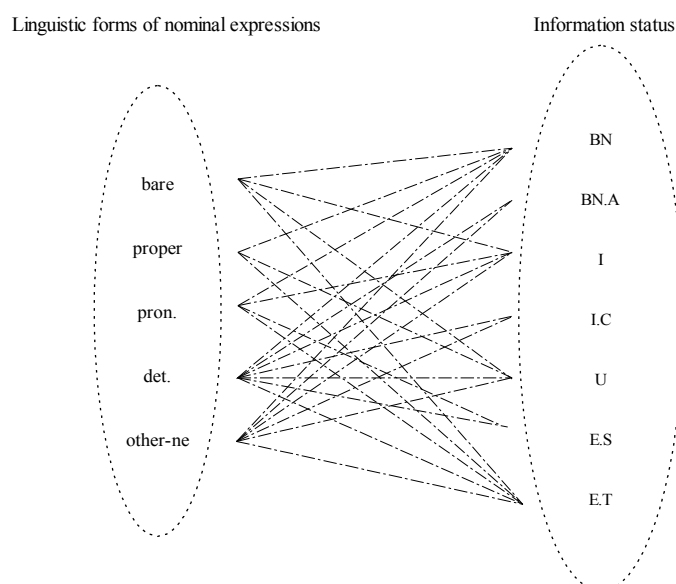


Figure 5–6: Multiple correspondences between linguistic forms of nominal expressions and information status in these texts

Second, some nominals could represent more types of information than others. For example, nominals with determiners can be used to represent all types of information, while proper expressions only represent Brand-new unanchored, Unused and Textually evoked information in the texts. The difference is closely related to their conceptual and referential features. Linguistic forms of nominal expressions to some extent are helpful to classify information status. Third, although different linguistic forms could represent

the same information status, their frequency distributions are clearly different in the texts. For example, all linguistic forms of nominals here could represent Brand-new unanchored information. However, the frequency of pronouns is the lowest and they are all indefinite.

More importantly, texts are different in the frequencies of varied linguistic forms of nominals expressing type of information. This finding suggests a close relation to the characteristics of their text types, with particular reference to communicative purposes and topical domains. For example, most proper expressions of the Travel guide texts are geographical names and represent Brand-new information as the texts aim to introduce places of interest that are unknown to the readers. Essay has the highest frequency of expressing Textually evoked information among the texts, since its topic is represented by a bare noun "*homosexuality*" which occurs more frequently than the other given discourse entities.

Besides, nominals expressing the same information status are different in informativeness based on the linguistic forms. Taking the Textually evoked entities as example, those represented by pronouns are more accessible, but they are less informative than nominals with determiners in terms of the description of the referents.

In conclusion, this chapter provided a detailed account of the relationship between linguistic forms of nominals and information status they represent in the texts. The next chapter presents the interplay between information status of nominals and thematic structures in the texts.

6 The interplay between information status of nominal expressions and thematic structures

In order to investigate the informational role of nominal expressions in English texts, Chapters 4 and 5 have already presented analysis results of the information distributions of nominal expressions in English texts and the relationship between linguistic forms of nominals and information status they represent. The main goal of this chapter is to examine the interplay between information status of nominal expressions and their thematic structures.

As pointed out in Chapter 2, the relationship between information status and clausal positions has been discussed by many scholars. However, most of them focused on the information only in the initial position of a sentence/clause²³ and did not provide empirical evidence to show whether texts have similarities and differences in terms of information distributions in both Theme and Rheme positions. With this in mind, this chapter provides a detailed account of these issues.

Specifically, the aim of the chapter is achieved by investigating information status of nominal expressions in Theme and Rheme positions. This investigation can be made from two perspectives. The first is to observe information distributions in Theme and Rheme positions by using nominals of the complete texts as raw frequencies. As previously presented in Chapter 4, information distributions of nominal expressions are different in the texts. This might influence the information distributions in their Themes and Rhemes. For example, the Travel guide texts have more Brand-new information than Essay, they are assumed to have higher frequencies of this category in Theme and Rheme positions. The second perspective is to observe information status in Theme and Rheme positions in isolation by using the number of nominals in each clausal position as raw frequencies. Information distributions of the Themes and Rhemes can reveal

²³ Here “the initial position” can either refer to Subject in the traditional grammar (Quirk *et al.* 1985: 123) or Theme in SFL (Halliday & Matthiessen 2014: 89).

different characteristics from those of the texts.

This chapter is organized as follows. Section 6.1 presents the relationship between the texts and their thematic structures in information distributions of nominal expressions. Section 6.2 explores information distributions of nominals in Theme and Rheme positions respectively. Finally, Section 6.3 brings the chapter to an end with a summary of the interplay between information status of nominal expressions and thematic structure in different texts.

6.1 Relationship between texts and thematic structures in information distributions of nominal expressions

This section aims to provide a detailed account of the relationship between the texts and their thematic structures in information distributions of nominal expressions. As presented in Chapter 4, information distributions of nominal expressions are different between the texts in terms of categories. These differences will in theory have an influence on the information distributions in Theme and Rheme positions of the texts, since the texts distribute all nominals to the two clausal positions. For example, the Travel guide and News report texts have more frequencies of nominals conveying Brand–new unanchored information than the others. They are expected to have more frequencies of nominals expressing the same information in both Theme and Rheme positions.

Exploring the relationship requires a comparison between the texts and their thematic structures in information distributions of all categories ranked in descending/ascending order. Since each category has the same procedure of comparison, it is not necessary to present all the details. Given the important role of Brand–new unanchored information in structuring the texts (see Section 4.1.1), this section will take the Brand–new unanchored distributions as an illustration.

Figure 6–1 displays the distributions of nominal expressions representing Brand–new unanchored information in the texts (BN-Txt) and their Theme (BN-Th) and

Rheme positions (BN-Rh). The horizontal axis of Figure 6–1 shows that although the BN-Txt frequencies become higher from Government document to Np6, the BN-Th and BN-Rh frequencies do not indicate the same tendencies. For example, Np2 has a higher BN-Txt frequency than Tg1, but its BN-Th frequency is unexpectedly lower. Np5 has a lower BN-Rh frequency than Np4 and is supposed to have a lower BN-Txt frequency. There is no complete positive relation between all texts and their Themes in Brand-new unanchored frequencies in Theme position. In Rheme positions, only four texts, namely Government document, Essay, Tg1 and Np6, show a positive relation. For instance, Figure 6–1 displays that Np6 has both the highest frequencies of both BN-Txt and BN-Rh. The above does not show a complete positive relation between texts and their thematic structures in distributions of nominals conveying Brand-new unanchored information.



Figure 6–1: Brand-new unanchored frequencies of the texts and their thematic structures

Figures illustrating the comparing results of the other categories are presented in Appendix B of this thesis. Table 6–1 illustrates the number of texts that show a positive relation to their thematic structures in information distributions of nominals. There are several main findings. First, four texts show a positive relation between the texts and

their Rhemes in Brand–new unanchored and Textually evoked information. Second, it is more likely to find the positive relation between texts and their Rhemes. Third, not many regular patterns were found between the texts and their thematic structures in information distributions of nominal expressions.

	BN	BN.A	I	I.C	U	E.S	E.T
Theme	0	0	3	0	0	0	3
Rheme	4	0	1	3	2	3	4

Table 6–1: The number of texts showing a positive relation

The relations are further examined by using average information distributions of the Travel guide and News report texts. There are four comparable texts in total, namely Travel guide-avg., News report-avg., Essay and Government document, examined under the comparison to their thematic structures in the information distributions of nominals. Table 6–2 shows the analysis results.

	BN	BN.A	I	I.C	U	E.S	E.T
Theme	0	2	2	2	0	0	4
Rheme	2	4	2	4	4	4	4

Table 6–2: The number of texts showing a positive relation in terms of average values of Travel guide and News report

The following patterns are identified from Table 6–2. First, Brand–new (both unanchored and anchored), and containing Inferrable illustrate the same pattern of showing the positive relation, with texts of two genres in Theme and four in Rheme. Second, Unused and Situationally evoked categories have the same pattern, with texts of four genres in Rheme showing the positive relation. Third, non–containing Inferrables have texts of two genres showing the positive relation in both Theme and Rheme positions. Furthermore, texts of four genres all show the positive relation under the category of Textually evoked.

In addition to the above patterns, Table 6–2 also displays that texts of four genres show the positive relation under five information categories in Rheme positions, namely Brand-new anchored, containing Inferrable, Unused, Situationally evoked and Textually evoked. This means the positive relation is more likely to be established between texts and their Rhemes in information distributions of nominal expressions. One possible explanation is the difference between thematic and rhematic zones. According to Gómez–González (2001: 132), the rhematic zone is usually larger in a text. More nominal expressions occur in Rheme position. It is more likely to see similarities between texts and their Rhemes in information distributions of nominals. Section 6.2 will provide a detailed account of the information distributions in Theme and Rheme positions.

6.2 Information distributions of nominal expressions in Theme and Rheme positions

This section presents the findings about information distributions of nominal expressions in Theme and Rheme positions. The analysis results presented in this section are based on the numbers of nominals in Theme and Rheme positions respectively, thus is different from Section 6.1 by using the number of nominal expressions in the texts as the raw frequency. Table 6–3 shows the frequency distributions of nominal expressions in Theme and Rheme positions per text.

Text	Theme		Rheme		Sum	
	Freq.	%	Freq.	%	Freq.	%
Tg1	63	33.33	126	66.67	189	100
Tg2	110	42.15	151	57.85	261	100
Np1	75	40.11	112	59.89	187	100
Np2	15	36.59	26	63.41	41	100
Np3	29	39.73	44	60.27	73	100
Np4	13	25	39	75	52	100
Np5	40	43.48	52	56.52	92	100
Np6	27	42.19	37	57.81	64	100
Gd	579	37.62	960	62.38	1539	100
Ey	227	39.07	354	60.93	581	100

Table 6–3: The distributions of nominal expressions in Theme and Rheme positions per text

Table 6–3 indicates that every text has more nominal expressions in Rheme than Theme. Taking into account the difference between Theme and Rheme, this section could reveal more characteristics of nominal expressions conveying type of information at the level of clause. This section is organized as follows. Section 6.2.1 focuses on how information values of nominal expressions are distributed in Theme positions of the texts. Subsequently, Section 6.2.1.1 and Section 6.2.1.2 further present the information distributions of nominals as Participants and Circumstances in Theme positions and show differences and similarities between the two different experiential roles in conveying information. Section 6.2.2 presents how information values of nominals are distributed in Rheme positions of the texts. Finally, Section 6.2.3 provides a contrastive analysis of Theme and Rheme in information distributions of nominal expressions.

6.2.1 Information distributions of Theme

This section presents information distributions of nominal expressions in Theme positions of all texts. Table 6–4 displays the distributions per text by percentages. The raw frequencies are presented as Table B1 in Appendix B of this thesis. The pink areas within Table 6–4 indicate that the texts have no frequency of certain categories of

information. For example, the complete text of Np3 does not have any containing Inferrables, thus cannot distribute any information of the category to Theme position.

Infor Text	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg1	17.46	0	33.33	0	14.29	1.59	33.33	100
Tg2	12.73	3.64	32.73	0.91	10.91	0	39.09	100
Np1	14.67	4	16	1.33	8	2.67	53.33	100
Np2	0	6.67	26.67	6.67	13.33	0	46.67	100
Np3	3.45	3.45	20.69	0	13.79	0	58.62	100
Np4	7.69	0	7.69	0	23.08	0	61.54	100
Np5	27.5	5	5	2.5	15	0	45	100
Np6	18.52	0	14.81	3.7	11.11	0	51.85	100
Gd	8.98	1.04	21.59	6.74	3.97	0.35	57.34	100
Ey	11.01	0.44	17.18	2.64	3.09	0	65.64	100

Table 6–4: Information distributions of nominals in Theme positions per text

Table 6–4 shows similarities between the ten texts in distributing information to Theme. Besides the zero values marked by the pink areas, the table also indicates that not every text includes frequencies of nominals expressing all types of information in Theme position. From a horizontal view, nominals conveying Textually evoked, non-containing Inferrable, Brand–new unanchored and Unused information occur more frequently. Apparently, nominals are more likely to express Textually evoked information in all texts, with the lowest frequency of 33.33% in Tg1 and the highest of 65.64% in Essay. With Tg1 and Tg2 as exceptions, the Textually evoked frequencies here are above 45% in texts. Compared with the frequencies of other categories, non-containing Inferrable is the second prominent feature of Theme in most texts, with Np4, Np5 and Np6 as exceptions. Np1 and Np6 are not clearly different in the frequency of nominals conveying non-containing Inferrables. However, in Np1, its frequency is only lower than Textually evoked, which is different from Np6 with non-containing Inferrables as its third most frequent information. Brand–new unanchored is also prominent in structuring Theme in texts. It is the third most frequent information in seven out of ten texts. Given the average values of Travel guide and News report texts

displayed in Table 6–5, Unused items play an important role in the Travel guide and News report texts, with values of 12.6% and 14.05% respectively. Combing the statistics given in Table 6–4, the average distributions of Travel guide texts, Essay and Government document have more nominals expressing unanchored information than Unused.

Infor Genre	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg-avr.	15.10	1.82	33.03	0.46	12.60	0.80	36.21	100
Np-avr.	11.97	3.19	15.14	2.37	14.05	0.45	52.84	100

Table 6–5: Average information distributions of Themes in the Travel guide and News report texts

The above results provide some empirical evidence to Halliday and Matthiessen’s (2014: 120) claim that “the Theme falls within the Given”. Since both Textually evoked and non-containing Inferrable nominals can be regarded as Given information based on Halliday’s (1967) interpretation of information values, a combination of the two categories of nominals contains the highest frequencies in each text. The information distributions of nominal expressions in Theme position are also consistent with the Scale of Information Distributions of nominals in text (see Section 4.2).

The frequency distributions in Table 6–4 also show differences between the texts. Compared with Essay and Government document, both Travel guide texts, Tg1 and Tg2, contain more information of Brand–new unanchored, non-containing Inferrable, Unused and Situationally evoked but less of the other categories in Theme position. News report texts have more frequencies of Unused nominals than Essay and Government document. They show clear differences under the same genre. For example, the largest gap of Brand–new unanchored frequencies between the News report texts is 27.5%, which is almost the same with the total frequencies of the Travel guide texts (30.19%) in the same category. Besides, Essay has a similar pattern with Government document in in information distributions in Theme position. Except for the category of

Textually evoked, gaps of the other categories are all less than 5%.

Taking into account the average values of the Travel guide and News report texts displayed by Table 6–5, the News report texts have similar frequencies of nominals expressing Brand–new unanchored and Inferrable information with Essay. However, the two are clearly different in the frequencies of nominals conveying Unused and Textually evoked information. The gaps between the two texts are 10.96% and 12.18% respectively. The News report and Government document texts are clearly different in information distributions in Theme position, except for the frequencies of nominals expressing Situationally evoked information, with a gap of 0.1%. The News report and Travel guide texts are similar in the distributions of Unused and Situationally evoked information but are different in the other categories. This suggests that the News report and Travel guide texts contain similar amount of information derived from things we take for granted and settings shared between the writers and readers.

The interaction between Theme and one type of information rather than another generates different characteristics. This finding reveals the communicative purposes of texts. When nominals expressing Brand–new information occur in Theme position, it contributes to an “unmotivated” thematic progression (Ventola and Mauranen 1991), since they do not contain anaphoric links to the preceding text. Besides, it also highlights the thematic function of orienting “the reader/listener to what is about to follow” (Fries 1995a: 5; cf. Ford 1993). Specifically, it does not serve as a sentence topic that needs further interpretation in the rhematic zone, but serves to introduce something new into the text. Examples from the texts of Travel guide and Government document are illustrated in (6–1):

(6–1) *Old customs* are still followed. (Tg1)

“*Old customs*” in (6–1) does not aim to convey information about the customs themselves, but to orient readers for the content to follow. This has also been discussed in terms of the topic acceptability scale by Lambrecht (1994) from the perspectives of

semantics and pragmatics. He (ibid: 169) indicates that it is acceptable for new referents to occur in the initial position of a sentence, if “the interpreter of the sentence does not feel the need to mentally identify the referent of the subject NP in order to assess the relevance of the information expressed in the predicate”. In (6–1), readers do not need to know what exactly old customs are in order to know whether they will be followed or not.

When Inferrable information occurs in Theme position, it is “less easily interpretable but still acceptable” (Lambrecht 1994: 166). Inferrable Themes contribute to a derived type of thematic progression (Daneš 1974), since they are related back to “a ‘hyper–Theme’²⁴ that establishes the topic for a longer stretch of text” (Williams 2007: 685). They also illustrate the results of selecting over all potential features of the hyper–Theme that might have been introduced explicitly to text. When a text has more Inferrables in the Theme position, it presents more specified features of the hyper–Themes to readers. The above is illustrated in example (6–2):

(6–2) This theory, which defined Urnings as males who turned to other males as sexual partners, was published in *twelve pamphlets by Ulrichs*, starting in 1864. ***The first***, Vindex, defended Urnings, while ***the second***, Inclusa, which followed shortly after, described the first scientific theory of homoerotic desire. (Ey)

In (6–2), nominals in bold italics “*the first*” and “*the second*” as ordinal numbers are typical indicators of the derived type of thematic progression. Their information values are inferred from the preceding expression “*twelve pamphlets by Ulrichs*”. The Inferrable Themes only specify two of the twelve pamphlets in the text.

Furthermore, Inferrable Themes reveal different types of background knowledge

²⁴ Hyper–theme is defined by Martin (1992: 437) as “an introductory sentence or group of sentences which is established to predict a particular pattern of interaction amongst strings, chains and Theme selection in the following sentences” and “what would be termed a Topic Sentence in school rhetoric”.

shared by the writers and readers. As indicated in Section 4.1.3, Inferrables express explicit inferences that are motivated by the mental models²⁵ of situational contexts (e.g. Johnson–Laird 1983; van Dijk and Kintsch 1983; van Dijk 1987). According to van Dijk (1987: 161), the situational models, also known as episodic models, are based on readers’ experiences stored in long–term memory. In the process of constructing texts, writers present their experiences as shared background knowledge as the point of departure of clauses, which are more familiar to the readers. Taking Tg1 and Tg2 as examples, both texts present nominal expressions expressing the situational categories of Visitor, Shop and Population as Inferrable Themes. For example, in Tg1, “*the impression of the visitor*” “*shopping*” and “*a population of nearly eight million*” can all reveal the situational categories.

When Evoked information occurs in Theme position, it forms the unmarked mapping between information status and thematic structure (Halliday and Matthiessen, 2014: 120). Using Evoked Themes has two main functions in the texts. First, in terms of cognitive preference (Chafe 1987; cf. Lambrecht 1994), the unmarked mapping is considered as the most acceptable pattern by starting a clause from information that the readers are already familiar with. It is also more helpful in interpreting the clauses, as it is “clearly addressee–friendly” (Williams 2007: 672). Second, it shows two types of thematic progressions (Daneš 1974). Evoked Themes establish either a constant progression by having the same reference with the Theme of the preceding clause, as “*the publishers in Monterrey*” and “*they*” in (6–3a), or a linear progression by referring back to the Rheme of the preceding clause, as the content shown only in italics and “*the Chinese people*” in (6–3b).

- (6–3) a. *The publishers in Monterrey* command no battalions, but **they** agreed to express their outrage with editorials in today’s editions. (Np1)
- b. Of Hong Kong’s population, 98 percent are *Chinese*. **The Chinese people**

²⁵ The term “model” is defined by van Dijk (1987: 161) as “a specific kind of knowledge structure in memory”.

have been described as hardworking and pragmatic, attitudes that have contributed to Hong Kong's success.

Based on the above data and the interaction between Theme and types of information expressed by nominal expressions, it can be concluded that information distributions in Theme position of the Travel guide texts indicate undue prominence in playing an orienting role and elaborating more features of the main topics with high frequencies of Brand-new and Inferrable information, but Themes in the Travel guide texts require more cognitive effort from the readers to process. Compared with the Travel guide texts, the information distributions of Theme in the News report texts play a less prominent role in presenting new information, but are easier to process for the readers with more Given information. The Themes of the News report texts do not present many specified features of the hyper-Themes by containing fewer Inferrables across texts of four genres. On the one hand, the Themes in Essay are similar to those of News report, which are also less likely to introduce new information or describe hyper-Themes with more details. They require less mental effort to process by containing the highest frequencies of Textually evoked information. The information distributions of Theme in Government document indicate that this text specifies fewer hyper-Themes than the Travel guide. Compared with the other texts, the Themes in Government document do not focus on introducing Brand-new information to readers.

To sum up, this section has analyzed similarities and differences between the texts selected from four genres in the information distributions of nominals in Theme position. Each text was found to show a tendency of containing more Given information that is derived from context and text. The Given information serves as coherent points, based on which more newsworthiness can be introduced to readers. The tendency to some extent reflects writers' purpose of orienting readers for successful text comprehension by making information in Theme position easy to process (cf. Martin 1992; Kilpert 2001; Thompson 2007). The frequencies of Textually evoked information of the texts provide evidence to support Halliday and Matthiessen's (2014: 120) claim

that “the Theme falls within the Given”. The choice of information status in Theme position is not only determined by recoverability, but also motivated by the informative nature of the written texts, in that it reflects one of the writers’ purposes to make full use of clausal spaces to introduce sufficient newsworthiness to the readers. Nominals mainly convey Brand–new unanchored, Textually evoked, non–containing Inferrable and Unused information in the texts. The selection of one information status rather than another especially from the four categories can be quantified as a determining factor in structuring different patterns of information distributions in Theme position. In addition, the interaction between Theme position and categories of information status reflects different characteristics of text types. Therefore, the information distributions of nominals in Theme position can be regarded as an observable indicator to distinguish text types.

6.2.1.1 Participant

This section focuses on the information distributions of Participants in Theme position of the texts. Table 6–6 displays the distributions per text. The shadow areas in pink within the table indicate that the complete texts have no frequency of certain categories of information and the blue areas indicate that the texts have no frequency of certain categories of information only in Theme position. For example, Tg1 does not have nominals conveying Brand–new anchored information in Theme position and it cannot distribute any of the information to nominals playing a Participant role in Theme.

Infor Text	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg1	9.52	0	26.98	0	9.52	1.59	23.81	71.42
Tg2	10	2.73	30	0.91	10.91	0	31.82	86.37
Np1	9.33	4	16	0	5.33	1.33	48	83.99
Np2	0	6.67	26.67	6.67	13.33	0	46.67	100
Np3	3.45	0	20.69	0	13.79	0	55.17	93.1
Np4	7.69	0	7.69	0	23.08	0	61.54	100
Np5	25	5	5	2.5	10	0	45	92.5
Np6	18.52	0	14.81	3.7	7.41	0	51.85	96.29
Gd	5.7	1.04	17.79	5.18	3.28	0.35	48.19	81.53
Ey	7.05	0.44	13.22	1.76	2.2	0	63	87.67

Table 6–6: Information distributions of nominals as Participants in Theme positions per text

Table 6–6 shows that not every text has information distributions of nominals as Participants in Theme position. For example, Np1 does not have frequencies of nominals expressing containing Inferrable information; and Np3 does not have frequencies of Brand–new anchored information. Both cases are marked by the green areas in Table 6–6. This is related to the information distributions of complete texts. As indicated in Chapter 4, frequencies of nominals conveying Brand-new anchored and containing Inferrable information are very low in all texts. It is not surprising we do not find any of them in Theme position functioning as participants. Another possible explanation for the green area is the textual characteristics of Np1 and Np3. This could be caused by a series of contextual variables, such as word order, communicative purpose, topic and so on, which cannot be further explored here.

Like the information distributions in Theme position, nominal expressions as Participants mainly express Textually evoked, non–containing Inferrable, Brand–new unanchored and Unused information here, while those of the remaining categories play a less important role. The finding is not surprising, as most nominals play a Participant role, which is “the most common type of Theme” (Halliday and Matthiessen 2014: 91). Among all categories of information, the frequencies of Textually evoked nominals are the highest in all texts, which also provides evidence to support Fries’ (1995a: 5) claim

that “experiential Themes are typically Given information”.

The statistics in Table 6–6 also illustrate differences and similarities from the information distributions of nominals in Theme position. Table 6–4 indicates seven of the ten texts have higher frequencies of nominals expressing Brand–new unanchored information than those expressing Unused information. But the number of texts decreases to five when nominals are mapped onto the role of Participant. The difference shows that compared with Brand–new unanchored information, nominals expressing Unused information are more likely to function as Participants in Theme position. The texts of Travel guide, Government document and Essay indicate almost the same tendencies in distributing categories of information to their Themes and Participants.

Infor Genre	BN (%)	BN.A (%)	I (%)	IC (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg-avr.	9.76	1.37	28.49	0.46	10.22	0.80	27.82	78.92
Np-avr.	10.67	1.95	15.14	2.15	12.16	0.22	51.37	93.66

Table 6–7: Average information distributions of Participants in Theme position in Travel guide and News report texts

Table 6–7 displays the average values of Travel guide and News report texts in information distributions of nominals functioning as Participants in Theme position. In terms of the average values of the News report texts, information distributions of nominals as Participants are not completely consistent with those in Theme position (see Table 6–5). Table 6–7 shows that the average value of News report texts is the highest when nominals convey Brand–new information across texts of four genres, both unanchored and anchored. This also indicates the highest degree of newsworthiness, which is different from the information distributions in Theme position with the average value of Travel guide text being more newsworthy. The News report and Essay texts are similar in the frequencies of Inferrables when nominals function as Participants and thematic elements. The average value of News report text has more Textually evoked Participants than Government document and the two have a smaller difference of

Inferrables compared with the frequencies in Theme position.

Given the typical role of Participant in Theme position, the occurrence of a referent is expected to be accessible to the readers. However, “if a constituent has a referent which is clearly not accessible in the context, in particular one that is unidentifiable, and if the sentence is nevertheless of normal acceptability, there is a good chance that the constituent is not a topic expression in the sentence” (Lambrecht 1994: 168). This issue has already been discussed in Section 6.2.1.

As the typical experiential role of Theme, the interaction between Participant and information status displays the same discourse functions that have been presented in Section 6.2.1. Therefore, the discussion here is more concerned with why the new Participants in Theme position are more highlighted in the average value of News report than texts of other genres. Examples are given in (6–4):

- (6–4) a. *A boy* was run over by a car. (Lambrecht, 1994: 169)
b. *One study* places the losses resulting from fraud on mortgage loans made between 2005 and 2007 at \$112 billion. (Gd)
c. *An ambitious expansion* has left Magna with excess capacity and a heavy debt load as the automotive industry enters a downturn. (Np6)

Like the referent in (6–1a), “*A boy*” “*one study*” and “*an ambitious expansion*” are all thematic Participants that do not form aboutness relation with their predicates as topics. Besides the features already mentioned, the three nominal expressions demonstrate a similar sentential pattern. The pattern is defined as the “Event–reporting type” and is closely related to text types (Lambrecht 1994: 133). Lambrecht (ibid: 196) also states that sentences belonging to this type basically answer the question “what happened”, so the whole proposition is construed as focus. The “Event–reporting type” is frequently used in newspaper writing.

Relating the “Event–reporting type” to the unidentifiable Participant in Theme position, it is not considered as a coincidence that the News report texts contain the

highest frequency among the four genres. Biber and Conrad (2014: 118) indicate that “newspapers focus more on current newsworthy events and have more emphasis on simple reporting”. The News report texts are thus structured in a way to answer the general question “what happened”. An unidentifiable Participant in Theme position clearly serves as an indicator of the event–reporting sentences. In comparison, the texts of the Essay and Government document genres aim to analyze and interpret certain phenomena, in which case reporting is not an emphasis. Based on the statistics given by Table 6–6 and Table 6–7, nominals in News report texts are more likely to be Participants in Theme than those in other texts. In comparison, texts of Travel guide, Essay and Government document tend to distribute more nominals to Circumstances in Theme position. This will be discussed in detail in Section 6.2.1.2.

