



Inter-clausal Anaphora in Chinese Conditionals: a Multi-factorial Analysis

Shunting Chen, Pascal Amsili, Yiming Liang

► To cite this version:

Shunting Chen, Pascal Amsili, Yiming Liang. Inter-clausal Anaphora in Chinese Conditionals: a Multi-factorial Analysis. 35th Pacific Asia Conference on Language, Information and Computation, Nov 2021, Shangai (virtual), China. hal-04029215

HAL Id: hal-04029215

<https://hal.science/hal-04029215v1>

Submitted on 14 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inter-clausal Anaphora in Chinese Conditionals: a Multi-factorial Analysis

Shunting Chen

Shanghai International
Studies University

shuntingchen@shisu.edu.cn

corresponding author

Pascal Amsili

Sorbonne Nouvelle &
Lattice (CNRS/ENS/SN)

Pascal.Amsili@ens.fr

Yiming Liang

University of Paris &
LLF (CNRS/U Paris)

flavie.liang@gmail.com

Abstract

Chinese inter-clausal anaphora, as in *Wo suiran xiang facai, keshi bugan maoxian*. ‘I although want to get rich, yet don’t dare to take risks’ is considered ‘rather striking’ (Chao, 1968) because of the way the subject can be fronted or not, which might come from reference patterns of the two clausal subjects. Previous corpus investigations (Xu (1995), Chen (2016), and others) have suggested a direct relationship between reference patterns (disjoint reference or co-reference between the two subjects) and the position of the subject. The present study explores other factors that might play a role in predicting the structure patterns.

Corpus data for 25 conditional subordinators are grouped according to logic relations from Huang and Liao (2002). 8007 sentences were subject to the final statistical analysis, performed employing logistic regression models. In this model, it is found out that in predicting structure, the genre Newspaper is less likely than Literature in getting fronted structure; disjoint reference pattern is less likely to indicate fronted structure than conjoint reference pattern; NP in the first clausal subject position is more likely to be associated with fronted structure than other pronouns or wh-pronouns or personal pronouns; NP in the second clausal subject position is more likely to result in fronted structure than personal pronouns but less likely than zero pronouns in the same place;

fronted subject in clause two is more likely to lead to fronted structure than un-fronted one. Motivations include topicality, accessibility, and others. Such model is possibly applicable to other types of inter-clausal anaphora.

1 Introduction

It is “a rather striking characteristic” that Chinese subordinate bi-clausal sentences are constructed in two ways (Chao 1968). One is similar to the English sentence pattern of IF...THEN; the other is to front the subject which might serve as a topic (Xu 1995). For example, with the if clause, Chinese can use two patterns: If I go to Beijing, I usually take trains. and I if go to Beijing, take trains. According to Chao, both patterns exist with different frequencies according to whether the NPs in the subject position in the two clauses co-refer or do not co-refer. His idea is that co-reference favours the second pattern while disjoint reference prefers the first one. In (1) and (3) the subject of the subordinate clause appears within the clause, after the main subordinator (let us call such cases CANnonical), while in (2) and (4) the subject appears on the left of the subordinator (let’s call those cases FRONTed subject). In (1) and (4) the subject of the subordinate clause and the subject of the matrix clause are disjoint (different, incompatible personal pronouns), while in (2) and (3) the subject of the matrix clause is elided, giving rise to a reading where both subjects are coreferential.

- (1) Ruguo wo qu Beijing, ta jiu buneng qu.
if I go Beijing he then not can go
‘If I go to Beijing, then he cannot go there.’

- (2) Wo ruguo qu Beijing, jiuhui qu chi Beijing kaoya.
I if go Beijing then will go eat Beijing roast duck
'If I go to Beijing, I will eat Beijing roast duck there.'
- (3) Ruguo wo qu Beijing, jiuhui qu chi Beijing kaoya.
if I go Beijing then will go eat Beijing roast duck
'If I go to Beijing, I will eat Beijing roast duck there.'
- (4) Wo ruguo qu Beijing, ta jiu buneng qu.
I if go Beijing he then not can go
'If I go to Beijing, then he cannot go there.'

