PREFERRED ARGUMENT STRUCTURE FOR DISCOURSE UNDERSTANDING

KA-WAI CHUI

Matsushita Electric Institute of Technology
(Taipei) Co., Ltd.

Abstract

The main purpose of communication is to exchange information. Any discourse understanding model should be able to process the flow of information throughout the entire text. According to Du Bois (1987)'s studies of information flow in discourse across a number of languages, information distribution among argument positions in clauses is by no means random, but certain grammatical patterns tend to recur consistently. He thus formulated a *Preferred Argument Structure* (PAS) for the preferential structural configurations of arguments. In our examination of Chinese narrative discourse, the language also displays PAS, yet the Chinese PAS challenges the universality of the one Du Bois proposed. Based on the quantity and distribution of lexical arguments and new referents across grammatical roles in discourse, it is realized that Chinese PAS also maintains one new argument at most within a basic information processing unit. Since new referents in Chinese have to be encoded in full NP form, it is thus less likely to have more than one lexical argument within a clause. Moreover, this single new argument appears preferentially in the *O* role, rather than the *A* and *S* roles Du Bois's PAS formulates.

Since the structure of information flow has a corresponding grammatical patterning, both grammatical and pragmatic processing can be carried out simultaneously, in that the information status of an argument can be identified by virtue of grammatical analysis. Although PAS is neither universal nor categorical, it can function in a discourse understanding model as heuristic device to process the information structure of a connected spoken discourse.

1. Introduction

The main purpose of communication is to exchange information. On the part of a speaker, he may employ various strategies to organize the information he intends to convey, in that some bear old information while others carry new information. Therefore, a discourse understanding model should be able to process the flow of information throughout the entire text. In this paper, our issue is focused on the referring arguments in Chinese narrative discourse, and our main concern is how they are structured in relation to information flow.

According to Du Bois (1987)'s studies of information flow in discourse across a number of languages, information distribution among argument positions is neither arbitrary nor random, but certain grammatical patterns are preferred over others, especially they tend to recur consistently in a connected spoken discourse. In other words, the structure of information flow has a corresponding grammatical patterning. Those recurrent patterns, which indeed reflect speakers' actual language use, are formulated as *Preferred Argument Structure* (PAS) The PAS he formulated comprises the following constraints: *One Lexical Argument Constraint* to avoid more than one referent in full NP form per clause, *Non-Lexical A Constraint* to have the single lexical referent not appearing in the *A* role, *One New Argument Constraint* to avoid more than one referent carrying new information per clause, *Given A Constraint* to have the new referent not appearing in the *A* role. However, in examining Chinese narrative discourse, it is discovered that the PAS that this particular type of discourse genre displays challenges the universality of Du Bois's. The idiosyncrasy of the Chinese PAS will be discussed in this paper.

In fact, from the computational point of view, no matter it is universal or language-specific, the existence of PAS has significant implication to discourse understanding. On the one hand, it enables grammatical and pragmatic processing being carried out simultaneously because the information status of a referent can be identified by virtue of grammatical analysis; on the other hand, PAS can function as heuristic device to process the information structure of a connected discourse.

2. Preferred Argument Structure in Chinese Narrative Discourse

Unlike the languages Du Bois has studied (1987), Mandarin Chinese is a typologically different language with no inflection and relatively free word order. Nevertheless, it still exhibits its own idiosyncratic PAS in spoken narrative discourse. The corpus for the present study comprises eight oral narratives as told by eight Mandarin native speakers of 20-25 years old. They were requested to describe the story about the popular movie *Ghost* to the interviewer in a speech laboratory. It portrayed

a young man who was killed accidentally in a robbery, and who tried to protect his girlfriend from the murderers and to take revenge on them in form of a spirit. The narratives were taped for later transcription.

To study the Chinese PAS, our examination is focused on the issues of quantity and role in distributing lexical arguments and new referents across grammatical positions at both the grammatical and pragmatic dimensions.

