

Pragmatics of classifier use in Chinese discourse

Ming-Ming Pu

University of Maine at Farmington, 270 Main Street, Farmington, ME 04938, USA
mingpu@maine.edu

1. Introduction

The present study examines a particular syntactic phenomenon in Chinese discourse, namely complex noun phrases (CNPs), and investigates the occurrence and distribution of the various forms of such constructions. The study focuses on the presence and absence of classifier phrases that modify CNPs, and explores, from a cognitive-functional perspective, what specific functions such modifiers in CNPs serve in discourse and how their positioning in CNPs manifests our cognitive constraints underlying discourse processing. The study aims to explain (i) why, of several possible CNP constructions, one occurs more frequently in discourse than do others, and (ii) what motivates speakers and writers to choose a particular type of CNP at a certain juncture of discourse processing.

2. Classifiers and CNPs

The classifiers discussed in the present study refer to classifier phrases (CL), because classifiers normally co-occur with demonstratives and/or quantifiers in Chinese discourse to modify and categorise noun phrases (NPs). Classifier phrases are pre-nominal, as shown in the underlined portion of 一个(人), *yi-ge (ren)* ("a-CL man") and 这匹(马), *zhe-pi (ma)* ("this-CL horse"). An NP with the structure [quantifier + classifier + N] is generally considered to be indefinite (e.g., 三间屋, *san-jian wu* ("three-CL rooms")), and one with the

structure [demonstrative (+ quantifier) + classifier + N] to be definite (e.g., 那两棵树, *na liang-ke shu* ("those two-CL trees")).

CNPs are those that are modified by relative clauses (RC). Chinese CNPs differ from those of English in the following ways: (i) a Chinese RC precedes its head NP while an English RC follows its head NP; (ii) a Chinese RC lacks a relative pronoun (RP) but is usually marked by the particle (PRT) *de* at the end of the RC while an English RC is usually headed by a RP; (iii) Chinese RCs are in general restrictive (Chu 1998: 226) while English RCs may be either restrictive or non-restrictive; and (iv) a pre-nominal RC in Chinese may separate a classifier phrase from its head noun while a post-nominal RC in English does not come between the head noun and its pre-modifiers such as articles, demonstratives and/or quantifiers. Moreover, classifiers may not always co-occur with a classifiable NP in Chinese discourse.

Hence, there are three CNP constructions in Chinese with respect to the presence and absence of a classifier: (i) one with a unified head when the classifier immediately precedes the head noun and follows the relative clause; (ii) one with a split head when the classifier precedes the relative clause and is separated from the head; and (iii) the other with a bare head when there is no classifier present in the CNP. The three structures are schematised with examples in (1) to (3).

(1) CNP with a unified head: **RC+CL+H** (Relative Clause + Classifier + Head NP)

[戴 眼镜 的] 那个 女孩子 是 小王 的 邻居。

dai yan.jing de na.ge nu.haizi shi Xiao.Wang de lingju
wear glasses PRT that-CL girl be Xiao.Wang's neighbor

"The girl (who) wore glasses is Xiao Wang's neighbor."

- (2) CNP with a split head: **CL+RC+H** (Classifier + Relative Clause + Head NP)

那个 [不 知 天高地厚 的] 女人 , 藏 在 何处 ?

na.ge bu zhi tian.gao.di.hou de nu.ren cang zai he.chu

that-CL not know sky.high.earth.thick PRT woman hide where

"Where does that women (who) overestimated herself hide?"

- (3) CNP without a classifier: **RC+H** (Relative Clause + Head NP)

卧 床 的] 女人 吃力地 转过身来。

wo chuang de nu.ren chi.li.di. zhuan.guo.shen.lai

lie.in.bed PRT woman strain turn.over

"(The) woman (who) lay in bed strained to turn over."

Although an important characteristic of Chinese NPs, classifiers are not syntactically required to modify classifiable NPs in discourse, neither is their positioning syntactically constrained in a CNP in relation to the RC. In fact, if we remove, add, or change the position of the classifiers in the above examples, the three utterances are still acceptable, at least as they are considered in isolation. At the level of semantics, on the other hand, the presence or absence of classifiers may not always be associated with definiteness or referentiality. Take for example the constructions in (2) and (3) above: the head noun *na.ge nu.ren* ("that woman") in (2) is classified and the head noun *nu.ren* ("(the) woman") in (3) is not, both of which are nevertheless definite NPs referring to the same woman that has been mentioned in the previous discourse. Further, although all of the CNPs in (1) to (3) are referential, referring to specific referents or entities in discourse, the CNPs in the constructions in (4) and (5) are not.

- (4) 她 企盼 一个 [又 英俊 又 有钱 的] 男人。

ta qi.pan yi.ge you ying.jun you you.qian de nan.ren

she yearn.for a-CL both handsome and wealthy PRT man

"She yearned to have a man (who) was both rich and handsome."

- (5) 当 医生 的 , 喜欢 [有点 小病 就 大叫大嚷 的] 病人。

dang yi.shen de xi.huan you.dian xiao.bing jiu da.jiao.da.rang de bing.ren

as doctor like have little.ailment just cry.and.shout PRT patient

"A doctor always likes those patients (who) would cry out loud when they feel slightly indisposed."

In both (4) and (5), the CNP is non-referential regardless of whether or not a classifier is present. Although the head noun *nan.ren* ("man") in (4) is modified by an indefinite classifier *yi.ge* ("a-CL"), it does not refer to a particular man in discourse or any particular man in the world, but to one from a class who fits the "class membership" designated by the relative clause *you ying.jun you you.qian* ("who are both handsome and rich"). Likewise, the unclassified head noun in (5) is non-referential, which does not refer to any particular *bing.ren* ("patient") but to one of a class of individuals who, as depicted by the relative clause *you dian xiao.bing jiu da.jiao.da.rang* ("cry out loud when they feel slightly indisposed").