According to his idea, versions (1) (canonical order, disjoint reference) and (2) (fronted order, coreference) should be preferred over the two other cases: "forms 1 and 2 are much more common than forms 3 and 4" (Chao 1968)¹. This prediction is also supported by Xu (1995), who considers in addition to Chao's examples (with only pronouns involved) data with full nouns, and uses a corpus approach, while it is argued against in Chen (2016). Chen (2016) studied 25 conditional subordinators in corpus and found that it is not the case that form 2 exceeds form 3 and form 1 exceeds form 4 in frequency, but that form 1 exceeds form 3 and form 2 exceeds form 4 in number. The subtle difference means that in comprehension data, it is always the structure that is more prominent. Referential variation only happens when one has already recognized the structure. However, it is not known whether it is the same case with the production process.

Besides corpus study, there are other descriptions of the same phenomenon from various theoretical perspectives. The pro-drop principle by C.-T.J. Huang (1995, 1998) says that a pro must be identified by its closest subject. In the case of bi-clausal reference, a pro that the subjects 'weakly c-command' is better than one that subjects do not.

¹Although Chao has used examples with only first person and second person, we suppose it can generalize to the third person to avoid the forced disjoint reference reading. What's more, in his examples, there is no fronted subject in the second clause. We will take this variation as a variable later.

- (5) a. Zhangsan_i suiran meiyou kong, e_i haishi lai-le.
though no time still come-ASP
'Though Zhangsan_i had no time, [he]_i came nevertheless.'
- b. ei suiran meiyou kong, Zhangsan_i haishi lai-le.
though no time still come-ASP
'Though [he]_i had no time, Zhangsan_i came nevertheless.'
- c. suiran Zhangsan_i meiyou kong, tai haishi lai-le.
though no time still come-ASP
'Though Zhangsan_i had no time, he_i came nevertheless.'
- d. *suiran Zhangsan_i meiyou kong, e_i haishi lai-le.
though no time still come-ASP
'Though Zhangsan_i had no time, he_i came nevertheless.'
- (C.-T.J. Huang 1982: 372-374)

Thus, (5a, 5b, 5c) are good cases while (5d) is ungrammatical in government and binding sense. However, (5d) does claim its existence in corpus and usage data. So such a prediction does not conform to real-life usages. What we can see from this analysis is that being licensed is important.

Tsao Feng-fu's (1990) proposes a topic-raising account for cross-sentential co-reference. Note that this is not cast within the generative syntax framework. Following Li & Thompson (1976)'s topic definition, he has proposed that subjects in clauses are topics which can raise to the main topic unless there is contrastive stress, leaving others to be the secondary topic. His account is irrelevant to disjoint reference patterns and their frequencies but a description of co-reference patterns and their way of movement. What is important is that contrastive stress has directed our attention to the second clausal subject, which, if fronted, can be a written sign of the stress pattern and disjoint reference.

A neo-Gricean pragmatics approach following Foley & Van Valin's (1984:269) hierarchy (IRH)² by Y. Huang (1991) says overt pronoun and co-reference in the second clause is usually in complementary distribution in bi-clausal anaphora. The reason why causal sentences did not follow the original hierarchy of Q[quantity]>M[anner]>I[nformativeness] is that in this case, I>M because implicatures in the higher constructions could win out over implicatures in lower constructions, so as to cancel them (Gazdar 1979, Levinson 1983). So we can have both:

- (6) a. Lao Li yinwei bing le, suoyi Ø buneng lai
Lao Li because ill CRS so not can come
'Because Li₁ is ill, (he₁) cannot come.'
- b. Lao Li yinwei bing le, suoyi ta bu neng lai
Lao Li because ill CRS so 3SG not can come
'Because Li₁ is ill, he₁ cannot come.'
- (Y. Huang 1991:323)

This draws attention to subordinator semantics, though in other sentences overt pronouns and zero both occur for co-reference in subject front cases as well. Causal may have a stronger effect in that but make no exception.

