A quantitative study of information structure and right dislocation

in Cantonese spoken discourse

by

LAI Choi Ting

M.A., B.A.(Hons)

A thesis submitted in partial fulfillment of the requirements for
the Degree of Master of Philosophy
at the University of Hong Kong
February 2015

Abstract of thesis entitled

A quantitative study of information structure and right dislocation in Cantonese spoken discourse

Submitted by

LAI Choi Ting

for the degree of Master of Philosophy
at The University of Hong Kong
in February 2015

This study investigated information structure in spoken Cantonese and the use of right dislocation (RD) in genres with different degrees of shared knowledge and planning time pressure. Previous studies based on written data have suggested that Chinese word order is greatly determined by information structure. The pre-verbal referent in a sentence is reserved for background or topical information; while focus information is expressed in the post-verbal or right-most position in a sentence. The preferred argument structure (PAS) account has proposed a pattern of information flow in spoken discourse. New referents are free to appear pre-verbally in intransitive clauses, but rarely found in the subject of transitive clauses. However, it has also been suggested that RD is a focus marking device motivated by limited planning time in conversation. In RD structures, less important information is dislocated to the end of a sentence.

The data of the present study were extracted from a recently developed corpus

of oral narratives in Cantonese Chinese based on approximately 8 hours of recordings. The corpus consisted of speech samples from 144 native Cantonese speakers evenly distributed in three age groups, two education levels and gender. Three genres representing different degrees of shared knowledge and planning time pressure were chosen for analysis: procedural description, story-telling and recount of personal event. A total of 2,227 simple active clauses were examined for information structure analysis, and 117 instances of right dislocations (RD) were identified.

The results of the analysis of information structure based on personal recounts conformed to the PAS account. The number of new referents was statistically significantly higher in intransitive than transitive clauses. Besides, higher information pressure has resulted in a higher frequency of pre-verbal new referents in intransitive clauses. Investigation of RDs revealed that 1) a lower rate of occurrence of RD was found in the monologue data of this study than previous studies based on conversations; 2) the highest proportion of RD occurred in personal event recounting, which is assumed to have the least shared knowledge and the highest information pressure; 3) the highest proportion of RD appeared near the end of a narrative and tended to co-occur with disfluency; 4) RD was often used for the expression of emotions. The results show that information distribution in spoken discourse and the

use of RD were influenced by genre type, information pressure and speakers' assumption of shared knowledge.

(392 words)

DECLARATIONS

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualification.

Signed ……

LAI Choi Ting

ACKNOWLEDGEMENTS

This thesis could not have been completed without the great support that I have received from so many people over the years. I wish to offer my most heartfelt thanks to the following people.

To my supervisor, Dr. Sam Po LAW, a respectable and professional scholar. Thank you for your responsible supervision, constant guidance and insightful comments. You have inspired me to view things from different perspectives and bought me to a higher level of thinking. Thank you for giving me precious opportunity to be your student. I have learnt a lot from you and never thank you enough.

To my favorite teacher, Dr. Roxana FUNG. You have encouraged my academic interests since I was an undergraduate student. Thank you for opening up the research path for me. Your patient guidance and constant encouragement throughout the years are very important to me.

To my examiners, Professor Kang Kwong LUKE and Dr. Anita WONG. Thank you for your valuable comments and inspiring advice.

To my beloved parents. Thank you for allowing me to stay in school as long as I like. I am deeply indebted to your love, care and understanding. I am so grateful to have parents like you.

To my husband, Sunny, who have confidence in me at all times and have always encouraged me in my study. Your support, tolerance, companionship and love have been essential to get me through all challenges.

To my daughter, Lok Zi, who motivates me to be a better person. Thank you for the joy you bring to me. I wish you will do whatever you feel passionate about in the future, just like I do with linguistics research.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviation

A	Subject in transitive clause
Adj	Adjective
Adv	Adverb
Asp	Aspect Marker
CL	Classifier
Int	Interjection
N	Noun
Neg	Negation
Num	Numeral
O	Object in transitive clause
PAS	Preferred Argument Structure
Poss	Possessive marker
Pro	Pro-form
Quan	Quantifier
S	Subject in intransitive clause
SFP	Sentence final particle
Strpt	Structural particle
V	Verb

Transcription Conventions

The transcription symbols employed for the speech extracts in this study are mostly based on the conventions of Codes for the Human Analysis of Transcripts format (CHAT: MacWhinney, 2000).

[/] repetition

[//] Retrace

[/-] Restart

(.) short pause

(..) long pause

xxx unintelligible speech

Chapter 1 Introduction

Information structure refers to the packaging of old and new information in a sentence that meets the immediate communicative needs of the interlocutors (Chafe, 1976). In addition to syntactic and morphological constraints, information structure has been regarded as a "determining factor in the formal structuring of sentences" (Lambrecht, 1994, p.3). It is suggested that the information status of a referent plays an important role in determining the word order of a sentence. It can be used to explain the difference of the following example (LaPolla, 1995, p.309):

(1) a. 客人 來 了

Guest come SFP

"The guest comes"

b.來 了 客人

Come Asp Guest

"The guest comes"

These two sentences conveyed essentially the same meaning, and are clearly different in word order. In terms of information structure, the "guest" before the verb in (1)a is old information; the post-verbal "guest" in (1)b is assumed to be new information by the speaker to the listener. It has been suggested that focus information always manifests in the post-verbal or right-most position in Chinese

sentences (Xu, 2004). Among different word orders that can express the same meaning, information structure is important in understanding why a speaker prefers to produce a sentence in that particular order.

Contrary to the typical information structure that places background information before focus information, right-dislocation in Cantonese is suggested to be a focus marking device by fronting the important information in a sentence (Law, 2003; Packard, 1986). Consequently, the focus information is placed in the sentence initial position in right-dislocation. An example of right dislocation is given below:

(1)c. 你　飲　　咩　呀？

you drink what SPF?

“What do you like to drink?”

咖啡　吖，我 要

“Coffee SFP, I want”

“I want coffee”

Information structure and the role it plays in the grammar of a language have been discussed mainly in European languages and to a certain extent in Mandarin Chinese. Despite its relevance to spoken communication, information structure in Cantonese, which is essentially colloquial and does not have a standardized written

form as Mandarin Chinese, and the influential factors that affect its information flow in spoken discourse have yet to be investigated. In addition, notwithstanding the common usage of right dislocation by Cantonese speakers in everyday communication, it has been studied almost exclusively in the mode of conversation. The present study drew on data from a large corpus of oral narratives in Cantonese. The different genres of narrative in the database allow one to examine factors that are likely to influence information flow and the occurrence of right dislocation but have not been considered, such as the degree of shared knowledge between the speaker and the listener, and planning time pressure. Moreover, contrary to most previous works, this investigation carried out a quantitative analysis to examine information structure focusing on the information status of pre-verbal referents, and the use of right dislocation in different Cantonese discourse genres. The findings of current study will contribute to the typology of information structure so that comparisons can be made between languages.

1.1. Theoretical background of information structure

The first major contribution to the study of information distribution was made by the Prague School linguists. The theory of Functional Sentence Perspective (FSP) (Danes, 1974; Mathesius, 1975) introduced the concepts of “theme” and “rheme”. A

sentence is divided into these two major parts. Theme is the point of departure for the transmission of information. It is the basis for the speaker to transmit information to the listener. In order to establish a common basis between the speaker and the listener, the theme should be the information known to both, while rheme is the explanation, description and comment part of the theme. As such, theme represents old information; rheme represents new information. The following example is cited from Dik (1978, p.141):

(2) As for Paris, the Eiffel Tower is really spectacular.

_____ _____

Theme Rheme

Since theme provides background information for the listener to receive new proposition, it is suggested that theme is always being placed in a sentence initial position while rheme is towards the sentence final position. Recognizing that information status has its impact on syntactic structure, researchers started to investigate the type of information status from different perspectives (e.g. Chafe, 1987; Givon, 1990; Halliday, 1985)

1.1.1. Information status. Different classifications of information status have been proposed. Givon (1990) adopted the terms of old and new information. The function of old information is to serve as background for new information. He

referred to old information as “assumed by the speaker to be accessible to the hearer” (p.897), and new information as assumed by the speaker to be inaccessible to the hearer. Halliday (1985) viewed the distinction of given and new in terms of recoverability. Given information is assumed to be recoverable by the listener while new information is not. Prince (1981) proposed a taxonomy of given and new information in terms of “assumed familiarity”, i.e. the hypothesis by the speaker about the hearer’s belief. She proposed the terms “new”, “inferrable” and “evoked” with further sub-classification to apply to discourse referents. In short, an evoked referent refers to entities that are already available in the context. It can be understood as old information. Inferable referents are assumed by the speaker that the hearer could have inferred logically or through the evoked entities, such as “the driver” is inferable from the mention of “a bus”. Along the same line, but in addition to the binary distinction of given and new information, Chafe (1987) put forth three dimensions of information status: given, new and accessible. These three information statuses correspond to three activation states respectively: active, inactive and semi-active. More specifically, given information is in the active state. Given information must be active in the mind of the speaker, and the speaker also judges that piece of information as active for the hearer as well. An example is

provided by Chafe (1987, p.26):

(3)… I do think that makes a difference...

The pronoun “I” is weakly stressed to refer to the speaker himself because the speaker assumes that he is already active in the consciousness of the hearer. The pronoun “*that*” refers to an idea that has just been expressed by another speaker in the discourse. The idea encoded by “that” is thus considered active in the consciousness of all conversational participants. Chafe has also observed that given information often involves pronominalization with an attenuated stress or omission from verbalization altogether. Accessible information refers to information in the semi-active state. It is usually associated with the given information in the discourse. In other words, accessible information can be inferred from an existing referent. For example, when the entity of “class” is given in the context, the concepts of “students”, “instructor”, “classroom” or “lecture” are inferable from the given referent as accessible information. An example is provided to illustrate this concept:

(4)…I can recall…a big undergraduate class that I had….where …students loved the instructor… (1987, p. 29-30)

Given the mention of “a big undergraduate class”, the referents “student” and “the instructor” are classified as accessible information since they are semantically related

to the concept of “class”. “New” refers to the information that is not in the active or semi-active state of the hearer’s mind, based on the speaker’s assumption. Chafe has significantly contributed to the study of information flow in discourse. He has drawn our attention to the relationship between activation state and language use. The proposal of the notion of degree of accessibility of information has also enriched our understanding of information status.

Given information has been referred to as “topic”, “presupposition” or “theme”; new information has also been termed as “comment”, “focus” or “rheme”. Lambrecht (1994) asserted that the grammatical system of all languages can distinguish focus from presupposition in sentences. He referred to the association of grammatical forms and the presentation of focus as “focus structure”. Although the concepts of presupposition and focus are not exactly equivalent to given and new information, they are highly related in terms of availability in the speaker’s and hearer’s minds. The discussion of how focus information is manifested in word order is presented next.

1.1.2. Focus structure and word order. Different languages have their grammatical system, including referential form, morphological rule or syntactic structure, for coding and contrasting given and focus information in a sentence

(Lambrecht, 1994). Given information is often coded with pronouns or sometimes unexpressed at all. New information is likely, if not always, presented in full lexical form. In addition to this referential form difference, a language like English also relies on pitch accent with stress to mark the focus information in sentences. Japanese makes use of a morphological system with marker *wa* and *ga* to mark non-focus subject and focus nominal phrase, respectively (Shimojo, 1995). In Chinese, information is suggested to be managed mainly through word order (Ho, 1993; LaPolla, 1995; Xu, 2004; Zhang & Fang, 1996).

Word order in Mandarin Chinese has become a topic of great interest since Li and Thompson (1975) started examining its semantic and pragmatic functions. They formalized the relationship between word order and the definiteness of the nouns of a sentence in Chinese. The following tendency was noted (Li & Thompson, 1975, p. 170):

Tendency A: Nouns preceding the verb tend to be definite, while those following the verb tend to be indefinite.

This tendency is modified by three refinements:

Refinement 1: The noun in post-verbal position will be interpreted as indefinite unless it is morphologically, inherently, or anaphorically definite.

Refinement 2: A sentence-initial noun must be interpreted as definite; it may not be interpreted as indefinite even if preceded by numeral yī[1] "one".

Refinement 3: The noun following 被 bèi, although pre-verbal, is immune to Tendency A.