6.2.1.2 Circumstance

This section concerns the information distributions of nominal expressions playing a circumstantial role in Theme position. Table 6–8 displays the distributions per text. The shadow areas in pink indicate that the complete texts have no frequency of certain categories of information, and the blue areas indicate that the texts have no frequency of certain categories of information in Theme position. The green areas show all zero values of nominals expressing categories of information as Circumstance in Theme position.

Table 6–8 shows that not every text has nominals that functions as Circumstance in Theme position. For example, no nominals play a circumstantial role in Np2 and Np4. The finding is not surprising, since circumstantial Theme is not considered to be a typical Theme in English declarative clauses (see Halliday and Matthiessen 2014: 92). Although nominals of the other texts play a circumstantial role in Theme position, they cannot express all categories of information.

Infor Text	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg1	7.94	0	6.35	0	4.76	0	9.52	28.57
Tg2	2.73	0.91	2.73	0	0	0	7.27	13.64
Np1	5.33	0	0	1.33	2.67	1.33	5.33	15.99
Np2	0	0	0	0	0	0	0	0
Np3	0	3.45	0	0	0	0	3.45	6.9
Np4	0	0	0	0	0	0	0	0
Np5	2.5	0	0	0	5	0	0	7.5
Np6	0	0	0	0	3.7	0	0	3.7
Gd	3.28	0	3.8	1.55	0.69	0	9.15	18.47
Ey	3.96	0	3.96	0.88	0.88	0	2.64	12.32

Table 6–8: Information distributions of nominals as Circumstances in Theme positions per text

Furthermore, most of the green areas were found in the texts of News report with zero values of nominals expressing type of information as circumstances in Theme position. Nominals are more likely to express Textually evoked and Brand–new unanchored information here. News report texts have no nominals expressing non–containing Inferrables. Based on the frequencies of zero values displayed by Table 6–8, the information categories for nominals functioning as Circumstances in Theme position tends to develop as: E.S < BN.A < I.C < U < BN < E.T. Most texts show that nominals as Circumstances in Theme position, though structurally marked, are more likely to convey Textually evoked information.

The analysis now shifts the focus to the differences between the texts. Table 6–8 displays that Tg1 has the highest frequency of nominals expressing Brand–new unanchored and Textually evoked information across the texts. Compared with Tg2, Tg1 is more likely to have nominals conveying all categories of information as Circumstances in Theme. Although Tg1 has a lower frequency of nominals expressing Brand–new anchored information, it has a small gap of only 0.91% in relation to Tg2. The News report texts do not indicate any clear patterns in having frequencies of nominals conveying information in Theme position, since not many texts have the frequencies. In addition, the Travel guide texts are clearly different from the texts of

Government document and Essay in the frequencies of nominals conveying categories of information.

Table 6–9 displays the average values of information distributions of nominals as Circumstances in Theme positions in the Travel guide and News report texts. The average value of the Travel guide texts shows a higher frequency of nominals conveying information of Brand–new, both unanchored and anchored, and non–containing Inferrable than Government document and Essay. Compared with Government document, Essay contains almost the same frequencies of nominals conveying Brand–new unanchored, non–containing and containing Inferrable information, but shows a difference of 6.61% in Textually evoked. Both Essay and Government document texts have no nominals expressing Brand–new anchored and Situationally evoked information when they play a circumstantial role in Theme position.

Infor Genre	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg-avr.	5.34	0.46	4.54	0	2.38	0	8.40	21.12
Np-avr.	1.31	0.58	0	0.22	1.06	0.22	1.46	4.85

Table 6–9: Average information distributions of Circumstances in the Travel guide and News report texts

To compare the News report with the other three genres in information distributions of Circumstances in Theme position, it is feasible to take the average values displayed by Table 6–9 into account. Apparently, the News report text has the lowest frequency of nominals expressing Brand–new unanchored, non–containing Inferrable and Textually evoked information. It is not clearly different from the average value of Travel guide texts with a little bit more Brand–new anchored and containing Inferrable information but less Brand–new unanchored, non–containing Inferrable, Textually evoked and Unused information. Furthermore, the News report texts are not similar with Government document and Essay in the frequencies displayed by Table 6–8 and Table 6–9, except for the low frequencies of Textually evoked information in News

report and Essay texts. As a whole, News report texts are less likely to have nominals conveying all categories of information when they play a circumstantial role in Theme position.

The interaction of Theme with one type of information rather than another generates different characteristics. Brand–new unanchored information in Theme position foregrounds the orienting function and reveals characteristics of text types.

- (6–5) a. *On **1 July, 1997*** the British Crown Colony of Hong Kong reverted to Chinese sovereignty as a Special Administrative Region of the People’s Republic of China. (Tg1)
- b. *By **2005***, the 10 largest U.S. commercial banks held 55% of the industry’s assets, more than double the level held in 1990. (Gd)
- c. *In **1935***, he furthered his claims when he wrote a now famous “Letter to an American Mother” of a homosexual... (Ey)
- d. *Over **the past 12 years***, at least 40 journalists have died there. (Np1)
- e. *Up **the hill*** from Temple Bar are Dublin Castle and Christ Church Cathedral. (Tg2)

From (6–5a) to (6–5d), expressions shown in bold italics, “*1 July, 1997*”, “*2005*”, “*1935*”, and “*the past 12 years*”, all refer to time and convey Brand–new unanchored information in the texts. Nominals playing a circumstantial role in Theme position indicate a strong preference for temporal references when they convey Brand–new unanchored information in the texts. Each nominal serves as a growth–point so that new information can be introduced to the texts afterwards.

Unlike the former instances, “*the hill*” in (6–5e) refers to spatial location in Tg2. Fries (1995a: 10) presented similar findings when exploring the experiential content of Themes in tour guide texts, claiming that “because of the purposes of tour guide texts, references to spatial location will play an orienting role. The orienting role will lead authors of tour guide texts to place a greater proportion of their references to spatial

location within the Themes of the text” (cf. Neumann 2003: 221). Halliday and Matthiessen (2014: 100; cf. Matthiessen 1992: 60–61; 1995: 37–39) have a similar comment that “locative Adjuncts and Complements are highly motivated as Themes since they enable speakers to guide their addressees in the development of the verbal map”, with “*Beyond the main complex is a lovely stream that bubbles under a wooden bridge*” as an example. Although previous literature did not further investigate information status of nominals that play a circumstantial role in Theme position, it has pointed out the particular function of orienting spatial location in texts of travel guide.

As indicated in Section 6.2.1, when nominal expressions convey Inferrable information in Theme position, they elaborate some discourse entities that occur in the preceding text and represent some activated categories of certain mental frames stored in our long-term memory. This feature is also displayed by nominals playing a circumstantial role in Theme positions. For instance:

- (6–6) a. Today Hong Kong remains a capitalist enclave with its laws and rights intact, and China has promised that Hong Kong will continue in this fashion for at least 50 years... *Around **the time of the transition*** there was much speculation about how things would change. (Tg1)
- b. New construction is everywhere, the streets buzz, traffic is increasingly congested, and *in **the frenetic pace of rush hour*** everyone in Dublin seems intent on changing places with everyone else. (Tg2)

In example (6–6), both nominal expressions “*the time of the transition*” and “*the frenetic pace of rush hour*” play a circumstantial role in Theme position and convey Inferrable information classified from the preceding clauses. The classification is straightforward, as the Inferrables have a short distance to their triggers, which make them dependent on the specific content of the texts. In contrast, nominal expressions in example (6–7) are not categorized from specific nominals or relations in the preceding texts, as their triggers are not instantiated as lexico-grammatical resources. Instead,

they are assumed to be part of the mental frames stored in our long-term memory and are activated under certain contextual grounds.

- (6–7) a. Now *to our major findings and conclusions*, which are based on the facts contained in this report: they are offered with the hope that lessons may be learned to help avoid future catastrophe. (Gd)
- b. However, *for the purposes of this paper*, homosexuality will be looked at solely in terms of men. (Ey)

In example (6–7), expressions like “*our major findings and conclusions*” and “*the purposes of this paper*” are the instantiation of typical categories of Purpose, Finding and Conclusion in the situational model of argumentations. They are expected to be expressed explicitly in the texts. Apart from texts from Government document and Essay, the Travel guide and News report texts do not contain such Inferreds playing a circumstantial role in Theme position. Besides, the prepositions “*for*” and “*to*” in example (6–7) also serve as clear signals to remind readers that they should focus on the following content.

The most common type of interaction between Theme position and information values is realized by nominals conveying Evoked information. Those nominals sometimes are embedded within certain phrases and play a circumstantial role as a whole in Theme position. They have different discourse functions in the texts. The first function of indicating arguments is illustrated by the examples as follows:

- (6–8) a. *In this report*, we detail the events of the crisis. But a simple summary, as we see it, is useful at the outset. (Gd)
- b. *In our inquiry*, we found dramatic breakdowns of corporate governance, profound lapses in regulatory oversight, and near fatal flaws in our financial system. (Gd)

In example (6–8), the expressions shown in bold italics “*this report*” and “*our inquiry*” share two things. The first is that they play a circumstantial role in Theme position when conveying Textually evoked information, and the second is that both refer to the texts themselves rather than specific nominals created by the writers like those illustrated in example (6–7).

The two expressions in example (6–8) actually are the self-referential Themes discussed by Francis (1987, 1989) (also see Halliday and Hasan 1976: 52). When distinguishing News and Editorials in thematic selection, she (Francis 1989) points out that expressions like *in this context*, *in that sense*, *under such circumstance* “function as labellers of chunks of previous text, and as such are an important feature of argument”. Although she did not explore the information status of the self-referential Themes, “*in this report*” is the same with “*in this context*” as the labellers of chunks of text from the angle of discourse function. Therefore, the expressions in example (6–8) also reveal the feature of argument.

As an indicator of arguments, the nominals “*this report*” and “*our inquiry*” are particularly noticeable. They occurred five and two times respectively, while most nominals playing a circumstantial role in Theme position only occurred once in the text. More surprisingly, such instances were only found in Government document. This may suggest that self-referential Themes do not occur randomly. Other things being equal, Government document reflects a salient feature of argument.

- (6–9) a. With a population of nearly eight million and a total area of just over 1,095 square km (423 square miles), *housing is one of Hong Kong’s perennial nightmares*. To alleviate ***the problem***, the government has become the city’s major landlord with the construction of massive apartment blocks that, though they have every modern facility, average only 9 square m (100 square ft) in size. (Tg1)
- b. Dr. Evelyn Hooker, a heterosexual psychologist, conducted *a ground-breaking study in the mid-1950s that went along similar reasoning as*

*Freud. In **this courageous experiment**, Hooker compared the psychological profiles of sixty men, half homosexual and half heterosexual. (Ey)*

- c. *And **just a month before Lehman's collapse, the Federal Reserve Bank of New York was still seeking information on the exposures created by Lehman's more than 900,000 derivatives contracts.** In addition, the government's inconsistent handling of major financial institutions during the crisis — the decision to rescue Bear Stearns and then to place Fannie Mae and Freddie Mac into conservatorship, followed by its decision not to save Lehman Brothers and then to save AIG — increased uncertainty and panic in the market. In making **these observations**, we deeply respect and appreciate the efforts made by Secretary Paulson, Chairman Bernanke, and Timothy Geithner... (Gd)*
- d. In some cases, they were obligated to use them, or regulatory capital standards were hinged on them. This crisis could not have happened without the rating agencies. Their ratings helped the market soar and their downgrades through 2007 and 2008 wreaked havoc across markets and firms... And you will see that *without **the active participation of the rating agencies***, the market for mortgage-related securities could not have been what it became. (Gd)

Besides indicating arguments, the second discourse function is the degree of cognitive complexity or abstractness when nominals convey Evoked information as Circumstances in Theme position. In example (6–9), four nominal expressions shown in bold italics have one thing in common: each represents chunks of information presented in the preceding text. However, they are different in the amount of information they integrate. In order to process these nominals, readers are required to form different integrated mental representations of their antecedents (also see Schmid 2000: 370). Abstract relations are built in the integration, which “are cognitively far

more complex than the representations of participants and temporal stages” (Schmid 2000: 373). The more chunks of information are integrated in the process, the more condensed an expression becomes. The cognitive complexity and abstractness of the instances in example (6–9) are regarded as important features for developing arguments (e.g. Francis 1989; Schmid 2000: 379).

In terms of cognitive complexity, “*these observations*” in (6–9c) and “*the active participation of the rating agencies*” in (6–9d) are more complex than “*this courageous experiment*” (6–9b) and “*the problem*” (6–9a), since the antecedents of the first two nominals are more complex with propositions represented by clauses rather than phrases. Besides, such instances were only found once in the texts of Travel guide and Essay, which is less frequent than those displayed by (6–9c) and (6–9d) in Government document. Among all texts, Government document contains the highest frequency of such nominals as Circumstances. Other things being equal, it displays a more salient feature of argument in Theme position.

- (6–10) a. Here we discuss three: *capital availability and excess liquidity, the role of Fannie Mae and Freddie Mac (the GSEs), and government housing policy*. First, as to ***the matter of excess liquidity***: in our report, we outline monetary policies and capital flows during the years leading up to the crisis. (Gd)
- b. As it happens, *the four countries cited, Colombia, Cuba, Panama and Nicaragua*, are not only where the press is under greatest attack but also are linked by the drug trade and left–wing politics. Noriega is close to Castro and may once have been his agent. Sandinistas Thomas Borge and the Ortega brothers are Castro proteges; he backed their takeover of Nicaragua. In ***Colombia***, the drug–financed guerrillas trying to seize the country and destroy democracy include M–19, which Castro has clearly backed. (Np1)

The third function is to avoid confusion when nominals convey Evoked information as Circumstance in Theme position. Classifying information status of “*the matter of excess liquidity*” and “*Colombia*” in example (6–10) is easy, since both have a short textual distance to their antecedents and they are repeated with full linguistic forms rather than pronouns. As each expression is selected from a list that contains several nominals, fully repeated forms indicate clear references, thus can avoid readers’ confusion in comprehending text.

Instances in example (6–10) were not found in Essay and Travel guide. However, to the best of our knowledge, previous research provides limited evidence to support that such instances are particularly sensitive to text types. Unlike the above-mentioned functions, selecting one nominal from a list is not closely related to the semantic content of text types. This may occur in any text when the writer aims to avoid confusion of references for readers. A larger scale of corpus will be needed to explore whether such nominals in Theme position only occur in particular text types.

- (6–11) a. Whole cities have been created in *the New Territories*, although the unimaginative architecture of these towns has been criticized..... Beyond, in *the New Territories*, are a mixture of high-rise suburban towns, ancient sites and walled villages, country parks, and farms with ducks and fish ponds. (Tg1)
- b. The slaughter in Colombia was very much on the minds of 450 editors and publishers from Latin America, the United States, the Caribbean and Canada attending the 45th general assembly of the Inter-American Press Association in *Monterrey*, Mexico, this week..... At *Monterrey*, publisher Luis Gabriel Cano, although shaken by the murders, issued a statement saying: “We will not cease our fight against drug trafficking. They want to terrify the press and in particular *El Espectador* because it has always been a torchbearer in this war.” (Np1)

The last discourse function discussed here is: **Present Given as New**, when nominals convey Textually evoked information with a circumstantial role in Theme position. Unlike the instances shown in example (6–10), “*the New Territories*” and “*Monterrey*” in example (6–11) as fully repeated forms have longer textual distances to their antecedents, with at least six sentences in between, which is not common for Textually evoked entities. Furthermore, without anaphoric links, the expressions are not closely related to the preceding texts, but serve as starting–points to introduce new information. They do not contribute much to topic continuity. This is an unusual way of presenting recoverable information and is contrary to the strategy of presenting New as Given proposed by Peng (2014). It is termed as **Present Given as New** in the current thesis based on the function in the texts.

In view of Peng’s (2014) claim that the pattern of presenting New as Given is a typical feature of classical English novels, it is worthwhile to explore whether the pattern of presenting Given as New correlates with text types. Taking a close look at the corpus, no instances were found in Essay and Government document. However, it is not appropriate to make a general claim that presenting Given as New is a feature of text types. The texts selected from four genres in the current thesis cannot be claimed representative and further explorations thus will be needed with a larger scale of corpus.

Based on the above, it can be concluded that the Travel guide texts are prominent in playing an orienting role and have the highest frequencies of nominals that refer to spatial location and time when conveying Brand–new unanchored information; both Essay and Government document texts are noticeable in revealing argumentative features by using typical categories of arguments that convey non–containing Inferrable information; Government document is particularly noticeable in the feature of arguments, since it has the highest frequency of self–referential and cognitively complex nominals expressing conveying Textually evoked information across the texts when they play a circumstantial role in Theme position; the News report texts did not reveal many features with the lowest frequencies of Textually evoked, non–containing Inferrable and Brand–new unanchored information.

To sum up, this section has presented findings of information distributions of nominals as Circumstances in Theme position. As structurally marked Themes in the texts, the findings indicate different characteristics in conveying categories of information, which are related to text types. More interestingly, most nominals convey Given information in the texts, though structurally marked. This provides some evidence to support Halliday and Matthiessen’s (2014: 120) claims that “the Theme falls within the Given”.

6.2.2 Information distributions of Rheme

Section 6.2.1 has so far explored the information distributions of nominal expressions in Theme position. This section aims to explore the information distributions of nominal expressions in Rheme position per text. Table 6–10 displays these distributions. The pink areas within the table indicate that the complete texts have no frequency of certain categories of information. The blue areas show that some texts have no frequency of a certain type of information in Rheme position, which all belong to the genre of News report.

Infor Genre	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg1	38.1	8.73	24.6	2.38	7.14	0	19.05	100
Tg2	36.42	1.99	16.56	3.31	16.56	1.32	23.84	100
Np1	33.93	8.04	19.64	0	13.39	1.79	23.21	100
Np2	46.15	0	15.38	0	7.69	3.85	26.92	100
Np3	34.09	4.55	18.18	0	9.09	0	34.09	100
Np4	43.59	5.13	7.69	7.69	12.82	0	23.08	100
Np5	48.08	1.92	11.54	0	1.92	0	36.54	100
Np6	62.16	2.7	8.11	0	2.7	0	24.32	100
Gd	22.71	2.08	28.65	4.27	3.54	0.21	38.54	100
Ey	24.58	1.98	25.14	3.11	5.08	0.56	39.55	100

Table 6–10: Information distributions of nominal expressions in Rheme positions per text

Table 6–10 illustrates similarities between the texts. Not every text has frequencies of nominal expressions conveying all categories of information in Rheme position. Among all the categories, containing Inferrable is less likely to occur in Rheme position. It has four blue areas displayed by Table 6–10, which is more than other categories. Although there are more zero values of Situationally evoked information (shaded areas in pink), they are caused by the distribution of complete texts rather than the Rheme position. As indicated in Section 4.1.6, Np3, Np4, Np5 and Np6 do not have any Situationally evoked instances in complete texts.

Most nominals express Brand–new unanchored, Textually evoked, non–containing Inferrable and Unused information, while those conveying the remaining categories of information are less likely to occur in Rheme position. Most texts have the highest frequency of nominals expressing Brand–new unanchored information. For example, Np6 has a frequency of 62.16% in the category of Brand–new unanchored. This finding provides some empirical evidence to Halliday and Matthiessen’s claim (2014: 120) that “the New falls within the Rheme”. Nominals expressing Textually evoked information also play an important role in Rheme position. Surprisingly, the Textually evoked frequencies are higher than the Brand–new unanchored in Essay and Government document, which indicates that the two texts do not focusing on bringing more newsworthiness to their readers. Essay is similar to Government document in the frequencies of nominals expressing all categories of information, with the biggest gap of 3.51% in non–containing Inferrable.

Table 6–10 also shows differences between the texts. First, Essay and Government document have less Brand–new unanchored but more Textually evoked and non–containing Inferrable information than the texts of Travel guide and News report. The finding is not surprising based on varied communicative purposes of the texts. As indicated before (see Section 6.2.1), the texts of Travel guide and News report aim to describe unknown things that are newsworthy to readers, while Essay and Government document focus on providing causes of certain phenomena. Second, texts under the same genre are different in distributing Brand–new anchored, Unused and both

containing and non-containing Inferrable information to to their Rhemes. For instance, in the News report texts, Np6 has a gap of 8.04% relative to Np2 in the frequency of nominals expressing Brand–new anchored information, containing a higher frequency than Np2, Np5 and Np6.

Infor Genre	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg-avr.	37.26	5.36	20.58	2.85	11.85	0.66	21.45	100
Np-avr.	44.67	3.72	13.42	1.28	7.94	0.94	28.03	100

Table 6–11: Average information distributions of Themes in the Travel guide and News report texts

Table 6–11 displays the average information distributions of Rhemes in the Travel guide and News report texts. The statistics in Table 6–11 show that the Travel guide text has the lowest frequency of nominals expressing Textually evoked information in Rheme position, while the News report text has more Brand–new unanchored information. Compared with the News report, the Travel guide has higher frequencies of nominals expressing Brand–new anchored, Inferrable of both subtypes and Unused information but lower frequencies of the remaining categories. Both the Travel guide and News report texts have higher frequencies of nominals expressing Brand–new anchored, Unused and Situationally evoked information than Essay and Government document texts (see Table 6–10), which reveals the characteristic of Travel guide and News report texts, namely newsworthy. As a whole, texts of the four genres display three patterns of distributing information to Rheme position. The first is represented by the Travel guide text; the second is displayed by the News report; and the third is found in the texts of Government document and Essay.

The interaction between nominals expressing categories of information and Rheme position in the texts generates characteristics that are closely related to text types. Example (6–12) displays nominals expressing Brand–new unanchored information in Rheme position:

- (6–12) a. On the Toronto Stock Exchange yesterday, Magna shares closed up **37.5 Canadian cents** to **C\$9.625**. (Np6)
- b. From 1978 to 2007, the amount of debt held by the financial sector soared from **\$3 trillion** to **\$36 trillion**, more than doubling as a share of gross domestic product. (Gd)

In example (6–12), both expressions shown in bold italics refer to the amount of money. Such instances were only found in the News report and Government document texts. This feature actually reflects the characteristic of particular text types. Taking a closer look at Np6 and the Government document text, both of their main topics are concerned with finance and economics, in which numbers are included as an important part. Previous studies show that numbers are “predominantly meant as signals of precision and hence of truthfulness” (van Dijk 1988: 88) and can therefore increase the credibility of messages. Numbers play an important role in both news/journalism (van Dijk 1988; Koetsenruijter 2011; McConway 2016) and political discourse (Lawson and Lovatt 2020; Steensberg 2021). They are described as a “number game” from a rhetorical perspective, on the basis that they make persuasive effects through the symbolic potency (see Merriam 1990). Using numbers is considered as a strategy of argumentation.

- (6–13) a. Hirschfeld also founded the Scientific Humanitarian committee, which was mostly homosexual, in **1897**. The committee published many books and other forms of literature, which gave Hirschfeld a great amount of prestige in his field. (Ey)
- b. By 2005, the 10 largest U.S. commercial banks held 55% of the industry’s assets, more than double the level held in **1990**. On the eve of the crisis in 2006, financial sector profits constituted 27% of all corporate profits in the United States, up from 15% in 1980. (Gd)

- c. Iraq says it has done so and has barred inspectors since *late 1998*. At least one other congressman has visited Iraq. Energy Secretary Bill Richardson went to Baghdad in 1995 while a representative for New Mexico. (Np3)

In example (6–13), expressions shown in bold italics all convey Brand–new unanchored information by referring to specific years. As nominals embedded within prepositional phrases, they play a circumstantial role in the clauses. Based on the classificatory nature of temporal expressions, instances displayed by example (6–13) also serve as ending points to indicate clear–cut boundaries between the current and following clauses. The following content clearly does not elaborate the former propositions. Furthermore, instances in example (6–13) only occur once in the texts, which play no role in topic continuity. The above findings are different from the characteristics of temporal adverbials in Theme position previous literature has discussed (e.g. Firbas 1972; Chafe 1976; Givón 1983; Thompson 1985; Virtanen 1992; Lambrecht 1994; Goutsos 1996; Hasselgård 2004).

Besides, instances displayed by example (6–13) support Hasselgård’s statement (2010: 129) that temporal adjuncts also “provide a specification of the action (Firbas 1986: 49; Horová 1976: 118), in which case they carry a higher degree of communicative dynamism (Firbas 1986: 50f) and are more aptly placed in end position”. Instances in (6–13) were only found in texts of Essay, News report and Government document but not in the Travel guide. This might indicate certain difference between text types. Hasselgård (2010: 207) indicates that the frequency of temporal adverbials varies across text types, from a low frequency of 38.3% in academic writing to a high frequency of 70.8% in news. Given the above findings, it is reasonable to claim that in the News report texts, the highest frequency of nominals conveying Brand–new unanchored information can be partially attributed to a high frequency of temporal adverbials in Rheme position.

- (6–14) a. This busy, modern European city sits on ***a thousand years of history*** — *history* is present everywhere, from *elegant Merrion Square* to *the bullet holes on the General Post Office*... *Dublin excels in packaging its past for the visitor. You can view artifacts from the Bronze Age, trace the history of the Easter Rising, or revisit Leopold Bloom’s odyssey in Ulysses.* (Tg2)
- b. It’s also a city of the imagination, reinvented and reappraised in ***the literature of its exiles***..... *Literature has always flourished in Dublin, the only city to have produced three Nobel Prize winners for literature — Yeats, Shaw, and Beckett. Joyce, the high priest of literary Modernism, imagined and interpreted Dublin for the world in Ulysses. However, sometimes it seems that the city produced artists of this stature by accident, even against its will. Beckett and Joyce, among others, had to leave their homeland to understand it — and to be understood.*
- (Tg2)

Example (6–14) is selected from a Travel guide text with *Dublin* as its heading. This text aims to introduce the city. Both expressions shown in bold italics are the last elements of the clauses and convey Brand–new information in Rheme position. “*A thousand years of history*” in (6–14a) suggests the following is about the history of the city. “*The literature of its exiles*” in (6–14b) is concerned with the culture of Dublin. Both nominals are elaborated by the following propositions. They establish textual links which involve more than one proposition. More importantly, they function as the elaborations of “*Dublin*”, which reflects the communicative purpose of Tg2. Instances in (6–14) provide some evidence to Fries’ (1995: 6) hypothesis that the content of the N–Rheme (the last element of a clause with New information) correlates with the goal of a text.

Now the analysis moves to the interaction between nominals conveying Inferrable

information and Rheme position in the texts. The instances are illustrated by example (6–15):

- (6–15) a. Timex is a major U.S. producer and seller of watches, including low-priced battery-operated watches assembled in the Philippines and *other developing nations* covered by the U.S. tariff preferences. (Np4)
- b. It's anyone's guess what may happen in the future, but for now Hong Kong bristles with energy and ambition, and for the visitor, this beautiful city with *its contrasts and variety* is an exhilarating experience. (Tg1)
- c. He went on to explain that while his nurture did not impact *his sexual orientation*, it did affect his view of it. (Ey)

The nominals shown in bold italics in example (6–15) convey Inferrable information in the texts. “*Other developing nations*” in (6–15a) and “*its contrasts and variety*” in (6–15b) occur with prepositions as embedded nominals of the clauses. The sentence in (6–15c) aims to reject the view that nurture determines homosexuality, in which “*his sexual orientation*” is not the information focus. The three nominals conveying Inferrable information in Rheme position are similar to some extent, since they do not express the core information of the clauses or bring in newsworthiness to the texts.