Summing up the previous accounts, we can see several factors that might be at play in cross-sentential anaphora in Chinese: subordinator semantics, subject position in the first clause, reference pattern, subject in front of second subordinator, c-command, and type of subjects in both clauses (zero as a special case of the second clause). All of these except structural factors will be accounted for in our multi-factorial analysis in predicting structure per se. Though there are two directions in the literature, Chao (1968) and Xu (1995) claiming the influence of reference patterns on structures while Chen (2016) focused on the

factor of structure on reference pattern, we would like to sort out factors predicting structure: 1) Surface syntactic structure prediction is the practice in Bresnan et. al (2007); 2) Structure prediction is easier than the other direction.

Please note that as a preliminary study, unlike dative construction research, verb meaning is not focused on in this study for two reasons. The first reason is from the previous psychological study carried out by Xu Xiaodong et al. (2013, 2017). While proving that FRONTed Subject has a stronger influence on reference than CANonical structure and forward-biased verbs have stronger influences on coreference of the subject in matrix clause to subject or object in subordinating clause, in 2017 study, they pointed out that topic or FRONTed Subject influence on coreference overrides the verb-based implicit causality influence. The second one is that from a frequency perspective, biased verbs are not a large number in corpus. They are often more influential in designed experiments than in natural conversations or writings.

2 Predicting from Multiple Variables

Corpus approaches in Chinese have suggested that subordinator position and reference pattern are highly related (Xu 1995, Chen 2016) but it is difficult to see which is the independent variable. More factors in separated models for production and comprehension may serve as a better solution. For this study, from the production side, we use a model to predict the structure. Factors are chosen according to previous literature. CCL (Center for Chinese Linguistics) corpus (modern Chinese)³ is a large corpus with good balance. The sentences generated from corpus are produced by users in the past, which can provide frequency information, while our annotation is a comprehension process with high inter-annotator agreement.

² From Weakest to Strongest: action-action/unspecified; sequential actions/non-overlapping; sequential actions/overlapping; simultaneous actions; conditions; temporal adverbial clause; indirect discourse complements; direct perception complements; jussive; psych-action; modality; causative.

³ This study is relying on a silent corpus (totally written, not transcribed from oral language). Patterns under contrastive stress influence are without corpus investigation until now. It can be more rule-based. And this part is another good topic indeed, with focus and information structure.

2.1 Linguistic factors

2.1.1 Subordinators

25 subordinators are categorized into 5 groups considering their logical relations in previous literature of Huang & Liao (2002). The groups are translated as follows:

Sufficient Condition in Conditional	Zhiyao (as long as), Zhixv (as long as), Yidan (once)
Necessary Condition in Conditional	Zhiyou (only if), Weiyou (only if), Chufei (unless)
Unconditional in Conditional	Wulun (no matter what, how...), Bulun (no matter), Buguan (no matter), Renpin (no matter)
Consistency in Hypothetical	Ruguo(if), Jiaru (if), Jiashi, Jiaruo (if), Jiashe (if), Tangruo(if), Tangshi (if), Ruoshi (if), Yaoshi (if), Wanyi (just in case, if by any chance)
Opposition in Hypothetical	Jishi (even if), Jiusuan (even if), Zongshi(even if), Zongran (even if), Napa (even if)

Table 1. Logical Relations of Conditional Subordinators

Group labels are ZHIYOU (ZHIYOU, CHUFEI, WEIYOU); ZHIYAO (ZHIYAO, ZHIXU, YIDAN); RUGUO (RUGUO, JIARU, JIASHI, JIARUO, JIASHE, TANGRUO, RUOSHI, YAOSHI, TANGSHI, WANYI); WULUN (WULUN, RENPING, BUGUAN, BULUN); NAPA (NAPA, JISHI, JIUSUAN, ZONGSHI, ZONGRAN). Later on, ZHIYOU group is dropped because CHUFEI in Chinese contains two senses: only and except which resulted in variation within this large quantity group while the remaining ZHIYOU and WEIYOU groups too small to show their statistical significance in the study.