2.1 Preliminaries for Analysis

Segmentation of the 120 minutes long narratives was subject to *intonation unit* being identified by ear as a stretch of speech uttered under a single coherent intonation contour and typically bounded by a pause. Chafe (1987) has hypothesized that intonation units representing linguistic expressions of focuses of consciousness are independent processing units typical of spoken discourse. In the present corpus, there were a total of 1433 intonation units, with a mean length of 6.69 words. The fact that the clause being defined as a verb and its arguments, and the intonation unit often coincide (Du Bois, 1987; Chafe, 1987, 1988) was further confirmed in this study, since 85.28% (1222) intonation units contained clausal elements. Those units comprised false starts, repetitions, filled pauses, as well as clause fragments such as conjunctions, adverbials, and particles would be excluded from further analysis. Therefore, the study of Chinese PAS is indeed based on clauses. Following is a sample of five clausal intonation units (a~e) produced by a female speaker:

(1) a. *jiouyitian ta genzhe ta nupengyou*
    one day   he follow his girlfriend
    'One day, he followed his girlfriend.'

   b. *zai ta nupengyou jiali deshihou*
    in his girlfriend home when
    'When (he was) in his girlfriend's home,'

   c. *ta nupengyou zai huan yifu*
    his girlfriend PROG change clothes
    'his girlfriend was changing clothes.'

   d. *ranhou you yi ge huairen*
    then   there-be one CL bad guy
    'Then, there was a bad guy.'

   e. *huairen chuang jinlai*
    bad guy break in
    'The bad guy broke in.'

Within a single clause, the morphological type of each referent, its grammatical role, as well as information status were all recorded. The morphological type of an overt referent in Chinese was either a lexical NP or a pronoun,

whose surface grammatical role would be classified as *A* (transitive subject), *S* (intransitive subject), *O* (transitive object), or *Oblique* (object of a preposition). Furthermore, Chafe (1987)'s three-way distinction of information for referents was adopted, mainly because his categories lay their foundation on the actual cognitive processing of information transfer by language users. They were *given information, accessible information*, and *new information*. A *given* referent referred to the entity mentioned previously, while a *new* referent was the one that had not yet been brought up in the prior context. Intermediate between these two was *accessible* information, either coming from the expectations associated with a schema or resulting from deactivation from an earlier state. Following Du Bois (1987), a referent constituted by deactivation should be at least twenty propositions away from its most recent appearance operationally.

2.2 The Grammatical Dimension of PAS

The purpose of studying PAS at the grammatical dimension is to examine whether there is a preferred surface configuration of arguments in the observed data. Therefore, we investigate both the number of lexical (NP) arguments and their distribution across the grammatical roles in clauses.

According to our tabulation shown in Table 1, of the 1127 clauses (excluding the equational type), those with zero or one lexical argument are the most common structure which constitute a distinct majority (94.15%).

Table 1. Frequency of clauses with 0, 1, and 2 lexical arguments

	frequency	percentage
0 lex arg	587	52.09
1 lex arg	474	42.06
2 lex arg	66	5.85
Total	1127	100

($X^2_{.99\,(2)} = 399.89$)

Since only transitive verbs can take more than one argument, it is necessary to seperate them from the intransitive ones for tabulation, in case the rarity of two-lexical-argument structures is simply due to the rarity of transitive clauses. The result in Table 2 shows clearly that even in transitive constructions, two-lexical-argument structures are still a minority (9.17%). The result indeed supports Du Bois's *One Lexical Argument Constraint* in that "there is a tendency for speakers to avoid more than one lexical argument per clause" (p.819).

Table 2. The frequency of lexical arguments
in transitive and intransitive clauses

	Transitive		Intransitive		Total	
	freq	%	freq	%	freq	%
0 lex arg	321	44.65	266	65.2	587	52.09
1 lex arg	332	46.18	142	34.8	474	42.06
2 lex arg	66	9.17	———		66	5.85
Total	719	100	408	100	1127	100

$$(X^2_{.99\ (2)} = 66.56)$$

Since speakers incline to use one lexical argument at most in a single clause, it is necessary to study whether this single lexical referent is randomly distributed across the grammatical roles. According to our tabulation as shown in Table 3, it is realized that O (84.3%) and *Oblique* (92.31%) each contain an overwhelming proportion of lexical arguments, whereas A and S contain a smaller portion of them.

Table 3. Grammatical roles and morphological types of arguments

	lexical		pronominal		Total
	n	%	n	%	n
A	155	38.08	252	61.92	407
S	132	55.46	106	44.54	238
O	306	84.3	57	15.7	363
OBL	192	92.31	16	7.69	208
Total	785	64.56	431	35.44	1216

$$(X^2_{.99\ (3)} = 265.09)$$

Since 64.56% of all referents are lexical, if they are randomly distributed across the grammatical positions, 38.98% of them will appear in the O role, while the A and S roles are restricted to include lexical referents, as indicated by Table 4.