Finally, the analysis focuses on the interaction of nominal expressions conveying Evoked information and Rheme position. Example (6–16) displays the cases in point:

- (6–16) a. In an interview in 1903, *he* professed his beliefs: “I am... of the firm conviction that homosexuals must not be treated as sick people... Homosexual persons are not sick. They also do not belong in a court of law!” (Ey)
- b. Our task was first to determine what happened and how *it* happened so

- that we could understand why it happened. (Gd)
- c. The universities attract students from all over the world, and this influx helps to make ***Dublin*** a busy, buzzing international city. (Tg2)
- d. The White House said President Bush has approved duty-free treatment for imports of certain types of watches that aren't produced in "significant quantities" in ***the U.S., the Virgin Islands*** and other U.S. possessions. (Np4)

In example (6-16), expressions shown in bold italics illustrate four types of nominals conveying Evoked information in Rheme position. In (6-16a), "*he*" plays a participant role and occurs right after the marked Theme expressed by a prepositional phrase, which is typical Given information in Rheme position in declarative clauses; in (6-15b), "*it*" occurs in an interrogative clause after the interrogative adverb "*how*" that requests particular information; in (6-16c), "*Dublin*" is a complement of the clause; and in (6-16d), "*the U.S., the Virgin Islands*" are embedded within a prepositional phrase that functions as the locative adjunct of the clause. The nominals here contribute to cohesion by repeating some information in the preceding texts.

To sum up, this section has presented similarities and differences between the texts selected from four genres in information distributions of nominals in Rheme position. The Travel guide and News report texts show a tendency of containing more Brand-new unanchored information. This tendency indicates two particularly important points. First it provides evidence to support Halliday and Matthiessen's (2014: 120) claim that "the New falls within the Rheme", and second it reflects the communicative purposes of Travel guide and News report that focus on describing newsworthy things. The texts selected from Essay and Government document are clearly different from those of Travel guide and News report, with most nominals expressing Textually evoked and non-containing Inferrable information in Rheme position. The difference also reflects the communicative purposes, since the texts of Essay and Government document aim to present the causes of certain phenomena rather than bring more newsworthiness to

readers. All the texts distribute most nominals to convey Brand–new unanchored, Textually evoked, non–containing Inferrable and Unused information. The selection of one information status rather than another especially from the four categories can be quantified as a potential factor in structuring different patterns of information distributions in Rheme position. More importantly, the interaction between Rheme position and categories of information status reflects different characteristics of text types. Therefore, the information distributions of nominals in Rheme positions can be regarded as an observable indicator to distinguish text types here.

Infor Text	BN (%)		BN.A (%)		I (%)		I.C (%)		U (%)		E.S (%)		E.T (%)		Sum (%)	
	Th	Rh	Th	Rh	Th	Rh	Th	Rh	Th	Rh	Th	Rh	Th	Rh	Th	Rh
Tg1	17.46	38.1	0	8.73	33.33	24.6	0	2.38	14.29	7.14	1.59	0	33.33	19.05	100	100
Tg2	12.73	36.42	3.64	1.99	32.73	16.56	0.91	3.31	10.91	16.56	0	1.32	39.09	23.84	100	100
Np1	14.67	33.93	4	8.04	16	19.64	1.33	0	8	13.39	2.67	1.79	53.33	23.21	100	100
Np2	0	46.15	6.67	0	26.67	15.38	6.67	0	13.33	7.69	0	3.85	46.67	26.92	100	100
Np3	3.45	34.09	3.45	4.55	20.69	18.18	0	0	13.79	9.09	0	0	58.62	34.09	100	100
Np4	7.69	43.59	0	5.13	7.69	7.69	0	7.69	23.08	12.82	0	0	61.54	23.08	100	100
Np5	27.5	48.08	5	1.92	5	11.54	2.5	0	15	1.92	0	0	45	36.54	100	100
Np6	18.52	62.16	0	2.7	14.81	8.11	3.7	0	11.11	2.7	0	0	51.85	24.32	100	100
Gd	8.98	22.71	1.04	2.08	21.59	28.65	6.74	4.27	3.97	3.54	0.35	0.21	57.34	38.54	100	100
Ey	11.01	24.58	0.44	1.98	17.18	25.14	2.64	3.11	3.09	5.08	0	0.56	65.64	39.55	100	100

Table 6–12: Information distributions of nominal expressions in Theme and Rheme positions per text

6.2.3 Contrastive analysis of Theme and Rheme in information distributions of nominal expressions

Section 6.2.1 and Section 6.2.2 have so far presented findings about information distributions of nominal expressions in Theme and Rheme positions in the texts. This section aims to provide a contrastive analysis of Theme and Rheme in information distributions of nominal expressions. More specifically, it examines the similarities and differences of nominal expressions expressing the same information value in different clausal positions. Table 6–12 displays the contrastive statistics of the texts. Like the tables presented before, the shadow areas in pink indicates some texts have no frequency of nominals conveying certain types of information.

Table 6–12 shows that nominals of all the texts mainly convey Brand–new unanchored information in Rheme position, but Textually evoked information in Theme position. This finding provides evidence to Halliday and Matthiessen’s (2014: 120) claim that “the Theme falls within the Given, while the New falls within the Rheme”. Nominals expressing non–containing Inferrable, Unused and Textually evoked information occur in both Theme and Rheme positions. However in some texts, nominals expressing Brand–new unanchored and anchored, containing Inferrable and Situationally evoked information only occur in Theme or Rheme position. For example, Np2 distributes all nominals expressing Brand–new unanchored information to the Rheme position. Similar cases are more likely to be found from the category of Containing Inferrable. In addition, most zero values were found in Theme position when nominals convey Brand–new unanchored and anchored, and Situationally evoked information.

There are similarities and differences between the texts in information distributions of nominals in different clausal positions. The similarities are not only based on the gaps between the frequencies of nominals conveying the same information in the same clausal position. They also require the same tendency of distributing nominals to both clausal positions. For example, Government document and Essay texts are similar in

the frequencies of nominals conveying Unused information in Theme and Rheme positions. However, the similarity cannot be established, since Government document has a higher frequency of the nominals in Theme position, which is in contrast to Essay. Tg2 and Government document are similar in having frequencies of nominals conveying Situationally evoked information in both Theme and Rheme positions. However, Tg2 does not have any of the frequency in Theme and its tendency of distributing nominals to Theme and Rheme positions is different from Government document. Based on gaps and tendencies of distributing nominals to Theme and Rheme positions, some texts illustrate similarity in conveying categories of information.

When nominals express Brand–new unanchored information, two sets of texts are similar in distributing the nominals to both Theme and Rheme positions. The first set is Tg1, Tg2 and Np1 and the second is Essay and Government document. For example, the differences of nominals expressing Brand–new unanchored information in Theme and Rheme positions between Tg2 and Np1 are 1.94% and 2.49% respectively. When nominals express Brand–new anchored information, two groups of texts Tg1 and Np4, Essay and Government document, are similar in distributing nominals to Theme and Rheme positions. Tg1 and Np4 distribute all nominals expressing Brand–new anchored information to Rheme position and the difference between their distributions is only 3.6%. When nominals express non–containing Inferrable information, four sets of texts, namely Tg1 and Tg2, Np2 and Np3, Np4 and Np5, and Essay and Government document, are similar in the frequencies of the nominals in Theme and Rheme positions. For instance, both Essay and Government document texts have more nominals conveying non–containing Inferrable information in Rheme position and the gaps in Theme and Rheme positions are 4.41% and 3.51% respectively. When nominals express containing Inferrables, Np5 and Np6 are similar in having all the nominals in Theme positions, with a gap of 1.2%. When nominals express Unused information, three sets of texts, Tg1 Np2 and Np3, Tg2 and Np1, and Np5 and Np6 are similar in distributing the nominals to Theme and Rheme positions. Taking the group of Np5 and Np6 as an example, both have higher frequencies of nominals expressing Unused

information in Theme position and their gaps of the Unused category are 3.89% and 0.78% in Theme and Rheme positions respectively. When nominals express Situationally evoked information, Tg2 and Essay are similar. Both texts distribute all of the nominals in Rheme position and they have a gap of 0.76%. When nominals express Textually evoked information, three sets of texts, Tg1 and Tg2, Np1 Np2 and Np6, Np3 Essay and Government document, are similar in distributing the nominals to Theme and Rheme positions. For instance, both Np3 and Government document have higher frequencies of the nominals in Theme position and their gaps are 1.28% and 4.45% in Theme and Rheme positions respectively.

The above findings indicate that in the four main categories, Brand–new unanchored, non–containing Inferrable, Unused and Textually evoked, two sets of texts, Essay and Government texts and the texts of Travel guide and News report, are more likely to show similarity in information distributions of nominals both in Theme and Rheme positions. This is not surprising, since the two sets of texts are similar in textual characteristics. The Essay and Government document texts aim to present the causes of some phenomena and develop arguments, while the Travel guide and News report texts focus on describing newsworthy things.

In addition, nominal expressions conveying the same information status generate different characteristics when they occur in different clausal positions. Example (6–16) displays nominals expressing Brand–new unanchored information in both Theme and Rheme positions:

(6–16) a. *A boy* was run over by *a car*. (Lambrecht 1994: 169)

In example (6–16), “*a boy*” in Theme position and “*a car*” in Rheme position are not specified with particular identities. According to Lambrecht (1994: 169), neither “*a boy*” nor “*a car*” is the information focus of the clause. In fact, the focus is the relational newness represented by the complete clause, which forms an event–reporting statement and describes what happened.

- (6–17) a. *In 1935*, he furthered his claims when he wrote a now famous “Letter to an American Mother” of a homosexual... (Ey)
- b. Hirschfeld also founded the Scientific Humanitarian committee, which was mostly homosexual, in *1897*. (Ey)

As illustrated in Section 6.2.1 and Section 6.2.2, nominals expressing Brand–new information sometimes refer to time and place both in Theme and Rheme positions in the texts. In example (6–17), both expressions in bold italics indicate temporal information. “*1935*” in (6–17a) in Theme position is the beginning of an event so that more things happened on that year can be introduced subsequently. In contrast, “*1897*” in (6–17b) in Rheme position indicates the current topic of the clause will not be continued. Fries (1995a: 5) indicates “that initial adverbial clauses perform different discourse functions from final adverbial clauses” (cf. Thompson 1985: 55; Ford 1993; Diessel 2008; Wiechmann and Kerz 2013). Such a difference also exists between initial and final adverbial nominal expressions (also see Hasselgård 2010).

- (6–18) a. *An ambitious expansion* has left Magna with *excess capacity* and *a heavy debt load* as the automotive industry enters *a downturn*. (Np6)
- b. *This busy, modern European city* sits on *a thousand years of history*..... Historically and culturally this north–south distinction has always been significant, and it still is today, with a dose of good–humored rivalry between the two areas. (Tg2)

In (6–18a), nominals shown in bold italics convey Brand–new unanchored information. However, compared with “*an ambitious expansion*” in Theme position, “*excess capacity*”, “*a heavy debt load*” and “*a downturn*” in Rheme position are the information focus of the clause and are more closely related to the main topic of Np6, which is concerned with how an entrepreneur would help the company Magna turn

around. In example (6–18b), “*this busy, modern European city*” conveys Textually evoked information in Theme position, while “*a thousand years of history*” is brand–new to readers in Rheme position, which is also the information focus of the clause. Their thematic structure and information values form the unmarked mapping pattern, i.e. Given ^ Theme and New ^ Rheme (Halliday 1994: 299). The above shows that although nominals convey Brand–new information in Theme position, they are less likely to be the information focus of the clause. Nominals conveying new information in Rheme position are more closely related to the main topics of the texts.

Now the analysis moves to the nominal expressions conveying Inferrable information in Theme and Rheme positions. The Inferrables elaborate discourse entities and propositions occurred in the preceding text, which contributes to textual cohesion and coherence. Example (6–19a) is a case in point. “*The time of the transition*” here is a non–containing Inferrable and is a coherent point to link the preceding and current clauses, which serves as the ground to introduce “*much speculation about how things would change*” in Rheme position, which is also the information focus of the clause.

- (6–19) a. Today Hong Kong remains a capitalist enclave with its laws and rights intact, and China has promised that Hong Kong will continue in this fashion for at least 50 years... *Around **the time of the transition*** there was *much speculation about how things would change*. (Tg1)
- b. He went on to explain that while *his nurture* did not impact *his sexual orientation*, it did affect his view of it. (Ey)
- c. *We* are keenly aware of *the significance of our charge*, given the economic damage that America has suffered in the wake of the greatest financial crisis since the Great Depression. (Gd)

In (6–19b), “*his sexual orientation*” also conveys non–containing Inferrable. Although it occurred in Rheme position, it is not the focus of the clause that presents newsworthiness to readers. In fact, it contributes to the relational newness of the clause

(Gundel 2003), by linking to “*his nurture*” expressing Textually evoked information in Theme position. The instance displayed by example (6–19c) is similar to example (6–19b), with “*we*” expressing Textually evoked information in Theme and “*the significance of our charge*” expressing Inferrable information in Rheme position. The information focus of the clause is realized by “*are keenly aware of*”, rather than the two nominals in Theme and Rheme positions.

Finally the analysis moves to the nominal expressions expressing Evoked information in Theme and Rheme positions. Instances are displayed by example (6–20):

- (6–20) a. A couple thousand years after Plato and Homer, Sigmund Freud still believed homosexuality to be a natural behavior. *In an interview in 1903*, he professed ***his beliefs***: “I am... of the firm conviction that homosexuals must not be treated as sick people... Homosexual persons are not sick. They also do not belong in a court of law!” (Ey)
- b. This influx helps to make ***Dublin*** *a busy, buzzing international city*. (Tg2)

“*An interview in 1903*” in (6–20a) conveys Brand–new unanchored information as part of a prepositional phrase that functions as the marked Theme of the clause and “*his belief*” is Textually evoked in Rheme position based on the content of the previous sentence. However, “*his belief*” in (6–20a) is not the information focus. The focus is realized by the verb “*professed*” in Rheme position. Similarly in (6–20b), “*Dublin*” conveys Textually evoked information in Rheme position, but the information focus of the clause is located on the following nominal “*a busy, buzzing international city*”. The two instances show that although Textually evoked nominals occur in Rheme position, they are less likely to be the information focus in the texts.

6.3 Summary

The goal of this chapter was to explore the interplay between information status of nominal expressions and thematic structures in the texts. Unlike previous literature, this chapter provided a detailed account of information distributions of nominals both in Theme and Rheme positions in ten texts selected from four comparable genres, thereby addressing the third research question identified in Chapter 1.

Specifically, Section 6.1 has examined the relationship between the texts and their Themes and Rhemes in information distributions by using the total sum of nominal expressions of each complete text. Section 6.2 has presented the information distributions in Theme and Rheme positions respectively by using the number of nominal expressions of each clausal position. Within this section, Section 6.2.1 and Section 6.2.2 have further examined the information distributions of nominals as Participants and Circumstances in Theme positions. Section 6.2.3 has provided a contrastive analysis of Theme and Rheme in information distributions.

There are several important findings presented by this chapter. First, not many texts indicate a positive relation to their Themes and Rhemes in information distributions of nominal expressions by using the total sum of nominals of a complete text as the quantity. Taking into account the average frequency distributions of the Travel guide and News report texts, texts of four comparable genres reveal a positive relation to their Rhemes in containing all categories of information except non-containing Inferrables. The information distributions of Rhemes are more similar to those of the complete texts, with more nominals occurring in the rhematic zone. Second, information distributions of nominal expressions in Theme and Rheme positions reveal the communicative purposes of the texts. Four main categories of information, namely Brand–new unanchored, Textually evoked, non–containing Inferrable and Unused contribute most to structure Themes and Rhemes of the texts and each frequency distribution of the categories can be regarded as an observable variable to distinguish text types. Third, nominal expressions indicate different discourse functions in the same clausal position

when they represent different categories of information. Those representing the same information status have different functions when they occur in different clausal positions.

In conclusion, this chapter has provided a detailed account of the interplay between information status of nominal expressions and thematic structures in English texts. The next chapter, Chapter 7, will explore the interaction between information values of complex nominal expressions and integration in meaning.

7 The interaction between information status of complex nominal expressions and integration in meaning

As stated in Chapter 1, the main goal of the present study is to explore the informational role of nominal expressions in English texts. Chapters 4, 5 and 6 have provided detailed descriptions of the information distributions of nominal expressions; the relationship between linguistic forms of nominal expressions and information status they express in the texts; and the interplay between information status and thematic structure. This chapter aims to address the final one that investigates information values of postmodifiers of complex nominal expressions in terms of integration in meaning.

As stated in Chapter 2, previous literature has so far mostly focused on the grammatical complexity of nominal expressions, the interaction between information status of complex nominal expressions (or nominal complexes in SFL) and integration in meaning in the texts has remained relatively understudied. With this in mind, this chapter provides a detailed account of this interaction.

Specifically, the aim of the chapter is achieved by presenting the analysis results of information distributions of postmodifiers in the texts and the degree of tightness in integration in meaning based on their functions in logico-semantic relations. As indicated by Halliday and Matthiessen (2014: 430), the effect of combining clauses into a clause complex is one of tighter integration in meaning. The same principle can be applied to nominal complexes, in that the effect of combining nominals into a nominal complex is tighter integration in meaning. Integration in meaning is compatible with the integration in information from the linguistic sense. Furthermore, as explained in Chapter 2, Section 2.5, types of expansion can be regarded as a measure of the degree of tightness. Elaboration, extension and enhancement are different in how they bring informational elements to nominal complexes.

This chapter is organized as follows. Section 7.1 outlines the distributions of postmodifiers in types of expansion in the texts. Section 7.2 mainly focuses on the

information status of postmodifiers within the same expansion type. Finally, Section 7.3 offers a summary of the interaction between information status of complex nominals and integration in meaning.

7.1 Outline of postmodifiers of the corpus dataset

This section outlines the distributions of postmodifiers involved in types of expansion. As illustrated in Table 7–1, the overall number of postmodifiers per text is very small. For example, Np6 has only one instance. The analysis results here cannot be claimed representative. It can be seen that most postmodifiers express an elaborative function, with a frequency of more than 50% in eight texts. The texts show clear similarities. Tg2 (59.38%), Np1 (64.71%), Essay (58.21%) and Government document (63.85%) are similar in the frequencies of postmodifiers expressing the elaborative function and Np3 (40%) and Np5 (45%) have a small gap of 5% in the frequencies. In contrast, texts of the same genre are obviously different. This is caused by the size of the dataset. The numbers of postmodifiers are really small, which makes it difficult to observe any regular patterns. Regardless of the individual differences, the average values of the Travel guide and News report texts are 66.36% and 60.35% respectively when their postmodifiers play an elaborative role.

Expansion Text	elaboration		extension		enhancement		expansion_other		Sum	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tg1	22	73.33	0	0	0	0	8	26.67	30	100
Tg2	19	59.38	0	0	2	6.25	11	34.38	32	100
Np1	22	64.71	1	2.94	0	0	11	32.35	34	100
Np2	6	60	0	0	0	0	4	40	10	100
Np3	4	40	0	0	0	0	6	60	10	100
Np4	11	52.38	0	0	0	0	10	47.62	21	100
Np5	9	45	0	0	0	0	11	55	20	100
Np6	1	100	0	0	0	0	0	0	1	100
Gd	189	63.85	0	0	9	3.04	98	33.11	296	100
Ey	39	58.21	3	2.99	3	4.48	23	34.33	67	100

Table 7–1: Distributions of postmodifiers in types of expansion per text

Table 7–1 also shows that the frequencies of postmodifiers expressing an extensive function are very low and they were only found in Np1 and Essay. Each text has only one instance. According to Halliday and Matthiessen (2014: 497), the only way for postmodifiers of nominal complexes to express the extensive function in embedded clauses is through possession, which is typically introduced by *whose* or *of which* or a contact relative ending with *of*, for example *one whose name has long been forgotten*. The relationship of possession is used to denote ownership and various kinds of abstract or concrete associations (Halliday and Matthiessen 2014: 497).

The finding here indicates that it is anomalous to find postmodifiers that only express the relationship of possession in the texts. The occurrence of possession can be attributed to many factors, such as the conceptual relations between head nouns and the postmodifiers of nominal complexes, the general probability of possession to be realized as nominal complex, the complexity of nominal expressions in different text types and the relation between extension and other types of expansion. However, a reliable conclusion requires an in–depth analysis of the possessive postmodifiers and this is beyond the scope of the current research.

Now the analysis moves to the frequencies of postmodifiers that express an enhancement function. Table 7–1 indicates that they were only found in three texts, which are Tg2, Government document and Essay. Although the three texts have postmodifiers that express an enhancement function, the frequencies are very low, with the highest being 6.25% in Tg2. Based on Halliday and Matthiessen (2014: 497), enhancement represents a circumstantial relation, which includes time, place, manner, cause or condition. The statistics in Table 7–1 suggest that nominal complexes are less likely to realize circumstantial relationships by themselves in these ten texts.

As previous literature (e.g. Biber and Conrad 2014: 116–117) indicates, information of time and place is regarded as a prominent feature of newspapers, which makes circumstantial adverbials common in text. For this reason, News report in the current study is expected to contain more circumstantial relations, regardless of the terminological difference from newspapers. Table 7–1 does not provide straightforward

evidence for the frequencies of circumstantial adverbials of the News report texts. However, all the zero values of postmodifiers expressing an enhancement function suggest that the News report texts do not prefer to convey circumstantial information through nominal complexes. As was indicated by Halliday and Matthiessen (2014: 222), circumstance is typically realized by prepositional phrases in clauses. The statistics in Table 7–1 actually imply that the News report texts prefer to realize enhancement through separate clause elements. This is consistent with previous findings of newspapers. In addition, postmodifiers are not as important and independent as separate clause elements in structuring text, since they do not form a separate tone group or characterize particular participants represented by the head nouns (Halliday and Matthiessen 2014: 494).

Finally, the analysis moves to the remaining types of expansion, which were labelled as “expansion_other” in the corpus annotation. As indicate before (see chapter 3), some postmodifiers can be involved in more than one type of expansion or the same type of expansion more than once. Table 7–1 displays that the frequencies of postmodifiers in the remaining types of expansion are higher than those in extension and enhancement, but are lower than those in elaboration. Except for Np6 (0), the frequencies of all the other texts are more than 30%. The finding suggests that types of expansion of nominal complexes are related and they tend to co-occur in text to some extent. Second, the frequencies indicate both similarities and differences between the texts. Four texts, Tg2 (34.38%), Np1 (32.35%), Government document (33.11%) and Essay (34.33%) are clearly similar in the frequencies, with the biggest gap of 2.03% between Np1 and Tg2. In contrast, Np6 has a gap of 60% with Np3 (60%) in the frequencies, which is strikingly different. The individual texts of Travel guide and News report show clear differences in the frequencies of postmodifiers that express multiple types of expansion. Tg1 (26.67%) and Tg2 (34.38%) have a difference of 7.71%, and the smallest gap of the News report texts is 7.38% between Np4 (47.62%) and Np5 (55%). Regardless of the individual differences, the average values of the Travel guide and News report texts are 30.53% and 39.16% respectively. Postmodifiers of the News

report texts are more likely to be involved in multiple types of expansion compared with the other texts.

As stated at the beginning of the chapter, one of the effects of combining nouns into a nominal complex is the tightness of integration in meaning (cf. Halliday and Matthiessen 2014: 430). Since nominal complexes are expanded in different ways, the degree of integration in meaning is expected to be different. However, postmodifiers of nominal complexes all define the head nouns as embedded elements. The three types of expansion within nominal complexes, namely elaboration, extension and enhancement, may not differ as clearly as those within clause complexes in the degree of integration in meaning. Postmodifiers in the remaining types of expansion are different from the three regular types. They are involved in more than one type of expansion, which might suggest a tighter degree of integration in meaning than those only involved in regular types.

This section has presented the distributions of postmodifiers in types of expansion and discussed the degree of integration in meaning. Section 7.2 will examine the informational features of the postmodifiers, with particular reference to the tendencies of conveying types of information in the texts when expressing different types of expansion.

7.2 Information status of postmodifiers in types of expansion

This section explores information status of postmodifiers that are involved in different types of expansion to see their similarities and differences. As presented in Section 7.1, the frequencies of postmodifiers are clearly different between the types of expansion. Most postmodifiers express an elaborative function and some are involved in the remaining types of expansion, with very few instances found in extension and enhancement in the texts. For this reason, this section first focuses on comparing information status of postmodifiers involved in elaboration and the remaining types of expansion and then provides an account of the instances in extension and enhancement.

Table 7–2²⁶ displays information status of the postmodifiers in elaboration and the remaining types of expansion and the two types of expansion are abbreviated to as “Elab.” and “Ep.–o”. The raw frequencies are presented as tables in the Appendix of this thesis. As indicated in Section 7.1, the size of the dataset in this chapter is very small. Percentages presented in Table 7–2 are based on few instances. The analysis results cannot be claimed representative.

²⁶ The shaded areas in Table 7–2 mean no nominal expressions of a certain information category were found in the complete texts. For example, the shaded areas of Np3 indicate that the complete text does not contain frequencies of nominals that convey containing Inferrable and Situationally evoked information.

Expansion Text	BN (%)		BN.A (%)		I (%)		I.C (%)		U (%)		E.S (%)		E.T (%)		Sum (%)	
	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o	Elab.	Ep.-o
Tg1	22.73	50	13.64	25	36.36	25	0	0	4.55	0	0	0	22.73	0	100	100
Tg2	15.79	36.36	0	0	31.58	18.18	0	0	21.05	36.36	5.26	0	26.32	9.09	100	100
Np1	18.18	9.09	0	18.18	4.55	9.09	0	0	22.73	54.55	9.09	0	45.45	9.09	100	100
Np2	33.33	25	0	0	0	0	0	0	16.67	50	16.67	0	33.33	25	100	100
Np3	25	0	0	0	0	0	0	0	0	50	0	0	75	50	100	100
Np4	54.55	10	0	0	9.09	20	0	0	18.18	30	0	0	18.18	40	100	100
Np5	11.11	36.36	0	0	22.22	9.09	0	0	0	18.18	0	0	66.67	36.36	100	100
Np6	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100	100
Gd	10.05	11.22	1.06	1.02	24.34	26.53	1.06	1.02	5.82	4.08	1.59	0	56.08	56.12	100	100
Ey	15.38	17.39	2.56	0	28.21	13.04	2.56	0	0	30.43	2.56	0	48.72	39.13	100	100

Table 7–2: Information status of postmodifiers in elaboration and the remaining types of expansion per text

Table 7–2 reveals three clear tendencies. First, most postmodifiers in both types of expansion mainly express Brand–new unanchored, non–containing Inferrable, Textually evoked and Unused information in the texts. The finding is not surprising, since the four categories of information are expressed by most nominal expressions of the texts (see Chapter 4, Section 4.2). Second, postmodifiers of the Travel guide and News report texts in both types of expansion do not convey containing Inferrable information. This can be explained by the frequencies of nominals expressing containing Inferrables in the texts. As indicated in Chapter 4, Section 4.4, the frequencies are very low and they were not found in all texts. Besides, another possible explanation is the linguistic features of containing Inferrables. According to the categorical criteria in the current study, each containing Inferrable is required to have at least two nominals, within which a prototypical relationship can be established (see Chapter 3, Section 3.2.2). The requirement makes the linguistic forms of containing Inferrable more complex than those of other categories of information. Most containing Inferrables are expressed by *of*–NPs in the texts (see Chapter 5, Section 5.2.4). Compared with Government document and Essay, the Travel guide and News report texts are easy to read and are expected to contain fewer complex nominals. They are less likely to contain more *of*–NPs, since the *of*–NPs are more complex than other types of nouns by having a larger number of words and phrasal nodes (cf. Berlage 2014: 5). For this reason, it is even more difficult to find them in the function of postmodifiers. Third, postmodifiers of the Government document text indicates a complex pattern of information distribution, which is in clear contrast to the News report texts. Except for Situationally evoked, they convey all other types of information when they are involved in both types of expansion.

The News report texts do not contain postmodifiers that express several categories of information status in both expansion types. Apart from the zero frequency of containing Inferrables, the News report texts do not have postmodifiers expressing Brand–new anchored information, except for Np1 with a frequency of 18.18% in the remaining types of expansion. Unlike the explanation for containing Inferrables, the

low Brand–new anchored frequencies of postmodifiers in the News report texts cannot be explained by the distributions of nominals in the complete texts.