2.1.2 Genre

In this study, two genres are focused on, namely newspaper and literature. Both have been considered in previous studies as representative forms of written Chinese. By sampling 2000 sentences from both parts of different subordinators, we build a raw corpus and annotate the qualified cases.

2.1.3 Reference type and structure

According to Chao (1968), different reference types are highly relevant to structure: disjoint reference patterns appear mostly in CANnonical, in contrast, coreference patterns frequent in FRONTed subject. In a model which predicts structure, we would count in reference patterns and vice versa. In a model where reference patterns or structures are concerned, the factor itself is not counted.

2.1.4 Types of NP1 and NP2

NP1 refers to the kinds of words in the syntactic slot of clausal 1 subject or what one labels ‘topic’. Since it is a solid presence, we intentionally ignore zero pronoun group in NP1 because it is a contextual anaphoric phenomenon. NP2 is relevant to clausal 2 subjects. Altogether NP1 has 4 groups while NP2 has 5. Names and nouns are counted as a subgroup because they stand separate from pronouns. Judging from the previous literature, pronoun types are of more importance, so we have a more fine-grained classification. There are four groups. Personal pronouns are separated from other parts, with a special subgroup of wh-pronouns whose special status is highly emphasized in Chinese. Zero pronouns or what one labels as ‘pro’ is the fourth group because ‘null’ most often point to coreference patterns. Other pronouns are what is left after these groupings, which covers pronouns like ‘it’. The division of nominal aspects can vary according to different research purposes. We choose the present one for it satisfies our demand to a high degree.

2.2 The annotation method

We have annotated the variables: Name of subordinator/GENRE/Semantic group/Reference type/Structure/NP1/NP2.

Type of Nouns	Code	Structure and Reference	Code
Names and nouns	N	CONJ NP1	A
Personal Pronouns	Y	NP1 CONJ	B
Other Pronouns	O	Conjoint Reference	C
Zero Pronoun	Z	Disjoint Reference	D
Wh-pronouns	W		

Table 2. Annotation System: basic model

Two genres are selected: literature and newspaper which are representative of narrative text and formal text. For each subordinator in Huang & Liao (2002)⁴, we have generated 2000 sentences, half from literature and half from newspaper to annotate the qualified sentences. The categorization of nominal expressions is decided according to previous literature which focus more on the types of pronouns instead of types of nouns and names. So we have only one label N for the latter and four labels representing personal pronouns, wh-pronouns, zero pronouns, and other pronouns for the former.

The sentences that are included have these features:

a. The conditional clause is a full clause that has its subject; b. The conditional clause precedes the matrix clause; c. The subject of the conditional clause can be the nominal type of name, pronoun which acts as a subject if it is behind the conditional subordinator and a topic if sentence-initial; d. The second clause can have its subject or zero subject; e. It does not embed another subordinate clause initiated by another conjunction. Similar to Xu's study (1995), the intersection of these rules is the type of sentence to be targeted.

3 Logistic Regression Model

In this part, we provide our model for predicting fronted structures. Genre 2 is newspaper; name_comb1 is Zhiyao Group; name_comb 2 is Ruguo Group; name_comb 3 is Wulun group; name_comb 4 is Napa group. Below is the formula and result summary. Then different factors and their influences on the structure are illustrated in figures from 3.2 to 3.6. Examples are listed to facilitate understanding.

3.1 Model A: Predicting Fronted Structures

glm(formula = structpattern ~ genre + refpattern + NP1 + NP2 +
name_comb2, family = binomial, data = d)

⁴This system of conditional subordinators is better than others in having a larger amount and being stated in an authoritative grammar textbook.