It is obvious that the occurrence and distribution of classifiers do not seem to be a syntactic or semantic phenomenon on the clause-level, but the alternative CNP constructions nonetheless occur in Chinese discourse in a systematic pattern. Several studies have argued for some pragmatic functions of classified NPs in Chinese discourse. While most research focuses on classifiers in simple NPs (cf., *inter alia*, Sun 1989; Li 2000), Chao (1968: 9) and Hashimoto (1971: 24) tried to explain the use of classifiers and their positions in CNP constructions. Chao (1968), for example, distinguishes between the two alternative structures in terms of information status of the head modified by the relative clause (cf. the examples in (6) and (7)).

- (6) 那个 [戴 眼镜 的] 人 是谁 ?

na.ge dai yan.jing de ren shi shui

that-CL wear glasses PRT man be who

"Who is the man (that) wears glasses?"

- (7) [戴眼镜的]那个人是谁?

dai yan.jing de na.ge ren shi shui

wear glasses PRT that-CL man be who

"Who is the man (that) wears glasses?"

Chao (1968) argues that the relative clause that follows the classifier, as in (6), is descriptive, which modifies a given head NP, i.e., known to both speaker and listener, and the referent of the head can be identified without the modifying relative clause. The relative clause that precedes the classifier, as in (7), on the other hand, is restrictive; the head of such a structure is new information that cannot be identified by the listener without the modifying relative clause. Hashimoto (1971) agrees with Chao (1968) and explains further that while a descriptive RC as in (6) may modify an indefinite head, a restrictive one as in (7) may only modify a definite head. Neither account, however, seems to have provided an adequate explanation for the alternative classifier positions of CNPs found in Chinese discourse. Consider the example in (8).

- (8) 我一定 能 为 您 买到 那种 [有 黑色 大理石 花纹 的] 扣子

wo yi.ding neng wei nin mai.doa na.zhong you hei.se da.li.shi hua.wen de kouzi

I certainly can for you buy that-CL have black marble design PRT button

"I will certainly get you the kind of buttons that have black marble designs."

- (9) 秋仪 扫了 一眼 [切 羊糕 的] 那个 男人。

Qiu.Yi sao.le yi.yan qie yang.gao de na.ge nan.ren

name glance an.eye cut mutton.cake PRT that-CL man

"Qiu Yi glanced at the man who was cutting mutton cakes."

In both utterances, the head nouns of *na.zhong kou.zi* ("that-CL button") and *na.ge nan.ren* ("that-CL man") are given information with descriptive modifying RCs, regardless of the alternative positions of the classifiers. Both relative clauses are descriptive since the referent of the head can easily be identified in the discourse context without the modifying relative clause. It is thus difficult to see the correlation among the position of the classifier, the

restrictiveness of the relative clause and the information status of the head noun in a CNP construction. It is equally difficult, on the other hand, to see the correlation between the definiteness of the head noun and the position of the classifier in a CNP, as argued for by Hashimoto (1971). Consider the examples in (10) to (12).

- (10) 五月 的一个早晨, [从 营队 开来的] 一辆 卡车

wu.yue de yi.ge zaocheng cong ying.dui kai.lai de yi.liang ka.che

May one morning from barracks came PRT a-CL truck

停在 翠云 坊 的 巷口

ting.zai Cui.yun.fang de xiang.kou

stop.at Cui.yun house PRT alley.mouth

"One morning in May, a truck (that) came from the barracks stopped at the mouth of the alley where Cuiyun House was located."

- (11) 旁观的人 包括 [在巷口 摆摊 的] 几个 小贩。

pang.guan.de.ren bao.kuo zai xiang.kou ban.tan de ji.ge xiao.fan

bystander include at alley.mouth setup.stall PRT several-CL peddler

"The bystanders included several peddlers (who) set up stalls at the mouth of the alley."

- (12) 这一定 是 [同 她 相好 的] 一个男人。

zhe yi.ding shi tong ta xiang.hao de yi.ge nan.ren

this must be with her on.intimate.terms PRT a-CL man

"This must be a man (who) was having an affair with her."

Contrary to Hashimoto's (1971) claim that "restrictive" relative clauses may only modify a definite head, the relative clause in each of the above three utterances is restrictive (in Chao's (1968) sense that the referent of the head noun cannot be identified without the modifying RC), yet it modifies an indefinite head.

The above discussion shows that the occurrence of classifiers in the CNP construction cannot be accounted for solely by certain isolated syntactic, semantic or discourse factors.

What, then, are the differences between classified and non-classified CNPs on the one hand, and between unified- and split-head CNPs on the other? What triggers the use of a particular structure at a certain juncture of discourse? And what motivates speakers to choose one structure over the others? The present study aims to uncover factors underlying the use of CNP constructions and attempts to offer an adequate explanation for their distributional patterns found in Chinese discourse, based on the analysis of spoken and written data obtained from an experiment and several narrative texts.

3. Cognitive Strategies and CNPs

The present study argues that the occurrence and distribution of the alternative CNP constructions may be best explained in terms of cognitive strategies that speakers adopt in processing particular structural units in order to facilitate language processing. Given that the human mind is capable of dealing with only a limited number of units (letters, words, digits, etc.) at a time (Van Dijk and Kinstch 1983; Just and Carpenter 1992), this limitation must manifest itself in our language production and comprehension. Consequently language units that do not tax our cognitive resources would be easier and faster to process.