The low frequencies of postmodifiers expressing Brand–new anchored information in the News report texts might be related to communicative purpose. The News report texts aim to describe newsworthy events that happen most recently, which is mainly achieved by containing higher frequencies of Brand–new information, both unanchored and anchored (see Section 4.1.1 and Section 4.1.2). However, the higher frequencies of Brand–new information were also found in the Travel guide texts. Under this circumstance, how can the News report texts highlight their newsworthiness? One plausible way is to realize Brand–new information as separate clause elements. When new information is expressed by postmodifiers, it is semantically and grammatically downgraded as dependent elements of the head nouns. For this reason, the postmodifiers conveying Brand–new anchored information texts have a lower frequency in the News report so that more new information could be realized as separate clause elements with a higher syntactic status to highlight the newsworthiness.

The statistics in Table 7–2 show important differences but also similarities between the two expansion types in information distributions of postmodifiers in these ten texts. First, except for Np4, Np6 and Government document, postmodifiers of the other texts contain higher frequencies of Textually evoked information when expressing an elaborative function. Similarly, postmodifiers in elaboration tend to express Situationally evoked information in these ten texts. The above suggests that postmodifiers playing an elaborative role are more likely to convey Evoked information. Furthermore, Government document is clearly different from the other texts. It has similar frequencies of postmodifiers in the elaboration and the remaining types of expansion when expressing the same category of information. The biggest gap of the frequencies between the two types of expansion is only 2.19% when postmodifiers express non–containing Inferrables.

Now the analysis moves to the differences between the texts by looking at the frequencies of postmodifiers when expressing categories of information. As indicated

above, most postmodifiers convey the four main categories of information, i.e. Brand–new unanchored, non–containing Inferred, Unused and Textually evoked. Therefore, the analysis mainly focuses on those categories with specific examples.

Table 7–2 illustrates three patterns in the category of Brand–new unanchored across the texts. The postmodifiers of Tg1 and Tg2 that express an elaborative function have lower frequencies of Brand–new unanchored information than those in the remaining types of expansion; most News report texts develop in a contrastive way by having higher frequencies of postmodifiers when expressing an elaborative function; and Essay is similar to Government document, with almost the same frequencies of Brand–new unanchored postmodifiers in the two types of expansion. Specific examples are given in (7–1):

- (7–1) a. There are many stories of *refugees who arrived with **nothing** in their pockets, set up a small sidewalk stall, worked diligently until they had their own store, and then expanded it into a modest chain.* (Tg1)
- b. The number of suspicious activity reports — *reports of possible financial crimes filed by **depository banks** and their affiliates* — related to mortgage fraud grew 20–fold between 1996 and 2005 and then more than doubled again between 2005 and 2009. (Gd)

In example (7–1), all the nominals in bold italics convey Brand–new unanchored information in the texts, but they are different in expressing functions of expansion. In (7–1a), “*nothing*” as a postmodifier is part of an expression that elaborates the head noun “*refugees*”. In (7–1b), the expression “*depository banks*” as a postmodifier plays a multifunctional role. It is not only part of an expression that elaborates “*reports of possible financial crimes*”, but also extends “*their affiliates*” in the text.

In addition, both “*nothing*” and “*depository banks*” only occur once in the texts and they are not closely related to the main topics based on the meaning of the expressions. Although both nominals convey Brand–new unanchored information, they

do not play an important role in conveying newsworthiness compared with the other Brand–new unanchored nominals that function as head nouns with higher frequencies of repetition in the texts. For example, “*homosexuality*” in Essay is categorized as Brand–new unanchored. As the heading of the text, it is repeated more frequently than the other expressions, which contributes more to structuring the text.

Another finding is more interesting. The head nouns of “*nothing*” and “*depository banks*” in example (7–1) are different in information status. “*Refugees*” in (7–1a) conveys Brand–new unanchored information, while “*reports of possible financial crimes*” in (7–1b) is a non–containing Inferrable. However, the nominals following “*nothing*” and “*depository banks*” have the same information status. “*Their pockets*” in (7–1a) and “*their affiliates*” (7–1b) both express Brand–new anchored information in the texts. The information distributions of the first three nominals of the nominal complex in (7–1a), i.e. “*refugees*”, “*nothing*” and “*their pockets*” is Brand–new unanchored + Brand–new unanchored + Brand–new anchored. In (7–1b), the information distributions of the three nominals of “*reports of possible financial crimes filed by depository banks and their affiliates*” are non–containing Inferrable + Brand–new unanchored + Brand–new anchored. The expression “*depository banks*” in (7–1b) is connected with a more activated entity compared with “*nothing*” in (7–1a). One possible explanation for the information distributions is the influence of expansion types. As has been shown, “*depository banks*” in (7–1b) play multiple roles in elaboration and extension. The complete nominal complex is more closely structured, both grammatically and semantically. The tighter degree of integration in meaning might make the information status of the involved discourse entities more activated.

In the category of non–containing Inferrable, Table 7–2 displays that postmodifiers that express an elaborative function in Tg1, Tg2 and Essay have higher frequencies than those in the remaining types of expansion. Three News report texts, Np2, Np3 and Np6, do not contain frequencies of postmodifiers that express both functions of expansion in the category of non–containing Inferrable. Besides, Np1 and Np4 have higher frequencies of postmodifiers in the remaining types of expansion than those in

elaboration, which contrasts with Np5. As indicated above, Government document has a nearly even distribution of postmodifiers that express both types of expansion in all categories of information. Typical examples are provided in (7–2):

- (7–2) a. The constantly crowded and busy Grafton Street is *the most visible center for shopping*, but there are shops all over that carry an international array of goods as well as the Irish crafts and souvenirs you expect. (Tg2)
- b. From financial firms to corporations, to farmers, and to investors, derivatives have been used to hedge against, or speculate on, *changes in prices, rates, or indices or even* on events such as the potential defaults on debts. (Gd)

Both “*shopping*” in (7–2a) and “*prices*” in (7–2b) as postmodifiers express an elaborative function in the texts when they convey non-containing Inferrable information. In comparison, “*prices*” in (7–2b) also expresses an extending function with the other two postmodifiers “*rates, or indices*” and the three nominals form a co-classification relationship. The relationship among the nominals facilitates the categorization of the information status to some extent, which makes it easy to infer “*rates, or indices*” based on the background knowledge of finance and economics. In addition, both the head nouns of “*shopping*” in (7–2a) and “*prices*” in (7–2b), “*the most visible center*” and “*changes*”, convey Brand–new unanchored information. Unlike “*shopping*” in (7–2a), “*prices*” in (7–2b) are closely followed by non-containing Inferrables.

In the category of Unused, Table 7–2 displays that most texts have lower frequencies of postmodifiers that express an elaborative function, with Tg1, Np6 and Government document as exceptions. Specific examples are given in (7–3):

- (7–3) a. Mr. Stronach, founder and controlling shareholder of Magna, resigned

as chief executive officer last year to seek, unsuccessfully, *a seat in Canada's Parliament*. (Np6)

- b. The slaughter in Colombia was very much on the minds of *450 editors and publishers from Latin America, the United States, the Caribbean and Canada attending the 45th general assembly of the Inter-American Press Association in Monterrey, Mexico*, this week. (Np1)

In (7–3), both “*Canada's Parliament*” and “*the United States*” as postmodifiers play an elaborative role and convey Unused information in the texts. The head nouns of the two postmodifiers, “*a seat*” and “*450 editors and publishers*”, both convey Brand–new unanchored information. However, different from “*Canada's Parliament*” in (7–3a), “*the United States*” in (7–3b) is involved in more than one expansion type, since it also extends “*Latin America*”, “*the Caribbean*” and “*Canada*”. Among them, “*Latin America*” expresses Textually evoked information in the text. Compared with “*Canada's Parliament*” in (7–3a), “*the United States*” in (7–3b) is closely followed by a known expression.

In the category of Textually evoked, Table 7–2 shows that most texts have higher frequencies of postmodifiers that express an elaborative function, with Np4, Np6 and Government document as exceptions. Examples in (7–4) illustrate nominals in the two expansion types:

- (7–4) a. *Sightseeing in Hong Kong* starts at sea level with the enthralling water traffic — a mix of freighters, ferries, tugs, junks, and yachts. (Tg1)
- b. During the same year, *68% of 'option ARM' loans originated by Countrywide and Washington Mutual* had low– or no–documentation requirements. (Gd)

In (7–4), “*Hong Kong*” and “*Countrywide*” as postmodifiers are part of the expressions that elaborate “*Sightseeing*” and “*68% of 'option ARM' loans*” respectively.

Although both postmodifiers convey Textually evoked information in the texts, “*Countrywide*” in (7–4b) occurs less frequently. The frequency actually is influenced by the topics of the texts. “*Hong Kong*” in (7–4a) is the heading of Tg1 and expresses the main topic, while “*Countrywide*” in (7–4b) as a name of company is not closely related to the theme of Government document, which is mainly concerned with the causes of the global financial crisis in 2008. In addition to the elaborative role, “*Countrywide*” in (7–4b) also has an extending function with another postmodifier “*Washington Mutual*”. Their head noun, “68% of *option ARM*”, is more familiar to readers as a containing Inferrable than the head noun of “*Hong Kong*” in (7–4a), which is a non-containing Inferrable.

Now the analysis moves to explore the information status of postmodifiers that express an extending function in the texts. As has been shown in Section 7.1, only three instances were found in the corpus dataset and they are illustrated in example (7–5):

- (7–5) a. This comes from *a man whose brother, Guillermo, was murdered in 1986*.
(Np1)
- b. Kertbeny derived this word from the Greek word for “same” and the Latin word for “sex,” whereas a heterosexual is *a person “whose feelings of sexual attraction are for the opposite sex”* (Marcus 1). (Essay)

In (7–5a), “1986” as a postmodifier only plays an extending role and conveys Brand–new unanchored information in the sentence. In (7–5b), “*whose feelings of sexual attraction*” and “*the opposite sex*” both express an extending function and are categorized as non-containing Inferrables in the text. The three instances have two things in common. Their head nouns, “*a man*” in (7–5a) and “*a person*” in (7–5b) express Brand–new unanchored information and they only occur once in the texts. It seems that the head nouns that express an extending function do not play an important role in continuing the main topics of the texts.

Infor Text	BN (%)	BN.A (%)	I (%)	I.C (%)	U (%)	E.S (%)	E.T (%)	Sum (%)
Tg2	0	0	0	0	0	0	100	100
Gd	22.22	0	22.22	0	0	0	55.56	100
Ey	33.33	0	0	0	0	0	66.67	100

Table 7–3: Information status of postmodifiers in enhancement per text

Finally, the analysis explores the information status of postmodifiers that express an enhancing function in the texts. As indicated in Section 7.1, very few instances were found from the dataset. Only three texts have postmodifiers playing an enhancing role, namely Tg2, Government document and Essay. Table 7–3 illustrates them in terms of information categories. The postmodifiers expressing an enhancing function can be used to convey Brand–new unanchored, non–containing Inferrable and Textually evoked information in the texts. Although most postmodifiers convey Textually evoked information in Tg2, Government document and Essay, the texts are different in the information distributions. Government document has frequencies of the postmodifiers expressing all three categories of information; Tg2 only has frequencies in Textually evoked; and the postmodifiers in Essay only express Brand–new unanchored and Textually evoked information. Specific examples are given in (7–6):

- (7–6) a. He says that if he married *a female with which he shared **a strong trust***, maybe he could work it out. (Essay)
- b. The CRA requires banks and savings and loans to lend, invest, and provide services to *the communities from which they take **deposits***, consistent with bank safety and soundness. (Gd)
- c. Dublin’s food has undergone a metamorphosis. There was *a time when **you might have apologized for it***, but no longer. (Tg2)
- d. I grew up in *a place where **homosexuality** is accepted*, so that’s why I’m so open about being gay; I accept myself... (Essay)

In (7–6a), “*a strong trust*” as a postmodifier is part of an expression that enhances

the head noun “*a female*” and it conveys Brand–new unanchored information. In (7–6b), “*deposits*” is part of an expression that enhances the head noun “*the communities*” and it is a non–containing Inferrable. The two postmodifiers only occur once in the texts. In comparison, instances in bold italics in (7–6c) and (7–6d) express Textually evoked information, which occur more than once in the texts. “*Homosexuality*” in (7–6d) is repeated more frequently than “*you*” and “*it*” in (7–6c), since it is the heading of Essay, which expresses the main topic. Interestingly, their head nouns, “*a time*” in (7–6c) and “*a place*” in (7–6d), convey Brand–new unanchored information. However, they only occur once in the texts, which is less important in structuring the texts in terms of frequency.

7.3 Summary

The goal of this chapter was to explore the interaction between information status of complex nominal expressions and integration in meaning. Unlike the previous literature that mainly focused on either the structural complexity of nominals or logico-semantic relations of clauses, this chapter has provided a detailed account of information distributions of postmodifiers and different roles they play in logico-semantic relations, thereby addressing the last research question identified in Chapter 1. Specifically, Section 7.1 has shown the frequencies of postmodifiers that express different functions of expansion in the texts and the difference in degree of tightness of integration in meaning; and Section 7.2 has analyzed the information status of postmodifiers that express different functions of expansion.

There are several important findings presented by this chapter. First, the frequencies of postmodifiers are clearly different in serving different functions of expansion in these ten texts. Most postmodifiers express an elaborative function. Some can be used to express more than one type of expansion and the frequencies are lower than those playing an elaborative role. In contrast, very few instances were found to express extending and enhancing functions.

Second, although postmodifiers are used to express elaborative, extending and enhancing functions, they are similar in the tightness of integration in meaning. The postmodifiers do not express single messages and they all serve to define the head nouns. In comparison, those expressing more than one function of expansion indicate a tighter integration in meaning, since they also elaborate, extend or enhance other nominals besides defining the head nouns. Within the nominal complexes, expressions are more closely related to each other, both semantically and grammatically.

Third, postmodifiers show clear similarities and differences in information status when expressing different functions of expansion in the texts. They mainly convey Brand–new unanchored, non–containing Inferrable, Textually evoked and Unused information. The finding is consistent with the information distributions of the texts (see Chapter 4, Section 4.2). Postmodifiers show varied tendencies of conveying categories of information between the elaborative and the multiple roles of expansion. The postmodifiers expressing an elaborative function tend to convey more Unused and Evoked information, both textually and situationally. Besides, postmodifiers in the Travel guide texts tend to convey less Brand–new but more non–containing Inferrable information when they express an elaborative function; postmodifiers in the News report texts contain more Brand–new unanchored but less Unused information when they play an elaborative role; in Essay, the elaborative postmodifiers have more non–containing Inferrable but less Unused frequencies; and in Government document, postmodifiers in the two expansion types indicate an even distribution in all categories of information.

Another interesting finding is the characteristics of postmodifiers in expressing different functions of expansion. Those in the remaining types of expansion are more likely to occur with more activated expressions under the same information status as postmodifiers that express an elaborative role. Although some postmodifiers convey Brand–new unanchored information, their newsworthiness is not relatively important in structuring the texts. Postmodifiers that convey Brand–new unanchored and Inferrable information only occur once in the texts and they are not closely related to

the main topics. When postmodifiers represent the headings of texts and convey Textually evoked information, their head nouns are more likely to be new to readers. In theory, the combination of information values will make the nominal complexes less difficult to process, since new information is attached to Given, both semantically and grammatically. Further research is needed to explore this aspect.

In conclusion, this chapter has explored the interplay between information status of complex nominals and integration in meaning in English texts. Chapter 8 provides a conclusion for the thesis. It will indicate how the previous chapters have contributed to examine the informational role of nominal expressions in English texts.

8 Conclusion

As stated in Chapter 1, this research aimed to explore the informational role of nominal expressions in English text. This thesis comes to a close after the discussion of integration in meaning and information status of complex nominals presented in Chapter 7. This chapter will summarize the main contributions of this research, discuss the limitations of this work and point out some directions for further research.

8.1 Main contributions of this thesis

This study as a whole has mainly contributed to three areas. The first area is the area of descriptive linguistics. It offers a full account of the frequencies of nominal expressions conveying types of information values at varied levels of text, clause and group/phrase. The account sheds light on how nominal expressions function as a hub through which information flows, how they establish semantic and referential relations throughout a text and how they are integrated into complex information units.

The second area in which this study contributes is theoretical linguistics. This study proposes a new indicator to observe variations of text types. The findings of this study show that the informational role of nominal expressions at various levels of text, clause and group/phrase are closely related to communicative purposes, topics and degrees of shared background knowledge of English written text types. The use of information values to develop our understanding of nominal expressions allows us to gain insights into differences between text types.

The third area is contributed by the methodology. This study provides a full description of the annotation scheme developed for analyzing the informational role of nominal expressions in text. The scheme contains an adapted framework from Prince (1981) to classify information values and a wide range of linguistic features of nominal expressions, such as modification, linguistic form, logical-semantic relations, referential and semantic relations. It not only further defines avoids ambiguities

between information categories, but also provides us with insights into types of nominals in text. The methodology is highly practicable to be applied to other information studies of nominals.

8.2 Limitations of the research

This research filled four gaps in exploring the informational role of nominal expressions in text, it however still has a number of limitations.

The first limitation is concerned with the challenges in developing the methodology. As indicated in Chapter 3, fuzzy boundaries exist between certain linguistic categories of the annotation scheme of this research. Although some criteria have been proposed to avoid potential misclassifications, more consideration needs to be given to those fuzzy boundaries. Some linguistic categories could be further classified into subtypes, such as the remaining types of nominal expressions, the multiple types of expansion and *of*-NPs. The annotation scheme developed in the methodology did not provide an in-depth analysis of these linguistic categories.

The second limitation is with respect to the corpus established for this research. The corpus contains a total of 3095 nominal expressions gathered from ten texts of four comparable genres, namely Travel guide, News report, Government document and Essay. The data could be regarded as substantial based on the amount of nominals and a variety of linguistic features manually analyzed here. However, the corpus is a mixed one. Each genre is composed of a different number of texts and the texts are different in length (detail can be seen in Chapter 3, Section 3.1). Information distributions of nominals in the four genres cannot be claimed to be representative, since the texts of Travel guide and News report show differences that need to be taken into further account, and Government document and Essay contain only one text. Therefore, a more balanced corpus with similar numbers and lengths of texts to represent text types would be needed to further explore the informational role of nominal expressions.

Another limitation is the analysis of statistics. This research did not use any tests

of significance to compare statistics, since the size of our dataset is really small. Further research in the area of the informational role of nominal expressions in text would benefit from tests of significance with a larger and more balanced corpus.

A final limitation is concerned with the range of phenomena addressed related to the actual linguistic features of nominals analyzed. This research is constrained to address four specific gaps in the informational role of nominal expressions in English texts. Corpus analysis conducted in this research covered a wide range of features of nominals and produced a vast amount of information. However, not all findings were presented and discussed here, such as the continuity of complex nominal expressions, textual distances between antecedents and nominals conveying Textually evoked information, and information values of nominals that are involved in varied semantic relations. Besides, as suggested in Chapters 4, 5, 6 and 7, differences between the texts of the same genre in information distributions of nominals also need further explorations at the levels of text, clause and group/phrase.

These limitations indicate that this research is more of a starting point than an end point to explore the informational role of nominal expressions in English texts. Section 8.3 will present directions for further research.

8.3 Directions for further research

This thesis has opened up several directions for further research. This section will focus on the main ones.

The first follows from one of the limitations described in the previous section. Further research need to use a more balanced corpus that cover similar numbers and lengths of texts selected from comparable text types. More descriptive and comparative statistics can be collected to observe the informational role of nominal expressions in text via significance tests.

The second area of further research is concerned with a more detailed analysis of certain linguistic features of nominal expressions. Although this research covers a wide

range of linguistic features, some of them require more consideration. For instance, the analysis of non-containing Inferrables in Chapter 4 has shown different types in terms of the antecedents and deserves further categorization; the remaining types of nominal expressions and determiners of nominal expressions investigated in Chapter 5 could be further categorized in several different types; the information distributions of nominals in Chapter 6 would require a contrastive analysis of those functioning as Participants and Circumstances in Theme position; the information values of nominal expressions as postmodifiers presented in Chapter 7 are worthy of more consideration in terms of the subcategories of expansion types defined by Halliday and Matthiessen (2014).

The third area of further research is with respect to theoretical implications for related fields of study. For example, the difference in information values could reflect changes of readers' mental status in theory. Some nominal expressions, though expressing the same information value, may still require different processing time due to varied linguistic forms and textual distances to their antecedents. Psycholinguistic experiments can be designed to test the processing time and information processing of varied nominal expressions in text.

The research of the informational role of nominal expressions can also be extended to the other languages, such as Chinese, Spanish or German. This will provide a contrastive analysis to find similarities and differences between English and the other languages. Corpora could also be selected from different periods of time, so that the findings of corpus analysis allow us to observe language change.

References

- Albrecht, Jason E. and Edward J. O'Brien. 1993. Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19: 1061–1070.
- Allerton, David J. 1978. The notion of 'Givenness' and its relation to presupposition and theme. *Lingua*, 44(1): 133–168.
- Anderson, Richard C. 1978. Schema-directed processes in language comprehension. In Alan M. Lesgold, James W. Pellegrino, Sipke D. Fokkema, and Robert Glaser (eds.), *Cognitive Psychology and Instruction*, pp. 67–82. Springer: Boston.
- Anderson, Richard C. 2019. Role of the reader's schema in comprehension, learning, and memory. In Donna E. Alvermann, Norman J. Unrau, Misty Sailors, and Robert B. Ruddell (eds.), *Theoretical Models and Processes of Literacy*, pp. 136–145. Routledge: New York and London.
- Anderson, Richard C., and David P. Pearson. 1988. A schema-theoretic view of basic processes in reading comprehension. In Patricia L. Carrell, Joanne Devine, and David E. Eskey (eds.), *Interactive Approaches to Second Language Reading*, pp. 37–55. Cambridge: Cambridge University Press.
- Ariel, Mira. 1985a. Givenness Marking. PhD thesis, Tel-Aviv University.
- Ariel Mira. 1985b. The discourse functions of Given information. *Theoretical Linguistics*, 12: 99–113.
- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics*, 24(1): 65–87.
- Ariel, Mira. 2001. Accessibility theory: an overview. In Ted Sanders, Joost Schilperoord, and Wilbert Spooren (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects*, pp. 29–88. Amsterdam/Philadelphia: John Benjamins.
- Ariel, Mira. 2014. *Accessing Noun-phrase Antecedents*. London/New York: Routledge.
- Arts, Anja, Alfons Maes, Leo G. M. Noordman, and Carel Jansen. 2011a. Overspecification in written instruction. *Linguistics*, 49(3), 555–574.
- Arts, Anja, Alfons Maes, Leo G. M. Noordman, and Carel Jansen. 2011b.

- Overspecification facilitates object identification. *Journal of Pragmatics*, 43: 361–374.
- Asher, Nicholas and Alex Lascarides. 1998. Bridging. *Journal of Semantics*, 15 (1): 83–113.
- Bach, Kent. 1999. The semantics–pragmatics distinction: what it is and why it matters. In Ken Turner (ed.), *The Semantics–Pragmatics Interface from Different Points of View*, pp. 65–84. New York: Elsevier Science.
- Barsalou, Lawrence W. 1983. Ad–hoc categories. *Memory and Cognition*, 11: 211–227.
- Bartlett, Frederik C. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Baumann, Stephan. 2006. *The Intonation of Givenness: Evidence from German*. Niemeyer: Tübingen.
- Baumann, Stephan, and Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta and Pilar Prieto (eds.), *Prosody and Meaning: Interface Explorations 25*, pp. 119–61. Berlin/New York: De Gruyter Mouton.
- Becker, Angelika, Renate Musan, and Valentina Cristante. 2020. The comprehension of coherence relations in expository texts at the age of 10 and 12. In Natalia Gagarina and Renate Musan (eds.), *Referential and Relational Discourse Coherence in Adults and Children*, pp. 11–40. Berlin/Boston: De Gruyter Mouton.
- Berlage, Eva. 2014. *Noun Phrase Complexity in English*. Cambridge: Cambridge University Press.
- Berlin, Brent. 1976. The concept of rank in ethnobotanical classification. *American Ethnologist*, 3: 381–399.
- Berlin, Brent. 1978. Ethnobotanical classification. In Eleanor Rosch and Barbara B. Lloyd (eds.), *Cognition and Categorization*, pp. 1–26. New Jersey: Lawrence Elbaum Associates.
- Berlin, Brent, Dennis E. Breedlove, and Peter H. Raven. 1974. *Principles of Tzeltal Plant Classification*. New York: Academic.

- Berlin, Brent, and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas, and Susan Conrad. 2014. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
- Bresnan, Joan. 2001. *Lexical-functional Syntax*. Oxford: Blackwell.
- Brown, Gillian and George Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Chafe, Wallace L. 1970. *Meaning and the Structure of Language*. Chicago: University of Chicago Press.
- Chafe, Wallace L. 1973. Language and memory. *Language*, 49: 261–281.
- Chafe, Wallace L. 1974. Language and consciousness. *Language*, 50: 111–113.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In Charles N. Li (ed.), *Subject and Topic*, pp. 25–55. London/New York: Academic Press.
- Chafe, Wallace L. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *On Coherence and Grounding in Discourse*, pp. 21–51. Amsterdam: John Benjamins.
- Chafe, Wallace L. 1994. *Discourse, Consciousness and Time: the Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.
- Clancy, Patricia M. 1980. Referential choice in English and Japanese narrative discourse. In Wallace L. Chafe (ed.), *The Pear Stories: Cognitive, Cultural, and*

- Linguistic Aspects of Narrative Production*, pp. 127–202. Norwood, NJ: Ablex.
- Clark, Herbert H. 1977. Bridging. In Philip N. Johnson–Laird and Peter C. Wasan (eds.), *Thinking: Reading in Cognitive Science*, pp. 411–420. Cambridge: Cambridge University Press.
- Clark, Herbert H. 1978. Inferring what is meant. In Willem J. M. Levelt and Giovanni Flores d’Arcais (eds.), *Studies in the Perception of Language*, pp. 295–321. London: John Wiley and Sons.
- Clark, Herbert H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press and Center for the Study of Language and Information.
- Clark, Herbert H. and Susan E. Haviland. 1974. Psychological Processes in Linguistic Explanation. In David Cohen (ed.), *Explaining Linguistic phenomena*, pp. 91–124. Washington, D.C.: Hemisphere.
- Clark, Herbert H. and Susan Haviland, E. 1977. Comprehension and the given–new contrast. In Roy O. Freedle (ed.), *Discourse Production and Comprehension*, pp.1–40. Norwood, New Jersey: Ablex Publishing Corporation.
- Clark, Herbert H. and Deanna Wilkes–Gibbs. 1986. Referring as a Collaborative Process. *Cognition*, 22: 1–39.
- Clark, Herbert H. and Catherine Marshall R. 1981. Definite reference and mutual knowledge. In Aravind K.Joshi, Bonnie L. Webber, and Ivan A. Sag (eds.), *Elements of Discourse Understanding*, pp. 10–63. Cambridge: Cambridge University Press.
- D’Andrade, Roy G. 1992. Schemas and motivation. In Roy G. D’Andrade and Claudia Strauss (eds.), *Human Motives and Culture Models*, pp. 23–44. Cambridge: Cambridge University Press.
- Daneš, František. 1974. Functional sentence perspective and the organization of the text. In František Daneš (ed.), *Papers on Functional Sentence Perspective*, pp. 106–128. Czecholovakia: Academia, Publishing House of the Czecholovak Academy of Sciences.
- Davies, Catherine and Jennifer Arnold. 2019. Reference and informativeness: How

- context shapes referential choice. In: Chris Cummins and Napoleon Katsos (eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics*, pp. 474–493. Oxford: Oxford University Press.
- Davies, Catherine and Napoleon Katsos. 2013. Are speakers and listeners ‘only moderately Gricean’? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, 49 (1): 78–106.
- de Haan, Pieter. 1989. *Postmodifying clauses in the English noun phrase: A corpus-based analysis*. Amsterdam: Rodopi.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17: 463–489.
- Diessel, Holger. 2008. Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, 19: 457–482.
- Dik, Simon C. 1997a. *The Theory of Functional Grammar 1: The Structure of the Clause*. Berlin: De Gruyter Mouton.
- Dik, Simon C. 1997b. *The Theory of Functional Grammar 2: Complex and Derived Constructions*. Berlin: De Gruyter Mouton.
- Doherty, Monika. 2006. *Structural Propensities: Translating Nominal Word Groups from English into German*. Amsterdam/Philadelphia: John Benjamins.
- Downing, Angela. 1995. Thematic layering and focus assignment in Chaucer’s General Prologue to The Canterbury Tales. In Mohsen Ghadessy (ed.), *Thematic Development in English Texts*, pp. 147–163. London/New York: Pinter Publishers.
- Fauconnier, Gilles. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.
- Fillmore, Charles J. 1975. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, pp. 123–131.
- Fillmore, Charles J. 1977. Scenes–and–frames semantics. In Antonio Zampolli (ed.), *Linguistic Structures Processing*, pp. 55–81. Amsterdam, New York and Oxford: North-Holland Publishing Company.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di*

Semantica, 6(2): 222–254.