Partially agreeing with the observations made by Li and Thompson (1975), LaPolla (1995) and Chu (1998) argued that it was not definiteness that affected word order. Instead, it was "focus structure" (LaPolla, 1995) or "information management" (Chu, 1998) that determined word order. Chu (1998) regarded word order as a means of information management in Chinese. Given information is presented in sentence initial position to serve as scene setting; high informative value or new information is placed in the sentence final position to receive focus. Given information is often expressed by definite forms while new information is expressed by indefinite form. LaPolla (1995) explained tendency A of Li and Thompson (1975) in terms of information structure. It is suggested that when referents are newly introduced to the discourse, it is new information so it always occurs in the sentence final post-verbal position. New referents are always coded as indefinite. Therefore,

[1] The transcription system in this thesis used for annotating Mandarin is Pinyin; for Cantonese annotation is jyutping.

post-verbal position becomes associated with indefiniteness. The sentence initial position usually contains a topic, as topic is most often identifiable in the context; thus, the pre-verbal position becomes associated with definite noun phrases. LaPolla (1995, p.306) illustrated this point with the following example:

(5) 我 最近 買 了 一雙鞋子， 可是 穿 了 一 次，

I recently buy Asp one pair shoes but wear Asp one time

它 就 破 了。

It then break SFP.

"I bought a pair of shoes recently, but only wore (them) once and (they) broke."

The new referent is introduced as a noun preceded by a numeral and a classifier 一雙鞋子 "a pair of shoes". It is new information so it occupies the post-verbal position. The pronoun 它 "it" is used when the referent is active in the context and therefore ready to be presented as given information appearing before the verb.

In summary, the word order in Chinese sentences has been suggested to signify information structure rather than definiteness. The pre-verbal position is restricted to given or background information, but not necessarily definiteness. The post-verbal position is reserved for new or focus information instead of indefinite noun. In the

next section, how different sentence constructions in Mandarin Chinese are used to represent given and focus information is described in detail.

1.1.3. Focus manifestation in Mandarin Chinese. LaPolla (1990, 1995) has argued rigorously that information structure is the key determiner in Chinese word order. As a pragmatically dominant language, the order of elements in a Chinese sentence reflects the flow of information. Syntactic constructions in Chinese are used to grammaticalize information status, such as topic and focus, rather than syntactic roles, such as subject and object. In particular, LaPolla has proposed that focus information is always manifested in the post-verbal position in a Chinese sentence while topic appears in a sentence initial position. Syntactic constructions in Chinese reserved the post-verbal position for non-focal elements. However, this generalization was made largely by examining written data. It remains to be seen whether the tendency is strictly followed in the other modality, i.e. spoken discourse.

Xu (2004) investigated the focus position in a variety of Chinese syntactic constructions on the basis of LaPolla (1995). The canonical word order SVO, the alternative order SOV, as well as sentences with locative and duration expressions were examined. Similar to LaPolla, Xu (2004, p. 281) found that focus elements always take the “rightmost” location in a sentence.

(6) 你剛才喝了什麼？

You just now drank what?

'What did you drink just now?'

6a. 我喝了咖啡

I drank Asp coffee

6b. 我把咖啡喝了

I prep coffee drank sfp

I drank coffee'

(7) 你怎麼處理那些咖啡呢？

You how deal-with that coffee?

'What did you do to the coffee?'

6a. is the canonical word order SVO in which the object appears in the post verbal position. 6b is an alternative word order SOV in which the object is being placed before the verb. Xu (2004) argued that the less easily recoverable element will occupy the sentence final position. As such, 6a. is a better answer to the question (6) and 6b. is more appropriate for the question in (7).

Further demonstration of the relationship between word order and information structure comes from locative expressions in Chinese, which may appear before or

after the verb. Xu (2004) attributed the different word orders to be focus motivated. If the location is considered important information, it will occupy the post-verbal position.

(8) 有　什麼 新聞？

Have what news?

'Is there any news?'

8a. 台灣 發生　　地震

Taiwan took-place earthquake

'An earthquake took place in Taiwan'

8b. 地震　　發生　　在 台灣

Earthquake take-place in Taiwan

"An earthquake took place in Taiwan"

(9)　在 台灣 什麼 地方?

In Taiwan what place

"In what place of Taiwan?"

9a. 地震　　發生　　在 台中

Earthquake took-place in Taizhong

"The earthquake took place in Taizhong"

9b. 台中　　發生　　地震

Taizhong took-place earthquake

“The earthquake took place in Taizhong”

Both 8b. and 9b. in isolation are good in terms of grammaticality. However, since the focus information for the question is not located in the sentence final position, they are not appropriate answers for the questions (8) and (9), respectively, in terms of information structure.

Despite the attempt of Xu (2004) to enrich the discussion of grammaticalization of information structure in Chinese, some of the observations are questionable. The sentences in the examples are made up to illustrate the researcher’s claim. They might not accurately reflect actual language use, especially in a dynamic spoken discourse. As pointed out by Du Bois (2003), some of the sentence structures are totally acceptable in an isolated sentence level, but speakers rarely use them in a spontaneously spoken discourse due to pragmatic or information management reasons. Therefore, a spoken corpus would be preferred for the study of information structure in order to obtain a proper understanding in language use. However, it is not the common type of data in previous research in Chinese information structure for two reasons. Firstly, it requires a huge amount of time and manpower to establish

a large spoken language database. From data collection to transcribing, it is highly time and resource demanding. Secondly, there is a lack of an objective, applicable, and generalizable framework to describe, testify or quantify a large amount of spoken data. Most of the analyses used in previous studies have been descriptive. It is difficult for other researchers to generalize the observations to other sets of language sample. As a consequence, quantitative and generalizable results are rare. In order to fill this gap, a systematic quantifiable framework, known as Preferred Argument Structure (PAS) (Du Bois, 1987), will be adopted in this study to investigate the language data in a recently developed large scale spoken corpus in Cantonese (Kong, Law, & Lee, 2010-2014). The outcomes of this study will contribute to a more realistic or possibly accurate picture of actual use of language with quantitative evidence.

1.2. Preferred Argument Structure

The theory of preferred argument structure (PAS) was proposed by Du Bois (1987) based on spoken narratives of Sacapultec, a Mayan language of Guatemala. PAS explicates the relationship between discourse and grammar. In particular, it correlates word order and information flow in spoken discourse. PAS puts forth a hypothesis that certain argument structure is more systematically preferred over

other alternatives in spontaneous speech. Following Dixon (1979), Du Bois denoted the subject of an intransitive clause as S, the subject of a transitive clause as A, and the direct object of a transitive clause as O[2].

1.2.1. The PAS constraints. There are both grammatical and pragmatic dimensions of PAS, expressed as "constraints". The grammatical aspect can be summarized in two constraints – (1) the one lexical argument constraint, and (2) non-lexical A constraint. The one lexical argument constraint refers to a tendency that no more than one lexical NP is present in a clause. Additional arguments tend to be expressed as pronouns or zero form. The non-lexical A constraint suggests the tendency of extremely rare number of lexical form found in the A role, while the roles S and O are free to accommodate lexical noun. The pragmatic dimensions are derived from these two grammatical constraints. Assuming that lexical NP carries new information, the one lexical argument constraint reflects the tendency of clauses to contain no more than one piece of new information. The non-lexical A constraint suggests that new information tends to appear in the O or S role, rarely in the A role. The constraints are summarized in Table 1.1.

Table 1.1. ***Preferred Argument Structure Constraints***

[2] In the rest of the thesis, S, A, and O will be used interchangeably with the subject of intransitive, the subject of transitive, and the object, respectively.

	Grammatical	Pragmatic
Quantity	One lexical argument constraint	Avoid more than one new information in a clause
Role	Non- lexical A Constraint	Non-new information in A role

Source: Adapted from Du Bois (1987)

Du Bois (2003, pp.35-37) elaborated the constraints with English transitive clauses:

(10) he named like half a dozen viruses, …

(11) …but I still miss my grandmother.

In these clauses, one argument is presented with a lexical noun phrase ("viruses" and "grandmother"), and one is realized with a pronoun ("he" and "I"). It is argued that speakers strongly prefer to avoid using two full lexical arguments in a transitive clause. However, it is acceptable with clauses showing only pronouns or zero form in the two arguments:

(12) …and she admired him, …

There is a tendency that only one lexical noun phrase is preferred in one clause. When there are two arguments in the clause, speakers do not treat them equally as potential slots for lexical noun phrase realization. They tend to avoid placing lexical referent in the position of transitive subject; the position of O is more preferable for

lexical argument. Different from the typical information structure in which new information is unfavored in the pre-verbal subject position, Du Bois claimed that speakers are open to having new lexical form in the subject role, as long as it is an intransitive clause. In other words, The S role is generally unconstrained.

PAS has been applied extensively to analyze different languages, including English, Korean, Finnish, Portuguese and Chinese. Besides, it has also been attested in a number of different linguistic genres, such as classroom discourse (Kumpf, 2003), aphasic speech (Kohn & Cragnolino, 2003) and discourse of autistic children (Weber, 2003). These studies have generally confirmed the proposed constraints.

1.2.2. Information pressure. Du Bois (1987) proposed a notion of "information pressure" (p.834), which can be estimated by the ratio of new referents to clauses. Information pressure is high when a number of new referents are introduced within a short span of discourse. Du Bois (1987) observed that information pressure affects the distribution of new referents in a discourse. As information pressure increases, intransitive subjects and objects will be filled with more lexical or new arguments – but not the transitive subject. It was summarized that "A tends to be consistently low in new and lexical mention content and O tends to be fairly consistently high; but S seems to be the most responsive to increasing or decreasing information pressure.

As information pressure rises, intransitive verbs are more often invoked for carrying out the introduction of new protagonist(s); this results in a higher frequency of new and lexical mentions in S (Du Bois, 1987, p.836).” The researcher further pointed out that information pressure is correlated with discourse genre. Telling a story to strangers demonstrated a higher information pressure than a conversation between friends or family members, where speakers share large amounts of currently active background information. Du Bois emphasized that “while conversation may well be the more frequent genre, narrative is especially likely to display conditions of relatively high information pressure; for this reason, it is especially valuable for effective testing of proposed quantity and role constraints (p.836)”. Chui (2005) and Long (1990), the only two studies that applied PAS in Mandarin Chinese, were mostly based on conversational data which may be inadequate in testing the constraints given the low information pressure. Following Du Bois’ suggestion, language samples of a personal recount narrative were employed for PAS analysis in the present study. Information pressure in the tested discourse is assumed to be high since there is minimal shared knowledge between the speaker and the listener (the interviewer).

1.2.3. PAS studies in Mandarin Chinese. Long (1990) was the first investigation that applied PAS in Mandarin Chinese conversational discourse. It adopted a

quantitative approach to examine the validity of all of the four constraints. The language sample was taken from two daily conservations with multiple participants. The results in Long (1990) supported the PAS in all four constraints. For the one lexical argument constraint, there were about 95% of the data containing zero or one lexical argument in one clause. The statistics showed a clear tendency of avoiding more than one lexical argument per clause in Mandarin Chinese. The constraint of avoiding more than one piece of new information per clause was also supported. There was no single clause containing more than one new information argument in the data. The majority of clauses (84%) carried no new argument. The remaining 16% of the clauses had one new argument. The data also showed that the largest proportion of them (43%) appear in the O role. The lowest rate of occurrence of new information (9%) was found in the A role, as predicted by PAS.

There are two interesting differences between the result of Long (1990) and Du Bois (1987). The first was a higher frequency (84%) of non-new argument clauses found in Long. It means the majority of the data contain no new argument. The second was a lower proportion (7.5%) of new referents found in role S compared to that in Du Bois (22.5%). Long attributed these differences to the shared knowledge of the participants in the conversations. Since the language sample was drawn from

casual conversations over the dining table, it can be assumed that the speakers knew each other quite well. The high degree of shared knowledge among them makes some of the referents more accessible in the context. He also observed that the deictic context played a significant role. Speakers might point at the existing referents instead of verbalizing them. These indicate that shared knowledge and the physical setting of the communicative environment in which sample data were collected are important. They directly affect the use of different referential forms. The manipulation of degree of shared knowledge and the choice of topics in the data collection process will be necessary in obtaining a more comprehensive picture. Recalling that in the PAS framework, Du Bois proposed that high information pressure would lead to high lexical S in a discourse. A relaxing conversational environment may indicate relatively low information pressure, which would lead to a lower number of lexical forms in the S role. In summary, Long (1990) suggested that PAS was applicable to Mandarin Chinese. The information status of referents is often related to argument roles and word order in Chinese.