Table 4. Distribution of lexical arguments
across grammatical roles

	frequency	percentage
A	155	19.75
S	132	16.82
O	306	38.98
Obl	192	24.45
Total	785	100

$$(X^2_{.99\ (3)} = 91.17)$$

Unlike Du Bois's *Non-Lexical A Constraint* to avoid lexical referents appearing in the A position, Chinese speakers would not prefer the A and S roles to mention a referent lexically. It is the position O (or *Oblique*) that preferentially favors lexical arguments. The *Lexical O Constraint* is thus proposed to characterize this particular phenomenon in Chinese narrative discourse. In short, the *One Lexical Argument Constraint* and the *Lexical O Constraint*, which are indeed the constraints on quantity and role respectively, constitute the Chinese PAS at the grammatical dimension. The quantity of lexical argument within a clause is usually one at most, and this single argument preferentially appears in the O role. Although they are not categorical rules, they do represent a statistically significant tendency of actual language use.

2.3 The Pragmatic Dimension of PAS

In the preceding section, it has been shown that in narrative discourse different argument positions of a clause have distinct morphological preferences. This section aims at studying the pragmatic dimension of PAS by examining the quantity of new arguments, as well as their distribution across the grammatical roles.

Firstly, it is found that transitive and intransitive clauses either contain zero or one new referent, with the former predominating (81.06%), as indicated by Table 5. Significantly, not a single clause contains two new referents. The result supports Du Bois's *One New Argument Constraint* to "avoid more than one new argument per clause" (p.826).

Table 5. The frequency of new arguments
in transitive and intransitive clauses

	Transitive		Intransitive		Total	
	freq	%	freq	%	freq	%
0 new arg	442	77.27	217	90.04	659	81.06
1 new arg	13	22.73	24	9.96	154	18.94
Total	572	100	241	100	813	100

$$(X^2_{.99\ (1)} = 18.01)$$

To understand whether the single new referent is randomly distributed across A, S, O, and *Oblique*, it is necessary to examine the distribution of information across these positions. As indicated in Table 6, a substantial proportion of A and S carry old information, and new referents preferentially occur in O and *Oblique*.

Table 6. Grammatical roles and information
status of argument

	new		accessible		given		Total
	n	%	n	%	n	%	n
A	12	2.95	10	2.46	385	94.59	407
S	24	10.08	11	4.62	203	85.3	238
O	122	33.61	34	9.37	207	57.02	363
OBL	81	38.94	25	12.02	102	49.04	208
Total	239	19.65	80	6.58	897	73.77	1216

$$(X^2_{.99\ (6)} = 229.02)$$

Of 239 new referents found in the corpus, a large portion occur in the O role (51.05%) as shown in Table 7, while only a small portion appear in the A and S roles which overwhelmingly convey old information. Since Chinese speakers disfavor both the A and S roles to mention a new referent, Du Bois's *Given A Constraint*, which "avoids introducing a new referent in the A-role argument position" (p.827), is inappropriate to Chinese narrative discourse. The *New O Constraint* is then proposed for Chinese to characterize the free occurrence of new referents in the O role, as well as the high restriction in the A and S roles.

Table 7. Distribution of new arguments
across grammatical roles

	frequency	percentage
A	12	5.02
S	24	10.04
O	122	51.05
Obl	81	33.89
Total	239	100

$$(X^2_{.99\ (3)} = 131.96)$$

Comparing the frequency distribution of A and S, it is even rare for A to code new referents. This can be explained by the fact that Chinese includes a type of presentative construction which "performs the function of introducing into a discourse a noun phrase naming an entity" (Li & Thompson, 1981). Verbs of this sentence type are usually intransitive, and the following arguments usually carry new information. Since speakers do not necessarily use presentative constructions to introduce a new entity, they merely constitute a minority (20 clauses) in our corpus, as exemplified in (2) and (3).

(2) *turan   you yi ge huairen pao chulai qiangjie*
suddenly exist one CL bad guy run out   rob
'Suddently, there is a bad guy running out to rob.'

(3) *jie   shang you yi ge zhaopai   ya*
street on   exist one CL signboard PART
'On the street, there is a signboard.'

In short, the *One New Argument Constraint* and the *New O Constraint* constitute the Chinese PAS at the pragmatic dimension. There is a strong tendency in discourse to limit the number of new argument in a clause to a maximum of one. This single new referent tends to be introduced in the O (or *Oblique*) role and the second occurrences preferentially appear in the A and S roles. It is of course this preponderance of old information found in the {A, S} alignment that gives Chinese the distinction of being a topic-prominent language.