- Fillmore, Charles, Nancy Ide, Daniel Jurafsky, and Catherine Macleod. 1998. An American national corpus: A proposal. *Proceedings of the First International Language Resources and Evaluation Conference (LREC)*, pp. 965–970.
- Firbas, Jan. 1964. From Comparative Word–order studies: (thoughts on V. Mathesius’ conception of the word–order system in English compared with that in Czech). *Brno Studies in English*, (4): 111–128.
- Firbas, Jan. 1972. On the interplay of prosodic and non–prosodic means of functional sentence perspective. In Vilém Fried (ed.), *The Prague School of Linguistics and Language Teaching*, pp. 77–94. London: Oxford University Press.
- Firbas, Jan. 1986. On the dynamics of written communication in the light of the theory of Functional Sentence Perspective. In Charles R. Cooper and Sydney Greenbaum, (eds.), *Studying Writing: Linguistic Approaches*, pp. 40–71. Beverly Hills, CA: Sage Publications.
- Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
- Fontaine, Lise, Katy Jones and David Schönthal. 2023. *Referring in Language: An Integrated Approach*. Cambridge: Cambridge University Press.
- Ford, Cecilia. 1993. *Grammar in Interaction: Adverbial Clauses in American English Conversations*. Cambridge: Cambridge University Press.
- Francis, Gill. 1987. *Anaphoric Nouns*. Birmingham: University of Birmingham.
- Francis, Gill. 1994. Labelling discourse: an aspect of nominal–group lexical cohesion. In Malcolm Coulthard (ed.), *Advances in Written Text Analysis*, pp. 83–101. London: Routledge.
- Francis, Gill. 1989. Thematic selection and distribution in written discourse. *Word*, 40 (1–2): 201–221.
- Francis, Gill, Elizabeth Manning, and Susan Hunston. 1998. *Grammar Patterns II: Nouns and Adjectives*. London: Harper Collins.
- Fries, Peter H. 1995a. A personal view of Theme. In Mohsen Ghadessy (ed.), *Thematic*

- Development in English Texts*, pp. 1–19. London: Pinter.
- Fries, Peter H. 1995b. Themes, methods of development and texts. In Ruqaiya Hasan and Peter H. Fries (eds.), *On Subject and Theme: From the Perspective of Functions in Discourse*, pp. 317–359. Amsterdam: John Benjamins.
- Garvin, Paul L. (ed.) 1964. *A Prague School Reader on Esthetics, Literary structure, and Style*. Washington, D.C.: Georgetown University Press.
- Ghadessy, Mohsen. (ed.) 1995. *Thematic Development in English Text*. London: Pinter.
- Givón, Talmy. 2001. *Syntax: An Introduction. Volume I*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross–Language Studies*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 1984. Universals of discourse structure and second language acquisition. In William Rutherford (ed.), *Language Universals and Second Language Acquisition*, pp. 109–136. Amsterdam/Philadelphia: John Benjamins.
- Gómez–González, María Ángeles. 2001. *The Theme–Topic Interface: Evidence from English*. Amsterdam/Philadelphia: John Benjamins.
- Graesser, Arthur C., Peter Wiemer–Hastings and Katja Wiemer–Hastings. 2001. Constructing inferences and relations during text comprehension. In Ted Sanders, Joost Schilperoord and Wilbert Spooren (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects*, pp. 249–271. Amsterdam/Philadelphia: John Benjamins.
- Graesser, Arthur C., Murray Singer, and Tom Trabasso. 1994. Constructing inferences during narrative text comprehension. *Psychological Review*, 101: 371–395.
- Grice Paul H. 1975. Logic and conversation. In: Peter Cole and Jerry L.Morgan (eds.), *Syntax and Semantics Vol. 3: Speech Acts*, pp. 41–58. London: Academic Press.
- Goutsos, Dionysis. 1996. A model of sequential relations in expository text. *Text*, 16(4): 501–533.
- Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21: 203–225.

- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2): 274–307.
- Gundel, Jeannette K. 2003. Information structure and referential givenness/ newness: how much belongs in the Grammar. *Proceedings of the HPSG03 Conference, Michigan State University, East Lansing, CSLI Publications*. <https://esli-publications.stanford.edu>.
- Halliday, Michael A. K. 1967. Notes on Transitivity and Theme in English, Part II. *Journal of Linguistics*, 3: 199–244.
- Halliday, Michael A. K. 1979. Modes of meaning and modes of expression: types of grammatical structure and their determination by different semantic functions. In: David J. Allerton, Edward Carney, and David Holdcroft (eds.), *Function and Context in Linguistic Analysis: A Festschrift for William Haas*, pp. 57–79. Cambridge: Cambridge University Press.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar (Second Edition)*. London: Edward Arnold.
- Halliday, Michael A. K. and Christian M. I. M. Matthiessen. 2014. *An Introduction to Functional Grammar (Fourth Edition)*. London: Routledge.
- Halliday, Michael A. K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. and Ruqaiya Hasan. 2012. *Language, Context, and Text: Aspects of Language in a Social–Semiotic Perspective*. Shanghai: World Publishing Corporation.
- Harris, James. 1771. *Hermes: or a Philosophical Inquiry concerning Language and Universal Grammar*. London: Printed for J. Nourse and P. Vaillan.
- Hasan, Ruqaiya. 1999. Speaking with reference to context. In Mohsen Ghadessy (ed.), *Text and Context in Functional Linguistics*, pp. 219–328. Amsterdam/Philadelphia: John Benjamins.
- Hasan, Ruqaiya, and Peter H. Fries. 1995. *Reflections on Subject and Theme: An Introduction*. Amsterdam/Philadelphia: John Benjamins.

- Hasselgård, Hilde. 2004. Temporal and spatial adjuncts as elements of texture. In David Banks (ed.), *Text and Texture, Systemic Functional Viewpoints on the Nature and Structure of Text*, pp. 141–158. Paris: L'Harmattan.
- Hasselgård, Hilde. 2010. *Adjunct Adverbials in English*. Cambridge: Cambridge University Press.
- Haviland, Susan E., and Herbert H. Clark 1974. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13: 512–521.
- Hawkins, John. 1978. *Definiteness and Indefiniteness*. London: Croom Helm.
- Hawkins, John. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, Roger. 1981. Towards an account of the possessive Constructions: NP's N and N of NP. *Journal of Linguistics*, 17: 247–269.
- Heim, Irene R. 1982. The Semantics of Definite and Indefinite Noun Phrases. PhD dissertation. University of Massachusetts.
- Hengeveld, Kees, and Lachlan J. Mackenzie. 2006. Functional discourse grammar. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics (Second Edition) Vol 4*, pp. 668–76. Oxford: Elsevier.
- Herbst, Thomas. 1988. A valency model for nouns in English. *Journal of Linguistics*, 24: 265–301.
- Holland, Dorothy, and Naomi Quinn (eds). 1987. *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press.
- Horová, Eva. 1976. On position and function of English local and temporal adverbials. *Brno Studies in English*, 12: 93–123.
- Huang, Yan. 2012. Relevance and neo-Gricean pragmatic principles. In Hans-Jörg Schmid (ed.), *Cognitive Pragmatics*, pp. 25–46. Berlin/New York: De Gruyter Mouton.
- Huddleston, Rodney D. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.

- Huddleston, Rodney, and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hyland, Ken. 2001. Bringing in the reader: Addressee features in academic articles. *Written Communication*, 18(4): 549–574.
- Ide, Nancy. 2008. The American National Corpus: Then, Now, and Tomorrow. In Michael Haugh, Kate Burridge, Jean Mulder and Pam Peters (eds.), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages*, pp. 108–113. MA: Cascadilla Proceedings Project, Somerville.
- Ide, Nancy, Randi Reppen, and Keith Suderman. 2002. The American National Corpus: More than the web can provide. *Proceedings of the Third Language Resources and Evaluation Conference (LREC)*, pp. 839–844. Spain: Las Palmas, Canary Islands.
- Ide, Nancy, and Keith Suderman. 2004. The American National Corpus First Release. *Proceedings of the Fourth Language Resources and Evaluation Conference (LREC)*, pp. 1681–1684. Portugal: Lisbon.
- Ide, Nancy, Collin Baker, Christiane Fellbaum, Rebecca Passonneau. 2010. The Manually Annotated Sub-Corpus: A community resource for and by the people. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 68–73. Sweden: Uppsala.
- Ivanič, Roz. 1991. Nouns in search of a context: a study of nouns with both open- and closed system characteristics. *International Review of Applied Linguistics in Language Teaching*, 2: 93–114.
- Johnson-Laird, Philip N. 1983. *Mental Models*. Cambridge, MA: Harvard University Press.
- Jones, Katy S. 2014. Towards an Understanding of the Use of Indefinite Expressions for Definite Reference in English Discourse. PhD Thesis, Cardiff University.
- Kaplan, David. 1989. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, pp. 481–563. Oxford:

Oxford University Press.

Kay, Paul, and Chad K. McDaniel. 1978. The linguistic significance of the meaning of basic color terms. *Language*, 54: 610–646.

Kecskes, Istvan. 2000. A cognitive–pragmatic approach to situation–bound utterances. *Journal of Pragmatics*, 32(6): 605–625.

Kecskes, Istvan. 2002. *Situation–Bound Utterances in L1 and L2*. Berlin/New York: De Gruyter Mouton.

Kecskes, Istvan. 2010. Situation–bound utterances as pragmatic acts. *Journal of Pragmatics*, 42(11): 2889–2897.

Kecskes, Istvan. 2012. Encyclopaedic knowledge and cultural models. In Hans–Jörg Schmid (ed.), pp. 175–197. *Cognitive Pragmatics*, Berlin/New York: De Gruyter Mouton.

Keenan, Janice M., Susan D. Baillet, and Polly Brown. 1984. The effects of causal cohesion on comprehension and memory. *Journal of Memory and Language*, 23: 115–126.

Keizer, Evelien. 2004. Postnominal PP complements and modifiers: a cognitive distinction. *English Language and Linguistics*, 8(2) 323–350.

Keizer, Evelien. 2007. *The English Noun Phrase: the Nature of Linguistic Categorization*. Cambridge: Cambridge University Press.

Kilpert, Diana. 2001. Seeing the patterns: systemic functional grammar and method of development in texts. *SAALT Journal for Language Teaching*, 35(1): 52–70.

Kintsch, Walter. 1970. *Learning, Memory and Conceptual Processes*. New York: Wiley.

Kintsch, Walter, and Teun van Dijk. 1978. Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

Kintsch, Walter. 1998. *Comprehension: A Paradigm for Cognition*. Cambridge, MA: Cambridge University Press.

Koetsenruijter, A. Willem M. 2011. Using Numbers in News Increases Story Credibility. *Newspaper Research Journal*, 32 (2): 74–82.

Kopple, William J. 1991. Themes, thematic progressions, and some implications for

- understanding discourse. *Written Communication*, 8: 311–347.
- Kuno, Susumu. 1972. Functional sentence perspective. *Linguistic Inquiry*, 3: 269–320.
- Kuno, Susumu. 1974. Lexical and contextual meaning. *Linguistic Inquiry*, 5: 469–477.
- Kuno, Susumu. 1978. Generative discourse analysis in America. In Wolfgang U. Dressler (ed.), *Current Trends in Textlinguistics*, pp. 275–294. Berlin/New York: Academic Press.
- Kunz, Kerstin. 2010. Variation in English and German Nominal Coreference: A Study of Political Essays. PhD thesis, Saarland University.
- Kunz, Kerstin, and Ekaterina Lapshinova–Koltunski. 2018. English vs. German from a textual perspective: Looking inside chain intersection. *Bergen Language and Linguistics Studies*, 9(1): 21–42.
- Labov, William. 2004. The boundaries of words and their meanings. In Bas Aarts, David Denison, Evelien Keizer, and Gergana Popova (eds.), *Fuzzy Grammar: A Reader*, pp. 67–80. Oxford: Oxford University Press.
- Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Vol 2, Descriptive Applications*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2004. Remarks on nominal grounding. *Functions of Language*, 11(1), 77–113.
- Lawson, Bredan T., and Matthew Lovatt. 2020. Towards a rhetorical understanding of statistics in politics: Quantifying the National Health Service ‘Winter Crisis’. *European Journal of Communication*, 36: 110–124.
- Levinson, Stephen C. 1989. A review of Relevance. *Journal of Linguistics*, 25: 455–472.
- Liu, Lijin, and Tucker Gordon. 2015. Thematic choice and progression in English and

- Chinese radio news texts: a systemic functional analysis. *Text and Talk*, 35: 481–504.
- Loock, Rudy. 2013. Extending further and refining Prince's taxonomy of given/new information: A case study of non-restrictive, relevance-oriented structures. *Pragmatics*, 23 (1): 69–91.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Martin, James R. 1992. *English Text: System and Structure*. Amsterdam/Philadelphia: John Benjamins.
- Mathesius, Vilém. 1975. *A Functional Analysis of Present Day English on a General Linguistic Basis*. The Hague: Mouton.
- Matthiessen, Christian M.I.M. 1995. *Lexicogrammatical Cartography: English Systems*. Tokyo: International Language Sciences Publishers.
- Matthiessen, Christian M.I.M. 1992. Interpreting the textual metafunction. In Martin Davies, and Louise Ravelli (eds.), *Advances in Systemic Linguistics: Recent Theory and Practice*, pp. 37–82. London: Pinter.
- Matthiessen, Christian M.I.M. 1995. THEME as an enabling resource in ideational 'knowledge' construction. In Mohsen Ghadessy (ed.), *Thematic Development in English Texts*, pp. 20–55. London/New York: Pinter.
- Matthiessen, Christian M.I.M., Kazuhiro Teruya, and Marvin Lam. 2010. *Key Terms in Systemic Functional Linguistics*. London: Continuum.
- Mauranen, Anna. 1996. Discourse Competence – Evidence from Thematic Development in Native and Non-native Texts. In Eija Ventola and Anna Mauranen (eds.), *Academic Writing Intercultural and Textual Issues Vol 41*, pp. 195–230. Amsterdam/Philadelphia: John Benjamins.
- McGloskey, Michael E., and Sam Glucksburg. 1978. Natural categories: well-defined or fuzzy sets? *Memory and Cognition*, 6: 462–472.
- McConway, Kevin. 2016. Statistics and the media: A statistician's view. *Journalism*, 17: 49–65.
- Merriam Allen H. 1990. Words and numbers: mathematical dimensions of rhetoric. *The*

- Southern Communication Journal*, 55 (4): 337–354.
- Minsky, Marvin. 1975. A framework for representing knowledge. In Patrick H. Winston (ed.), *The Psychology of Computer Vision*, pp. 211–277. New York: McGraw-Hill.
- Müller, Christopher, and Michael Strube. 2006. Multi-Level Annotation of Linguistic Data with MMAX2. In: Sabine Braun, Kurt Kohn, and Joybrato Mukherjee (eds.), *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*, pp. 197–214. Frankfurt: Peter Lang.
- Neumann, Stella. 2003. *Textsorten und Übersetzen: Eine Korpusanalyse englischer und deutscher Reiseführer*. Frankfurt: Peter Lang Verlag.
- Neumann, Stella. 2010. Quantitative register analysis across languages. In Elizabeth Swain (ed.), *Thresholds and Potentialities of Systemic Functional Linguistics: Multilingual, Multimodal and Other Specialised Discourses*, pp. 85–113. EUT: Trieste cop.
- Neumann, Stella. 2014. *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. Berlin: de Gruyter Mouton.
- O' Brien, Edward J., and Jason E Albrecht. 1991. The role of context in accessing antecedents in text. *Journal of Experimental Psychology: Learning, Memory and Cognition*, (17): 94–102.
- Peng, Xuanwei. 2014. Functional Stylistics in the TEXTUAL respect: message organization patterns for literary discourse. *Linguistics and the Human Sciences*, (10)3: 263–294.
- Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication, *Proceedings of the National Academy of Sciences*, (108)9: 3526–3529. <https://doi.org/10.1073/pnas.1012551108>.
- Poesio, Massimo. 2004. The MATE/GNOME Proposals for Anaphoric Annotation, Revisited. *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL*, pp. 154–162. <https://aclanthology.org/W04-2327>.
- Poesio, Massimo, and Renata Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2): 183–216.

- Prince, Ellen F. 1981. Toward a taxonomy of given–new information. In Peter Cole (ed.), *Radical Pragmatics*, pp. 223–255. New York: Columbia University Press.
- Prince, Ellen F. 1992. The ZPG Letter: Subjects, definiteness, and information status. In William C. Mann, and Sandra A. Thompson (eds.), *Discourse Description: Diverse Linguistic Analysis of a Fund–raising Text*, pp. 295–325. Amsterdam/Philadelphia: John Benjamins.
- Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rambow, Owen. 1993. *Pragmatic Aspects of Scrambling and Topicalization in German: A Centering Approach*. MS: University of Pennsylvania.
- Riester, Arndt. 2008a. A semantic explication of ‘information status’ and the underspecification of the recipient’s knowledge. In Atle Grønn (ed.), *Proceedings of Sinnund Bedeutung 12*, pp. 508–522. Oslo.
- Riester, Arndt. 2008b. *The Components of Focus and their Use in Annotating Information Structure*. Ph.D Thesis, Universität Stuttgart..
- Riester, Arndt, and Stephan Baumann. 2017. The RefLex Scheme – Annotation Guidelines. Vol 14 of SinSpeC. *Working Papers of the SFB 732*, pp. 1–31. University of Stuttgart.
- Rijkhoff, Jan. 1989. The identification of referents. In: Simon C. Dik and John Connolly (eds.), *Functional Grammar and the Computer*, pp. 229–246. Dordrecht: Foris.
- Rijkhoff, Jan. 2002. *The Noun Phrase*. Oxford: Oxford University Press.
- Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology*, 4, 328–50.
- Rosch, Eleanor H. 1978. Principles of categorization. In Eleanor Rosch and Barbara B. Lloyd (eds.), *Cognition and categorization*, pp. 27–48. Hillsdale NJ: Lawrence Erlbaum.
- Rosch, Eleanor H. 1983. Prototype classification and logical classification: the two systems. In Ellin K. Scholnick (ed.), *New Trends in Cognitive Representation: Challenges to Piaget’s Theory*, pp. 73–86. Hillsdale, NJ: Lawrence Erlbaum.
- Rumelhart, David E. 1975. Notes on a Schema for Stories. In Daniel G. Bobrow and

- Allan M. Collins (eds.), *Representation and Understanding: Studies in Cognitive Science*, pp. 211–36. New York/San Francisco/London: Academic Press.
- Rumelhart, David E. 1980. Schemata: the building blocks of cognition. In Rand J. Spiro, Bertram C. Bruce and William F. Brewer (eds.), *Theoretical Issues in Reading Comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education*, pp. 33–58. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rumelhart, David E., and Andrew Ortony. 1977. The representation of knowledge in memory. In Richard C. Anderson, Rand J. Spiro and William E. Montague (eds.), *Schooling and the Acquisition of Knowledge*, pp. 99–135. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sanders, Ted J. M., 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, 24: 119–148.
- Sanders, Ted J. M., and Anneloes R. Canestrelli. 2012. The processing of pragmatic information in discourse. In Hans-Jörg Schmid (ed.), *Cognitive Pragmatics*, pp. 201–231. Berlin/New York: De Gruyter Mouton.
- Sanders, Ted J. M., Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes*, 15: 1–36.
- Sanford, Antony J., and Simon C. Garrod. 1981. *Understanding Written Language: Explorations of Comprehension beyond the Sentence*. Chichester: John Wiley and Sons.
- Saux, Gaston, Mary A. Britt, Vibert Nicolas, Jean-François Rouet. 2021. Building mental models from multiple texts: How readers construct coherence from inconsistent sources. *Language and Linguistic Compass*, 15(3): 1–19.
- Schank, Roger C., and Robert P. Abelson. 1977. *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin/Boston: De Gruyter Mouton.

- Schober, Michael F., and Herbert H. Clark. 1989. Understanding by Addressees and Overhearers. *Cognitive Psychology*, 21: 211–32.
- Schönthal, David. 2016. On the Multifaceted Nature of English *of*-NPs: A Theoretical, Corpus, Cotextual and Cognitive Approach. PhD Thesis, Cardiff University.
- Sperber, Dan, and Deirdre Wilson. 1986/95. *Relevance: Communication and Cognition (2nd edition)*, Oxford: Blackwell.
- Stensberg, Morten T. 2021. In numbers we trust – the use of statistics in political news. *Journalism Practice, Latest Articles*.
<https://doi.org/10.1080/17512786.2021.1987298>.
- Strube, Michael, and Udo Hahn. 1999. Functional Centering: Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(3): 309–344.
- Sun, Chao-Fen, and Talmy Givón. 1985. On the so-called SOV word-order in Mandarin Chinese: A quantitative text study and its implications. *Language*, 61(2): 329–351.
- Tannen, Deborah. 1979. What's in a frame? Surface evidence for underlying expectations. In Roy O. Freedle (ed.), *New Directions in Discourse Processing*, pp. 137–181. Norwood: ALEX Publishing Company.
- Thompson, Sandra A. 1985. Grammar and written discourse: Initial vs final purpose clauses in English. *Text*, 5(1–2): 55–84.
- Traugott, Elizabeth C. 2008a. The Grammaticalization of NP of NP Patterns. In Alexander Bergs and Gabriele Diewald (eds.), *Constructions and Language Change*, pp. 23–45. Berlin: De Gruyter Mouton.
- Traugott, Elizabeth C. 2008b. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In Regine Eckardt, Gerhard Jäger and Tonjes Veenstra (eds.), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, pp. 219–250. Berlin: De Gruyter Mouton.
- Thompson, Geoff. 2007. Unfolding theme: the development of clausal and textual perspectives on theme. In Ruqaiya Hasan, Christian M.I.M Matthiessen, and

- Jonathan Webster (eds.), *Continuing Discourse on Language: A Functional Perspective Vol II*, pp. 671–696. London: Equinox Publishing Ltd.
- Trnavac, Radoslana, and Maite Taboada. 2016. Cataphora, backgrounding and accessibility in discourse. *Journal of Pragmatics*, (93): 68–84.
- van Dijk, Teun A. 1977. *Text and Context*. London: Longman.
- van Dijk, Teun A. and Walter Kintsch. 1983. *Strategies of Discourse Comprehension*. New York: Academic Press.
- van Dijk, Teun A. 1987. Episodic models in discourse processing. In Rosalind Horowitz and Samuels S. Jay (eds.), *Comprehending Oral and Written Language*, pp. 161–196. London: Academic Press.
- van Dijk, Teun A. 1988. *News as Discourse*. Hillsdale/New Jersey/Hove/London: Lawrence Erlbaum Associates.
- Ventola, Eija, and Anna Mauranen. 1991. Non–native writing and revising of scientific articles. In Eija Ventola (ed.), *Functional and Systemic Linguistics: Approaches and Uses*, pp. 457–492. Berlin: De Gruyter Mouton.
- Vieira, Renata, and Massimo Poesio. 2000. Processing definite descriptions in corpora. In Simon Botley and Tony McEnery (eds.), *Corpus–based and Computational Approaches to Anaphora*, pp. 189–212. London: UCL Press.
- Virtanen, Tuija. 1992. *Discourse functions of adverbial placement in English*. Abo: Abo Akademi University Press.
- von Heusinger, Klaus. 2002a. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics*, 19: 209–211.
- von Heusinger, Klaus. 2002b. The cross–linguistic implementations of specificity. In Katarzyna Jaszczolt and Ken Turner (eds.), *Meanings through Language Contrast, Vol. 2*, pp. 405–421. Amsterdam: Benjamins.
- Wannagat, Wienke, Gensine Waizenegger, and Gerhild Nieding. 2021. Coherence formation during narrative text processing: a comparison between auditory and audiovisual text presentation in 9– to 12–year–old children. *Cognitive Process*, 22: 299–310.

- Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change*, 9: 81–105.
- Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann and Yakov Malkiel (eds.), *Directions for Historical Linguistics*, pp. 95–195. Austin: University of Texas Press.
- Wiechmann, Daniel, and Elma Kerz. 2013. The positioning of concessive adverbial clauses in English: assessing the importance of discourse-pragmatic and processing-based constraints. *English Language and Linguistics*, 17(1): 1–23.
- Wilson, Deirdre. 2010. Relevance theory. In Louise Cummings (ed.), *The Pragmatics Encyclopaedia*, pp. 394–398. London: Routledge.
- Winston, Morton E., Roger Chaffin and Douglas Herrmann. 1987. A taxonomy of part–whole relations. *Cognitive Science*, (11): 417–444.
- Winter, Eugene O. 1992. The notion of unspecific versus specific as one way of analyzing the information of a fund–raising letter. In William C. Mann and Sandra A. Thompson (eds.), *Discourse Descriptions: Diverse Analysis of a Fund–raising Text*, pp. 131–170. Amsterdam/Philadelphia: John Benjamins.
- Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Translated by Gertrude Elizabeth Margaret Anscombe. Oxford: Blackwell.
- Xu, Hongying, and Xuanwei Peng. 2020. Why such word order as “on the stage were sitting X”/“Y died his father”? A functional account of message–salient constructions in modern Chinese. *Journal of Foreign Languages*, (6): 52–63.
- Xu, Hongying, and Xuanwei Peng. 2021. Information value system introduced by *Dui(yu)* in mandarin. *Modern Foreign Languages*, 44 (6): 767–778.
- Yule, George. 1981. New, current and displaced entity reference. *Lingua*, 55: 41–52.