The present study proposes, along the line of psycholinguistic research (Slobin 1973; Frazier 1979; Prideaux and Baker 1986; Andersen 1989; Prideaux 1999), that certain cognitive strategies interact with grammatical characteristics of Chinese CNP structures, especially their modifying relative clauses, yielding a certain structural type easier to process than others. The two cognitive strategies that are particularly relevant to relative clause processing are CLOSURE and MARKEDNESS, both of which address the ease with which our minds deal with non-disruptive (rather than disruptive) units, and normal, familiar patterns (rather than novelty and deviation). Consider first the strategy of CLOSURE, according to (Prideaux 1999: 9).

CLOSURE: A constituent (e.g., clause) which is internally interrupted by another instance of the same constituent (e.g., another clause) will require more processing resources (i.e., will be more difficult to process) than that same constituent which is not internally interrupted.

According to Prideaux (1999), CLOSURE is a generalised Gestalt principle, which hinges on working-memory limitations. Since we only have limited cognitive resources to allocate to a processing task at hand, CLOSURE would facilitate our assembly of one processing unit as quickly as possible so we can clear our active (verbatim or syntactic) buffer in order to deal with the next unit. In contrast, a unit containing another embedded unit does not allow for early closure, and we would have to suspend processing one unit in order to attend to the other, hence increasing demands on working memory.

MARKEDNESS, on the other hand, addresses the ease with which our minds deal with normal, familiar patterns rather than novelty and deviation (Givón 1993: 178; Prideaux 1999:11; Fenk-Oczlon 2001: 432).

MARKEDNESS: The unmarked member of a set of forms is the general norm. It tends to be structurally simpler, more frequent in discourse, and should be easier to process than the corresponding marked form.

Givón (1993: 178) suggests three criteria for the theoretical construct of MARKEDNESS, namely structural complexity, discourse distribution, and cognitive complexity. He argues that the unmarked case, the general norm, should be structurally simpler or neutral, more frequent in discourse, and easier to process, whereas the marked case, the counter-norm, should be structurally more complex, less frequent in discourse, and harder to process. Take English RCs for example. Subject RCs with the normal word order of [RPsVO] (RPs = subject relative pronoun) is the unmarked form, and the object RC with the non-normal order of [(RPo)SV] (RPo = object relative pronoun) the marked. Prideaux and Baker (1986) found in a series of experiments that subject RCs were judged more natural, were read and responded to faster, and were used more frequently by subjects in various production and comprehension tasks.

The present study argues that both strategies of CLOSURE and MARKEDNESS interact with the specific grammatical properties of Chinese CNP constructions, especially those of RCs, rendering some forms more frequent in discourse than others. In Chinese discourse, RCs in CNPs are found to be mostly subject or object clauses. The former contains a relativised (zero) subject in the RC, and the latter a relativised (zero) object in the RC, and both RCs

frequently modify the grammatical subject or object of a clause. There are therefore four relative clause structures that commonly occur in the CNPs found in our data: SS, SO, OS and OO, where the first letter stands for the grammatical role of the head NP (S for subject and O for object), and the second for the gap (i.e., relativised NP) in RC (S for subject gap and O for object gap). The four RC constructions are schematised with examples in (13) to (16).

- (13) SS: Subject head NP modified by a subject RC

[_{RC} Ø_S V (O)] S V (O)

[Ø 值班 的] 士兵 啪的 开了 灯。

zhi.ban de shi.bing pa.de kai.le deng

on.duty PRT soldier onomatopoeia turn.on light

"The soldier (who) is on duty turned on the light."

- (14) SO: Subject head NP modified by an object RC

[_{RC} S V Ø_O] S V (O)

[她车 出来 Ø 的] 活计 象 她的衣服 一样 清洁 合体。

ta che chu.lai de huo.ji xiang ta.de yi.fu yi.yang qing.jie.he.ti

she lathe out PRT item like her clothes same clean-fit

"The items (that) she lathed were as clean and fit as her clothes."

- (15) OS: Object head NP modified by a subject RC

S V [_{RC} Ø_S V O] O

她们 齐声 叱骂 那个 [Ø 吐吐沫 的] 人。

ta.men qi.sheng chi.ma na.ge tu.tu.mo de ren

they(fem.) in.unison curse that-CL spit.spit PRT man

"They were all bawling at the man (who) spat. "

(16) OO: Object head NP modified by an object RC

S V [_{RC} S V Ø_o] O

那 两个 孩子 穿了 [我 织 Ø 的] 线衣。

na liang.ge hai.zi chuan.le wo zhi de xian.yi

that two-CL child wear I knit PRT cotton.sweater

"The two children wore the cotton sweaters that I knitted"

The present study considers subject RCs to be unmarked in form, and object RCs to be marked, as the former contains a relativised/zero subject ($[\text{RC } \emptyset_S \text{V}(\text{O})]$), and the latter a relativised/zero object ($[\text{RC } \text{SV}\emptyset_O]$). A clause with a zero subject is the most common clause type in Chinese, as often found in topic chains, which are regarded as basic functional units in Chinese discourse (Tsao 1990; Xu 1995; Chu 1998; Li 2005). Many quantitative studies of Chinese discourse have shown that zero subjects occur much more frequently than zero anaphora in any other grammatical positions. Pu (1997: 289), for example, demonstrates with data culled from written narratives that sentences with zero subjects appear to be a grammaticalised phenomenon, accounting for about 93% of all zero anaphora, but that sentences with zero objects are only marginal (about 5%). Similarly, Li (2005: 112) reports that topic chains with zero subjects are the most typical, most frequently occurring pattern in the Chinese discourse (92% of all topic chains) in her data. An unmarked structure that has a high frequency of occurrence in discourse would be processed easier than a marked one, as research suggests (Hawkins 2004).