Contrary to what has been suggested by Long (1990), Chui (2005) argued against the universality of PAS. The information status of referents in her data was found to be less associated with argument roles. The language data of Chui (2005) came from

one conversation and two oral narratives. The conversation was recorded over a meal among four colleagues. The narratives were about recounting of two films to a third-person. Although two types of discourse were adopted in the study, the author did not compare or contrast the data. For the constraint of non-new information in A role, Chui's data supported the claim with an average of 98% of the referents in the A role considered given information. Regarding the roles of S and O, Chui argued that "taking the roles as unified categories may fail to give a coherent account of information structuring (p.45)". Therefore, the roles S and O were further classified into pre-verbal and post-verbal positions. The post-verbal S mainly referred to the S referent appears after existential verb "have":

(13) 有　一 對　戀人，男　的 在銀行工作， 女　　的　是一個藝術家

Have one CL lover , Male Stprt at bank work, Female Stprt is one CL artist

"There is a couple, the man works in a bank; The woman is an artist " (p.54)

Chui treated the 戀人"the couple" as post-verbal S since it followed the verb 有 "have". However, Chui has failed to note that the structure in (13), known as presentative sentence, has the main function of introducing new referent into the discourse. Given this background knowledge, it is not surprising to find 69% of new referent appearing in the post-verbal S role. To what extent PAS is applicable to

describe presentative sentences is worth investigating. Simply combining this structure into investigating the post-verbal S would be misleading. The role O was further sub-classified into “clause-initial O”, “marked O” and “post-verbal O”; the first two were then combined into pre-verbal O type. An example of “clause-initial O” is provided as following (p.48):

(14) 她 寫 的 字, 你 看 了 會 吐血 的

She write Stprt letter, you see Asp will feel miserable

“ Her writing, you will feel miserable after seeing”

When we examine the example closely, the object 她寫的字 “her writing” here is actually a topic of the sentence. Some have suggested that topic is a pragmatic notion, rather than a syntactic property. Topicality has been a controversial subject matter in Chinese linguistic studies. There are three major views about the relationship between topic and subject in Chinese. The first maintain that there is only subject, but no topic in Chinese and that subject-predicate is the fundamental relationship between VP and preverbal in Chinese sentences. The second view claims that syntactic notions like subject and object are not grammaticalized in Chinese, based on the argument that information structures rather than syntactic structures are used in Chinese to convey information and that the relevant grammaticalized

notions in Chinese are topic and focus (LaPolla, 1990, 1993). There could be more than one topic in a sentence, if there are two or more NPs in front of the verb. The third view recognizes that both topic and subject exist in Chinese as separate grammatical notions and the two can appear in the same sentence.

Long (1990) excluded topic sentences from his analysis. The present study has, therefore, adopted Long's position that it would not be appropriate to merge pragmatic notions and syntactic roles in an analysis. The "marked O" refers to the object in passive construction (also known as *ba* construction). In passive construction, the object is being fronted to before the verb:

(15) 還是 把 錢 給 捐走 了

Still *Ba* money Stprt donate SFP

" the money was still being donated"

In the canonical word order, 錢 "money" would be the object that normally occurs after the verb 捐走 "donate". It is being fronted to the pre-verbal position in this passive construction. Pre-verbal O has a greater percentage of given information than post-verbal O. There is nearly 80% of pre-verbal O occupied by given information, while about 60% of post-verbal O is given information. Chui (2005) concluded "a significantly high percentage of given O in the pre-verbal position, the

constraint on “given” in Chinese discourse is better characterized in terms of word order rather than by argument role (p.57)”.

A close examination of Chui (2005) reveals that there are two major problems. Firstly, the statistics shown in the paper were not in line with the claims made by the researcher. The extremely small number of new referent (2%) appearing in the transitive subject position in fact confirmed the PAS tendency of non-new information in A role. The statistical difference in the pre-verbal and post-verbal S and O might not be significant to support the researcher’s claim. Secondly, Chui did not take into account the characteristics in different syntactic constructions. Sentences with canonical word order, topic-comment structure, presentative sentences as well as passive constructions were combined during analysis. In light of the underlying problem of merging different syntactic structures for analysis in Chui (2005), the present study has focused on canonical simple active clauses in the spoken corpus.

1.3. Right Dislocation

Right dislocation is an interesting construction in terms of information structure, see examples (16) and (17) below:

(16) 唔 拎　　遮　　喇， 我

Not bring umbrella sfp , I

“I don’t bring umbrella”

(17) 我 唔 拎 喇 , 把 遮

I not bring sfp , CL umbrella

“I don’t bring umbrella”

In (16), the referent 我 “I” is the subject of the sentence which is usually placed in a clause initial position in canonical word order, but it is dislocated to the sentence final position after the sentence final particle (SFP). (17) is an example of object dislocation, where the object referent “umbrella” is dislocated to the right-most position after the SFP. In addition to noun phrases that can be dislocated, a wide range of constituents, such as verbs, adverbs, complements and adverbial clauses can also be found in this dislocated slot (Cheung, 1997; 2005). Contrary to the given before new information distribution, the focus information is suggested to be dislocated to the initial position in a sentence preceding the given information. As pointed out by Luke (2012), this grammatical phenomenon has been given many different names. It was initially proposed by Chao (1968) as “afterthought”, referring to the unplanned content that was added to a completed sentence. Packard (1986) has argued that it is actually a left dislocation in which the focus information in the

sentence has undergone a left dislocation movement to the sentence initial position. Other terms for this construction include “transposition” (Lu, 1980), “dislocation” (Liang, 2002), “focus-fronting” (Luke & Zhang, 2007; Zhang & Fang, 1996), and “incremental sentences” (Luke, 2004; 2012).

Liang (2002, p. 85) proposed a set of criteria for identifying Cantonese right dislocation as follows:

i. The SP (sentence particles) occurs between the preceding part and the dislocated unit.

ii. No noticeable pause occurs between the preceding part and the dislocated unit, but a pause stands to the right of the dislocated unit.

iii. There is unified rising-falling contour of energy trace in spectrogram.

iv. There is semantic dependency between the dislocated unit and preceding part.

Two main approaches can be found to study right dislocation. The formal grammatical approach has been the focus in previous works, which studied the syntactic relations between constituents in dislocation sentences, as well as the structural constraints of the grammatical process of right dislocation (Cheung, 1997; 2005). Alternatively, the functional approach focuses on the discourse functions and

pragmatic motivations of right dislocation in verbal communication.

Law (2003) briefly suggested that right dislocation in Cantonese is a device to mark narrow focus. It is a result of leftward movement of focus constituent; the fronted phrase preceding the sentence final particle (SFP) must be interpreted as focus.

In terms of the size and scope of data, the most detailed study of Cantonese right dislocation from a functional perspective is Liang (2002). Twenty-five conversations and one story-telling in the Corpus of Spoken Cantonese (Luke & Nancarrow, 1997) were selected for analyzing instances of right dislocation. Most of the speakers aged between 21 and 26 with high education level. Liang reported that 99 % of the 600 instances of right dislocation identified were from conversation. However, this result may not be meaningful since conversations outnumbered story-telling by 25 times in the data samples. Some of the examples in Liang (2002, pp. 160-161) are cited below, in which the right dislocated element is underlined:

(18) Speaker A: 我哋 不如 講吓 買 車 吖

we why-not talk buy car SFP

"Why don't we talk about buying car"

Speaker B: 好 啊

good SFP

“why not”

→ Speaker A : 諗住　邊　個 俾 啊 , 架 車

thinking who CL pay sfp ,　CL car

“Who are going to pay for the car?”

Speaker B: 一 人 一 半 囉

one people half sfp

“Each pays half the price. ”

(19) Speaker A：諗住 去 邊度　旅行　啊

thinking go where travel　sfp

“Where do you wish to tour around?

→ Speaker B : 我 想 睇雪　　唔係 睇　櫻花　　啊 , 最 正

I want see-snow not be see cherry-blossom SFP , most

fantastic

“I’d like to see snow-not so it's the most fantastic to see

flowering cherry.”

Speaker A :　瞰 即係 要 日本

so that-is need Japan

"So the choice must be Japan."

→ Speaker B: 櫻花 係 幾 月 開 啊，日本？

cherry-blossom be which month blossom SFP , Japan?

"When will flowering cherry be blossoming in Japan?"

1.3.1. The motivations of using right dislocation. Liang (2002) agreed with Zhang and Fang (1996) that due to the lack of planning time, speakers in conversational discourse tend to place the most important information (i.e. the focus) in the initial position of a sentence, and then fill in the less important information at the end. Thus, it is suggested that right dislocations mainly occurred in conversation. Although Liang (2002) represented the largest scale of work on Cantonese right dislocation related to information structure to date, there are several limitations regarding the methodology. Firstly, the speech data were almost exclusively drawn from conversations. It was hypothesized that due to the lack of planning time, right dislocation mainly occurred in conversations. Such a hypothesis was not properly assessed with the data. It is not clear whether other discourse types, such as story-telling or picture description may also contain right dislocation. If the lack of planning time is the primary reason for using right dislocation, then the occurrence of this structure should be minimal in monologue discourse with familiar topics. In

order to understand the motivation behind right dislocation, a greater variety of discourse types other than conversation should be studied. Secondly, the sociolinguistic background of Liang's participants is quite restricted, i.e. young speakers with relatively high education. The data cannot be said to be representative of use of right dislocation in Cantonese.

A conversation analytic account was adopted in Luke (2012) in analyzing right dislocation (incremental sentences in his term). By examining naturally occurring conversational data, Luke concluded that right dislocation performs a variety of functions: intensification, emphasis, backgrounding, qualification, clarification, and disambiguation. An example of intensification (p.347) is cited here:

(20) Ian: …但係 都 凍，好似話，啲 天氣 好似 琴日 嚟 囉

But Adv cold, I hear, CL weather like yesterday Pro SFP

"… but still cold, I hear, the weather's like yesterday"

→ Joe: 有 冇 check 過 呀 , 你 ？

Have not check Stprt SFP , you

" Have you checked? "

The second person pronoun 你 "you" is dislocated to the sentence final position. In terms of information load, this pronoun carries very little information since the

addresser knew it well that the question is for him. The speaker intensifies the addressee 你 “you” by placing it at the sentence final position with a slight increase in length and loudness. The intensification effect can also be achieved through repetition:

→ (21) 佢 而家 真係 勁 呀 , 佢 而家

He now really great SFP , He now

“He’s really great now, he really is now” (p.350)

The third person pronoun 佢 “he”, together with the time phrase 而家 “now” are repeated. Luke claimed this repetition gives the utterance a particularly animated and emphatic character. Supplementing background information is another major function of right dislocation (Luke, 2012). Luke observed that the dislocated slot is often used for providing background or secondary information. In a conversation, before a speaker ends his/her turn, right dislocation provides a last opportunity for secondary information to be introduced.

(22) Ian: 去 過 邊度 呀?

Go Asp where SFP?

“Where have you been?”

→ Joe: …我 諗 小學四年班，同 媽咪 去。兩 個 人之嘛 , 去旅行

I think primary four, perp mother go. Two CL people SFP , go travel.

“When I was in Primary 4. Went with mum, just the two of us went, went traveling”

The dislocated phrase 去旅行 “go travel” belongs to accessible information since the speakers were talking about travelling in Europe. However, this information is not redundant given that the mother could have gone to Europe for other purpose rather than travelling. It is regarded as secondary information that the speaker intended to supplement before ending his turn. The last two suggested functions, revision and disambiguation, appeared to be quite similar. Luke (2012) explained that revision is “to supply additional information in order to revise or clarify a reference or to avoid potential misunderstanding” (p.355). The function of disambiguation is to “add detail to a reference made in the main part of the sentence for the purpose of clarification or disambiguation” (p.358). Although some of the functions suggested in Luke are not so distinctive, the discussion provides important insights that the discourse context does influence the purpose of using right dislocation. Hence, it is important to investigate right dislocations with consideration of the context in which they are uttered.

1.3.2. The affective functions of right dislocation. In Mandarin Chinese, Guo (1999) examined the use of right dislocation of a discourse produced by children at play. He discovered that right dislocation often served an affective function to focus the addressee's attention and convey the speaker's negative feeling. By analyzing the dislocated noun phrase, he classified three main types of right dislocation: (i) zero anaphoric, (ii) elaborations, and (iii) reduplications. Examples (Guo, 1999, p.1111) are provided to illustrate these three types, respectively:

(23) 都 在 哪兒 ， 我 那 個?

All at where SFP , I that CL?

"Where is (it), that one of mine?"

The above example does not have explicit co-referent in the sentence. Therefore, it is an example of zero anaphoric right dislocation.

(24) 這 不 是 做操 ， 這 音樂

this not be do exercise, this music

"This is not (for) doing exercises, this music"

(24) is an example of elaboration. In this category, the co-referent in the main clause is usually a pronoun and the dislocated part is a lexical form.

(25) 這 剪刀 怎麼 兩 半 了, 這 剪刀?

this scissors how-come two half SFP , this scissors

"How come this pair of scissors is broken, this pair of scissors?"

(25) is an example of reduplication. The dislocated form being exactly the same as its co-referent without information value added.

The types of zero anaphoric and elaboration of right dislocation are found to be mainly used for statement purpose, while reduplications extensively serve the functions of questions, ridicules and reprimands. Based on these findings, he argued that right dislocation in Mandarin Chinese has developed an emphatic function to express the speaker's focus of attention. He further suggested a hypothesis of Mandarin Chinese right dislocation development. Right dislocation has originally been a repair device in spoken discourse as afterthoughts. It has then become grammaticalized for managing information. Subsequently, it has taken up the emphatic function. Finally, it has developed the affective function to express the speaker's negative attitude.