Appendix

Appendix A: Raw frequencies of varied linguistic forms of nominal expressions representing types of information in English texts

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	13	1	1	32	12	59
Tg2	13	9	1	36	10	69
Np1	10	11	0	26	2	49
Np2	1	2	0	6	5	14
Np3	4	5	0	2	7	18
Np4	3	2	0	5	8	18
Np5	8	4	0	14	10	36
Np6	6	4	0	10	8	28
Gd	48	18	0	129	76	271
Ey	15	21	4	55	18	113

Table A1: Raw frequencies of varied forms of nominals representing Brand-new unanchored information per text

N-type Text	pre-de.	dem.	article- <i>a</i>	article- <i>the</i>	other	Sum
Tg1	0	0	13	8	11	32
Tg2	0	1	19	14	2	36
Np1	0	0	11	5	10	26
Np2	0	0	1	5	1	7
Np3	0	0	1	1	0	2
Np4	0	0	3	1	1	5
Np5	0	0	5	0	9	14
Np6	0	0	7	0	3	10
Gd	2	2	37	32	56	129
Ey	0	0	29	12	14	55

Table A2: Raw frequencies of nominals with varied determiners representing Brand-new unanchored information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	0	0	0	9	2	11
Tg2	0	0	0	6	1	7
Np1	0	0	0	7	5	12
Np2	0	0	0	1	0	1
Np3	0	0	0	3	0	3
Np4	0	0	0	2	0	2
Np5	0	0	0	3	0	3
Np6	0	0	0	1	0	1
Gd	0	0	0	14	12	26
Ey	0	0	0	4	4	8

Table A3: Raw frequencies of varied forms of nominals representing Brand-new anchored information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	18	0	1	26	8	53
Tg2	18	0	3	26	15	62
Np1	4	0	1	21	8	34
Np2	4	0	1	2	1	8
Np3	2	0	0	10	2	14
Np4	0	0	0	2	2	4
Np5	0	0	0	4	4	8
Np6	0	0	0	5	2	7
Gd	56	4	10	203	127	400
Ey	14	0	15	78	21	128

Table A4: Raw frequencies of varied forms of nominals representing non-containing Inferrable information per text

N-type Text	pre-de.	poss.	dem	article- <i>a</i>	article- <i>the</i>	other-det.	Sum
Tg1	0	7	0	7.69	19	0	26
Tg2	0	0	0	19.23	24	2	26
Np1	0	7	0	4.76	11	3	21
Np2	0	1	0	0	1	0	2
Np3	0	0	0	20	9	1	10
Np4	0	0	0	0	0	2	2
Np5	0	1	0	0	1	2	4
Np6	0	2	0	20	3	0	5
Gd	1	41	1	10.84	133	27	203
Ey	0	17	0	19.23	43	18	78

Table A5: Raw frequencies of nominals with varied determiners representing non-containing Inferrable information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	0	0	0	2	1	3
Tg2	0	0	0	5	1	6
Np1	0	0	0	0	1	1
Np2	0	0	0	1	0	1
Np3	0	0	0	0	0	0
Np4	0	0	0	3	0	3
Np5	0	0	0	1	0	1
Np6	0	0	0	0	1	1
Gd	0	0	0	58	22	80
Ey	0	0	0	11	6	17

Table A6: Raw frequencies of varied forms of nominals representing containing Inferrable information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	0	17	0	1	1	19
Tg2	1	32	0	6	1	40
Np1	0	21	0	0	0	21
Np2	0	5	0	0	0	5
Np3	0	8	0	0	0	8
Np4	0	8	0	0	0	8
Np5	0	7	0	0	0	7
Np6	0	4	0	0	0	4
Gd	1	31	0	22	3	57
Ey	0	20	0	4	1	25

Table A7: Raw frequencies of varied forms of nominals representing Unused information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	0	0	1	0	0	1
Tg2	0	0	1	1	0	2
Np1	0	0	1	3	0	4
Np2	0	0	0	1	0	1
Np3	0	0	0	0	0	0
Np4	0	0	0	0	0	0
Np5	0	0	0	0	0	0
Np6	0	0	0	0	0	0
Gd	0	0	1	3	0	4
Ey	0	0	1	1	0	2

Table A8: Raw frequencies of varied forms of nominals representing Situationally evoked information per text

Infor Text	bare	proper	pron.	det.	other	Sum
Tg1	4	16	12	12	1	45
Tg2	13	31	16	22	0	82
Np1	0	20	19	27	0	66
Np2	2	4	2	6	0	14
Np3	1	15	11	4	1	32
Np4	0	7	0	5	5	17
Np5	2	4	9	19	3	37
Np6	1	12	3	7	0	23
Gd	83	83	184	270	82	702
Ey	66	26	120	61	16	289

Table A9: Raw frequencies of varied forms of nominals representing Textually evoked information per text

N-type Text	possessive	personal	demonstrative	reflexive	reciprocal	other-pron.	Sum
Tg1	0	12	0	0	0	0	12
Tg2	1	13	1	1	0	0	16
Np1	0	16	3	0	0	0	19
Np2	0	2	0	0	0	0	2
Np3	0	11	0	0	0	0	11
Np4	0	0	0	0	0	0	0
Np5	0	8	0	0	0	1	9
Np6	0	3	0	0	0	0	3
Gd	2	160	14	7	0	1	184
Ey	0	101	8	4	0	7	120

Table A10: Raw frequencies of varied forms of pronouns representing Textually evoked information per text

N-type Text	pre-de.	poss.	dem	article- <i>a</i>	article- <i>the</i>	other-det.	Sum
Tg1	1	0	3	1	7	0	12
Tg2	0	2	5	1	13	1	22
Np1	0	0	1	0	26	0	27
Np2	0	0	0	1	5	0	6
Np3	0	0	0	0	4	0	4
Np4	0	0	1	0	2	2	5
Np5	0	4	1	1	12	1	19
Np6	0	0	0	0	7	0	7
Gd	0	36	63	15	152	4	270
Ey	1	9	11	18	14	8	61

Table A11: Raw frequencies of nominals with varied determiners representing Textually evoked information per text

Appendix B: Information distributions of nominal expressions and thematic structures

1. The information distributions of nominal expressions in clausal positions and texts for Section 6.1:

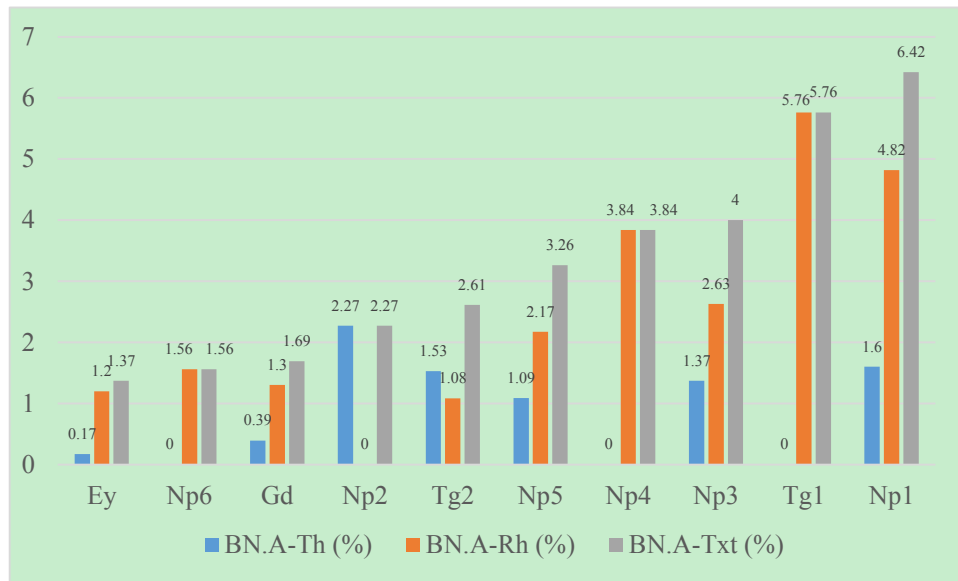


Figure B1: Brand-new anchored frequencies of the texts and their thematic structures

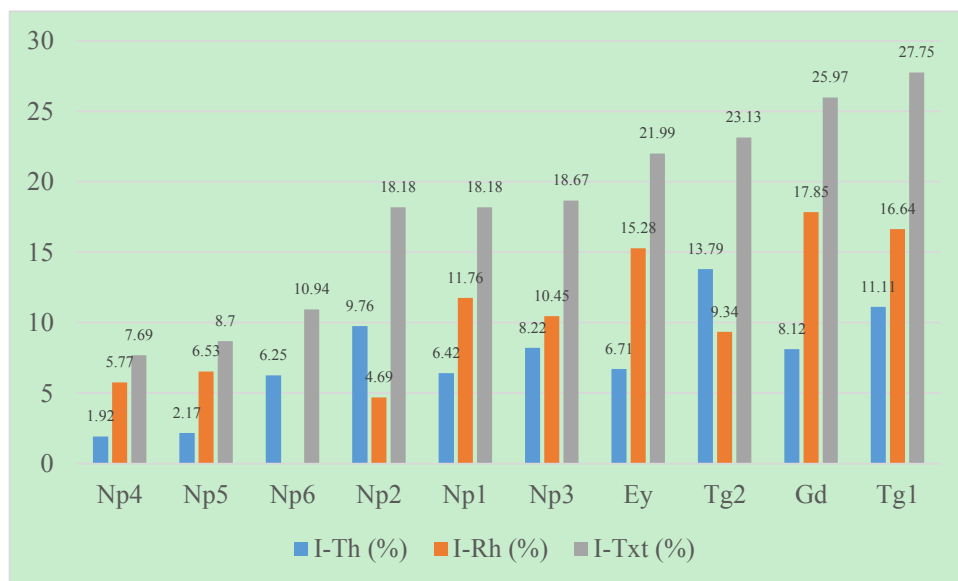


Figure B2: Non-containing frequencies of the texts and their thematic structures

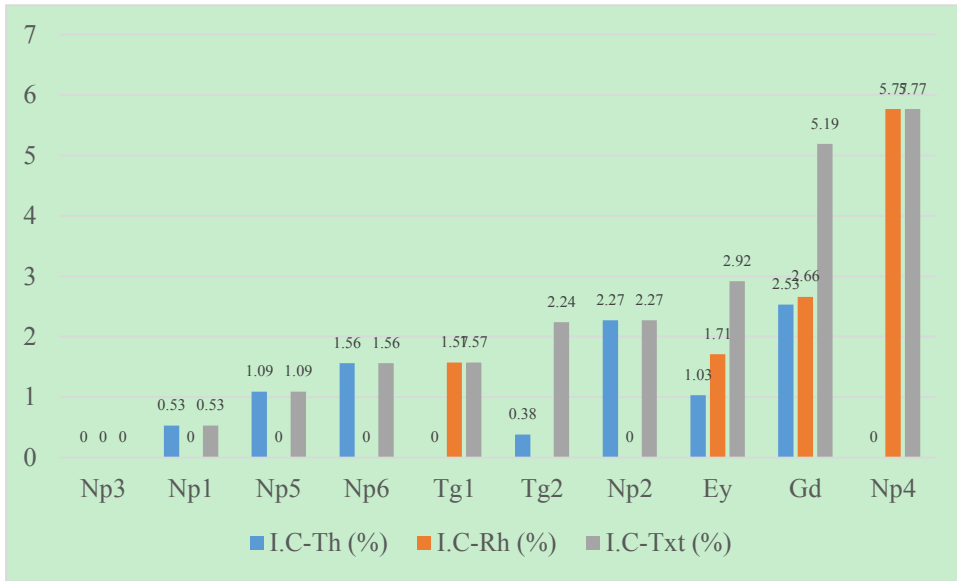


Figure B3: Containing Inferrable frequencies of the texts and their thematic structures

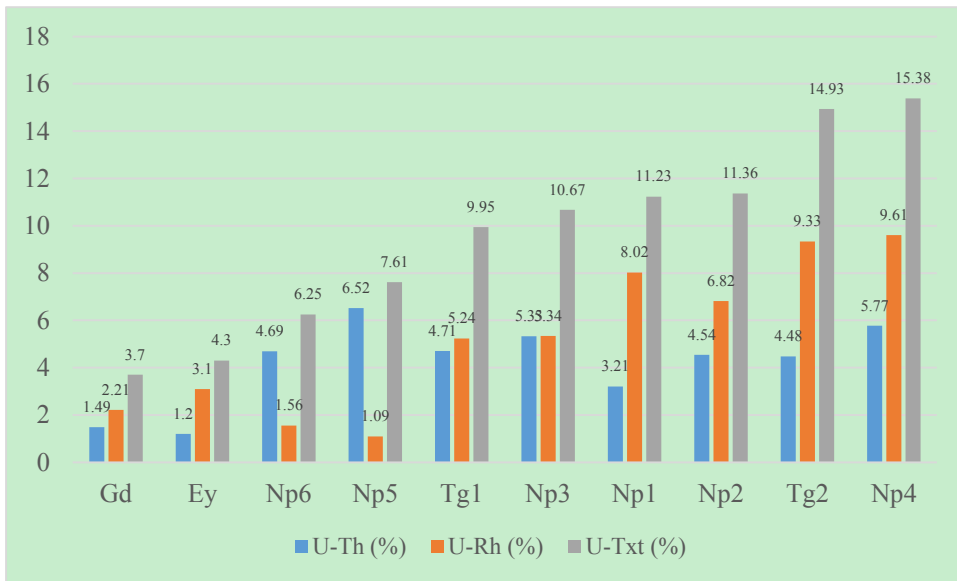


Figure B4: Unused frequencies of the texts and their thematic structures

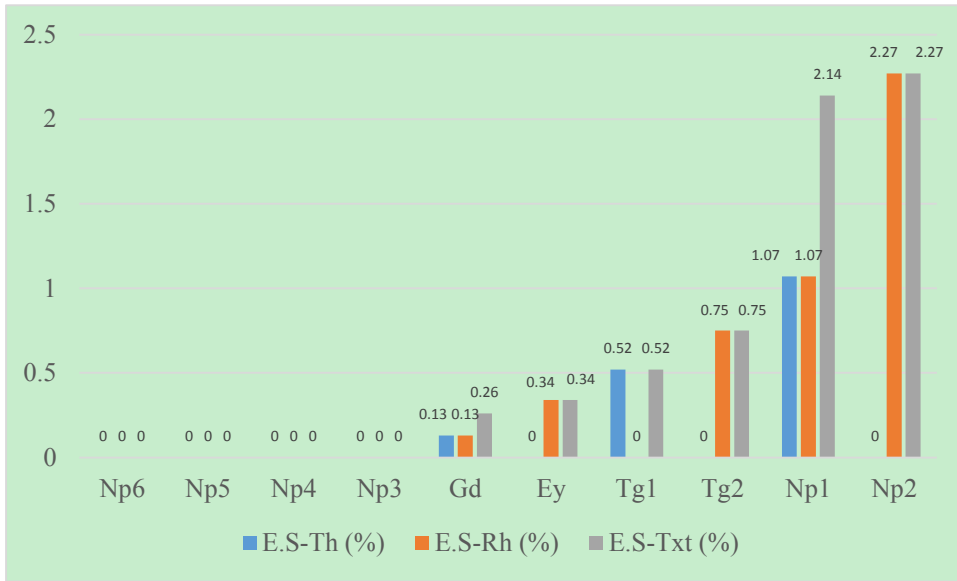


Figure B5: Situationally evoked frequencies of the texts and their thematic structures

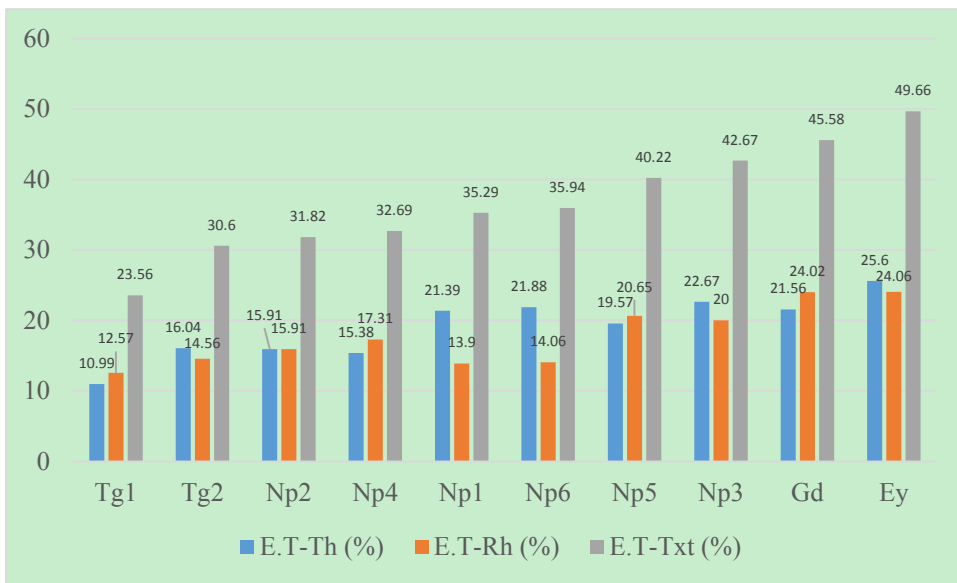


Figure B6: Textually evoked frequencies of the texts and their thematic structures

2. Raw frequencies of nominal expressions in Theme and Rheme positions for Section 6.2:

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	11	0	21	0	9	1	21	63
Tg2	14	4	36	1	12	0	43	110
Np1	11	3	12	1	6	2	40	75
Np2	0	1	4	1	2	0	7	15
Np3	1	1	6	0	4	0	17	29
Np4	1	0	1	0	3	0	8	13
Np5	11	2	2	1	6	0	18	40
Np6	5	0	4	1	3	0	14	27
Gd	52	6	125	39	23	2	332	579
Ey	25	1	39	6	7	0	149	227

Table B1: Raw frequencies of nominal expressions representing types of information in Theme position per text

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T
Tg1	6	0	17	0	6	1	15
Tg2	11	3	33	1	12	0	35
Np1	7	0	12	0	4	1	36
Np2	0	1	4	1	2	0	7
Np3	1	0	6	0	4	0	16
Np4	1	0	1	0	3	0	8
Np5	10	2	2	1	4	0	18
Np6	5	0	4	1	2	0	14
Gd	33	6	103	30	19	2	279
Ey	16	1	30	4	5	0	143

Table B2: Raw frequencies of nominal expressions representing types of information as Participants in Theme position per text

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T
Tg1	5	0	4	0	3	0	6
Tg2	3	1	3	0	0	0	8
Np1	4	0	0	1	2	1	4
Np2	0	0	0	0	0	0	0
Np3	0	1	0	0	0	0	1
Np4	0	0	0	0	0	0	0
Np5	1	0	0	0	2	0	0
Np6	0	0	0	0	1	0	0
Gd	19	0	22	9	4	0	53
Ey	9	0	9	2	2	0	6

Table B3: Raw frequencies of nominal expressions representing types of information as Circumstances in Theme position per text

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	48	11	31	3	9	0	24	126
Tg2	55	3	25	5	25	2	36	151
Np1	38	9	22	0	15	2	26	112
Np2	12	0	4	0	2	1	7	26
Np3	15	2	8	0	4	0	15	44
Np4	17	2	3	3	5	0	9	39
Np5	25	1	6	0	1	0	19	52
Np6	23	1	3	0	1	0	9	37
Gd	218	20	275	41	34	2	370	960
Ey	87	7	89	11	18	2	140	354

Table B4: Raw frequencies of nominal expressions representing types of information in Rheme position per text

Appendix C: Raw frequencies of postmodifiers representing types of information in different types of expansion

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	5	3	8	0	1	0	5	22
Tg2	3	0	6	0	4	1	5	19
Np1	4	0	1	0	5	2	10	22
Np2	2	0	0	0	1	1	2	6
Np3	1	0	0	0	0	0	3	4
Np4	6	0	1	0	2	0	2	11
Np5	1	0	2	0	0	0	6	9
Np6	0	0	0	0	1	0	0	1
Gd	19	2	46	2	11	3	106	189
Ey	6	1	11	1	0	1	19	39

Table C1: Raw frequencies of postmodifiers representing types of information in elaboration

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	0	0	0	0	0	0	0	0
Tg2	0	0	0	0	0	0	0	0
Np1	1	0	0	0	0	0	0	1
Np2	0	0	0	0	0	0	0	0
Np3	0	0	0	0	0	0	0	0
Np4	0	0	0	0	0	0	0	0
Np5	0	0	0	0	0	0	0	0
Np6	0	0	0	0	0	0	0	0
Gd	0	0	0	0	0	0	0	0
Ey	0	0	1	0	0	0	0	1

Table C2: Raw frequencies of postmodifiers representing types of information in extension

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	0	0	0	0	0	0	0	0
Tg2	0	0	0	0	0	0	2	2
Np1	0	0	0	0	0	0	0	0
Np2	0	0	0	0	0	0	0	0
Np3	0	0	0	0	0	0	0	0
Np4	0	0	0	0	0	0	0	0
Np5	0	0	0	0	0	0	0	0
Np6	0	0	0	0	0	0	0	0
Gd	2	0	2	0	0	0	5	9
Ey	1	0	0	0	0	0	2	3

Table C3: Raw frequencies of postmodifiers representing types of information in enhancement

Infor Text	BN	BN.A	I	I.C	U	E.S	E.T	Sum
Tg1	4	2	1	0	0	0	1	8
Tg2	4	0	2	0	4	0	1	11
Np1	1	2	1	0	6	0	1	11
Np2	1	0	0	0	2	0	1	4
Np3	0	0	0	0	3	0	3	6
Np4	1	0	2	0	3	0	4	10
Np5	4	0	1	0	2	0	4	11
Np6	0	0	0	0	0	0	0	0
Gd	11	1	26	1	4	0	55	98
Ey	4	0	3	0	7	0	9	23

Table C4: Raw frequencies of postmodifiers representing types of information in multiple expansion relations

Appendix D: Source texts

Travel guide

Tg1: Hong Kong and Its People

Exciting, mysterious, glamorous — these words have described Hong Kong for at least a century. With its vibrant atmosphere and night-and-day activity it is an intoxicating place. Hong Kong is crowded — it has one of the world's greatest population densities. But it is also efficient, with one of the best transportation systems anywhere, and for such a crowded place, quiet — you don't hear voices raised in anger, motorists sitting on their horns, or loud boomboxes. Shopping never ends — there's always another inviting spot just down the street. You'll find Hong Kong easy to get around, the people helpful, English spoken everywhere, and food that lives up to its reputation. On 1 July, 1997 the British Crown Colony of Hong Kong reverted to Chinese sovereignty as a Special Administrative Region of the People's Republic of China. Today Hong Kong remains a capitalist enclave with its laws and rights intact, and China has promised that Hong Kong will continue in this fashion for at least 50 years. Beijing's announced policy of maintaining Hong Kong's prosperity and stability makes sense. Hong Kong has long been China's handiest window the West, and the city is unrivaled in its commercial know-how and managerial expertise. Around the time of the transition there was much speculation about how things would change. But in fact, once news of the handover vanished from the front pages, the people of Hong Kong returned to their usual topics of conversation: the economy and the price of housing. The impression of the visitor today will be that very little has changed. Establishments are no longer called "Royal," Queen Elizabeth has vanished from the coinage, and the Union Jack has been replaced by the flag of China and the new Hong Kong flag with its bauhinia flower. But in fact, there have been changes, many of them due to economic progress, new construction, and other factors that influence cities all over the world. Others are more subtle. British social customs are still evident in the kind of polite service you get in hotels and in the long lines of people waiting for buses at rush hour.

The British population has decreased; today there are as many American and Australian ex-pats as there are British. With a population of nearly eight million and a total area of just over 1,095 square km (423 square miles), housing is one of Hong Kong's perennial nightmares. To alleviate the problem, the government has become the city's major landlord with the construction of massive apartment blocks that, though they have every modern facility, average only 9 square m (100 square ft) in size. Whole cities have been created in the New Territories, although the unimaginative architecture of these towns has been criticized. Of Hong Kong's population, 98 percent are Chinese. The majority are Cantonese, born in Hong Kong, or from South China, but there are immigrants from all over China. The Chinese people have been described as hardworking and pragmatic, attitudes that have contributed to Hong Kong's success. There are many stories of refugees who arrived with nothing in their pockets, set up a small sidewalk stall, worked diligently until they had their own store, and then expanded it into a modest chain. Old customs are still followed: Fate and luck are taken very seriously, and astrologers and fortune-tellers do a steady business. Before a skyscraper can be built, a feng shui investigation must take place to ensure that the site and the building will promote health, harmony, and prosperity. You'll also notice that gambling is a passion, whether it be cards, mah-jong, the lottery, or the horses. Hong Kong has two major racetracks as well as an intensive off-track betting system, and on weekends the ferries to Macau are crowded with people on their way to the casinos. Sightseeing in Hong Kong starts at sea level with the enthralling water traffic — a mix of freighters, ferries, tugs, junks, and yachts. Views of the city and the harbor are panoramic. From Victoria Peak, Hong Kong's highest point, or from skyscrapers and hotels, they are especially exciting at night when the lights are on. The business and financial center and the signature soaring architecture are on Hong Kong Island. Across Victoria Harbor, connected by ferry and the MTR rail line, is the Kowloon peninsula with its hotels, nightlife, and almost non-stop shopping. Beyond, in the New Territories, are a mixture of high-rise suburban towns, ancient sites and walled villages, country parks, and farms with ducks and fish ponds. Hong Kong's other, less developed islands, Lantau, Lamma, and Cheung Chau,

provide getaways. You can also take a ferry to Macau to find an entirely different kind of city, a unique blend of Chinese and Iberian culture. It's anyone's guess what may happen in the future, but for now Hong Kong bristles with energy and ambition, and for the visitor, this beautiful city with its contrasts and variety is an exhilarating experience.

Tg2: Dublin and the Dubliners

As capital of Europe's most explosive economy, Dublin seems to be changing before your very eyes. New construction is everywhere, the streets buzz, traffic is increasingly congested, and in the frenetic pace of rush hour everyone in Dublin seems intent on changing places with everyone else. At night the streets are crowded with people bent on having a good time. Prosperity is in the air; the roar of the "Celtic Tiger" can clearly be heard. But this is not the whole picture. The proverbial hospitality and warm welcome are still here. This busy, modern European city sits on a thousand years of history — history is present everywhere, from elegant Merrion Square to the bullet holes on the General Post Office. It's also a city of the imagination, reinvented and reappraised in the literature of its exiles. And the old Dublin is with us, too — the irreverent city of wit and charm and that peculiar magic possessed by Ireland and the Irish. Prosperity has brought with it a new emphasis on historic preservation. Dublin excels in packaging its past for the visitor. You can view artifacts from the Bronze Age, trace the history of the Easter Rising, or revisit Leopold Bloom's odyssey in *Ulysses*. Old buildings are being recycled; for example, the 17th-century Royal Hospital now holds the Museum of Modern Art. And Dublin, a city large in expectations, is still small enough for the visitor to see most of its sights on foot.

City on the Liffey

The River Liffey flows from west to east through the center of the city to Dublin Bay. The river forms a natural line between the north and south sections of the city. This geography is important in understanding Dublin. Historically and culturally this north-south distinction has always been significant, and it still is today, with a dose of good-

humored rivalry between the two areas. “I never go north of the Liffey,” one man remarked. Farther out, both north and south, are the sweeping curves of the Royal and Grand Canals. The occasional cry of gulls and unexpected distant vistas will remind you that Dublin is by the sea, and the Wicklow mountains, which hold Dublin closely to the coast, are visible from everywhere. Dublin is an intimate city, physically small but tightly packed, a perfect place for walking. College Green, the home of Trinity College, provides a natural focus just south of the O’Connell Street Bridge. O’Connell Street, the city’s grand boulevard, leads north to Parnell Square. To the south and east is St. Stephen’s Green and Georgian Dublin where the national museums are located. Along the Liffey to the west is Temple Bar, center of nightlife and home to many of Dublin’s cutting-edge artists and artistic endeavors. Up the hill from Temple Bar are Dublin Castle and Christ Church Cathedral. It won’t rain on you in Dublin all the time. The climate here can best be described as “changeable” and yet the sudden shifts from light to dark, sunshine to shower, are part of the city’s magic. Buildings seem to transform themselves depending on the light; Dublin under a lowering sky is a different place from Dublin in sunshine.