Hence, the strategy of early CLOSURE would render SS and SO easier and faster to process than OS and OO, as the former contains a subject head noun, and the modifying relative clause does not interrupt the main clause (cf. the examples in (13) and (14)), while the latter contains an object head noun and the modifying relative clause is disruptive to the main clause processing (cf. the examples in (15) and (16)). However, if a CNP has a split head, the relative clause would always be disruptive to the main clause processing regardless of the grammatical role of the head noun. According to the strategy of MARKEDNESS, on the other hand, CNPs that contain subject relative clauses (cf. the examples in (13) and (15)) are assumed to be easier to process than those that contain object relative clauses (cf. the examples in (14) and (16)).

Given that even a non-embedding, subject-modifying RC may interrupt a head NP if it comes between the classifier and its head in a CNP construction, i.e., CL + RC + H, Chinese RCs in general would be more of disruptive types than the non-disruptive ones, which seems to run counter to the strategy of early CLOSURE. The present study argues, nonetheless, that early CLOSURE still operates at the global level of discourse processing, while the strategy of MARKEDNESS predicts the processing ease of subject RCs at local levels of discourse. In other words, the overall frequency of RC occurrence would be much lower in cognitively demanding tasks such as oral, real-time production than planned speech or written texts that are cognitively less taxing. On the other hand, relative clauses that are actually produced in discourse would be more of the unmarked subject RC type than the marked object RC type (cf. Pu 2007).

Based on the discussion of how cognitive strategies interact with RCs in CNP constructions, the present study suggests that classifiers may serve a multitude of functions in oral and written discourse and that their positioning may be strongly influenced by the RC type in a CNP construction. Since a Chinese RC lacks a leading relative pronoun or any other marker, it looks just like a main clause until the particle *de* is reached at the end of the RC, which would very likely result in a so-called "garden-path" sentence, e.g., [_{RC} Ø_SVO *de*] SVO, and risk being reprocessed at the end of the RC. A classifier placed before a RC in a CNP, for example, would lead the listener/reader to expect the forthcoming head NP and resist closure at the end of the RC that immediately follows the classifier. The present study suggests that a prominent function of classifiers is to signal to the listener/reader its upcoming head NP and hence avoid ambiguity and reprocessing, although this signaling function would be of primary importance in cognitively more constrained oral discourse because listeners are less able to afford reprocessing than readers.

With regard to the occurrence and distribution of classifiers in CNPs, the study proposes the two-part working hypothesis in (17).

- (17) In Chinese discourse, unclassified CNPs would occur more frequently than classified CNPs because the unclassified is the neutral, unmarked structure in discourse; split-head CNPs would occur more frequently than unified-head CNPs because the classifier placed before the RC in a CNP serves to signal the upcoming head to avoid

ambiguity in clause processing.

In the remainder of this paper, evidence will be provided to show that both cognitive strategies of CLOSURE and MARKEDNESS interact with RC structures to determine the distributional patterns of CNPs. While the ease with which speakers and listeners process different CNP constructions may be attributed to the type of RCs (subject vs. object) in a CNP, the decision to classify a head NP is influenced largely by pragmatic and discourse factors.

4. Distributional patterns of CNPs

4.1 Data and focus

The data used in this study came from two sources, namely (i) a narrative study in which native speakers of Mandarin Chinese were asked to watch a four-minute video-clip, *The New Doorbell*, and then describe it in either spoken or written form; and (ii) four short stories written by well-known contemporary Chinese authors, two of which are female. The short stories are similar in style (narratives focusing on human characters) and comparable in length (24 to 35 pages).

The video clip is a (Chinese) cartoon about a man who installs a new doorbell in his apartment and then waits anxiously for others to ring it. It is a silent color movie with background music; no written language appears on the screen at any time except the title, *The New Doorbell*, which is shown at the beginning of the video-clip in both Chinese and English. Thirty undergraduate students of various majors from the Central China University of Finance and Economics, sixteen males and fourteen females who were between the ages of 18 and 21, participated in the narrative study. They were randomly assigned to two gender-balanced groups of equal number: one group recounted the video-clip in oral language, and the other in written form. The oral data were tape-recorded and later transcribed, and the written data were collected immediately after subjects finished writing. The three target CNP constructions were collected for analysis from the two types of narratives and the four short stories. The analysis focused on CNP constructions with distributional significance in our narrative data, i.e., CNPs that function either as grammatical subject or object, because 63 out of the total of 69 CNPs in the narrative data

are of these two kinds. Hence, in the short story data, only those CNPs were culled for analysis.

4.2 Distribution of CNPs

We shall first examine the oral and written narrative data elicited from participants who watched the same video clip and performed the same task under the same condition (except the production mode). Indeed, the participants produced the two types of narratives with similar discourse density, i.e., the number of propositions per narrative. On average, speakers produced 87 propositions and writers 76 per narrative; there is no statistically significant difference between the two counts. Nevertheless, the tally results summarised in table 1 indicate that CNPs occur much more frequently in the written narratives than in the oral ones, regardless of the fact that the latter is in general longer than former. In other words, the written narratives contained 0.046 CNP per proposition, six times more than did the oral narratives (0.0077 CNP per proposition).

	Unified Head		Split Head		Bare Head		Total	
	N	%	N	%	N	%	N	%
Oral narratives	0	0.00	8	80.00	2	20.00	10	100
Written narratives	0	0.00	21	39.62	32	60.38	53	100
Short stories	14	6.73	61	29.33	133	63.94	208	100
Total	14	5.17	90	33.21	167	61.62	271	100

Table 1. Classified versus non-classified CNPs

The difference in the production rate of CNPs between oral and written narratives are striking, though not surprising, because of the operation of the cognitive strategy of early CLOSURE in a broader sense. In English discourse, the strategy works at the local, sentence level, while it operates at a global, discourse level in Chinese, especially in the production of oral narratives. Due to the fact that the majority of RCs in Chinese are of the embedding kind, speakers would try to avoid such clauses because they themselves and their listeners are more severely constrained by cognitive activities of memory and attention than are their counterparts in writing, and hence the on-line processing of embedded RCs is cognitively more costly to them. On the one hand, a RC precedes its head; this complex and relatively long modifier increases memory load and denies rapid access to the head. On the other hand, since an embedded RC disrupts its main clause, its processing works against the closure of a language unit before the advent of another; it is hence cognitively more demanding than a

non-disruptive unit. Consequently, speakers would in general try to avoid RC production in oral narrative since the unique structural characteristics of Chinese RCs seem to tax cognitive resources.