Deviance Residuals Summary:

	Min	1Q	Median	3Q	Max
	-1.8166	-0.7167	-0.5447	0.8593	2.5221
Predictor	Coefficients	Estimate	z value	Pr (> z)	
		Std. Error			
(Intercept)	-0.05818	0.09704	-0.600	0.548793	
genre2-1	-0.43456	0.06011	-7.229	4.86e-13	***
refpatternD-1	2.5409	0.07365	-17.029	< 2e-16	***
NP1O	-0.98736	0.23110	-4.272	1.93e-05	***
NP1W	-0.99342	0.21772	-4.563	5.05e-06	***
NP1Y	-0.80053	0.06256	-12.795	< 2e-16	***
NP2O	-0.20494	0.15689	-1.306	0.191477	
NP2W	0.96331	0.35091	2.745	0.006048	**
NP2Y	-0.31529	0.08704	-3.622	0.000292	***
NP2Z	0.60112	0.08923	6.737	1.62e-11	***
name_comb22-1	-0.97640	0.07320	-13.339	< 2e-16	***
name_comb23-2	-0.12278	0.09497	-1.293	0.196076	
name_comb24-3	0.46849	0.10766	4.352	1.35e-05	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Table 3. Result summary: coefficient estimates, estimate standard errors, z vlaue and significance level p for all predictors in model A.

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 9561.7 on 8007 degrees of freedom
Residual deviance: 8088.0 on 7995 degrees of freedom
(280 observations deleted due to missingness)
AIC: 8114

The model predicts structure B (fronted). Other things being equal:

- 1) newspaper is unlikely to favour structure B than literature;
- 2) coreference pattern favours structure B than disjoint reference pattern;
- 3) for NP1, when NP1 is a wh-pronoun, it is more likely to be structure A than when NP1 is N;
- 4) when NP1 is a personal pronoun, it is more likely to be structure A than when NP1 is N;
- 5) for NP2, personal pronoun as NP2 favours structure A than N as NP2; zero pronoun as NP2 favours structure B than N as NP2;
- 6) for the semantics of subordinators, Ruguo group (Consistency) is more unlikely to favour structure B than Zhiyao group (Sufficient); Wulun (Unconditional) is less likely to favour structure B than Ruguo group ; Napa group (Opposition) favours structure B than Wulun group.

3.2 Genre

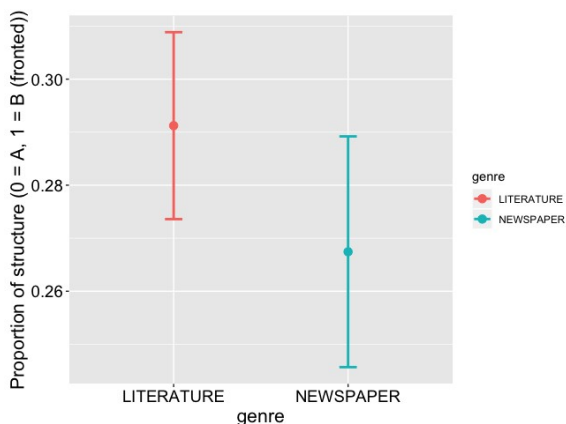


Figure 1. Effects of genre on structures

Literature compared with newspaper resulted in coefficient (-0.43456), z value (0.06011) and pr (4.86e-13), when predicting fronted pattern. It is possible that although both forms are formal, newspaper is more formal with fewer topics while literature is full of names of characters which needs to be in the front of sentences. This is with significance.

For example, sentence (8) from the literature genre is a FRONTed with coreference while sentence (7) from the newspaper, if we take the part after the comma, is a CANonical with coreference. In data, the newspaper sentences go less likely with fronted structure.

(7) 所以，小伙子 and 姑娘们，[如果] 你们患同样症状，不要悲观失望。【文件名:\当代\报刊\读者\读者（合订本）.txt】
so boys and girls if you suffer same symptom no want sad disappointed
'So, boys and girls, if you suffer from the same symptom, don't lose heart.'

(8) 你比我大一岁，你[如果]不嫌弃，就认我做你的妹妹罢。【文件名:\当代\文学\大陆作家\巴金.txt】
you compare me big one year you if no despise then identify me be your sister BA
'You are one year older than me, if you don't mind, please accept me as your sister.'