1.4. Interim summary

The foregoing review of the literature has reflected our current understanding of how information status influences sentence structure and word order. However, there are limitations in terms of the size of dataset and

methodologies. Some of the studies were based solely on written materials (LaPolla, 1995, 2009; Xu, 2004). The observations made might fail to account for patterns of usage in spoken discourse. When spoken discourse was adopted, conversational data were almost exclusively focused on (Liang, 2002; Luke, 2012). Nearly all of the conversational data were recorded casually among speakers who were friends with supposedly a great deal of shared knowledge. In such cases, the researcher would have little idea of which referent available in the shared knowledge among the speaker and listener. A newly introduced referent in a discourse is not necessarily new in the minds of the conversational partners. Without being able to determine what is or is not shared knowledge, the researcher would not accurately know the information status of a referent in a discourse. When a referent is mentioned in the discourse for the first time, the researcher may simply categorize it as new information. However, the referent might implicitly exist in the shared knowledge among the speaker and the listener. As a consequence, the word order or lexical coding of a referent might have been affected.

The present study takes into consideration shared knowledge and information pressure as factors that influence information structure. Genres other than

conversation are chosen for analysis to represent different degrees of shared knowledge and information pressure in order to examine how these factors interact with information status of referents.

1.5. Research questions and hypotheses

In an attempt to fill the gap by supplementing normative and quantitative data of information structure in spoken discourse, we will address the following research questions using a large scale Cantonese spoken corpus (Kong, Law, & Lee, 2010-2014) :

1. Are the PAS constraints applicable to Cantonese spoken discourse?

It is predicted that the majority of clauses do not contain two new arguments, and the subject role of transitive clauses is least likely to contain a new referent.

2. Is the distribution of new pre-verbal referent sensitive to the degree of information pressure associated with the discourse genre?

It is predicted that the chance of new referent occurring in the intransitive subject position increases when the information pressure of the genre is high.

Regarding right dislocation, the following questions will be addressed:

3. How extensive is the use of right dislocation in Cantonese monologues?
4. Is the use of right dislocation similar across speakers differing in age, gender,

and education level?

5. How do shared knowledge and information pressure influence the frequency of occurrence of right dislocation?

It is hypothesized that genre with less shared knowledge and higher information pressure will yield greater use of right dislocation.

Chapter 2 Methodology

Taking advantage of the recent development of a large scale Cantonese spoken corpus, a quantitative approach is adopted in this study to answer the research questions stated in Chapter 1. The details of the corpus, the analytic methods and a demonstration of data analysis will be presented.

2.1. Data corpus

The data of this study were drawn from a spoken Cantonese corpus of oral narratives (CANON). CANON is part of a large scale research project with an overarching goal to improve the planning of assessment and remediation procedures for Chinese speakers with aphasia worldwide (Kong, Law, & Lee, 2010-14). CANON serves as a control database for a parallel corpus of aphasic discourse production. These corpora will form a component of the AphasiaBank consortium (MacWhinney et al.,2008; http://talkbank.org/AphasiaBank/), which is a shared database of orthographic transcriptions, audio and/or video recordings from language databases in 15 laboratories across the United States and other English speaking countries, including Australia and United Kingdom.

2.1.1. Participants. CANON is composed of language samples from neurologically unimpaired right-handed native Cantonese speakers. Aiming to

include participants with different demographic backgrounds, CANON consists of 12 sub-groups classified by gender (male and female), age (18-39 years old; 40-59 years old and above 60), and education level (higher or lower than secondary school for the two younger groups; and higher or low than primary school for the oldest group). Each sub-group contains at least 12 participants. Altogether there are 144 participants who are balanced in gender, age and education level in the corpus. Their demographic information is summarized in Table 2.1.

Table 2.1. Demographic background of normal speakers in CANON

	Young (18-39 years old)		**Middle age (40-59 years old)**		**Old (60 or above)**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Low Education level (F.5 or below for younger and middle age group / 6 years of education or below for older age group)	12	12	12	12	12	12
High Education level (F.5 above for younger and middle age group / more than 6 years of education for older age group)	12	12	12	12	12	12
Total	24	24	24	24	24	24

2.1.2. Procedure of data collection. A face-to-face interview was conducted for each participant following the AphasiaBank protocol. Connected speech samples were elicited by means of four narrative tasks, including descriptions of picture stimuli, procedural descriptions, story-telling, and recount of a personally important event. Picture stimuli consisted of three tasks- Broken Window, Refused Umbrella, and Cat Rescue. The participant was asked to describe the stimulus pictures. Procedural description involved describing the procedure of making a ham and egg sandwich with a photo of bread, egg and ham. Story-telling included two highly familiar stories- “The Hare and the Tortoise” and “The Boy who Cried Wolf”. Story cards without words were provided to the speaker before they started telling the story (see Appendix 1). They were required to tell the stories without the picture cues. Additionally, the participant was asked to recount an important event in their life. In other words, each participant provided eight monologue samples in total. During the interview, the examiner’s input was kept to a minimum. At the time of data collection, the participants’ speech was videotaped and audio-taped.

All language samples were orthographically transcribed. Inter-rater reliability of orthographic transcription was computed by randomly selecting 10% of the samples and double-checked against the audio recordings by two native speakers, and was

found to have an agreement of greater than 99%. The transcriptions were formatted according to the Codes for the Human Analysis of Transcripts (CHAT) (MacWhinney, 2000). They were linked to audio and video recordings through a computerized analytic program called Child Language Analyses (CLAN) (MacWhinney, 2003). CLAN allows one to carry out a variety of linguistic analyses, such as frequency count, lexical diversity, mean length of utterance, as well as searches for user-specified combinations of words, character strings, or words in context.

2.2. Data Analyses

2.2.1. Preferred Argument Structure analyses.

2.2.1.1.Unit of analysis. All narratives were orthographically transcribed. The transcriptions were segmented into clauses for analyses. As proposed in Givon (1983), the basic information processing unit in discourse is a clause. Some discourse researchers suggested the smallest unit in spoken discourse should be an intonation unit, which is an utterance under a single coherent intonation contour (cf. Chafe, 1980, 1987). According to Du Bois (1987), a clause and an intonation unit are often coincided in spoken discourse. Following Long (1990) and Du Bois (1987), the working definition of a clause in this study is a “stretch of speech containing a verb and its arguments that is uttered under a single coherent intonation contour”

typically bounded by a pause. To illustrate, consider the following excerpt, which is a sample of the first few clauses from a participant:

(26)

(i) 有 一 次 e6 去 旅行 呢.

Have one CL filler go travel SFP

"Once we were going to travel"

(ii) 同 我 先生 去 旅行

With my husband go travel

"Traveling with my husband"

(iii) 去 嗰 個 瑞士 嗰 度 呀

Go that CL Swiss that place SFP

"go to the Switzerland"

(iv) 嗰 度 好 靚

That place very beautiful

"It is very beautiful"

(v) 即係 嗰 啲 風景 呢 好 值得 (.) 懷念

That is that CL scene SFP very worth yearning

"The scene there is very yearning"

Each line represented one clause that was produced under a single intonation contour. A short pause was also noted between clauses. In the present study, the PAS analysis was restricted to simple active clauses containing one verb. Serial verb construction[3], *zoeng1* construction and *bei2* construction were not included for analysis. Inter-rater reliability of clause segmentation was found to have an agreement of 96.4%. The selected clauses were then divided into intransitive and transitive clauses. A typical type of intransitive clause takes an intransitive verb (underlined in (27)) or adjective (underlined in (28)):

(27) 結果 佢哋 都 能夠 生還.

Consequently they Adv able survive

"They are able to survive at last"

(28) 我 老細 就 好 好人 嘅.

My boss adv very nice SFP

"My boss is very nice"

As the verb "survive" and adjective "nice" do not take any object argument, the examples in (27) and (28) belong to the intransitive clause type. Another type of intransitive clause contains a copula verb followed by a predicative complement:

[3] Serial verb construction constitutes about 18.5% of the total number of clauses in the present database.

(29) 我 個 女 係 中醫師 嚟嘅

I CL daughter be Chinese medicine doctor SFP

“My daughter is a Chinese medicine doctor”

The non-referential noun phrase “Chinese medicine doctor” is not the object of the clause but serves as a complement to characterize or add information to the subject referent.

In a transitive clause, the verb takes an object argument. Semantically, the agent is doing something to or directing some behavior to the object referent as in (30).

(30) 我 要 更新 公司 電腦 嘅 系統

I need update office computer POS system

“I need to update the office computer’s system”

公司電腦嘅系統 “office computer’s system” is the object in this clause. Another type of transitive clause identified in the corpus involved an evaluative verb and a predicative complement:

(31) 我 覺得 哩 次 意外 裡面 充滿 神 嘅 恩典

I feel this CL accident inside full of God POS grace

“I feel that this accident was full of God’s grace”

The noun phrase 哩次意外"this accident" was treated as an object in example (31) although it did not undergo any action taken by the agent physically.

2.2.1.2. Referential form & information status of arguments. The analysis for this section was only done on the recounts of personal experience. The pre-verbal referent in intransitive clauses was classified as S; the pre-verbal and post-verbal referents in transitive clauses were coded as A and O, respectively. The referential forms and information status of all occurrences of S, A, and O were examined. The referential forms included 1) NP, e.g. 公司電腦嘅系統 "office computer's system" in (30) and 哩次意外 "this accident" in (31); 2) pronoun, and 3) zero form, e.g. the subject 我 "I" in (32) was not expressed in the second clause since it was mentioned in the first clause.

Three types of information status are recognized, following Chafe (1987), namely, new, given, and accessible. The category "new" is assigned to the referent that has not yet been mentioned in the discourse and the interviewer cannot infer it from the immediate context. "Given" refers to a referent that has been overtly mentioned in the previous discourse context. A referent is "accessible" when it is inferable from the previously mentioned context, or generally known to the

interviewer. The excerpt below shows an example of accessible information which is underlined:

(32) 我 份 工 返 clinic 嘅

I CL job go clinic SFP

"I work at a clinic"

有時 都 幾 清閒

Sometimes Adv quite free

"it is quite free sometimes"

因為 個 醫生 成日 周圍 飛.

Because CL doctor always around fly

"it is because the doctor is always out of town"

The NP 個醫生 "the doctor" is assigned as accessible information since it can be referred from the term "clinic" in the previous clause.

The following is another excerpt with detailed explanations of the analysis of each clause.

(33)

	Clause	Analysis
i.	ze1 (我) 返去 中心 做 義工 呢	It is a serial verb construction and

	(I) go back centre do volunteer SFP “ I was volunteering at a centre”	therefore is excluded from analysis.
ii.	(我) 搞 下午茶聚 嘅. (I) organize tea gathering SFP “ I organized tea gathering”	It is a transitive clause because the verb 搞 “organize” takes the 下午茶聚 “tea gathering” as the object. A covert pronoun 我 “I” is classified as an A(gent) in this clause and it is considered given information. The NP 下午茶聚 “Tea gathering” is an O(bject) and new information since it is newly mentioned in the discourse.
iii.	好 多 老友記 參加 Very much elderly buddies join “There are many elderly buddies joining”	It is an intransitive clause. The NP 老友記 “elderly buddies” is a S. It has not been mentioned before, so it is new information.
iv.	突然間 個 婆婆 暈 咗 Suddenly CL old lady faint Asp “And old lady fainted suddenly”	It is an intransitive clause. The NP 婆婆 “old lady” is a S. It is classified as accessible information since it can be inferred from the NP 老友記 “elderly

		buddies” in the previous clause.
v.	我 馬上 aa6 去 叫 啲 姑娘 I immediately go call CL staff “I went to find the staff immediately”	It is a transitive clause. The pronoun 我 “I” is an A and given information. The NP 姑娘 “the staff” is new information.
vi.	之後 呢 e6 同 佢 xxx. After SFP with she	Xxx stands for unintelligible speech. It is an incomplete clause thus excluded from analysis
vii.	後尾 呀 十字車 嚟到 呢 later SFP ambulance arrive SFP “Ambulance arrived”	It is an intransitive clause. The NP 十字車 “ambulance” is the subject. It cannot be directly inferred from the previous context, so it is new information .
viii.	就 將 佢 送 咗 Adv zeong1 she send Asp 醫院 aa6 同埋 急救 hospital fil and first-aid “ she was sent to hospital for	*Zeong1* construction is excluded from analysis.

	rescue”	
ix.	好彩 我 發覺 得 早 Luckily I find out strpt early “Luckily I found out early”	It is an intransitive clause. The pronoun is the subject and given information.
x.	(我) 覺得 哩 件 事 (I) think this CL incident 幾 得意 quite interesting “I found this incident pretty interesting”	It is a transitive clause with a covert subject 我“I”. The NP 哩件事 “this incident” is the object and belongs to the type of given information

2.2.1.3. Statistical analysis. The rates of occurrence of S and A with accessible, given and new information of each participant were calculated, respectively. Wilcoxon signed-rank test with Bonferroni correction ($p < 0.05/3 = 0.0167$) would be used to compare the rates of occurrence of S and A in each information status if the assumption of normality was violated. In addition, to assess the prediction of the PAS account, the number of new referent(s) in transitive clauses was counted. A transitive clause could have zero, one, or two new referents. Chi-square goodness-of-fit test was employed to see whether the numbers of transitive clauses

containing different numbers of new referents were equally distributed.