The processing constraints underlying spoken language affect writing and reading to a much lesser extent, because writers/readers are not pressed for on-line production, and therefore are relatively freed from the cognitive constraints of short-memory effects (Chafe 1992: 172). Writers usually plan a clause/sentence ahead before they actually write it down, and they have time to edit and revise their writing; readers can always stop when they have comprehension difficulties, and reread and reprocess at a more leisurely pace. The slower, editable, and retraceable written form allows writers to construe more complex sentences at ease and they can thus afford to use disruptive relative clauses more frequently.

Comparing the written narratives with the four short stories, we see that the stories, although much longer and more complex in plot and content, do not contain proportionally more CNP constructions than do the written narratives. On average, each written narrative (about one hand-written page long) contains 3.5 CNPs, whereas each short story (at least 24 pages long in print) contains 52 CNPs (excluding a smaller number of CNPs modified by RCs other than the subject and object types). We therefore assume that the interaction of cognitive activities and syntactic properties of RCs (or the production mode) rather than the length or complexity of narratives, results in the overall difference in the distribution rate of CNPs between oral and written narratives.

4.3 Bare-head CNPs

Table 1 shows further that 62% of the 271 CNPs are non-classified (or bare head NPs) and only 38% are classified. The frequency difference between the two is statistically significant ($p < 0.001$ in a X^2 -test), even though the pattern is reversed in the oral narrative data, where classified CNPs greatly outnumber unclassified ones. The difference in the two distributional rates between the oral and written data (short stories included) is very likely due to cognitive constraints underlying the two modalities, which will be discussed in detail in section 4.4. The general results, nevertheless, lend support to the first part of the hypothesis in (17), that the neutral, unmarked structure (i.e., unclassified CNPs) would occur more frequently than the biased, marked ones (i.e., classified CNPs) in discourse, or

written discourse, to be more accurate.

Classified NPs in Chinese have generally been considered to be thematically more prominent than bare head NPs, given the iconicity principle that more important information is coded by more material. Several studies on Chinese classifiers argue for the pragmatic and discourse functions of classifiers. Li (2000: 344), for example, proposes that classifiers serve to foreground their NPs, while Sun (1989: 298) claims that classified NPs are used to code thematically important referents. Neither consideration seems to be able to account for the CNPs in our data, which, regardless of whether they are classified, are foregrounded and encode thematically important referents in discourse. Consider the examples in (18) to (21).

- (18) [Ø 卧 床 的] 女人 转过身来。

wo chuang de nu.ren zhuan.guo.shen.lai

lie.in bed PRT woman turn.over

"The woman who was lying in bed turned over."

- (19) 她们 齐声 叱骂 那个 [Ø 吐 吐沫 的] 人。

ta.men qi.shen chi.ma na.ge tu.tu.mo de ren

they (fem.) in.unison curse that-CL spit spit PRT person

"They all cursed the man who spat."

- (20) 一个 [Ø 拍 皮球 的] 小 女孩 走 上楼来。

yi.ge pai pi.qiu de xiao nu.hzi zou shang.lou.lai

a-cl bounce ball PRT little girl walk upstairs

"A girl who was bouncing a ball walked upstairs."

- (21) [Ø 戴着 瓦片帽 的] 护士 小姐 走 进来。

dai.zhe wa.pian.mao de hu.shi xiao.jie zou jin.lai

wear tile.cap PRT nurse miss walk in

"A nurse who wore a cap walked in."

In each of four constructions in (18) to (21), the relative clause foregrounds the head noun or brings its referent to saliency or temporary saliency. The unclassified head noun *nu.ren* ("woman") in (18) encodes a given referent, which is the protagonist of a narrative (the short story data) and thus thematically central to the narrative, while the classified head noun *na.ge ren* ("that man") in (19), despite carrying given information and being brought to salience by the RC at the moment, is not mentioned again in the discourse and hence thematically less important in the narrative. The head nouns of (20) and (21), despite one being classified and the other not, are new referents, both of which serve as the topic for the subsequent span of the discourse and are hence thematically important for the particular episode of the discourse.

The present study argues that the tendency for bare-head CNPs to occur more frequently in written discourse is due to the weakening of the pragmatic function of classifiers, contrary to the claims made by Li and Sun. Since the RC in a CNP construction functions to highlight and/or provide further information for its head NP (Givón 1993), it diminishes and obliterates the functions that are commonly assumed by classifiers in simple NPs. In other words, the grounding status and thematic importance of the referent coded in the head NP do not seem to be based solely on the existence of classifiers: topical or thematically important head referents in CNP constructions are not always classified, while non-newsworthy referents often take classifiers in discourse, as illustrated in the above examples. Consequently, unclassified CNPs are the neutral, unmarked structure with a "default" value of being foregrounded, and thus occur much more frequently in discourse than do classified CNPs, the marked structure, to whose specific functions we turn in section 4.4.

4.4 Classified CNPs

Table 1 also shows that, of the two classified constructions, one with a unified head and one with a split head, the latter occurs much more frequently than the former (33% vs. 6%), and the difference between them is statistically significant ($p < 0.001$ in a X^2 -test). These results lend support to the second part of the hypothesis in (17), namely that split-head CNPs would occur more frequently than unified-head ones in discourse production.