3.3 Semantics of the subordinators

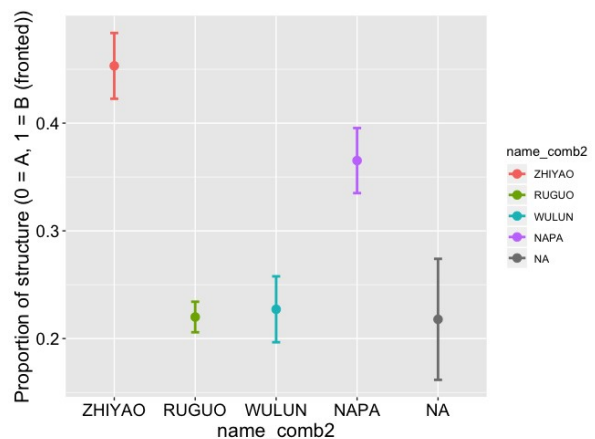


Figure 2. Effects of different logical groups on structures

From this figure, we can roughly conclude that one tendency shows that Consistency is less likely than Sufficient to predict fronted structure. In Sufficient Group, subject is usually fronted because the subordinators are mostly verb-like such as Zhiyao and Zhixv. Whether verb-like feature is the cause or the consequence of talking about sufficiency needs more research. Another tendency is that Unconditional is less likely to favour FRONTed than Opposition. It is possible that in Opposition, contrastive stress mentioned by Tsao (1990) is needed to emphasize different topics. So a subject raises to the beginning part to be focused.

3.4 NPs in Subject 1 Position

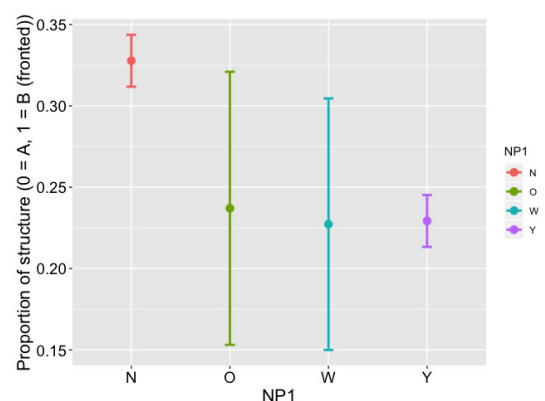


Figure 3. Effects of NPs in Subject 1 position on structures

When NP is contrasted with Other pronouns, it has a significant pr (1.93e-05), a coefficient of -0.98736 and a z (-4.272). It is more likely to predict fronted patterns than other pronouns. The same tendency also applies to the comparison between NP and Wh-pronouns with coefficient (-0.99342), z value (-4.563), and pr (5.05e-06), another significant effect. Still, NP is more likely than personal pronouns to favour FRONTed with a coefficient (-0.80053), z value (0.06256), and pr (2e-16). In all cases, topicality (Xu 1995) is supported because names, be they personal, objective, or event-related, are better topics than pronouns. Pronouns, on the other hand, indicate a textual continuation of topics. Here we list a wh-pronoun as NP1 sentence, two subjects in this sentence are the same, and the structure is a FRONTed.

(9) 任何人[如果] 不发扬吃苦精神, 不努力拼搏, 就有被社会淘汰的危险。【文件名:\当代\报刊\人民日报\1993年人民日报\10月份.txt】

anyone if no develop eat hard spirit no struggle strive then have BEI-PASSociety discard DE danger

‘Anyone who does not strive hard will be left behind by society.’

tend to go with FRONTed with a coefficient (0.60112), z value (6.737) and pr (1.62e-11). Such tendency is in line with the accessibility trend in Ariel (1988). Here are four examples:

(10) 北京是文化中心, 教育[如果]搞不好, 区长、区委书记, 县长、县委书记就没当好, 市长就没当好。【文件名:\当代\报刊\人民日报\1994年人民日报\第1季度.txt】

Beijing be culture center education if do no good district mayor district party secretary county mayor county secretary then no be good mayor then no be good

‘Beijing is the centre of culture. If education is not promoted well, district mayor, district party secretary, county mayor, county secretary and mayor are not working satisfactorily.’

(11) 他妻子冷静地告诉他, 公寓是她以她的名义从学校租来的。[如果]他不走的话, 她只好叫警察了。【文件名:\当代\报刊\作家文摘\1996\1996B.TXT】

he wife calmly tell he apartment be she use her name from school rent DE

‘His wife told him calmly, this apartment is rented from school in her name. If he does not leave, she has to call the police.’