Data from 15 subjects (10%) were selected randomly and the information status was re-analyzed by the author and two trained independent raters to obtain intra-rater and inter-rater reliabilities, respectively.

2.2.2. Right dislocation analyses.

***2.2.2.1. Genres for comparison*.** For right dislocation analysis, language samples from three genres were elicited for comparison. These were procedural description-egg and ham sandwich, story-telling –The Hare and the Tortoise and The Boy who Cried Wolf, and recount of a personal event. These three genres represented different degrees of shared knowledge between the speaker and listener. For instance, in the genre of procedural description, the referents of egg, ham and bread were supposed to be given information shared by the speaker and listener since the picture remained in sight during narration. In story-telling, the tortoise, hare, shepherd boy and wolf belong to rather familiar or accessible information since they were highly familiar stories with story cards presented to the speaker before the narrative began. The entities in personal event were believed to be all new in the discourse since the interviewer did not know what the speakers would say about their personal experience. Apart from the degree of shared knowledge, the three

genres were supposed to differ in the amount of planning time. The planning time of making the sandwich should be the minimum since the content was rather straightforward. Recounting of a personal event was expected to require the longest planning time. Many participants expressed that they had never been asked to recount a personal event to a stranger (interviewer) before. They had to organize the content and introduce entities spontaneously during the discourse. The planning time of telling the two stories is expected to be in between. The content of the stories was rather fixed although speakers might spend more time in planning the introduction of entities in the discourse. Zhang and Fang (1996) suggested that the motivation of using right dislocation was under the pressure of planning time in conversation. If this hypothesis is correct, right dislocation should appear more frequently in the less familiar discourse, i.e. the recount of personal event. The frequency of occurrence of right dislocation in these genres enabled us to understand whether shared knowledge and planning time play a role in using right dislocation in spoken Cantonese. The degree of shared knowledge and amount of planning time in the three genres are summarized in Table 2.2.

Table 2.2. Shared knowledge and planning time of selected genres

Genres	Making sandwich	Story- telling	Personal Event

Shared Knowledge	More	←	less
Planning time pressure	Less	→	More

2.2.2.2. Identification of instances of right dislocation. Following Cheung (1997) and Liang (2002), there are several criteria for identifying dislocation. Structurally, dislocation takes the form of [α (SFP), β], where α and β refer to components of a clause. If there was a sentence final particle (SFP), it must appear between the α and β strings. SFP is not obligatory although it is commonly found in right dislocation (Liang, 2002). Semantically, the β string should be able to form a complete clause with the α string when the clause is in canonical word order. Phonetically speaking, there is no noticeable pause between the main clause and the dislocated element; and the dislocated element is always said in a fast tempo (Liang, 2002). An example is provided in (34). The subject 哩個小朋友 "this kid" should appear before the verb 玩"play" when it is in the canonical word order.

(34) 玩 之嘛，哩 個 小朋友

Play SFP , this CL kid

"This kid is kidding only"

2.2.2.3. Statistical analysis. After obtaining the frequencies of occurrence of

right dislocations in each genre, the rates of occurrence with reference to the duration of each of the narrative samples were calculated. This is important as the duration of each production was different. Non-parametric tests were used for statistical analyses as the Shapiro-Wilk test revealed that the distributions of frequency of occurrence failed to meet the normality assumption. The Kruskal-Wallis H test was performed to determine if the different genre types had an effect on the rate of usage of right dislocation. Post hoc analyses using Mann-Whitney U test were conducted for pairwise comparisons in case of significant difference. The results can answer the question whether right dislocation is more commonly used in personal recount than the other two genres.

2.2.2.4. Position within narrative. In order to examine if right dislocation was equally likely to occur over the course of narration, each narrative was segmented evenly into three sections, i.e. beginning, middle and final, according to its number of utterances. The Chi-square goodness-of-fit test was employed to determine whether right dislocations were evenly distributed across the three sections.

Chapter 3 Results

3.1. Preferred Argument Structure (PAS)

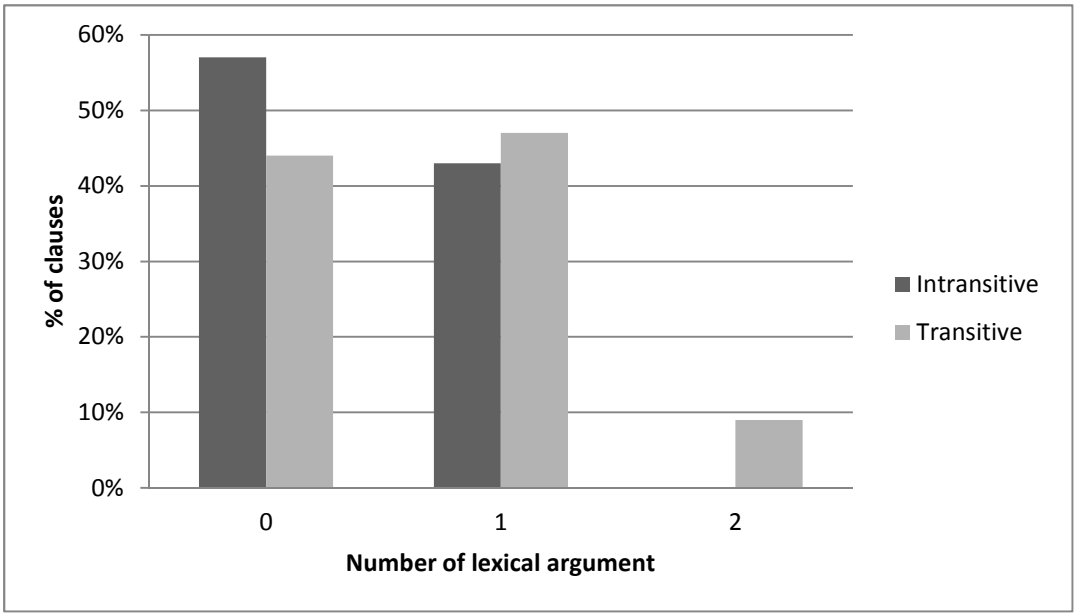
The PAS constraints mentioned in Chapter 1 suggest that argument roles are correlated with information status in spoken discourse. To recap, they are (i) one lexical argument constraint, and (ii) non-lexical A constraint in the grammatical dimension. The parallel constraints on pragmatics are (iii) one new argument constraint, and (iv) Given A constraint. To testify the PAS account in the present data, a total of 2,227 simple active clauses were identified from the database. About 68% of the clauses were intransitive while 32% were transitive.

Intra-rater and inter-rater reliabilities were obtained on coding information status. The intra-rater agreement was 95.4%. The inter-rater agreements between the author and each of the two independent raters were 85.3% and 88.1%, respectively. The inconsistencies were mainly from the judgment of accessible information.

3.1.1. The one lexical argument constraint. Of the total relevant clauses identified, over half of them did not contain any lexical argument; 44% contained one lexical argument; only 3% of the clauses were found to have two lexical arguments. Since only transitive clauses can have two arguments, the two types of clauses were

examined separately. The occurrence of two lexical arguments is still rare. Figure 3.1 presented the percentages of intransitive and transitive clauses by the number of lexical argument, respectively. The results have supported the one lexical argument constraint.

Figure 3.1. Percentage of intransitive and transitive clauses with zero, one or two lexical arguments

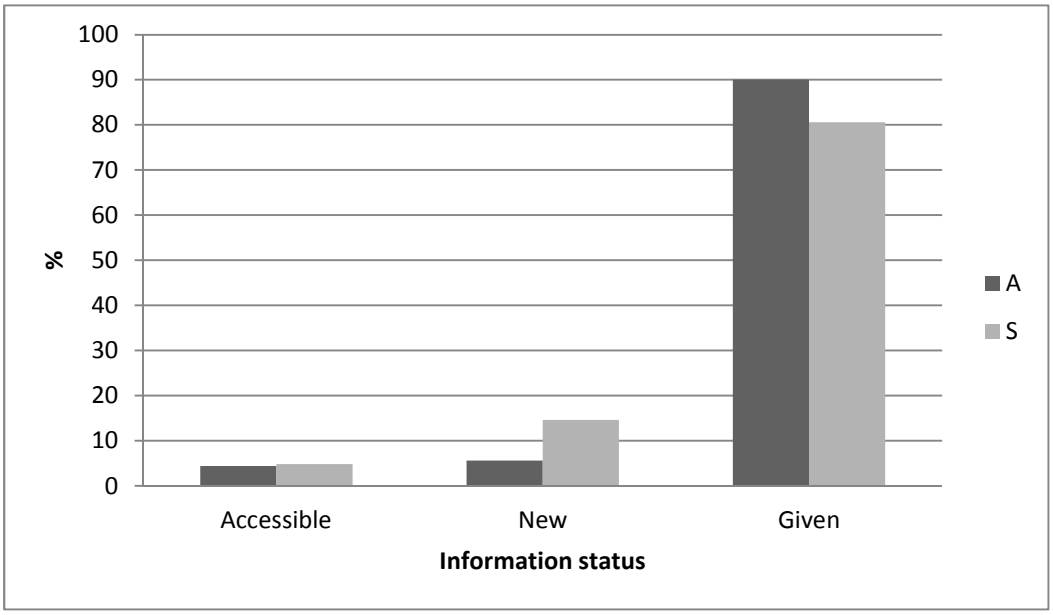


There are a total of 1,014 lexical referents in the data. The highest proportion of them was in the object role (54%), followed by 35% in the intransitive subject role, and the lowest was in the A role of transitive clauses with only 11%. The results conformed to the grammatical constraints of PAS. The pragmatic dimensions of PAS are assessed with inferential statistical tests.

3.1.2. Pragmatic constraints of PAS. The overwhelming majority of A and S arguments expressed given information, 90% and 80%, respectively. For new

information, higher percentage is found in the S role (15.1%) than in the A role (5.3%). The percentages of accessible information in S and A were similar (4.8% and 4.4% respectively), see Figure 3.2.

Figure 3.2. Distribution of information status in A and S

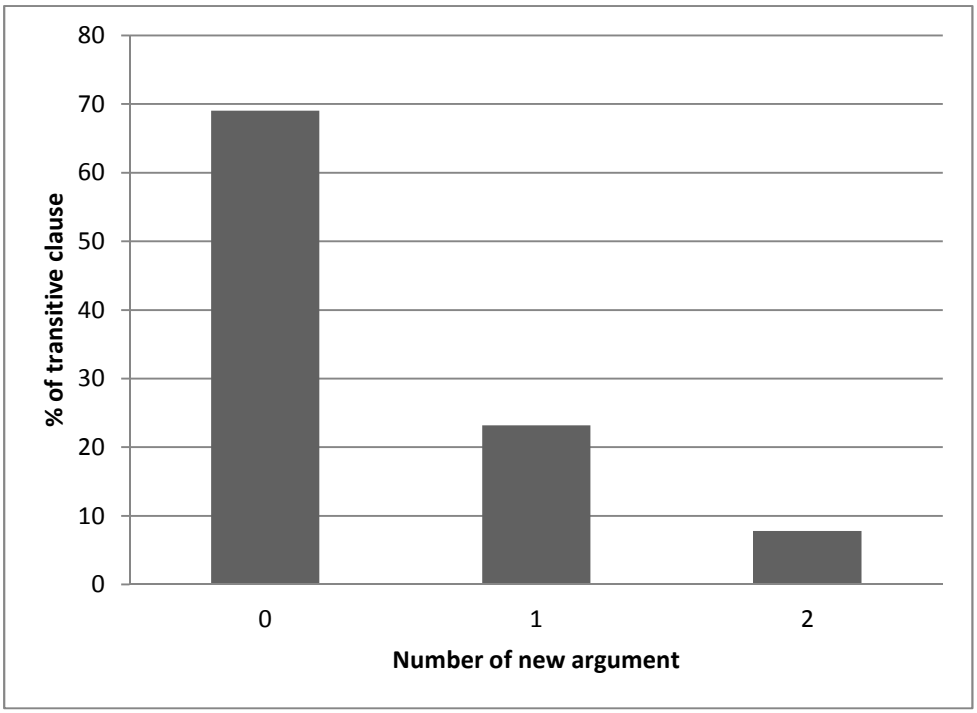


Wilconxon signed-rank test was run to determine whether there is a difference of S and A in the three information statuses. The test results showed that the proportion of given information was significantly higher in A (*Mdn*=100) than S (*Mdn*=83.8), *T*=28, *p*<.016, *r* = -0.41. For new information, S (*Mdn*=11.1*)* was significantly higher than A (*Mdn*=0.0), *T*=25, *p*<.016, *r* = -0.43. There was no significant difference between S and A in terms of rate of occurrence of accessible information, *T*=28, *p*=.25, *r* = -0.97.