Obviously when a classified CNP is used, speakers prefer one with a split head to one with a unified head. The most striking phenomenon is the avoidance of a unified head in the

narrative (experimental) data where, if a classifier is ever used to modify a CNP, it is placed before the relative clause away from the head noun. Moreover, in oral narratives, CNPs with split heads greatly outnumber those with bare heads, a trend that is different from the other two sets of written data.

It seems counter-intuitive that classifiers should be placed overwhelmingly frequently before the relative clause rather than next to its head, since a CNP with a split head always contains an interruptive relative clause, regardless of the grammatical role of the head NP. Why did speakers prefer the split structure when they had available the unified one which appears to be cognitively less demanding?

The overwhelming preference for the split-head CNP, this paper argues, to a large extent manifests the working of the cognitive strategy of MARKEDNESS at the sentence level in Chinese discourse processing. The results become clearer when we re-tabulate CNPs in terms of relative clause distributions, as reflected in table 2.

	CNP with subject RC				CNP with object RC				Total
	SS		OS		SO		OO		
	N	%	N	%	N	%	N	%	
Oral narratives	5	50	3	30	0	0	2	20	10
Written narratives	26	49.06	13	24.53	2	3.77	12	22.64	53
Text	93	44.71	60	28.85	14	6.73	41	19.71	208
Subtotal	124	45.76	76	28.04	16	5.90	55	20.30	271
Total	N=200		73.80%		N=71		26.20%		

Table 2. Distribution of the four relative clauses constructions

Table 2 shows that the distributions of the four types of RCs are markedly similar across the three sets of data. They can be ranked in a hierarchy of frequency from the highest to the lowest: SS > OS > OO > SO. In general, subject RCs greatly outnumber object RCs with a ratio of almost 3:1, and the difference between the two is statistically significant ($p < 0.001$ in a X^2 -test). The RC results also support, to a certain extent, Keenan and Comrie's (1977) Noun Phrase Accessibility Hierarchy, an implicational universal proposing that a grammatical subject is more accessible for relativisation than is a direct object or any other grammatical role.

The present study proposes that the preponderance of subject RCs results from the

interaction of MARKEDNESS and the structural characteristics. As discussed in section 3, subject RCs with zero subjects are unmarked in form (i.e., [_{RC}ØsVO *de*]) whereas object RCs with a zero object are marked (i.e., [_{RC}SVØo *de*]). Since the unmarked case, the general norm, is structurally simpler (or neutral) and easier to process than the marked case, subject RCs are used more frequently in Chinese discourse.

The tendency for subject RCs to occur more frequently than object RCs explains the overwhelming preference for split-head CNPs over unified-head ones, especially in cognitively more demanding spoken tasks. Since a subject RC assumes the form of a main clause, it is prone to confusion when it modifies a subject head that appears sentence-initially. It is very likely to be taken as the main clause and would have to be re-processed when the RC marker 的 (PRT *de*) is encountered at the end of the clause. Reprocessing is costly both time-wise and cognitively, and needs to be avoided. An effective way to avoid such a structure is to place a classifier before the RC, which serves as a signaling device to indicate to the hearer/reader that if what follows is not a head noun, it must be part of the entire NP and the head noun is to be expected. Consider the examples in (22) to (24).

(22) 那个 [Ø 早 在 门 后 等着 的] 中年 人 很 失望。

na.ge zao zai men hou deng.zhe de zhong.nian ren hen shi.wang
that-CL early at door back wait PRT middle.age man very disappoint

(23) [Ø 早 在 门 后 等着的] 那个 中年 人 很 失望。

early at door back wait PRT that-CL middle.age man very disappoint

(24) [Ø 早 在 门 后 等着的] 中年 人 很 失望。

early at door back wait PRT middle.age man very disappoint

"(That) middle-aged man (who) had long waited behind the door was very disappointed."

The construction in (22) is taken from the oral narrative data, where the grammatical subject is coded by a split-head CNP instead of an alternative unified-head CNP (as in (23)) or a bare-head CNP (as in (24)). The referent coded by any of the three CNPs is the same

protagonist that is given, topical and salient in the video clip, and in addition, the referent is being foregrounded by the RC at the juncture of discourse. Thus the classifier may be redundant if its sole purpose is to foreground the head NP (Li 2000: 344), to signal thematic importance of the referent (Sun 1989: 298), or to indicate definiteness and given-ness of the referent (Chao 1968: 9; Hashimoto 1971: 24). However, the speaker here chose to use a classifier and place it before the RC because the classifier would signal to the listener that the head noun is forthcoming, and anything between the classifier and the head would be part of the NP, not an independent clause. If, on the other hand, the speaker had chosen either (23) or (24), the initial RC may have resulted in structural ambiguity: the listener may take it as a main clause and need a "second pass" to re-analyse it as the end of the RC is reached.

Indeed, the results show the significance of the signaling function served by the split-head CNP in the oral production task, where cognitive demand is greater and structural ambiguity is to be avoided. Table 3 presents the distributional patterns of the classified CNPs with regard to SS, SO, OS and OO structures in the three sets of data.

	Oral	Written	Text	Total	
				N	%
SS	5	15	23	43	47.78
OS	2	5	25	32	35.56
SO	0	0	4	4	4.44
OO	1	1	9	11	12.22
Total	8	21	61	90	100.00

Table 3. Distribution of split-head CNPs

Of the ten CNPs found in the oral narratives, eight (80%) were of the split-head construction, and five (50%) took subject RCs. In the written narrative, however, 40% of CNPs (21 out of 53 tokens) were split-head cases and 28% (15 tokens) took subject RCs. It was expected that the occurrence of the split-head construction would decrease in the written narratives because in cognitively less demanding production tasks, the tolerance for structural ambiguity would be higher, since both writers and readers are not as pressed for time and as constrained in working-memory as are speakers. Moreover, in the short stories that were composed and edited over a longer period of time, the signaling function of the classifier was still less compelling, and hence the proportion was further reduced: 61 of the 208 CNPs (29%) had split heads, of which 27 (13%) occurred with subject RCs.