Enjoying Dublin

Literature has always flourished in Dublin, the only city to have produced three Nobel Prize winners for literature — Yeats, Shaw, and Beckett. Joyce, the high priest of literary Modernism, imagined and interpreted Dublin for the world in *Ulysses* (you’ll see references to it all over). However, sometimes it seems that the city produced artists of this stature by accident, even against its will. Beckett and Joyce, among others, had to leave their homeland to understand it — and to be understood. Dublin theater is legendary, and no visitor should miss seeing a performance at the Abbey Theatre or Gate Theatre. The city’s impact on the rock and pop music scene with the likes of U2 and Bob Geldof is well known — there’s even a self-guided tour of their haunts. Traditional Irish music is also alive and well, especially in the pubs, and there has been a revival of storytelling, poetry reading, and traditional dancing. And in this city, where literature and theater have historically dominated the scene, visual arts are finally

coming into their own with the new Museum of Modern Art and the many galleries that display the work of modern Irish artists. The constantly crowded and busy Grafton Street is the most visible center for shopping, but there are shops all over that carry an international array of goods as well as the Irish crafts and souvenirs you expect. And while multinational chains have made inroads, they seem less blatant here than elsewhere. Many shops, and also hotels and guest houses, have been owned and managed by the same families for years, and theirs is the welcome of traditional Dublin hospitality. Dublin's food has undergone a metamorphosis. There was a time when you might have apologized for it, but no longer. Dublin has international restaurants galore, and the New Irish Cuisine is built upon fresh products of Ireland's seas, rivers, and farms. Coffee has replaced the ubiquitous tea — Dublin is now almost as much a coffee city as Vienna or Seattle.

City and Countryside

In a city of such human proportions it is not surprising that parks and gardens abound for recreation and relaxation. Phoenix Park in the northwest is the largest open space, but squares like St. Stephen's Green are the garden oases of the city. On the coast, Sandymount, Dollymount, and Killiney strands are the places to go. The beautiful Wicklow Mountains, and the Wicklow Mountains National Park provide a more rugged countryside, and the area has breathtaking houses and gardens such as Castletown, Mount Usher, and Powerscourt. To the north and west are the ancient sites of Ireland: Malahide Castle, the evocative hill of Tara, and the long barrows of Knowth and Newgrange. The DART (Dublin Area Rapid Transit) runs north and south along the coast. It's an ideal way for the visitor to reach outlying sights and villages. There are many guided bus tours to sights outside the city, and some are accessible by city bus.

Young at Heart?

Dublin is a young city. Almost half of Ireland's population is under twenty-five, and with its universities and professional schools, Dublin also has a large student population. The universities attract students from all over the world, and this influx helps to make Dublin a busy, buzzing international city. However, young and old, stranger and

Dubliners rub shoulders quite happily. Religion and respect for parents has not yet gone out of fashion. And young graduates are not leaving now — multinational corporations and European Union investment mean there are plenty of opportunities for them at home. Unemployment is at an all-time low. The Irish are actually beginning to come home.

News report

Np1 (wsj_2465)

We don't know who is winning the drug war in Latin America, but we know who's losing it -- the press. Over the past six months, six journalists have been killed and 10 kidnapped by drug traffickers or leftist guerrillas -- who often are one and the same -- in Colombia. Over the past 12 years, at least 40 journalists have died there. The attacks have intensified since the Colombian government began cracking down on the traffickers in August, trying to prevent their takeover of the country. The slaughter in Colombia was very much on the minds of 450 editors and publishers from Latin America, the United States, the Caribbean and Canada attending the 45th general assembly of the Inter-American Press Association in Monterrey, Mexico, this week. On Tuesday the conference got word of another atrocity, the assassination in Medellin of two employees of El Espectador, Colombia's second-largest newspaper. The paper's local administrator, Maria Luz Lopez, was shot dead, and her mother wounded, while her car was stopped for a red light. An hour later, the paper's circulation manager, Miguel Soler, was shot and killed near his home. The drug lords who claimed responsibility said they would blow up the Bogota newspaper's offices if it continued to distribute in Medellin. They bombed the Bogota offices last month, destroying its computer and causing \$2.5 million in damage. El Espectador has been a special target because of the extraordinary courage of its publisher and his staff. At Monterrey, publisher Luis Gabriel Cano, although shaken by the murders, issued a statement saying: "We will not cease our fight against drug trafficking. They want to terrify the press and in particular El Espectador because it has always been a torchbearer in this war." This

comes from a man whose brother, Guillermo, was murdered in 1986. The publishers in Monterrey command no battalions, but they agreed to express their outrage with editorials in today's editions. Many will use a common editorial. A final statement yesterday said: "While some advances are being made in nations throughout the hemisphere, the state of press freedom in the Americas still must be regarded as grim as long as journalists and their families are subject to the crudest form of censorship: death by assassination." The report charged that Panama's Manuel Noriega is not only in league with the drug traffickers but also is bullying the press as never before. "Noriega has closed every independent newspaper, radio and television station and arrested, tortured or forced into exile a long list of reporters," the statement declared. It added: "In Cuba, public enemy No. 1 of press freedoms in the hemisphere, repression of journalists both Cuban and foreign is worse than ever." And in Nicaragua, promises of press freedom by the Sandinistas "have not materialized." As it happens, the four countries cited, Colombia, Cuba, Panama and Nicaragua, are not only where the press is under greatest attack but also are linked by the drug trade and left-wing politics. Noriega is close to Castro and may once have been his agent. Sandinistas Thomas Borge and the Ortega brothers are Castro proteges; he backed their takeover of Nicaragua. In Colombia, the drug-financed guerrillas trying to seize the country and destroy democracy include M-19, which Castro has clearly backed. Robert Merkel, a former U.S. attorney handling drug indictments in Florida, doesn't think for a minute that Castro's much publicized trials of high officials engaged in the drug trade mean he has broken off with the Medellin drug cartel. "If the cartel succeeds in blackmailing the Colombian authorities into negotiations, the cartel will be in control and Fidel can exploit his past relationships with them," he told the Journal's David Asman recently. The struggle against the drug lords in Colombia will be a near thing. This week, the government arrested Jose Abello Silva, said to be the fourth-ranking cartel leader. He will probably be extradited to the U.S. for trial under an extradition treaty President Virgilia Barco has revived. Later, another high-ranking trafficker, Leonidas Vargas, was arrested and 1,000 pounds of dynamite seized. Mr. Barco has refused U.S. troops or

advisers but has accepted U.S. military aid. President Bush has agreed to meet within 90 days with Mr. Barco, President Alan Garcia of Peru and President Jaime Paz Zamora of Bolivia to discuss the drug problem. It might not be a bad idea to do that sooner, rather than later. After the Panama fiasco, they will need some reassurance. Certainly, the Colombian press is much in need of that.

Np2 (A1.E1.-NEW)

Inauguration of free zone in Dubai for e-commerce

Dubai 10-28 (FP) - Dubai's Crown Prince Sheikh Mohamed Bin Rashid Al Maktoum inaugurated a free zone for e-commerce today, called Dubai Internet City. The preliminary stages of the project, the only one of its kind according to its designers, are estimated at \$200 million. Sheikh Mohamed, who is also the Defense Minister of the United Arab Emirates, announced at the inauguration ceremony that "we want to make Dubai a new trading center." The minister, who has his own website, also said: "I want Dubai to be the best place in the world for state-of-the-art technology companies." He said companies engaged in e-commerce would be able to set up offices, employ staff and own equipment in the open zone, including fully-owned foreign companies. The e-commerce free zone is situated in north Dubai, near the industrial free zone in Jebel Ali, the top regional and tenth international leading area in container transit. The inauguration of Dubai Internet City coincides with the opening of an annual IT show in Dubai, the Gulf Information Technology Exhibition (Gitex), the biggest in the Middle East.

Np3 (20000415-APW_ENG_NEW)

Ohio Congressman Arrives in Jordan

Amman, Jordan (AP)

U.S. Representative Tony Hall arrived in Jordan on Saturday en route to Iraq, where he is expected to look into the plight of Iraqis after nearly 10 years of U.N. trade sanctions. Hall, an Ohio Democrat and one of very few U.S. congressmen to visit Iraq since the

1991 Gulf War over Kuwait, is scheduled to embark Sunday the 12-hour overland trip to the Iraqi capital, Baghdad. He did not speak to reporters in Jordan, but he told The Associated Press before leaving the United States that he hopes to "separate the humanitarian work from the political issues." During his four days in Iraq, Hall said he wanted to investigate reports from relief agencies that a quarter of Iraqi children may be suffering from chronic malnutrition. He said he would pay particular attention to what happens to food and medicine entering the country under the U.N. oil-for-food program. If supplies are not reaching the people who need them, Hall said, he wanted to find out whether the United Nations or relief agencies needed to handle things differently, or whether "Iraq needs to get out of the way and let us do the job." The Iraqi government blames the embargo for the malnutrition, infant mortality and other hardships. The sanctions cannot be lifted until U.N. inspectors certify that Iraq has eliminated its weapons of mass destruction and the means to produce them. Iraq says it has done so and has barred inspectors since late 1998. At least one other congressman has visited Iraq. Energy Secretary Bill Richardson went to Baghdad in 1995 while a representative for New Mexico.

Np4 (wsj_0026)

The White House said President Bush has approved duty-free treatment for imports of certain types of watches that aren't produced in "significant quantities" in the U.S., the Virgin Islands and other U.S. possessions. The action came in response to a petition filed by Timex Inc. for changes in the U.S. Generalized System of Preferences for imports from developing nations. Previously, watch imports were denied such duty-free treatment. Timex had requested duty-free treatment for many types of watches, covered by 58 different U.S. tariff classifications. The White House said Mr. Bush decided to grant duty-free status for 18 categories, but turned down such treatment for other types of watches "because of the potential for material injury to watch producers located in the U.S. and the Virgin Islands." Timex is a major U.S. producer and seller of watches, including low-priced battery-operated watches assembled in the Philippines and other

developing nations covered by the U.S. tariff preferences. U.S. trade officials said the Philippines and Thailand would be the main beneficiaries of the president's action. Imports of the types of watches that now will be eligible for duty-free treatment totaled about \$37.3 million in 1988, a relatively small share of the \$1.5 billion in U.S. watch imports that year, according to an aide to U.S. Trade Representative Carla Hills.

Np5 (wsj_0158) 28.31

Reuters Holdings PLC said Michael Reupke resigned as general manager to pursue unspecified interests, a move the news organization termed an "amicable separation." Mr. Reupke, 52 years old and a 27-year Reuters veteran, had been the information-services company's general manager for only six months. His appointment to that post, which has senior administrative, staff and policy responsibilities, followed a several-year tenure as Reuters's editor in chief. No successor was named, and Mr. Reupke's duties will be split among three other senior Reuters executives, the company said. In a telephone interview, Mr. Reupke said his departure was for "personal reasons," which he declined to specify. "There is no business reason for my departure," nor any disagreement over policy, he added. He also rejected reports that his departure stemmed from disappointment the general manager's post hadn't also led to a board directorship at the London-based news organization. Mr. Reupke was one of three executives on Reuters's eight-person executive committee who didn't also serve on the company's board of directors. "If I were choosing the people of tomorrow, I would have chosen the people who are now on the board," he said. A Reuters spokesman said the departure reflects "no change in strategy or profits." Mark Shepperd, an analyst at UBS Phillips and Drew in London, said, "I suspect (the departure) will be fairly irrelevant for the company. I would be very surprised if his departure signals any change in strategy or change in profit expectations." On London's Stock Exchange, Reuters shares rose five pence to 913 pence (\$14.43). In the U.S. over-the-counter market, American depositary shares for Reuters, each representing three shares in the London market, closed unchanged at \$43.875. The senior of the three executives who will assume Mr. Reupke's

duties is Nigel Judah, 58, finance director and a Reuters board director. Peter Holland, 45, deputy general manager, becomes director of corporate affairs. And Patrick Mannix, 46, international technical manager, becomes director of group quality programs.

Np6 (wsj_0027)

Magna International Inc.'s chief financial officer, James McAlpine, resigned and its chairman, Frank Stronach, is stepping in to help turn the automotive-parts manufacturer around, the company said. Mr. Stronach will direct an effort to reduce overhead and curb capital spending "until a more satisfactory level of profit is achieved and maintained," Magna said. Stephen Akerfeldt, currently vice president finance, will succeed Mr. McAlpine. An ambitious expansion has left Magna with excess capacity and a heavy debt load as the automotive industry enters a downturn. The company has reported declines in operating profit in each of the past three years, despite steady sales growth. Magna recently cut its quarterly dividend in half and the company's Class A shares are wallowing far below their 52-week high of 16.125 Canadian dollars (US\$13.73). On the Toronto Stock Exchange yesterday, Magna shares closed up 37.5 Canadian cents to C\$9.625. Mr. Stronach, founder and controlling shareholder of Magna, resigned as chief executive officer last year to seek, unsuccessfully, a seat in Canada's Parliament. Analysts said Mr. Stronach wants to resume a more influential role in running the company. They expect him to cut costs throughout the organization. The company said Mr. Stronach will personally direct the restructuring, assisted by Manfred Gingl, president and chief executive. Neither they nor Mr. McAlpine could be reached for comment. Magna said Mr. McAlpine resigned to pursue a consulting career, with Magna as one of his clients.

Government document

Conclusions of the Financial Crisis Inquiry Commission

The Financial Crisis Inquiry Commission has been called upon to examine the financial and economic crisis that has gripped our country and explain its causes to the American

people. We are keenly aware of the significance of our charge, given the economic damage that America has suffered in the wake of the greatest financial crisis since the Great Depression. Our task was first to determine what happened and how it happened so that we could understand why it happened. Here we present our conclusions. We encourage the American people to join us in making their own assessments based on the evidence gathered in our inquiry. If we do not learn from history, we are unlikely to fully recover from it. Some on Wall Street and in Washington with a stake in the status quo may be tempted to wipe from memory the events of this crisis, or to suggest that no one could have foreseen or prevented them. This report endeavors to expose the facts, identify responsibility, unravel myths, and help us understand how the crisis could have been avoided. It is an attempt to record history, not to rewrite it, nor allow it to be rewritten. To help our fellow citizens better understand this crisis and its causes, we also present specific conclusions at the end of chapters in Parts III, IV, and V of this report. The subject of this report is of no small consequence to this nation. The profound events of 2007 and 2008 were neither bumps in the road nor an accentuated dip in the financial and business cycles we have come to expect in a free market economic system. This was a fundamental disruption — a financial upheaval, if you will — that wreaked havoc in communities and neighborhoods across this country. As this report goes to print, there are more than 26 million Americans who are out of work, cannot find full-time work, or have given up looking for work. About four million families have lost their homes to foreclosure and another four and a half million have slipped into the foreclosure process or are seriously behind on their mortgage payments. Nearly \$11 trillion in household wealth has vanished, with retirement accounts and life savings swept away. Businesses, large and small, have felt the sting of a deep recession. There is much anger about what has transpired, and justifiably so. Many people who abided by all the rules now find themselves out of work and uncertain about their future prospects. The collateral damage of this crisis has been real people and real communities. The impacts of this crisis are likely to be felt for a generation. And the nation faces no easy path to renewed economic strength. Like so many Americans, we

began our exploration with our own views and some preliminary knowledge about how the world's strongest financial system came to the brink of collapse. Even at the time of our appointment to this independent panel, much had already been written and said about the crisis. Yet all of us have been deeply affected by what we have learned in the course of our inquiry. We have been at various times fascinated, surprised, and even shocked by what we saw, heard, and read. Ours has been a journey of revelation. Much attention over the past two years has been focused on the decisions by the federal government to provide massive financial assistance to stabilize the financial system and rescue large financial institutions that were deemed too systemically important to fail. Those decisions—and the deep emotions surrounding them — will be debated long into the future. But our mission was to ask and answer this central question: how did it come to pass that in 2008 our nation was forced to choose between two stark and painful alternatives—either risk the total collapse of our financial system and economy or inject trillions of taxpayer dollars into the financial system and an array of companies, as millions of Americans still lost their jobs, their savings, and their homes? In this report, we detail the events of the crisis. But a simple summary, as we see it, is useful at the outset. While the vulnerabilities that created the potential for crisis were years in the making, it was the collapse of the housing bubble — fueled by low interest rates, easy and available credit, scant regulation, and toxic mortgages — that was the spark that ignited a string of events, which led to a full-blown crisis in the fall of 2008. Trillions of dollars in risky mortgages had become embedded throughout the financial system, as mortgage-related securities were packaged, repackaged, and sold to investors around the world. When the bubble burst, hundreds of billions of dollars in losses in mortgages and mortgage-related securities shook markets as well as financial institutions that had significant exposures to those mortgages and had borrowed heavily against them. This happened not just in the United States but around the world. The losses were magnified by derivatives such as synthetic securities. The crisis reached seismic proportions in September 2008 with the failure of Lehman Brothers and the impending collapse of the insurance giant American International Group (AIG). Panic fanned by a lack of

transparency of the balance sheets of major financial institutions, coupled with a tangle of interconnections among institutions perceived to be “too big to fail,” caused the credit markets to seize up. Trading ground to a halt. The stock market plummeted. The economy plunged into a deep recession. The financial system we examined bears little resemblance to that of our parents’ generation. The changes in the past three decades alone have been remarkable. The financial markets have become increasingly globalized. Technology has transformed the efficiency, speed, and complexity of financial instruments and transactions. There is broader access to and lower costs of financing than ever before. And the financial sector itself has become a much more dominant force in our economy. From 1978 to 2007, the amount of debt held by the financial sector soared from \$3 trillion to \$36 trillion, more than doubling as a share of gross domestic product. The very nature of many Wall Street firms changed — from relatively staid private partnerships to publicly traded corporations taking greater and more diverse kinds of risks. By 2005, the 10 largest U.S. commercial banks held 55% of the industry’s assets, more than double the level held in 1990. On the eve of the crisis in 2006, financial sector profits constituted 27% of all corporate profits in the United States, up from 15% in 1980. Understanding this transformation has been critical to the Commission’s analysis. Now to our major findings and conclusions, which are based on the facts contained in this report: they are offered with the hope that lessons may be learned to help avoid future catastrophe. We conclude this financial crisis was avoidable. The crisis was the result of human action and inaction, not of Mother Nature or computer models gone haywire. The captains of finance and the public stewards of our financial system ignored warnings and failed to question, understand, and manage evolving risks within a system essential to the well-being of the American public. There was a big miss, not a stumble. While the business cycle cannot be repealed, a crisis of this magnitude need not have occurred. To paraphrase Shakespeare, the fault lies not in the stars, but in us. Despite the expressed view of many on Wall Street and in Washington that the crisis could not have been foreseen or avoided, there were warning signs. The tragedy was that they were ignored or discounted. There was an

explosion in risky subprime lending and securitization, an unsustainable rise in housing prices, widespread reports of egregious and predatory lending practices, dramatic increases in household mortgage debt, and exponential growth in financial firms' trading activities, unregulated derivatives, and short-term "repo" lending markets, among many other red flags. Yet there was pervasive permissiveness; little meaningful action was taken to quell the threats in a timely manner. The prime example is the Federal Reserve's pivotal failure to stem the flow of toxic mortgages, which it could have done by setting prudent mortgage-lending standards. The Federal Reserve was the one entity empowered to do so and it did not. The record of our examination is replete with evidence of other failures: financial institutions made, bought, and sold mortgage securities they never examined, did not care to examine, or knew to be defective; firms depended on tens of billions of dollars of borrowing that had to be renewed each and every night, secured by subprime mortgage securities; and major firms and investors blindly relied on credit rating agencies as their arbiters of risk. What else could one expect on a highway where there were neither speed limits nor neatly painted lines? We conclude widespread failures in financial regulation and supervision proved devastating to the stability of the nation's financial markets. The sentries were not at their posts, in no small part due to the widely accepted faith in the self-correcting nature of the markets and the ability of financial institutions to effectively police themselves. More than 30 years of deregulation and reliance on self-regulation by financial institutions, championed by former Federal Reserve chairman Alan Greenspan and others, supported by successive administrations and Congresses, and actively pushed by the powerful financial industry at every turn, had stripped away key safeguards, which could have helped avoid catastrophe. This approach had opened up gaps in oversight of critical areas with trillions of dollars at risk, such as the shadow banking system and over-the-counter derivatives markets. In addition, the government permitted financial firms to pick their preferred regulators in what became a race to the weakest supervisor. Yet we do not accept the view that regulators lacked the power to protect the financial system. They had ample power in many arenas and they chose not to use it. To give just

three examples: the Securities and Exchange Commission could have required more capital and halted risky practices at the big investment banks. It did not. The Federal Reserve Bank of New York and other regulators could have clamped down on Citigroup's excesses in the run-up to the crisis. They did not. Policy makers and regulators could have stopped the runaway mortgage securitization train. They did not. In case after case after case, regulators continued to rate the institutions they oversaw as safe and sound even in the face of mounting troubles, often downgrading them just before their collapse. And where regulators lacked authority, they could have sought it. Too often, they lacked the political will — in a political and ideological environment that constrained it — as well as the fortitude to critically challenge the institutions and the entire system they were entrusted to oversee. Changes in the regulatory system occurred in many instances as financial markets evolved. But as the report will show, the financial industry itself played a key role in weakening regulatory constraints on institutions, markets, and products. It did not surprise the Commission that an industry of such wealth and power would exert pressure on policy makers and regulators. From 1999 to 2008, the financial sector expended \$2.7 billion in reported federal lobbying expenses; individuals and political action committees in the sector made more than \$1 billion in campaign contributions. What troubled us was the extent to which the nation was deprived of the necessary strength and independence of the oversight necessary to safeguard financial stability. We conclude dramatic failures of corporate governance and risk management at many systemically important financial institutions were a key cause of this crisis. There was a view that instincts for self-preservation inside major financial firms would shield them from fatal risk-taking without the need for a steady regulatory hand, which, the firms argued, would stifle innovation. Too many of these institutions acted recklessly, taking on too much risk, with too little capital, and with too much dependence on short-term funding. In many respects, this reflected a fundamental change in these institutions, particularly the large investment banks and bank holding companies, which focused their activities increasingly on risky trading activities that produced hefty profits. They took on enormous exposures in acquiring

and supporting subprime lenders and creating, packaging, repackaging, and selling trillions of dollars in mortgage-related securities, including synthetic financial products. Like Icarus, they never feared flying ever closer to the sun. Many of these institutions grew aggressively through poorly executed acquisition and integration strategies that made effective management more challenging. The CEO of Citigroup told the Commission that a \$40 billion position in highly rated mortgage securities would “not in any way have excited my attention,” and the cohead of Citigroup’s investment bank said he spent “a small fraction of 1%” of his time on those securities. In this instance, too big to fail meant too big to manage. Financial institutions and credit rating agencies embraced mathematical models as reliable predictors of risks, replacing judgment in too many instances. Too often, risk management became risk justification. Compensation systems — designed in an environment of cheap money, intense competition, and light regulation — too often rewarded the quick deal, the short-term gain — without proper consideration of long-term consequences. Often, those systems encouraged the big bet — where the payoff on the upside could be huge and the downside limited. This was the case up and down the line — from the corporate boardroom to the mortgage broker on the street. Our examination revealed stunning instances of governance breakdowns and irresponsibility. You will read, among other things, about AIG senior management’s ignorance of the terms and risks of the company’s \$79 billion derivatives exposure to mortgage-related securities; Fannie Mae’s quest for bigger market share, profits, and bonuses, which led it to ramp up its exposure to risky loans and securities as the housing market was peaking; and the costly surprise when Merrill Lynch’s top management realized that the company held \$55 billion in “super-senior” and supposedly “super-safe” mortgage-related securities that resulted in billions of dollars in losses. We conclude a combination of excessive borrowing, risky investments, and lack of transparency put the financial system on a collision course with crisis. Clearly, this vulnerability was related to failures of corporate governance and regulation, but it is significant enough by itself to warrant our attention here. In the years leading up to the crisis, too many financial institutions,

as well as too many households, borrowed to the hilt, leaving them vulnerable to financial distress or ruin if the value of their investments declined even modestly. For example, as of 2007, the five major investment banks — Bear Stearns, Goldman Sachs, Lehman Brothers, Merrill Lynch, and Morgan Stanley — were operating with extraordinarily thin capital. By one measure, their leverage ratios were as high as 40 to 1, meaning for every \$40 in assets, there was only \$1 in capital to cover losses. Less than a 3% drop in asset values could wipe out a firm. To make matters worse, much of their borrowing was short-term, in the overnight market—meaning the borrowing had to be renewed each and every day. For example, at the end of 2007, Bear Stearns had \$11.8 billion in equity and \$383.6 billion in liabilities and was borrowing as much as \$70 billion in the overnight market. It was the equivalent of a small business with \$50,000 in equity borrowing \$1.6 million, with \$296,750 of that due each and every day. One can't really ask "What were they thinking?" when it seems that too many of them were thinking alike. And the leverage was often hidden — in derivatives positions, in off-balance-sheet entities, and through "window dressing" of financial reports available to the investing public. The kings of leverage were Fannie Mae and Freddie Mac, the two behemoth government-sponsored enterprises (GSEs). For example, by the end of 2007, Fannie's and Freddie's combined leverage ratio, including loans they owned and guaranteed, stood at 75 to 1. But financial firms were not alone in the borrowing spree: from 2001 to 2007, national mortgage debt almost doubled, and the amount of mortgage debt per household rose more than 63% from \$91,500 to \$149,500, even while wages were essentially stagnant. When the housing downturn hit, heavily indebted financial firms and families alike were walloped. The heavy debt taken on by some financial institutions was exacerbated by the risky assets they were acquiring with that debt. As the mortgage and real estate markets churned out riskier and riskier loans and securities, many financial institutions loaded up on them. By the end of 2007, Lehman had amassed \$111 billion in commercial and residential real estate holdings and securities, which was almost twice what it held just two years before, and more than four times its total equity. And again, the risk wasn't being taken on just by the big

financial firms, but by families, too. Nearly one in 10 mortgage borrowers in 2005 and 2006 took out “option ARM” loans, which meant they could choose to make payments so low that their mortgage balances rose every month. Within the financial system, the dangers of this debt were magnified because transparency was not required or desired. Massive, short-term borrowing, combined with obligations unseen by others in the market, heightened the chances the system could rapidly unravel. In the early part of the 20th century, we erected a series of protections — the Federal Reserve as a lender of last resort, federal deposit insurance, ample regulations — to provide a bulwark against the panics that had regularly plagued America’s banking system in the 19th century. Yet, over the past 30-plus years, we permitted the growth of a shadow banking system — opaque and laden with short-term debt — that rivaled the size of the traditional banking system. Key components of the market — for example, the multitrillion-dollar repo lending market, off-balance-sheet entities, and the use of over-the-counter derivatives — were hidden from view, without the protections we had constructed to prevent financial meltdowns. We had a 21st-century financial system with 19th-century safeguards. When the housing and mortgage markets cratered, the lack of transparency, the extraordinary debt loads, the short-term loans, and the risky assets all came home to roost. What resulted was panic. We had reaped what we had sown. We conclude the government was ill prepared for the crisis, and its inconsistent response added to the uncertainty and panic in the financial markets. As part of our charge, it was appropriate to review government actions taken in response to the developing crisis, not just those policies or actions that preceded it, to determine if any of those responses contributed to or exacerbated the crisis. As our report shows, key policy makers — the Treasury Department, the Federal Reserve Board, and the Federal Reserve Bank of New York — who were best positioned to watch over our markets were ill prepared for the events of 2007 and 2008. Other agencies were also behind the curve. They were hampered because they did not have a clear grasp of the financial system they were charged with overseeing, particularly as it had evolved in the years leading up to the crisis. This was in no small measure due to the lack of transparency

in key markets. They thought risk had been diversified when, in fact, it had been concentrated. Time and again, from the spring of 2007 on, policy makers and regulators were caught off guard as the contagion spread, responding on an ad hoc basis with specific programs to put fingers in the dike. There was no comprehensive and strategic plan for containment, because they lacked a full understanding of the risks and interconnections in the financial markets. Some regulators have conceded this error. We had allowed the system to race ahead of our ability to protect it. While there was some awareness of, or at least a debate about, the housing bubble, the record reflects that senior public officials did not recognize that a bursting of the bubble could threaten the entire financial system. Throughout the summer of 2007, both Federal Reserve Chairman Ben Bernanke and Treasury Secretary Henry Paulson offered public assurances that the turmoil in the subprime mortgage markets would be contained. When Bear Stearns's hedge funds, which were heavily invested in mortgage-related securities, imploded in June 2007, the Federal Reserve discussed the implications of the collapse. Despite the fact that so many other funds were exposed to the same risks as those hedge funds, the Bear Stearns funds were thought to be "relatively unique." Days before the collapse of Bear Stearns in March 2008, SEC Chairman Christopher Cox expressed "comfort about the capital cushions" at the big investment banks. It was not until August 2008, just weeks before the government takeover of Fannie Mae and Freddie Mac, that the Treasury Department understood the full measure of the dire financial conditions of those two institutions. And just a month before Lehman's collapse, the Federal Reserve Bank of New York was still seeking information on the exposures created by Lehman's more than 900,000 derivatives contracts. In addition, the government's inconsistent handling of major financial institutions during the crisis — the decision to rescue Bear Stearns and then to place Fannie Mae and Freddie Mac into conservatorship, followed by its decision not to save Lehman Brothers and then to save AIG—increased uncertainty and panic in the market. In making these observations, we deeply respect and appreciate the efforts made by Secretary Paulson, Chairman Bernanke, and Timothy Geithner, formerly president of the Federal Reserve Bank of