Figure 3.3 illustrates that, out of the 713 transitive clauses, the majority did not

contain any new referent. About 20% of the clauses had one new referent while less than 10% possessed two new arguments. This significantly biased distribution was confirmed by the chi-square goodness-of-fit test (X^2 (2) = 456, p=.000).

Figure 3.3. Percentage of transitive clauses with zero, one and two new arguments



In summary, the statistical test results confirmed the PAS account. New information is avoided in the A role of transitive clauses, while the S role is relatively free to accommodate new referent. Moreover, clauses with two new arguments are a rarity.

3.2. Right Dislocation

3.2.1. Demographic background of speakers. A total of 117 instances of right dislocation were identified. They were produced by 63 speakers. In other words,

more than half of the participants did not use any right dislocation in their narratives. The 63 speakers seemed to be distributed evenly across age groups and education levels. Young female speakers with low education level tended to rarely use right dislocation. The details are presented in Table 3.1.

Table 3.1. Demographic of speakers who produced right dislocation in monologues

	High Education Lev.		Low Education Lev.		
Age Group	**Male**	**Female**	**Male**	**Female**	**Total**
Young	4	7	8	1	21
Middle Age	6	4	5	6	21
Older	5	5	7	5	22
Total	15	17	20	12	64

3.2.2. Genres. In terms of genres, the lowest frequency of right dislocation was found in the procedural description task. The instances of right dislocation of this genre accounted for 10% of the total occurrences. The remaining 90% were found in story-telling and personal recount (see Figure 3.4). The frequencies of occurrence in these two genres were very similar, 52 and 56, respectively. However, the duration of story-telling narratives was more than two times longer than the personal recount. The total length of the story-telling narratives which contained right dislocation was about 6,100 seconds; the length of personal recounts was about 2,800 seconds. When the duration of narration was taken into account, a higher proportion of right dislocation occurred in the personal recount at the rate of one occurrence in every 50 seconds in average. The story-telling had the lowest frequency with an average of

one instance in every 118 seconds (see Table 3.2)

Figure 3.4. Frequency of right dislocation across narrative tasks

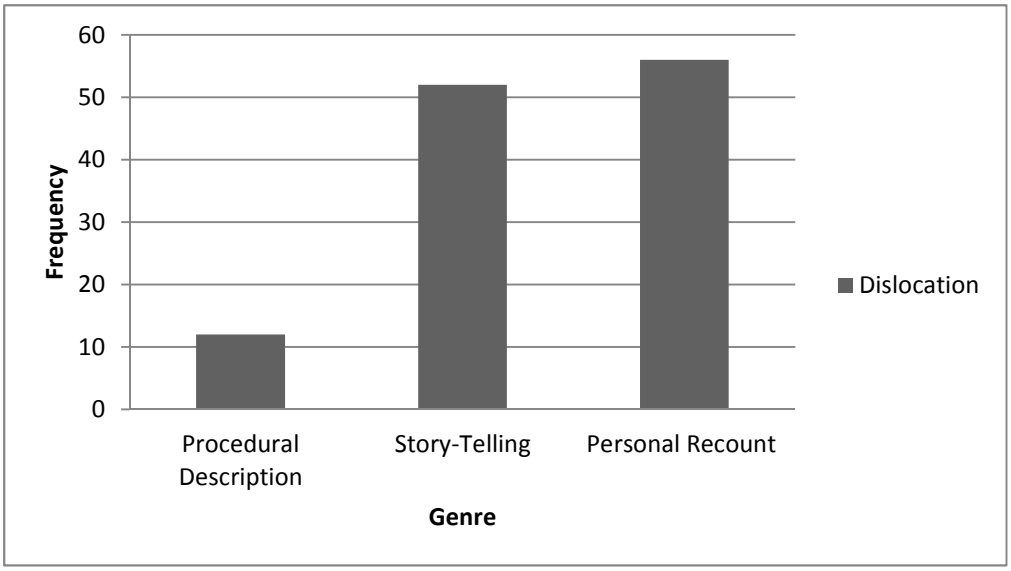


Table 3.2. Descriptive Statistics of right dislocation in genres

	Genre Length (sec)	Frequency	Ratio Mean	Ratio SD
Procedural Description	926	12	77.2	76.3
Story-telling	6136	52	118	113.5
Personal Recount	2809	56	50.2	45.3

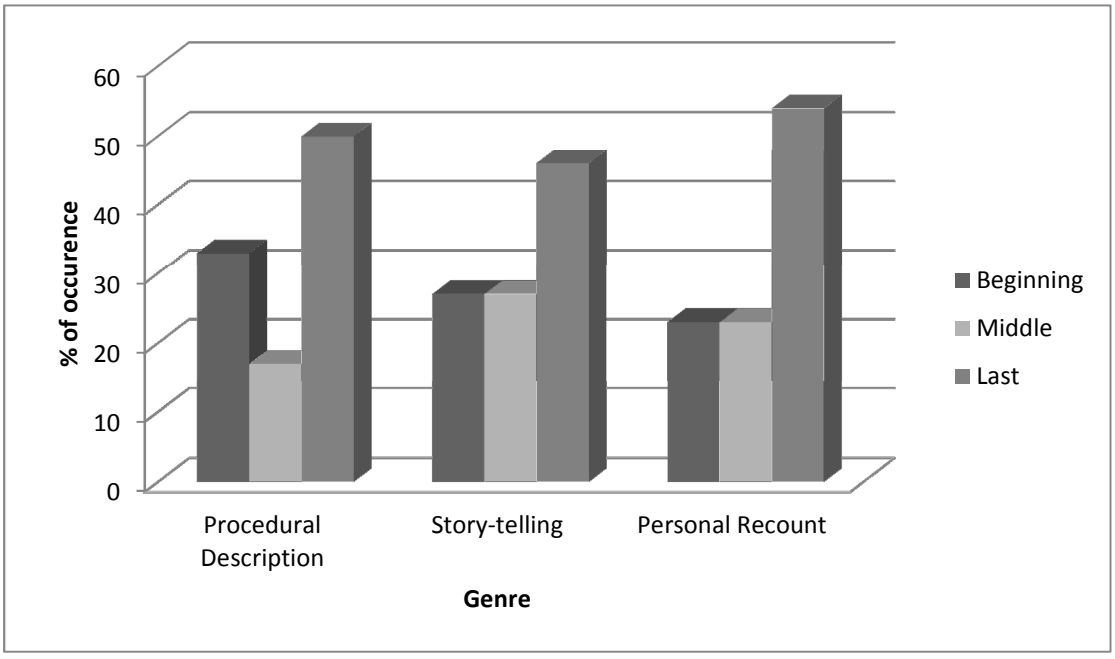
The Kruskal-Wallis test was employed to compare the distribution of right dislocations in the three genres. The rates of occurrence were significantly different across the genres, H (2) = 25.8, $p < .05$, with an effect size 0.14[4]. Mann-Whitney U tests were conducted with a Bonferroni correction (p = .05/3 = .0167) to investigate which two genres were different from each other. Right dislocation was produced significantly more frequently in personal recount than in procedural description (U =1198, z = -4.61, $p < .001$), and in story-telling (U =1211, z = -4.4, $p < .001$).There was

[4],The formula for calculating effect size is $r = \frac{z}{\sqrt{N}}$

no significant difference between story-telling and procedural description (U =1637, z = -1.77, $p > .05$).

3.2.3. Position within narrative. Regarding the position within a narrative in which right dislocations tended to occur, all three genres showed the same tendency that the majority of right dislocations appeared near the end of the narrative (see Figure 3.5). The remaining right dislocations distributed evenly in the beginning and middle of the narrative. The bias in distribution was confirmed by the chi-square goodness-of-fit test when the three genres were combined ($X^2 (2) = 6.29, p < .05$).

Figure 3.5. Position of occurrence of right dislocation in different genres



The results showed that instances of right dislocations are more commonly found in personal recount than story-telling and procedural description. Besides,

speakers tended to produced right dislocations near the end of a narrative.

Chapter 4 Discussion

4.1. PAS constraints

One of the aims of the present study was to verify whether the constraints of Preferred Argument Structure proposed by Du Bois (1987) were observed in Cantonese spoken discourse. These constraints include (i) avoiding lexical form and new referent in the subject of a transitive clause, and (ii) avoiding two new referents within one clause. To do so, this study carried out a quantitative analysis of the referential form and information status of transitive and intransitive subjects in simple active clauses extracted from a narrative task consisting of recounts of personal experience produced by 144 native Cantonese speakers evenly distributed in three age groups, genders and two education levels. The quantitative results of referential forms and information status of transitive and intransitive subjects revealed a pattern generally consistent with previous research on Chinese PAS (Chui,2005; Long,1990). The great majority of clauses contained no new referent. A small proportion of clauses possessed two new referents. Besides, new referents were found to be least frequent in the subject position of transitive clauses.

Although the PAS constraints were shown to hold in the speech samples of the current study, two main differences were found from the typical distribution of

argument structure in Chui (2005) and Long (1990), and also in other languages. First, the rate of new referents in intransitive subject position in this study was found to be twice as high as that reported in Long (1990), which investigated two sets of Chinese conversational material. Second, there was a significantly higher proportion of clauses containing two new referents in comparison with Long (1990) and the original PAS work in Du Bois (1987). In those studies, there was an absence of clauses with two new referents. In contrast, our finding showed that the rate of this type of clauses accounted for 7.8% of simple active clauses.

For the high rate of new intransitive subjects in our data, we propose that it may reflect different information pressure associated with genre types. As suggested in Du Bois (1987), information pressure was estimated by the ratio of new referents ("protagonists" in his term) to clauses. The more new referents needed to be introduced, the higher the information pressure. The speech samples in Long (1990) were drawn from two after-dinner conversations. The information pressure was assumed to be low due to the large amount of shared knowledge among participants in the conversation. In contrary, the data in this present study were extracted from monologues in which a participant was required to recount a personal experience to a stranger (interviewer). The speaker and listener did not share much background

knowledge in this setting. The higher information pressure in our narrative task has thus resulted in a higher percentage of new intransitive subjects compared with the conversational data in Long (1990). This explanation was also supported by the results reported in Du Bois (1987). His data samples were elicited from a high information pressure genre by asking participants to narrate a short film to strangers who had not seen the film. The percentage of new intransitive subjects was 22.5%, which was three times higher than Long (1990). The high rate of new intransitive subjects in this study is compatible with Du Bois (1987)'s observation that intransitive subject was "most responsive to increasing or decreasing information pressure (p.836)". When information pressure increases, intransitive verbs were used more for introducing new referents into the discourse. This would result in a higher frequency of new and lexical referent in intransitive subjects. One possible explanation for this pattern is related to the human-ness with the argument roles. The A role is highly agentive, and therefore more likely contains human referents. The S role, however, represents a heterogeneous category denoting both human and non-human referents. This mixed status of S might allow it to introduce new referents at a greater rate than the A role (Everett, 2009).

The findings in this study also support the claim that genre (narrative vs.

conversation) and information pressure affect the distribution of information flow in crucial ways (Du Bois, 1987). In other words, information pressure, information flow and types of discourse genre interact with one another. Therefore, it is important to bear in mind the factors of genre and information pressure when comparing across studies of information flow in the same language or between languages.

The second major difference found in current study was the high rate of clauses with two new referents. This result apparently contradicts the constraint of "one new argument per clause" (Du Bois, 1986). However, it should be noted that this type of argument structure was still of the lowest proportion in the data. The majority of the clause, conforming to the PAS account, contained no or only one new referent. As highlighted in Du Bois (1987), this constraint was not a categorical rule, but represented a tendency. It is not grammatically ill-formed for speakers to produce clauses with two new referents.

Nonetheless, it is interesting to examine why the occurrence of two-new-referents clauses was significantly higher than that in previous studies (Du Bois, 1987; Long, 1997). In the database of the present study, the two-new-referent clauses can be divided into two main types according to its referents' semantic meaning. The first type contained referents with kinship relations, as seen in

examples (35) and (36).

(35) 好　耐 之前 喎

Very long ago SFP

“it was a long time ago”

gam2 嗰 時　淨係 我 喺 香港 嘅嘛.