What other function(s), then, do classifiers perform in CNPs found in written discourse if the signaling function is less important? The present study suggests, with evidence from the narrative and story data, that classifiers in split-head CNP constructions serve two additional pragmatic and discourse functions: one is to introduce a new referent that is thematically important into the discourse, and the other is to draw momentary attention to a head noun in order to set off or emphasise certain attributes of its referent at a particular moment of discourse. The former is usually taken by classifiers that modify subject head NPs and the latter by classifiers that modify object head NPs. In a subject CNP that encodes an important new referent, the relative clause grounds or anchors the head NP cataphorically to the subsequent discourse or to the general knowledge of the hearer/reader (Givón 1993: 119), and the classifier placed before the RC puts the newness, as indicated by the classifier, in a prominent clause-initial position (cf. the examples in (25) to (27)).

(25) 一个 [Ø 戴 墨镜 的] 年轻 人 来到 他 门前。

yi.ge dai mo.jing de nian.qing ren lai.dao ta men.qian
 a-CL wear dark.glasses PRT young man came.to his door.front
 "A young man who wore dark-glasses came to his door."

(26) 一个 [Ø 住 在 顶西头 的] 媳妇 找到 我。

yi.ge zhu zai ding.xi.tou de xi.fu zhao.dao wo
 a-CL live at westmost PRT woman find me
 "A woman who lived at the west-most of the village came to see me."

(27) 一些 [Ø 说话 机智的] 人 在 我们 庄 享有 盛誉。

yi.xie shuo.hua ji.zhi de ren zai wo.men zhuang xiang.you shen.yu
 some-CL speak witty PRT man at our villagea enjoy great.honor
 "Some people (who) were witty enjoyed great prestige in our village."

The split-head CNP such as the ones in (25)-(27) above is typically the subject of a clause and introduces a new referent of thematic importance, which usually persists over the subsequent span of discourse. The RC provides distinctive and characterising information to

help the hearer/reader single out the new referent from among other possible candidates and quickly establish the identity of the referent, thus facilitating discourse comprehension.

The classifier and its following RC in a split-head construction, on the other hand, may not necessarily have to designate thematic importance to a new, topical referent, but may often put an otherwise non-topical referent in the spotlight just momentarily in discourse. Such a CNP construction usually functions as the object of a clause, and the referent of the head noun is typically indefinite and non-referential, as in (28) to (30).

(28) 他继续 等待 下一个 [Ø 来 按 门铃 的] 人。

ta ji.xu deng.dai xia.yi.ge lai an men.ling de ren
he continue wait next.CL come ring doorbell PRT person

"He continued to wait for the next person who would ring the doorbell."

(29) 她 企盼 一个 [Ø 又 英俊 又 有钱 的] 男人。

ta qi.pan yi.ge you ying.jun you you.qian de nan.ren
she yearn.for a-CL both handsome and wealthy PRT man

"She yearned to have a man (who) was both rich and handsome."

(30) 厂长 已经 开始 扮演 一个 [Ø 赖帐 的] 角色了。

chang.zhang yijing kai.shi ban.yan yi.ge lai.zhang de jiao.se.le
CEO already started play a-CL repudiating.a.debt PRT role

"The CEO started to play someone (who) was repudiating a debt."

Although the referent encoded in an object CNP such as in (28) to (30) above is non-topical and transient in discourse, the modifying RC sets off its temporary saliency and emphasises the importance of its attribute. The classifier before the RC, however, quantifies and marks the entity from the set and scope designated by the RC. It makes it more conspicuous that the referent encoded by the CNP in any of the three examples above, though indefinite, is not any person in the world, but *yi.ge* ("one+CL") from a class with the defining and qualifying information provided in its respective relative clause.

Of interest here is that the split-head CNPs discussed above, be they subject or object, mostly take subject RCs. This is so because a subject-gapped RC is not only unmarked in form but also typically modifies a head NP that encodes a human referent, i.e., as the subject of the clause. It is only natural that human referents, rather than animals or inanimate objects, are frequently focused on as topic or temporarily highlighted in narrative discourse (cf. Pu 2007). Cognitive constraints underlying language processing together with pragmatic and discourse factors conspire to account for the overwhelming preference for split-head CNPs taking subject RCs (75 out of the 90 split-head tokens).

CNPs with unified heads, on the other hand, were used quite rarely in the narratives: 14 tokens in total, all of which occur in the short story data and account for only about 13% of all classified CNPs and 5% of all CNPs. Most of these unified-head CNPs contain definite head nouns that carry given information, as in (31) and (32).

(31) [你说 \emptyset 的] 那种 扣子 很 贵。

Wni shuo de na.zhong kou.zi hen gui
you mention PRT that-CL button very expensive

"The kind of buttons (that) you mentioned is very expensive."

(32) [\emptyset 从 坟场 回来 的] 那群 女人 后来 聚到 秋仪 家里。

cong fen.chang hui.lai de na.qun nu.ren hou.lai ju.dao Qiuyi.jia.li
from cemetery return PRT that-cl women later gather.at Qiuyi home

"The group of women (who) returned from the cemetery later gathered in Qiuyi's home."