New York and now treasury secretary, and so many others who labored to stabilize our financial system and our economy in the most chaotic and challenging of circumstances. We conclude there was a systemic breakdown in accountability and ethics. The integrity of our financial markets and the public's trust in those markets are essential to the economic well-being of our nation. The soundness and the sustained prosperity of the financial system and our economy rely on the notions of fair dealing, responsibility, and transparency. In our economy, we expect businesses and individuals to pursue profits, at the same time that they produce products and services of quality and conduct themselves well. Unfortunately — as has been the case in past speculative booms and busts — we witnessed an erosion of standards of responsibility and ethics that exacerbated the financial crisis. This was not universal, but these breaches stretched from the ground level to the corporate suites. They resulted not only in significant financial consequences but also in damage to the trust of investors, businesses, and the public in the financial system. For example, our examination found, according to one measure, that the percentage of borrowers who defaulted on their mortgages within just a matter of months after taking a loan nearly doubled from the summer of 2006 to late 2007. This data indicates they likely took out mortgages that they never had the capacity or intention to pay. You will read about mortgage brokers who were paid “yield spread premiums” by lenders to put borrowers into higher-cost loans so they would get bigger fees, often never disclosed to borrowers. The report catalogues the rising incidence of mortgage fraud, which flourished in an environment of collapsing lending standards and lax regulation. The number of suspicious activity reports — reports of possible financial crimes filed by depository banks and their affiliates — related to mortgage fraud grew 20-fold between 1996 and 2005 and then more than doubled again between 2005 and 2009. One study places the losses resulting from fraud on mortgage loans made between 2005 and 2007 at \$112 billion. Lenders made loans that they knew borrowers could not afford and that could cause massive losses to investors in mortgage securities. As early as September 2004, Countrywide executives recognized that many of the loans they were originating could result in “catastrophic consequences.” Less

than a year later, they noted that certain high-risk loans they were making could result not only in foreclosures but also in “financial and reputational catastrophe” for the firm. But they did not stop. And the report documents that major financial institutions ineffectively sampled loans they were purchasing to package and sell to investors. They knew a significant percentage of the sampled loans did not meet their own underwriting standards or those of the originators. Nonetheless, they sold those securities to investors. The Commission’s review of many prospectuses provided to investors found that this critical information was not disclosed. These conclusions must be viewed in the context of human nature and individual and societal responsibility. First, to pin this crisis on mortal flaws like greed and hubris would be simplistic. It was the failure to account for human weakness that is relevant to this crisis. Second, we clearly believe the crisis was a result of human mistakes, misjudgments, and misdeeds that resulted in systemic failures for which our nation has paid dearly. As you read this report, you will see that specific firms and individuals acted irresponsibly. Yet a crisis of this magnitude cannot be the work of a few bad actors, and such was not the case here. At the same time, the breadth of this crisis does not mean that “everyone is at fault”; many firms and individuals did not participate in the excesses that spawned disaster. We do place special responsibility with the public leaders charged with protecting our financial system, those entrusted to run our regulatory agencies, and the chief executives of companies whose failures drove us to crisis. These individuals sought and accepted positions of significant responsibility and obligation. Tone at the top does matter and, in this instance, we were let down. No one said “no.” But as a nation, we must also accept responsibility for what we permitted to occur. Collectively, but certainly not unanimously, we acquiesced to or embraced a system, a set of policies and actions, that gave rise to our present predicament. This report describes the events and the system that propelled our nation toward crisis. The complex machinery of our financial markets has many essential gears — some of which played a critical role as the crisis developed and deepened. Here we render our conclusions about specific components of the system that we believe contributed significantly to the financial meltdown. We conclude

collapsing mortgage-lending standards and the mortgage securitization pipeline lit and spread the flame of contagion and crisis. When housing prices fell and mortgage borrowers defaulted, the lights began to dim on Wall Street. This report catalogues the corrosion of mortgage-lending standards and the securitization pipeline that transported toxic mortgages from neighborhoods across America to investors around the globe. Many mortgage lenders set the bar so low that lenders simply took eager borrowers' qualifications on faith, often with a willful disregard for a borrower's ability to pay. Nearly one-quarter of all mortgages made in the first half of 2005 were interest-only loans. During the same year, 68% of "option ARM" loans originated by Countrywide and Washington Mutual had low- or no-documentation requirements. These trends were not secret. As irresponsible lending, including predatory and fraudulent practices, became more prevalent, the Federal Reserve and other regulators and authorities heard warnings from many quarters. Yet the Federal Reserve neglected its mission "to ensure the safety and soundness of the nation's banking and financial system and to protect the credit rights of consumers." It failed to build the retaining wall before it was too late. And the Office of the Comptroller of the Currency and the Office of Thrift Supervision, caught up in turf wars, preempted state regulators from reining in abuses. While many of these mortgages were kept on banks' books, the bigger money came from global investors who clamored to put their cash into newly created mortgage-related securities. It appeared to financial institutions, investors, and regulators alike that risk had been conquered: the investors held highly rated securities they thought were sure to perform; the banks thought they had taken the riskiest loans off their books; and regulators saw firms making profits and borrowing costs reduced. But each step in the mortgage securitization pipeline depended on the next step to keep demand going. From the speculators who flipped houses to the mortgage brokers who scouted the loans, to the lenders who issued the mortgages, to the financial firms that created the mortgage-backed securities, collateralized debt obligations (CDOs), CDOs squared, and synthetic CDOs: no one in this pipeline of toxic mortgages had enough skin in the game. They all believed they could off-load their risks on a moment's notice to the next person in

line. They were wrong. When borrowers stopped making mortgage payments, the losses — amplified by derivatives — rushed through the pipeline. As it turned out, these losses were concentrated in a set of systemically important financial institutions. In the end, the system that created millions of mortgages so efficiently has proven to be difficult to unwind. Its complexity has erected barriers to modifying mortgages so families can stay in their homes and has created further uncertainty about the health of the housing market and financial institutions. We conclude over-the-counter derivatives contributed significantly to this crisis. The enactment of legislation in 2000 to ban the regulation by both the federal and state governments of over-the-counter (OTC) derivatives was a key turning point in the march toward the financial crisis. From financial firms to corporations, to farmers, and to investors, derivatives have been used to hedge against, or speculate on, changes in prices, rates, or indices or even on events such as the potential defaults on debts. Yet, without any oversight, OTC derivatives rapidly spiraled out of control and out of sight, growing to \$673 trillion in notional amount. This report explains the uncontrolled leverage; lack of transparency, capital, and collateral requirements; speculation; interconnections among firms; and concentrations of risk in this market. OTC derivatives contributed to the crisis in three significant ways. First, one type of derivative — credit default swaps (CDS) — fueled the mortgage securitization pipeline. CDS were sold to investors to protect against the default or decline in value of mortgage-related securities backed by risky loans. Companies sold protection — to the tune of \$79 billion, in AIG’s case — to investors in these newfangled mortgage securities, helping to launch and expand the market and, in turn, to further fuel the housing bubble. Second, CDS were essential to the creation of synthetic CDOs. These synthetic CDOs were merely bets on the performance of real mortgage-related securities. They amplified the losses from the collapse of the housing bubble by allowing multiple bets on the same securities and helped spread them throughout the financial system. Goldman Sachs alone packaged and sold \$73 billion in synthetic CDOs from July 1, 2004, to May 31, 2007. Synthetic CDOs created by Goldman referenced more than 3,400 mortgage securities, and 610 of them were

referenced at least twice. This is apart from how many times these securities may have been referenced in synthetic CDOs created by other firms. Finally, when the housing bubble popped and crisis followed, derivatives were in the center of the storm. AIG, which had not been required to put aside capital reserves as a cushion for the protection it was selling, was bailed out when it could not meet its obligations. The government ultimately committed more than \$180 billion because of concerns that AIG's collapse would trigger cascading losses throughout the global financial system. In addition, the existence of millions of derivatives contracts of all types between systemically important financial institutions—unseen and unknown in this unregulated market—added to uncertainty and escalated panic, helping to precipitate government assistance to those institutions. We conclude the failures of credit rating agencies were essential cogs in the wheel of financial destruction. The three credit rating agencies were key enablers of the financial meltdown. The mortgage-related securities at the heart of the crisis could not have been marketed and sold without their seal of approval. Investors relied on them, often blindly. In some cases, they were obligated to use them, or regulatory capital standards were hinged on them. This crisis could not have happened without the rating agencies. Their ratings helped the market soar and their downgrades through 2007 and 2008 wreaked havoc across markets and firms. In our report, you will read about the breakdowns at Moody's, examined by the Commission as a case study. From 2000 to 2007, Moody's rated nearly 45,000 mortgage-related securities as triple-A. This compares with six private-sector companies in the United States that carried this coveted rating in early 2010. In 2006 alone, Moody's put its triple-A stamp of approval on 30 mortgage-related securities every working day. The results were disastrous: 83% of the mortgage securities rated triple-A that year ultimately were downgraded. You will also read about the forces at work behind the breakdowns at Moody's, including the flawed computer models, the pressure from financial firms that paid for the ratings, the relentless drive for market share, the lack of resources to do the job despite record profits, and the absence of meaningful public oversight. And you will see that without the active participation of the rating agencies, the market for mortgage-

related securities could not have been what it became. THERE ARE MANY COMPETING VIEWS as to the causes of this crisis. In this regard, the Commission has endeavored to address key questions posed to us. Here we discuss three: capital availability and excess liquidity, the role of Fannie Mae and Freddie Mac (the GSEs), and government housing policy. First, as to the matter of excess liquidity: in our report, we outline monetary policies and capital flows during the years leading up to the crisis. Low interest rates, widely available capital, and international investors seeking to put their money in real estate assets in the United States were prerequisites for the creation of a credit bubble. Those conditions created increased risks, which should have been recognized by market participants, policy makers, and regulators. However, it is the Commission's conclusion that excess liquidity did not need to cause a crisis. It was the failures outlined above — including the failure to effectively rein in excesses in the mortgage and financial markets — that were the principal causes of this crisis. Indeed, the availability of well-priced capital — both foreign and domestic—is an opportunity for economic expansion and growth if encouraged to flow in productive directions. Second, we examined the role of the GSEs, with Fannie Mae serving as the Commission's case study in this area. These government-sponsored enterprises had a deeply flawed business model as publicly traded corporations with the implicit backing of and subsidies from the federal government and with a public mission. Their \$5 trillion mortgage exposure and market position were significant. In 2005 and 2006, they decided to ramp up their purchase and guarantee of risky mortgages, just as the housing market was peaking. They used their political power for decades to ward off effective regulation and oversight —spending \$164 million on lobbying from 1998 to 2008. They suffered from many of the same failures of corporate governance and risk management as the Commission discovered in other financial firms. Through the third quarter of 2010, the Treasury Department had provided \$151 billion in financial support to keep them afloat. We conclude that these two entities contributed to the crisis, but were not a primary cause. Importantly, GSE mortgage securities essentially maintained their value throughout the crisis and did not contribute to the significant financial firm losses

that were central to the financial crisis. The GSEs participated in the expansion of subprime and other risky mortgages, but they followed rather than led Wall Street and other lenders in the rush for fool's gold. They purchased the highest rated non-GSE mortgage-backed securities and their participation in this market added helium to the housing balloon, but their purchases never represented a majority of the market. Those purchases represented 10.5% of non-GSE subprime mortgage-backed securities in 2001, with the share rising to 40% in 2004, and falling back to 28% by 2008. They relaxed their underwriting standards to purchase or guarantee riskier loans and related securities in order to meet stock market analysts' and investors' expectations for growth, to regain market share, and to ensure generous compensation for their executives and employees — justifying their activities on the broad and sustained public policy support for homeownership. The Commission also probed the performance of the loans purchased or guaranteed by Fannie and Freddie. While they generated substantial losses, delinquency rates for GSE loans were substantially lower than loans securitized by other financial firms. For example, data compiled by the Commission for a subset of borrowers with similar credit scores -- scores below 600 -- show that by the end of 2008, GSE mortgages were far less likely to be seriously delinquent than were non-GSE securitized mortgages: 6.2% versus 28.3%. We also studied at length how the Department of Housing and Urban Development's (HUD's) affordable housing goals for the GSEs affected their investment in risky mortgages. Based on the evidence and interviews with dozens of individuals involved in this subject area, we determined these goals only contributed marginally to Fannie's and Freddie's participation in those mortgages. Finally, as to the matter of whether government housing policies were a primary cause of the crisis: for decades, government policy has encouraged homeownership through a set of incentives, assistance programs, and mandates. These policies were put in place and promoted by several administrations and Congresses — indeed, both Presidents Bill Clinton and George W. Bush set aggressive goals to increase home-ownership. In conducting our inquiry, we took a careful look at HUD's affordable housing goals, as noted above, and the Community Reinvestment Act (CRA).

The CRA was enacted in 1977 to combat “redlining” by banks — the practice of denying credit to individuals and businesses in certain neighborhoods without regard to their creditworthiness. The CRA requires banks and savings and loans to lend, invest, and provide services to the communities from which they take deposits, consistent with bank safety and soundness. The Commission concludes the CRA was not a significant factor in subprime lending or the crisis. Many subprime lenders were not subject to the CRA. Research indicates only 6% of high-cost loans — a proxy for subprime loans — had any connection to the law. Loans made by CRA-regulated lenders in the neighborhoods in which they were required to lend were half as likely to default as similar loans made in the same neighborhoods by independent mortgage originators not subject to the law. Nonetheless, we make the following observation about government housing policies — they failed in this respect: As a nation, we set aggressive homeownership goals with the desire to extend credit to families previously denied access to the financial markets. Yet the government failed to ensure that the philosophy of opportunity was being matched by the practical realities on the ground. Witness again the failure of the Federal Reserve and other regulators to rein in irresponsible lending. Homeownership peaked in the spring of 2004 and then began to decline. From that point on, the talk of opportunity was tragically at odds with the reality of a financial disaster in the making. When this commission began its work 18 months ago, some imagined that the events of 2008 and their consequences would be well behind us by the time we issued this report. Yet more than two years after the federal government intervened in an unprecedented manner in our financial markets, our country finds itself still grappling with the aftereffects of the calamity. Our financial system is, in many respects, still unchanged from what existed on the eve of the crisis. Indeed, in the wake of the crisis, the U.S. financial sector is now more concentrated than ever in the hands of a few large, systemically significant institutions. While we have not been charged with making policy recommendations, the very purpose of our report has been to take stock of what happened so we can plot a new course. In our inquiry, we found dramatic breakdowns of corporate governance, profound lapses in regulatory oversight, and near

fatal flaws in our financial system. We also found that a series of choices and actions led us toward a catastrophe for which we were ill prepared. These are serious matters that must be addressed and resolved to restore faith in our financial markets, to avoid the next crisis, and to rebuild a system of capital that provides the foundation for a new era of broadly shared prosperity. The greatest tragedy would be to accept the refrain that no one could have seen this coming and thus nothing could have been done. If we accept this notion, it will happen again. This report should not be viewed as the end of the nation's examination of this crisis. There is still much to learn, much to investigate, and much to fix. This is our collective responsibility. It falls to us to make different choices if we want different results.

Essay

Homosexuality

The debate over homosexuality has been one of the most long-lasting and controversial ones ever. What, exactly, causes homosexuality? Some would say it is a gene, passed on from parents to child. Others would argue that it is a result of a child's upbringing. Still more would claim that it is a mental illness that can and should be cured. Perhaps then, it is a combination of some of these? No one knows for sure, and it is possible no one ever will, but that surely does not stop everyone from coming up with their own theories and beliefs on the matter. Documented homosexuality dates as far back as ancient Greece and other cultures of the time, where it was considered to be a very normal and natural occurrence (Emond). In his book *The Symposium*, Plato wrote "Those who are halves of a man whole pursue males, and being slices, so to speak, of the male, love men throughout their boyhood, and take pleasure in physical contact with men" (qtd. in Isay 11). This shows that not only did Plato consider it normal for a male to be attracted to another male, he also believed that it began at a very young age, as the word "boyhood" signifies. In fact, Plato even considered love between two members of the same sex to be the only "real and lasting love" and necessary for democracy. Furthermore, there were many occurrences of homosexual behavior in

Greek mythology; Hercules is rumored to have had 14 male lovers, and Zeus himself partook in such behavior. Even Homer wrote about Achilles and Patroclus, who have been considered to be the perfect model of true love (Emond). But it was not until 1869 that the term “homosexual” was first used, to describe “a man or woman whose feelings of sexual attraction are for someone of the same sex” (Marcus 1). (However, for the purposes of this paper, homosexuality will be looked at solely in terms of men). At this point in time, Karl Maria Kertbeny used the word in a pamphlet which fought to repeal the current antihomosexual laws of Prussia. Kertbeny derived this word from the Greek word for “same” and the Latin word for “sex,” whereas a heterosexual is a person whose feelings of sexual attraction are for the opposite sex” (Marcus 1). It was also in the 1800s when the debate itself over the cause of homosexuality was started by Magnus Hirschfeld, a physician, sex researcher, leading sexologist, homosexual, and founder of the first gay rights movement in Germany, who believed that homosexuality was biological in nature (Marcus 10). Hirschfeld also founded the Scientific Humanitarian committee, which was mostly homosexual, in 1897. The committee published many books and other forms of literature, which gave Hirschfeld a great amount of prestige in his field. He became known as one of the founding fathers of sexology, and furthered this position when he opened the world’s first sexological institute, the institute for Sexual Science in Berlin, in 1919, which was destroyed by the Nazis 15 years later. Hirschfeld largely supported the Urning theory of Karl Ulrichs, with minor additions; he believed in some hormonal theories as a cause of homosexuality, but this only led to unsuccessful attempts to “cure” homosexuals with the use of hormone injections. This theory, which defined Urnings as males who turned to other males as sexual partners, was published in twelve pamphlets by Ulrichs, starting in 1864. The first, *Vindex*, defended Urnings, while the second, *Inclusa*, which followed shortly after, described the first scientific theory of homoerotic desire. Ulrichs believed that Urnings were attracted to other males because they were “hermaphrodites of the mind,” meaning while they may have been male in body, they were female in soul and mind, leading them to be naturally attracted to males in terms of sexual partners.

This, he claimed, made laws such as paragraph 175, a law adopted by King William I throughout the German Kingdoms at the time of their unification which forbade sex between males, and forced criminal penalties upon individuals partaking in such behavior, unfair and unreasonable. Ulrichs claimed that the origin of such a disposition was natural and inborn (Wikholm). A couple thousand years after Plato and Homer, Sigmund Freud still believed homosexuality to be a natural behavior. In an interview in 1903, he professed his beliefs: “I am... of the firm conviction that homosexuals must not be treated as sick people... Homosexual persons are not sick. They also do not belong in a court of law!” (Qtd. in Isay 3). In 1935, he furthered his claims when he wrote a now famous “Letter to an American Mother” of a homosexual, which stated that “Homosexuality is assuredly no advantage, but is nothing to be ashamed of, no vice, no degradation, it cannot be classified as an illness... Many highly respectable individuals of ancient and modern times have been homosexuals” (qtd. in Isay 3). However, the general public of the 1800s and early 1900s, including the medical professionals, regarded homosexuality as a curable mental illness. Treatments such as castration, hysterectomy, lobotomy, and electroshock therapy were used as attempted cures. By the mid-1900s, psychotherapy became the most common “cure,” and many homosexuals spent countless hours being analyzed in hopes of changing their sexual preference (Dudley 125). Dr. Evelyn Hooker, a heterosexual psychologist, conducted a ground-breaking study in the mid-1950s that went along similar reasoning as Freud. In this courageous experiment, Hooker compared the psychological profiles of sixty men, half homosexual and half heterosexual. She disagreed with the popular belief at the time that homosexuality was a mental illness, and concluded that there was no significant psychological difference between homosexual and heterosexual men; “gay” men were no more insane than their “straight” counterparts (Marcus 183). Fortunately, many prominent psychiatrists also believed that homosexuality was not an illness, and their lobbying, along with the innovative study performed by Hooker, who has been referred to as “the Rosa Parks of the gay movement” convinced the American Psychiatric Association Board of Trustees to vote to remove homosexuality from the

Diagnostic and Statistical Manual as a mental illness in December of 1973. Just over a year later, the American Psychological Association also removed homosexuality from their list of mental illnesses (Marcus 11). The American Psychological Association further reiterated this belief when an overwhelming majority of the Council of Representatives adopted the Resolution on Appropriate Therapeutic Responses to Sexual Orientation in 1997, which publicly chastised those who attempted to “cure” homosexuality by means of various forms of treatment. It stated, rather that homosexuality was merely a difference in the opinion and values of the common person, but must still be respected as an individual orientation, and thus treated as one (Resolution on Appropriate Therapeutic Responses to Sexual Orientation). Perhaps one of the most famous studies on this topic was concluded in 1991 by Michael Bailey, an assistant professor of psychology at Northwestern University, and Richard Pillard, an associate professor of psychiatry at Boston University School of Medicine, and found that sexual orientation in males is largely due to genetics. For two years, these men studied the number of occurrences of homosexuality in both monozygotic, or identical twins (a set of twins coming from the same fertilized egg and thus having identical DNA), and dizygotic, or fraternal twins (a set of twins from two separate zygotes, causing them to have similar DNA, but not any more so than that of two ordinary siblings), in addition to adoptive brothers of gay males (meaning that all of these males would have been raised in the same environment). In total, 110 pairs of twins (identical and fraternal) and 142 sets of male and their adopted brothers were studied, where at least one of the two had been classified as homosexual, either by self-identification or other means. Out of the 56 homosexual males who had identical twins, 29, or approximately 52 % of their identical twin brothers were also found to be homosexual, as compared to only 12, or approximately 22 % of the 54 non-identical twins of homosexual males, and 6, or approximately 11% of the 57 adopted brothers of homosexual males who were unrelated in terms of genetics. Strangely enough, the study also included pairs of biological brothers that were not twins, and out of the 142 homosexual males studied, only 14, or about 9 %, had homosexual brothers, which is

approximately the normal statistical incidence of homosexuality in the general population (Bailey and Pillard). More and more people are beginning to believe that homosexuality is not a “choice,” but rather a feeling that one is born with. As one grows older, they become aware of sexual feelings towards other persons. The only difference between homosexuals and heterosexuals is that while heterosexuals are attracted to members of the opposite sex, homosexuals are attracted to those of the same sex. Therefore, neither homosexuals nor heterosexuals really have a “choice” in the matter, and asking a homosexual “Why are you attracted to other members of your sex” is likened to asking a heterosexual “Why are you attracted to members of the opposite sex?” After all, why would anyone choose such a difficult lifestyle? As a homosexual, one risks horrifying their parents, other family members, and friends to the point of losing all contact with them, ruining their career, being condemned by their religion, being beat up for publicly displaying their sexuality, and much much more. Rather, the only actual choice is whether or not to be open with one’s sexuality and sexual preference; whether to act on one’s sexual desires, whether to tell others about such feelings, or whether to live a whole life as a lie and suppress these attractions (Marcus 9). To supplement my findings from research, I have conducted personal interviews with two adolescent homosexual males, James Dobbens and Daniel Woods. Both were asked how what they thought determined homosexuality (nature vs. nurture) and why, when they realized they were homosexual and how they knew, and similar questions. Both believed that homosexual was a result of nature, rather than nurture (Dobbens and Woods). Dobbens reasoned that most parents would not raise their children to be homosexual; “They’re not like ‘My child’s going to be gay!’” (Dobbens). Dobbens believes that he was born homosexual. When why he was a homosexual, he explained “It’s just the way you are, you can’t explain it, I was just born that way, it’s like asking how the world was made – no one really knows” (Dobbens). He went on to explain that while his nurture did not impact his sexual orientation, it did affect his view of it. When discussing the role of parents and upbringing in a child’s sexuality, he commented “They can bring you up [to be] open minded to [homosexuality], but they don’t bring

you up [to be a homosexual]... I grew up in a place where [homosexuality is accepted], so that's why I'm so open about being gay; I accept myself... [Whether or not you accept homosexuality and can be honest with yourself if you are homosexual] depends on how you're brought up" (Dobbens). Woods generally agreed with Dobbens, and justified his belief that homosexuality was caused by one's nature, rather than nurture, when he commented that "there's nothing in my upbringing that exposed me to anything like [homosexuality]" (Woods). When asked why he was gay, he explained that it is simply "something that I can't help... embedded in my head; nothing made me do it" (Woods). He went on to explain that it the same thing as the primitive attraction between males and females, only it was between males and other males. He added that "nothing happened [to make me homosexual]; it's always been there. I've always been attracted to guys" (Woods). What I find to be extraordinary about Woods' case is that he remarked how while he has always been sexually attracted to guys, he is romantically attracted to women in the sense that he has always wanted a girlfriend and to be close with other females, etc... but never in a sexual sense. Though he has "always kind of fantasized about getting married [to a female]," these feelings come from his "romantic side, which is different from the sexual side" (Woods). He says that if he married a female with which he shared a strong trust, maybe he could work it out. He then went back to talk about homosexuality as in innate characteristic; "It's like race. It's all something you can't help; it all comes to you... It's internal, you can't help it" (Woods). When asked if he thought homosexuality was caused by a gene, he commented that he thought that it was more of an instinct. He gave the example of birds, which are born with instincts such as knowing how to build a nest, or catch their prey. Woods also thought that maybe it was caused by something in the development of the embryo. All in all, Woods views homosexuality as "fascinating." His final thought: "I also believe everyone has at least one homosexual attraction. It's not black and white [homosexual vs. heterosexual]; it's a spectrum" (Woods). Yet there are arguments against homosexuality as caused by genetics. According to Darwin's theory of natural selection, the advantageous traits are passed on, while the disadvantageous ones eventually die

out. For instance, heterosexual males have an average of five times as many children as homosexual ones, as a female is required to reproduce children, yet homosexuals are, by definition, not sexually attracted to females. Therefore, from the evolutionary standpoint, homosexuality becomes one of the disadvantageous traits. Because consequently heterosexual men contribute five times as much genetic information to the next the gene pool, if homosexuality was indeed caused by a gene, it would have died out entirely by now, or at least been reduced immensely in the number of occurrences. Since neither of these events have yet taken place, it can be concluded that homosexuality is not caused by a gene (Fulkerson). My findings throughout my research have led me to conclude that there is no definitive answer on what causes homosexuality. No one knows for sure now (although many think they do), and perhaps no one ever will know the whole truth. Nonetheless, I am apt to believe that homosexuality is the product of some kind of combination of genetic and environmental causes. Perhaps one may have a genetic predisposition towards homosexuality that has been passed on from previous generations. However this does not necessarily determine a homosexual; rather this predisposition must be triggered by environmental factors. This is why Bailey and Pillard found in their study that while there was a greatly elevated chance that two males sharing the same genetic makeup would also share the same sexuality, this was not the case 100 % of the time. This seems to be the most likely explanation as of now, but even as I write this, new studies are being done. Perhaps the truth is right around the corner.