2.4 Correlation of PAS between Grammatical and Pragmatic Dimensions

We have already studied the quantity and the role constraints that constitute PAS for Chinese narrative discourse at both the grammatical and pragmatic dimensions. The correlation of PAS between these two dimensions is so strong that the grammatical *One Lexical Argument Constraint* and *Lexical O Constraint* are parallel to the pragmatic *One New Argument Constraint* and *New O Constraint* respectively, as shown in Table 8. In other words, the most preferred structure is to have one new argument at most within a single clause. Since new referents in Chinese have to be coded in full NP form, it is thus less likely to include more than one lexical argument within one discourse unit. Moreover, there is a strong tendency for the single new argument to appear in the O role, so that the lexical referent typically appear in this particular position. The flow of information does have a corresponding grammatical patterning.

Table 8. Dimensions and constraints of
Chinese PAS

	Grammar	Pragmatics
Quantity	*One Lexical Argument Constraint*	*One New Argument Constraint*
Role	*Lexical O Constraint*	*New O Constraint*

Comparing the Chinese PAS with the one Du Bois proposed for the languages he has studied such as Sacapultec Maya, as shown in Table 9, it is obvious that Du Bois's PAS cannot completely be generalized to Chinese, at least the narrative discourse genre is concerned. Their difference lies in the distribution of lexical

arguments and new referents across grammatical roles. As the PAS in Sacapultec avoids mentioning new lexical arguments in the *A* role, Chinese speakers disfavor both the *A* and *S* roles and strongly prefer *O*.

Table 9. Dimensions and constraints of PAS in Chinese and Sacapultec

	Grammar	
	Chinese	Sacapultec
Quantity	One Lexical Argument Constraint	
Role	Lexical O Constraint	Non-Lexical A Constraint
	Pragmatics	
	Chinese	Sacapultec
Quantity	One New Argument Constraint	
Role	New O Constraint	Given A Constraint

3 . Implication of PAS to Discourse Understanding

Chinese, like a number of other languages whose pattern of information flow in spoken narrative discourse has been investigated to date, also exhibits PAS. This suggests that there is a strong discourse pressure driving the various grammatical patterning in different languages, so that the universality of the PAS Du Bois proposed encounters challenge. However, from the computational point of view, no matter whether PAS is universal or language-specific, its existence has significant implication to discourse understanding, in that the flow of information throughout a connected discourse is highly structured with a corresponding grammatical patterning as far as quantity and role are concerned. Therefore, it is possible to identify the information status of an argument by virtue of grammatical analysis, so that both grammatical and pragmatic processing can be carried out simultaneously. Even though PAS is not categorical in nature, a discourse understanding model can still use it as heuristic device to process the information structure of a connected spoken discourse.

In short, a discourse understanding model employing PAS for information processing should take the following points into consideration:
(a) Clauses are the basic information processing units.
(b) Transitive and intransitive clauses should be seperated for analysis.
(c) The morphological type, grammatical role, and information status should be recorded for each argument position.

(d) The quantity and role constraints are the heuristic principles for information processing.

4. Conclusion

In this paper, we have demonstrated that Chinese narrative discourse also displays *Preferred Argument Structure* based on the quantity and distribution of lexical arguments and new referents across grammatical roles. However, the Chinese PAS challenges the universality of the one Du Bois proposed, because they differ in the distribution of lexical arguments and new referents across grammatical roles. In other words, the discourse pressure driving the various grammatical patterning in different languages reflects the underlying pragmatic preference of the different groups of language users.

From the computational viewpoint, no matter whether PAS is universal or language-specific, its existence has significant implication to discourse understanding. On the one hand, PAS can function in a discourse understanding model as a heuristic device to process the information structure of a connected spoken discourse; on the other hand, the information status of an argument can be identified by virtue of grammatical analysis since the flow of information has a corresponding grammatical patterning.

References

Chafe, Wallace L. 1987. Cognitive Constraints on Information Flow. In Russells S. Tomlin (ed.), Coherence and Grounding in Discourse. Amsterdam: Benjamins.

Chafe, Wallace L. 1988. Linking Intonation Units in Spoken English. In Haiman & Thompson (eds.), Clause Combining in Discourse and Grammar. Amsterdam: Benjamins.

Du Bois, John W. 1987. The Discourse Basis of Ergativity. Language 63, 805~855.

Givon, Talmy. 1983. (ed.). Topic Continuity in Discourse: a Quantitative Cross-language Study. Arnsterdam: Benjamins.

Li, Charles, & Thompson, Sandra. 1981. Mandaring Chinese: A Functional Reference Grammar. California: University of California Press.

Givon, Talmy. 1983. (ed.). Topic Continuity in Discourse: A Quantitative Cross-Language Stduy. Amsterdam: Benjamins.