That time only　I　at　HK　SFP

“There was only me in HK”

⟶ gam2 嗰 啲 細佬　就 喺 廣州　照顧　外婆　嘅

That CL brother then at Guangzhou take care grandmother SFP

“My brothers took care my the grandmother in Guangzhou”

⟶ (36) 我　細妹 呢 就 接觸　好多　兔唇嘅細路仔

I　sister SFP then contact very much harelip children

“ My sister has a lot of interaction with harelip children”

咁啱　呢　我 個　細仔　呢 就有 啲　問題　嘅…

Coincidently SFP　I　CL　younger son SFP have CL problem SFP

“My younger son has this problem”

The referents “my brother” and “my sister” were classified as new information. Following the working definitions adopted in Du Bois (1987), given information was a

referent that was mentioned in the previous context; new information was mentioned for the first time. The category of accessible information was an intermediate category. A referent was accessible if “it was part of a previously evoked entity-based frame… and previously unmentioned (p.816)”. The referents in the examples were new based on these definitions. However, the speaker may not treat them as new information to the listener as evidenced by the fact that they were often introduced as definite noun phrases even in the first mention. Since the narrative task was to recount a personal experience, kinship relations were quite commonly found in the narratives. Speakers may treat these kin referents as a normal implicature from their personal experience, and thus assume them to be easily accessible. Some participants were more sensitive to the fact that the interviewer did not share his/her family background. They would introduce their kin(s) explicitly to establish the shared knowledge first. The first utterance in (37) was not related to the main content but produced solely for the purpose of introducing the referent “my wife”. :

(37) 我上年 啱啱 結咗婚

I last year just married

“I am married last year”

我同我太太 呢 都 好 想 養 隻 狗 嘅

I and my wife SFP Adv very want raise CL dog SFP

“My wife and I always want to have a dog”

但 一直 都 未 有 遇到

But always Adv not yet have met

“But we haven’t met one yet”

England and Martin (2003) discussed issues arising from analyzing PAS in Mayan narratives, and noted the difficulty to categorize the information status of some referents. Some referents appeared to be new on the surface but in fact they could be considered accessible. One source was the narrative titles or explanatory materials that were given to the speaker before recording started. The referents associated with the materials should be treated as accessible information although they were mentioned for the first time in the recording. It is not always straightforward to determine “newness” in lexical referents; hence, special care must be taken in order to make proper identification of information status especially in less controlled speech. The narratives used by Du Bois were specifically elicited to reduce the effects of personal or cultural associations. On the other hand, the speech samples examined in this study were rather rich with such associations. Kinship

relations shown above constituted one source of two-new-referents clauses; another source was perhaps due to the way we recruited and tested some of the participants. Since some of the subject recruitment and interviews took place at a dental hospital, the speaker might have assumed referents such as “dentist” and “patient” as shared information with the interviewer and introduced them in their narratives without providing any background.

(38) 有　一　日

“Have one day”

→ 個 病人　拎住 個 大　果　籃

CL patient hold CL big fruit basket

“The patient held a big fruit basket”

企　喺　大　門口　等 我

Stand at main entrance wait me

“Waiting for me at the main entrance”

(39) 好　耐　之前 喇

Very ago before SFP

“it was a long time ago”

→ 我跟　嘅　牙醫　呢 就 炒咗　佢 工人 喎

I follow POSS dentist SFP fire Asp his helper SFP

“The dentist I was working with fired his helper”

變　　咗 呢 冇 人　　幫 佢 清潔 屋企

Become Asp SFP no people help his clean home

“As a result, No one clean his home”

The discussion thus far has demonstrated that although shared knowledge between the speaker and the listener was controlled to a certain extent in the present study, some universal relations such as kinship, and cultural associations seem difficult to avoid. As a result, new referents by definition might be better to be treated as accessible information. In fact, if we re-assigned this type of referents from new to accessible information, the proportion of two-new-referents clauses would then be greatly reduced to less than 1 %. In sum, the tendency to avoid placing two new referents in one clause was upheld in our results.

As mentioned in Chapter 1, the existing literature on Chinese information structure was mostly based on written data and qualitative analyses. The current study intends to enrich our understanding of information structure in Chinese by extending the investigation to a much more ecologically valid modality of communication- spontaneous spoken discourse. The findings offer a more in-depth

and fine-grained depiction of information structure in spoken Cantonese. The rule of "given before new information" is not categorically applied in the spoken language. The results have also shown that, in addition to word order, argument structure also influences Cantonese information structure. In intransitive clauses, the speaker is relatively freer to present new information before the verb, especially under high information pressure in the discourse. The factors of sentence word order, transitivity, information pressure, discourse genre as well as speaker's assumption should all be taken into consideration in information structure analyses.

4.2. Right Dislocation

The previous section has shown that information structure in spoken discourse is more dynamic than that in the written discourse. This study further investigated a construction that is very commonly used in spoken discourse – right dislocation. Recent studies of Cantonese right dislocation have suggested that this structure is mainly motivated by the pressure of limited planning time in conversation (Liang, 2002; Zhang & Fang, 1996). The pragmatic motivation of using right dislocation is to place the most important information to the initial position of a sentence due to the planning time pressure in conversational discourse. Right dislocation is used as a focus marking device for a speaker to highlight the important information. In order to

have a more precise understanding of right dislocation, quantitative analysis was conducted in this study to investigate a mode of communication that has been far less studied, i.e. monologue, in light of the hypothesis of planning time pressure.

Three genres of narratives from the corpus (Kong et al., 2010-14) were examined – procedural description, story-telling and personal experience recount. These genres can be said to represent different degrees of familiarity and shared knowledge between the speaker and the listener. Recount of personal experience, a genre with least amount of shared knowledge, yielded the highest rate of occurrence of right dislocation. The result partially confirmed the “lack of planning time” account proposed in previous studies. To recap the methodology in the present investigation, unlimited preparation time was given to all speakers to plan their narratives before starting. In this arrangement, the so-called “planning time pressure” should be more or less similar among the three genres. Then the question is why the rate of right dislocation was higher in the personal experience recount than the other two narrative tasks. Firstly, the term “planning time pressure” is somehow misleading because the pressure is not merely from time. In conversation, speaker experiences dual pressures from the need to plan utterances that suit the particular addressee and to do so with minimal delay (Swets, Jacovina, & Gerrig, 2013). While monologues

were chosen in this study, the participants still assumed the interviewer as a listener and tried to construct the most appropriate utterances quickly. In the genre of personal recount, the participant was required to tell his/her personal experience to a stranger without shared background knowledge. Similar to the motivation arisen from conversation, the high level of stress in the personal recount gave rise to the higher rate of right dislocations compared with the other narrative tasks. In contrast, the contents in procedural description and story-telling were rather standardized and assumed to be more familiar to the listener. The lower numbers of right dislocations may reflect less pressure in planning the output during narration in these tasks.

Moreover, we found that the majority of right dislocations were produced late in the course of narration. Preparation time prior to narration may be a cause for this pattern. Mehnert (1998) studied the effects of planning time on second language performance. It was found that advanced speech planning is an effective way to reduce the amount of online speech planning needed, and thus eases the burden of communicative load and processing pressure. Language performance was generally improved with pre-task planning time, in terms of sentence complexity and accuracy. Under a planned condition, speech was more fluent as reflected in shorter pauses or self-correction. Beattie (1983) investigated speech planning and discovered that the

general semantic content of utterances would have been worked out first, while the detailed syntactic and semantic planning was carried out online. Speakers typically begin to speak before they have all utterances planned. To apply these observations in the present study, the utterances in the early part of a narrative would be well-planned given the pre-task preparation time. Subsequent utterances would only be planned while narrating. The pressure would be higher from the dual tasks of planning and speaking. As a consequence, more right dislocations would be produced. A corroborating observation is that right dislocation utterances in this corpus were often accompanied by repetition, retracing and longer pauses, as illustrated in (40).

(40)

(i) 啲 農民 呢 聽 到 喎.

CL farmer SFP heard SFP

"The farmer have heard (the boy shouting)"

(ii) 即刻 停 嘅 啲 嘢.

At once stop Quan CL thing

"They stop everything at once"

(iii) 停 嘅 唔 做 耕田.

Stop Quan not do farming

"Stop farming"

(iv) gam2aa6 攞 嘅 鋤頭 呀 盛

Take Quan hoe SFP

"They took the hoe etc."

(v) 即刻 趕 上 山.

Immediately rush up hill

"They rushed up to the hill immediately"

→ (vi) 就 去 [/] 去 打 [/] 打 走 啲 [/] 啲 狼 啦 , 想 話.

Then go go hit hit away CL CL wolf SFP , intend

"They intended to go and hit the wolf away"

(40) is an excerpt from the task "A boy who cried wolf". In line (vi), there was more than one repetition, including the verbs 去 "go", 打 "hit" and classifier 啲 "some". The main verb 想話 "intend" was dislocated to the end. The dislocated verb served the function of adding supplementary information to the utterance. It seems that the content of "hitting the wolf away" was available in the speaker's mind, but the semantic and syntactic units were not fully planned, as evidenced by the repetitions. While uttering the sentence, the speaker might realize that "hitting the wolf away"

was not entirely correct because there was actually no wolf (it was a lie told by the shepherd boy). His adding the verb “intend” after the main clause would make the action of “hitting the wolf away” become the intention of the farmer - a depiction that fitted the story-line better. In this example, it was unlikely that the speaker deliberately dislocated any constituents for highlighting; it might instead be a result of pressure under time and communicative constraints. The co-occurrence of disfluency and right dislocation in relation to planning pressure merits further investigation. The findings may shed light on both Cantonese right dislocation and processes underlying spontaneous speech production.

In comparison with previous studies involving conversations, a lower rate of occurrence of right dislocation was found in this corpus of monologues. More specifically, in 25 conversations and one story-telling totaling about 5.5 hours, Liang (2002) identified 666 right dislocations. Luke (2012) analyzed a sample of four hours of conversations and collected about 300 examples. In the present study, approximately 8 hours of recordings was examined, but only 117 instances were found. These results strongly suggest that Cantonese speakers tend to use right dislocation far more frequently in conversation than in monologue. The account of planning time pressure seems insufficient to explain the situation. It has been argued

earlier that planning pressure exists in both types of discourse; the question is what characteristics in conversation that trigger the use of right dislocation. One factor might be turn-taking. Right dislocation has been proposed to play a role in turn-taking during conversation, i.e. to indicate the next turn to the addressee. Consider the following excerpt taken from Luke (2012, p.346):

(41) Ian: … 啲 天氣 好似 琴日 嘅 囉

CL weather like yesterday stprt SFP

"…the weather is like yesterday"

Joe: 有 冇 check 過 呀, 你?

Have not check Asp SFP , you ?

"Have you checked?"

Luke argued that the dislocated 你 "you" carried little value in term of information load. It was more like a term of address than a pronoun. The function of its existence is to indicate to an addressee for the next turn-taking in the conversation. Luke (2000) reported that 78% of the right dislocation utterances in his data co-occurred or was directly followed by a turn-transition. Right dislocation in turn-taking was also claimed to be a "post-completion device" (p.307). The speaker attempts to extend his/her turn of speaking by supplementing more information upon the completion of

the utterance. If turn-taking is one of the main motivations in using right dislocation, the absence of turn-taking in monologue may then be a reason for its lower rate of occurrence.

Another interesting observation of the present findings was the involvement of emotion in utterances containing right dislocation. Emotional involvement is signaled by explicit expression of personal feelings in a sentence, such as 難過 "sad", 遺憾 "regretful", 難忘 "unforgettable", 開心 "happy" as shown in (42)-(51), or presenting an incident strongly associated with personal emotions, such as 做咗一次錯事 "made a mistake" in (52) and 生咗個瘤 "had a tumor" in (53). Examples of right dislocations that are emotionally neutral are also given in (54) and (55).

(42) 係 e6 e6 好 難過 喇 , 依 樣 嘢.

Be very grieved SFP , this incident

" This incident made me feel very grieved"

(43) gam2aa6 覺得 一 生 最 遺憾 , 哩 樣 嘢.

Feel one life most regretful , this incident

"This incident is most regretful in my life"

(44) 所以 覺得 好 難忘 , 哩 件 事.

So feel very unforgettable , this incident"

“I found this incident is unforgettable”

(45) xx 但係 都 係 開心 㗎囉 ， 嗰 日.

But Adv be happy SFP , that day

“It was still happy on that day”

(46) 好 開心 囉 ， 哩 樣 嘢.

Very happy SFP , this incident

“This incident made me very happy”

(47) 好 麻煩 喎 ， 諗住.

Very troublesome SFP , think

“I thought it was troublesome”

(48) 好 [/] 好 輕鬆 嘞 ， 個 人

Very relax SFP , CL people

“ I feel relaxing”

(49) 哎吔 好 開心 呀 ， 同 佢.