The head noun in both utterances above encodes a given, known entity or referent that has been introduced earlier in the discourse, but the referential distance between the current and the previous is relatively large (longer than 3 clauses). The RC used to modify this type of head noun provides repeated information to remind the listener/reader that the referent of the head noun is known. In such cases, the definiteness and given-ness carried in the classifier such as *na.zhong* ("that kind") in (31) and *na.qun* ("that group") in (32) is entirely redundant, and can be omitted without altering the meaning of the utterance or making difficult the identification of its referent. It is clear why the unified-head CNPs are rarely used in

discourse: they do not seem to serve the function of introducing a new referent, nor do they seem to serve the function of quantifying and qualifying the head noun or emphasising its attribute. Therefore, the general principle underlying the use of classified CNPs may be that if a classifier does not serve its cognitive, pragmatic, or discourse functions, it is not preferred at all.

5. Conclusion

The present study takes a cognitive-functional approach to the analysis of a syntactic phenomenon, namely complex NPs, and aims to uncover the factors underlying the distribution of CNPs in Chinese discourse in general, and the occurrence and positioning of classifiers in such CNPs in particular. The analysis of the data obtained from both spoken and written narratives reveals that the distributional patterns of various CNP constructions can well be accounted for by the cognitive strategies of CLOSURE and MARKEDNESS, and the bias found toward a split-head CNP in the classified CNP construction, especially in spoken discourse, suggests the classifier placed before the RC may serve as a cue to avoid structural ambiguity and facilitate sentence processing.

The results of the present study also argue for multiple functions of classifiers in CNP constructions, which are different from the pragmatic functions of foregrounding the head noun, of granting thematic importance to its referent, or of indicating information status, as discussed in prior research. The primary function of classifiers, the present study suggests, is to provide a cue for listeners and readers to the forthcoming head of the complex CNP construction. Furthermore, two other important functions of classifiers are to make prominent the introduction of a new, topical referent (most likely a subject) encoded in the head noun on the one hand, and to designate certain unique attributes of the head NP referent (mostly an object) on the other.

The primary function of classifiers as a signaling device also explains the overall frequency of distribution of classified and unclassified CNPs in the data: the majority (80%) of CNPs in oral discourse are classified whereas more than 60% of CNPs in written discourse are unclassified. In cognitively more demanding oral discourse, where hearers are pressed for time and cognitive resources, a cue to a potentially ambiguous structure would greatly

facilitate comprehension, whereas such a cue would be less compelling in written discourse, where cognitive constraints are relaxed, and readers have more time to process language. Therefore in written discourse CNPs tend not to be classified. The use of classifiers in CNPs is very much in accord with Givón's (1983: 13) general psychological principle underlying language processing: "Expend only as much energy on a task as is required for its performance."

References

- Andersen, A.H. 1989. Markedness theory – The first 150 years. In O.M. Tomic (ed). *Markedness in synchrony and diachrony*. New York: Mouton de Gruyter. pp. 11-46.
- Chafe, W. 1992. The flow of ideas in a sample of written language. In W.C. Mann and S.A. Thompson (eds). *Discourse description: Diverse linguistic analysis of a fund-raising text*. Amsterdam: John Benjamins. pp. 268-94.
- Chao, Y.R. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chu, C.C. 1998. *A discourse grammar of Mandarin Chinese*. New York: Peter Lang Publishing.
- Fenk-Ozslon, G. 2001. Familiarity, information flow and linguistic form. In J. Bybee and P. Hopper (eds). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. pp. 431-448.
- Frazier, L. 1979. *On comprehending sentences: Syntactic parsing strategies*. Bloomington: Indiana University Press.
- Givón, T. 1993. *English grammar: A function-based introduction (I and II)*. Amsterdam: John Benjamins.
- Givón, T. 1983. Topic continuity and word order pragmatics in Ute. In T. Givón (ed). *Topic continuity in discourse: Quantitative cross-language studies*. Amsterdam: John Benjamins. pp. 343-363.
- Hashimoto, A.Y. 1971. Mandarin syntactic structures. *Unicorn* 8: 1-149.
- Hawkins, J. 2004. *Efficiency and complexity in grammar*. Oxford: Oxford University Press.
- Just, M.A. and P.A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99(1): 122-49.
- Keenan, E.L. and B. Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63-99.
- Li, W. 2000. Numeral-classifiers as a grounding mechanism in Mandarin Chinese. *Journal of*

- Chinese Linguistics* 28(2): 337-368.
- Li, W. 2005. Topic chains in Chinese: A discourse analysis and applications in language teaching. *LINCOM Studies in Asian Linguistics* 57.
- Prideaux, G.D. 1999. Discourse analysis and psycholinguistics: Breaking down barriers. In A. Lommel and A. Melby (eds). *LACUS Forum* 26: 1-16.
- Prideaux, G.D. and W.J. Baker. 1986. *Strategies and structures: The processing of relative clauses. Current Issues in Linguistic Theory* 46. Amsterdam: John Benjamins.
- Pu, M-M. 1997. Zero anaphora and grammatical relations in Mandarin. In T. Givón (ed). *Grammatical Relations; Typological Studies in Language* 35. Amsterdam: John Benjamins. pp. 283-322.
- Pu, M-M. 2007. The distribution of relative clauses in Chinese discourse. *Discourse Processes* 43(1): 25-53.
- Slobin, D.I. 1973. Cognitive prerequisites for the development of grammar. In C.A. Ferguson and D.I. Slobin (eds). *Studies of child language development*. New York: Holt, Rinehart and Winston. pp. 175-208.
- Sun, C. 1989. The discourse function of numeral classifiers in Mandarin Chinese. *Journal of Chinese Linguistics* 16(2): 298-322.
- Tsao, F. 1990. *Sentence and clause structure in Chinese: A functional perspective*. Taiwan: Student Book.
- Van Dijk, T. and W. Kintsch. 1983. *Strategies in discourse comprehension*. New York: Academic Press.
- Xu, Y. 1995. *Resolving third person anaphora in Chinese text: Toward a functional-pragmatic model*. Doctoral dissertation, Hong Kong Polytechnic University.