3.5 NP in Subject 2 Position

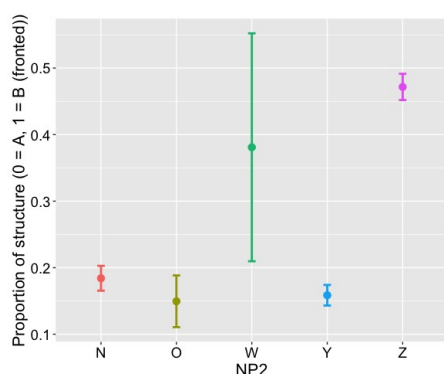


Figure 4. Effects of NP in Subject 2 Position on structures

When NP2 and personal pronouns are compared, we arrived at a coefficient of (-0.31529), z value (-3.622) and pr (0.000292) which indicates that personal pronouns are more likely to favour unfronted pattern. On the contrary, zero pronouns

(12) 我有个朋友喜欢剪报。我的名字不常见报, [如果]他在报上见到了我的名字, 就剪下来寄给我。【文件名:\当代\报刊\读者\读者(合订本).txt】

I have CL friend like cut paper my name no appear paper if he in paper see LE my name then cut down send me

‘I have a friend who likes to cut from newspapers. My name does not show up often in newspapers. If he saw my name there, he cut this part down and sent it to me.’

(13) [如果] 教师不合格或不努力工作, 校方会毫不客气地解聘他。【文件名:\当代\报刊\1994年报刊精选\01.txt】

if teacher no qualify or no hard work school will relentlessly fire he

‘If a teacher is disqualifying or not hardworking, the school will fire him relentlessly.’

Sentence (10) is a FRONTed with disjoint reference while sentence (11) is CANonical with disjoint reference. The statistical result points to the fact that if one uses NP in NP2 position, it is more likely to use FRONTed, while personal pronouns used in NP2 position are less likely to favour FRONTed. In sentence (12), zero is used in the second clausal subject and this sentence is with fronted structure and coreference. By contrast, sentence (13) is with an NP in clausal 2 subject position as disjoint reference, the structure is CANonical. Sentences like (12) will be more than sentences like (13).

3.6 Reference Pattern

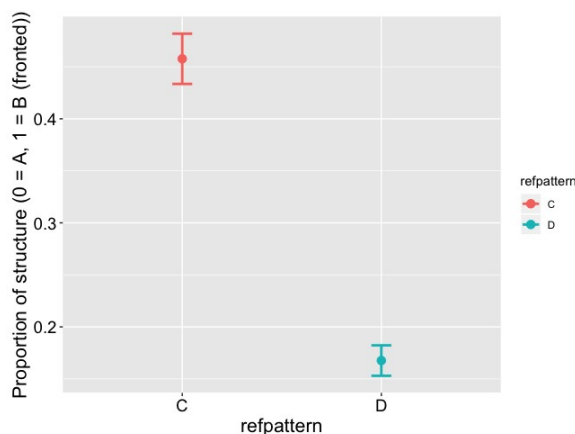


Figure 5. Effects of reference patterns on structures

When conjoint reference pattern is contrasted with disjoint reference pattern, it has a significant p ($2e-16$), a coefficient of -1.25409 and a z value (-17.029). The strong co-relevance of coreference with FRONTed and disjoint reference with CANonical supports Chao (1968). We exemplify in (14) and (15):

(14) 你要加意小心，伺候他们；他们要什么，你就给什么，——他们[如果]要你的女儿，你就立刻打轿子！【文件名:\当代\文学\大陆作家\欧阳山】

you want more small heart serve them they want what you give what they if want your daughter you then immediately beat sedan

‘Watch out when serving them. Give them whatever they want. If they want your daughter, you ask for a sedan immediately.’