Inj very happy SFP , with he

“ we both are very happy”

(50) 因為 嗰 時 係 e6 好 緊張 啦 ,個 心情.

Because that time be very nervous SFP , CL mood

" Because my mood was nervous at that time"

(51) 吖 哩 樣 嘢　　真係 難忘　　啊 ,真係.

Inj this type thing really unforgettable SFP, really.

"This is really unforgettable"

(52) 即係 做　咗 一 次　錯事　　啦，自己.

That is　do Asp one CL wrong thing SFP , myself

" I have done a wrong thing once"

(53) 生 咗 個 瘤 呀 (..)，自己

Birth Asp CL tumor SFP , myself

" I have a tumor"

(54) gam2 我哋 瞓著 咗 㗎喇，嗰 陣 時.

We sleep Asp SFP ,　that CL time

"We were asleep at that time"

(55) 我 而家 仔 女　　已經 廿 八 歲 嘞，我 個 女.

I now son daughter already　28　age SFP , I　CL daughter

" My daughter is already 28 years old now"

The present finding has revealed that many instances of right dislocation appear to involve emotional expressions. One plausible reason for this may be related to the

task of recounting personal experience. It was very common for the speakers to share their feelings and evaluations in the narrative. This observation, however, echoes the affective functions of right dislocation proposed in Guo (1999), although Guo only investigated dislocated noun phrases and pronouns in Mandarin Chinese. By examining discourse produced by children at play, Guo found that right dislocations often carried negative evaluations. He proposed that this sentence construction was developed from managing information to serving affective function of expressing negative feelings. He argued that " … Mandarin Chinese is one of the Asian languages that tend to grammaticalize social and affective notions. …Mandarin is very rich in sentence final particles that express notions such as surprise, suggestion, concern, etc. Mandarin word order also often codes affects. For example, the so-called passive *bei*-construction is mostly used to refer to events the speaker regards as unfortunate (p.1126)". In addition to Guo (1999)'s claim, the present data have further demonstrated that different kinds of emotion, including positive emotions, can be associated with the use of right dislocation.

As for the function of focus marking suggested in Law (2003), illustrated in (46)-(48) where the adjective conveying the speaker's feeling was moved from the later to an earlier position in order to receive focus, it is worth highlighting that the

account does not describe all the instances in the data. Luke (2012) insightfully pointed out that "…although the whole of the first sentence part may be the focus, often times it only *contains* the focus (p.343)". In (51), the definite NP 哩樣嘢 "this thing" is not the focus. Instead of treating the first part as focus that was moved to the front, it seems more plausible to regard the second 真係 "really" as being dislocated to the sentence final position for emphasizing. The present data are consistent with Luke's (2012) proposal that the functions of right dislocation actually span a spectrum of possibilities. No single account could characterize all the cases of right dislocation.

4.3. Conclusions

The current investigation has extended the study of Chinese information structure from written discourse in previous studies to spontaneous spoken data. The findings have shown that the general principle of given before new information must be refined when applying to spoken discourse. More specifically, the pre-verbal subject in intransitive clauses is relatively free to accommodate new referent. The quantitative nature of this study has also yielded results revealing that information pressure and genre type affect the way information is managed; that is, a higher rate of pre-verbal new referent is observed in those genres with higher information

pressure. This study has demonstrated that the factors of information pressure, planning time, discourse genre as well as speaker's assumption of shared knowledge should all be considered in analyzing information structure.

The examination of the use of right dislocation in spoken Cantonese as a focus marking device which may be sensitive to planning time pressure has been largely confirmed. The frequent co-occurrence of right dislocation with disfluency and emotional expressions is worthy of further investigation. This thesis explored the aspect of discourse level analysis of language sample. The findings in this thesis may have implications for studies of adult language disorders. In particular, further research could be conducted on investigating whether speakers with language disorder are sensitive to planning time and shared knowledge when producing RD.

Finally, we recognize that the scope of the present study was restricted to simple active clauses, and it would be desirable for future research to investigate complex sentence constructions and passive sentences. It is also important to have careful consideration of genre type and different information status classification, especially the accessible information. The quasi-experimental nature of the present study has obtained interesting observations showing the relationship between the use of right dislocation and planning time pressure. Future studies may take an

experimental approach with proper manipulation of this factor.

References

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation*. Milton Keynes, U.K.: Open University Press

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 25-55). New York: Academic Press.

Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (Ed.), The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production (pp.9-50). Norwood, NJ: Ablex.

Chafe, W. (1987). Cognitive constraints on information flow. In R.S. Tomlin (Ed.), *Coherence and grounding in discourse: Outcome of a symposium, Eugene, Oregon, June 1984* (pp. 21-52). Amsterdam: John Benjamins.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Cheung .L. (2009). Dislocation focus construction in Chinese. *Journal of East Asian Linguist, 2009(18), 197–232.*

Cheung, L. (1997). *A study of right dislocation in Cantonese.* Unpublished M.Phil. thesis, The Chinese University of Hong Kong, Hong Kong.

Cheung, L. (2005). *Syntax and semantics of dislocation focus construction in Cantonese*. Unpublished M.A. thesis, University of California Los Angeles, CA.

Chu, C. (1999). *A discourse grammar of Mandarin Chinese*. New York: Peter Lang.

Chui, K. (2005). Structuring of information flow in Mandarin Chinese. *Journal of Chinese Linguistics, 33*(1), 34-67.

Danes, F. (1974). Functional sentence perspective and the organization of the text. *Papers on functional sentence perspective*, *106*, 128.

Dik, S.C. (1978). *Functional Grammar*. Amsterdam: North-Holland.

Dixon, R. M. (1979). Ergativity. *Language, 55*(1), 59-138.

Du Bois, J.W. (1987). The Discourse Basis of Ergativity. *Language, 63* (4), 805-855.

Du Bois, J.W. (2003). Argument Structure: Grammar in use. In J.W., Du Bois, L.E., Kumpf, & W.J., Ashby. (Eds.), *Preferred Argument Structure* (pp. 11-60). Amsterdam: John Publishing Company.

England, N.C., & Martin L. (2003). Issues in the comparative argument structure analysis in Mayan narratives. In J.W., Du Bois, L.E., Kumpf, & W.J., Ashby. (Eds.), *Preferred Argument Structure* (pp.131-158). Amsterdam: John Publishing Company.

Everett, C. (2009). A reconsideration of the motivations for preferred argument structure. *Studies in Language, 33*(1), 1-24.

Givón, T. (1983). *Topic continuity in discourse*. Amsterdam: Benjamins.

Givón, T. (1990). *Sytax: a functional-typological introduction. Vol. II.* Amsterdam: John Benjamins.

Guo, J. (1999). From information to emotion: The affective function of right-dislocation in Mandarin Chinese. *Journal of Pragmatics, 31(9), 1103-1128.*

Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics, 3*(37-81), 199-244.

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

Ho, Y. (1993). *Aspects of discourse structure in Mandarin Chinese*. Lewiston, New York: Mellen University Press.

Kohn, S.E., & Cragnolina A. (2003). The role of Preferred Argument Structure for understanding aphasic sentence planning. In J.W., Du Bois, L.E., Kumpf, & W.J., Ashby. (Eds.), *Preferred Argument Structure* (pp.339-352). Amsterdam: John Publishing Company.

Kong, A. P. H., Law, S. P., & Lee, A. S. Y. (2010-2014). Cantonese Corpus of Oral Narratives (CANON).

Kumpf, L.E. (2003). Genre and Preferred Argument Structure: Sources of argument structure in classroom discourse. In J.W., Du Bois, L.E., Kumpf, & W.J., Ashby. (Eds.), *Preferred Argument Structure* (pp. 109-131). Amsterdam: John Publishing Company.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

LaPolla, R. J. (1990). *Grammatical relations in Chinese: Synchronic and diachronic considerations.* Doctoral dissertation, University of California, Berkeley.

LaPolla, R. J. (1995). Pragmatic relations and word order in Chinese. In P. Downing & M. Noonan (Eds.), *Word Order in Discourse* (pp. 297-329). Amsterdam: Benjamins.

LaPolla, R. J. (2009). Chinese as a Topic-Comment (Not Topic-Prominent and Not SVO). In J. Xing (Ed.), *Studies of Chinese Linguistics: Functional Approaches* (pp. 9-22). Hong Kong: Hong Kong University Press.

Law, A. (2003). Right dislocation in Cantonese as a focus-marking device. *UCL Working Papers in Linguistics, 15,* 243-275.

Li, C. N., & Thompson, S. A. (1975). The semantic function of word order in Chinese. In C.N. Li (Ed.), *Word Order and Word Order Change*, (pp. 163-195). Austin: University of Texas Press.

Liang, Y. (2002). *Dislocation in Cantonese: Sentence form, information structure, and discourse function.* Unpublished doctoral dissertation, University of Hong Kong, Hong Kong.

Long, Z.H. (1990). The Preferred Argument Structure in Mandarin Chinese and Its Cognitive Implications. *Colorado Research in Linguistic, 11*, 42-49.

Lu, J. (1980). 漢語口語句法裡的易位現象 [The phenomenon of inversion (of sentence parts) in syntax of spoken Chinese]. *Zhongguo yuwen (Chinese Language and Writing), 154*, 28-41.

Luke, K. K. & Wong, M.L.Y. (2015). The Hong Kong Cantonese Corpus: Design and Uses. In *Linguistic Corpus and Corpus Linguistics in the Chinese Context*, edited by Tsou, Benjamin and Kwong, Oi Yee, Journal of Chinese Linguistics Monograph Series, 309-330.Luke, K.K. (2000). 句子成份的後置與話輪交替機制中的話輪後續手段 [Postposed sentential consituents as post-completion devices in conversational turn-taking]. *Zhongguo yuwen (Chinese Language and Writing), 2000*(4), 303-310.

Luke, K.K. (2004). 說延伸句 [On incremental sentences]. In Editorial board of Zhongguo yuwen (Chinese Language and Writing), Institute of Linguistics, Chinese Academy of Social Sciences (Ed.), *Academic papers for celebrating the 50th anniversary of 'Zhongguo yuwen (Chinese Language and Writing)*, (pp. 39-48). Beijing: The Commercial Press.

Luke, K.K. (2012). Dislocation or Afterthought?—A Conversation Analytic Account of Incremental Sentences in Chinese. *Discourse Processes*, *49*(3-4), 338-365.

Luke, K.K., & Zhang, W. (2007). Retrospective turn continuations in Mandarin Chinese conversation. *Pragmatics: quarterly publication of the International Pragmatics Association*, *17*(4), 605-636.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2003). Child Language Analyses (CLAN). *Using CLAN*. Retrieved June 2013, from http://childes.psy.cmu.edu/clan/.

MacWhinney, B., Holland, A., Forbes, M., Spector, L., & Fromm, D. (2008). *AphasiaBank: An International Archive of Aphasic Language.* Paper presented at Meeting of Computer-based Intervention and Diagnostic Procedure-Applications for Language Impaired Persons, Vienna, Austria.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in second language acquisition*, *20*(1), 83-108.

Mathesius, V. (1975). *On information bearing structure of the sentence. Harvard Studies in Syntax and Semantics*. Massachusetts, Cambridge: Harvard University Press.

Packard, J. L. (1986). A left-dislocation analysis of 'afterthought' sentences in Peking Mandarin. *Journal of the Chinese Language Teachers Association, 21*(3), 1-12.

Prince, E.F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp.223-255). New York: Academic Press.

Shimojo, M. (1995). *Focus structure and morphosyntax in Japanese: wa and ga, and word order flexibility.* Doctoral dissertation, State University of New York.

Sung, K.Y. (2007). *On the prosodic and thematic properties of post-completion constituents in focus-first constructions in Cantonese.* Unpublished M.phil dissertation, University of Hong Kong, Hong Kong.

Swets. B., Jacovina M.E. & Gerrig, R.J. (2013). Effects of Conversational Pressures on Speech Planning. *Discourse Processes, 50* (1), 23-51.

Tai, J., & Hu, W. (1991). Functional Motivations for the So-Called" Inverted Sentences" in Beijing Conversational Discourse. *Journal of the Chinese Language Teachers Association*, *26*(3), 75-104.

Weber, E.G. (2003). Nominal information flow in the talk of two boys with autism. In J.W., Du Bois, L.E., Kumpf, & W.J., Ashby. (Eds.), *Preferred Argument Structure* (pp.339-352). Amsterdam: John Publishing Company.

Xu, L. (2004). Manifestation of informational focus. *Lingua*, *114*(3), 277-299.

Zhang, B. & Fang, M. (1996). *《漢語功能語法研究》[Studies in Chinese Functional Grammar]*. Nanchang: Jiangxi Education Publishing House.

Appendix I.

Story cards- “The boy who cried wolf”

Story cards- “The tortoise and the hare”