(15) 第二，左帅大人只是朝廷一个总兵，我们张帅[如果]投降，也只能向朝廷投降，由杨阁部代朝廷受降。【文件名:\当代\文学\大陆作家\姚雪垠】

the two Zuo General only be imperial court one CL general our Zhang General if surrender too only can toward imperial court surrender by Yang Cabinet Minister represent imperial court accept surrender

‘Number two, General Zuo is only a general of the imperial court. Our General Zhang will only surrender to the emperor if he surrenders, so it must be Cabinet Minister Yang who accepts this as a representative.’

For disjoint reference, one see sentences like (14). In (14), though the structure is fronted, it is less likely to appear than in (15), which has conjoint reference. This illustrates the tendency in this comparison.

4 General Discussion

This research investigates the effects of multiple linguistic factors mentioned in Chao (1968), Huang C-T. J. (1982), Tsao (1990), Huang Y. (1991), Xu (1995) and Chen (2012), etc. on the Conditional part of Chinese inter-clausal subject anaphora in written corpus. Fixed-effects logistic regression supports the high co-relevance between structure and reference and generated models to predict structure patterns. It is revealed that in predicting structure, zero pronouns in the second clausal subject contribute the most. Topicality by Xu (1995) is proved in our discussion of NP1 in prediction. Accessibility by Ariel (1988) is evidenced in our discussion of NP2 in subject2 position. The tendency that Opposition group favours fronted structure than Unconditional supports Tsao’s contrastive stress (1990). The factors of genre, semantics of subordinators deserve more explanations. Further studies are also needed to improve the accuracy of the model in the computational approach and test the relevance of each factor to the final result in the psycholinguistic approach. Other subordinator groups such as causal, concessive, transitional, etc. are worth future researching.

Acknowledgments

We would like to extend our sincere thanks to Yair Haendler for his help in coding the models and reviewers for their advices. This work is supported by LLF, Paris University; Innovative Research Team of Shanghai International Studies University (2020114050); Chinese National Major Social Science Project (17ZDA027); Chinese National Youth Social Science Project (17CYY017); the Fundamental Research Funds for the Central Universities in China (KY01X0222017095).

References

- Joan Bresnan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. Predicting the Dative Alternation. In *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts. Amsterdam: Royal Netherlands Academy of Science, pp. 69--94.
- CCL. (n.d.). In Center for Chinese Linguistics Corpus online. Retrieved from http://ccl.pku.edu.cn:8080/ccl_corpus/.
- Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: UC Press.
- Chen, S. T. 2012. *Inter-clausal Anaphora in Chinese: The Case of Conditionals [D]*. Doctoral Dissertation. Shanghai International Studies University.
- Chen, Y. (2016). *The Influence of Information Structure on Pronoun Resolution*. M.A. thesis. Shanghai International Studies University.
- Foley, W. and van Valin, R. D. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Gazdar, G. (1979). *Pragmatics: implicature, presupposition and logical form*. London: Academic Press.
- Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Doctoral dissertation, MIT.
- (1998). *Logical relations in Chinese and the theory of grammar*. New York: Garland.
- Huang, B.R. and Liang X.D. (2002). *Modern Chinese (3rd. Ed.)*. Volume II. Beijing: High Education Press.
- Huang, Y. (1991). A neo-Gricean pragmatic theory of anaphora. *Journal of Linguistics*, (27:2): 301-335.
- Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Levinson, S. C. *Pragmatics*. (1983). Cambridge: Cambridge University Press.
- Levinson, S. C. (1987a). Minimization and conversational inference. In Verschueren, J. & Bertuccelli-Papi, M. (Eds). *The Pragmatic Perspective* (pp.61-129). Amsterdam: John Benjamins,
- Levinson, S. C. (1987b). Pragmatics and the grammar of anaphora: a partial pragmatic reduction of binding and control phenomena. *JL* 23, (23): 379-434.
- Li, C. N. and Thompson, S. A. Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and Topic* (pp. 457-490). New York: Academic Press, 1976.
- Tsao, F.-F. (1990). *Sentence and Clause Structure in Chinese: A Functional Perspective*. Taipei: Student Book Co.
- Xu, X.D., Ni C.B. & Chen L.J. (2013). The influence of topic structure and verb-based implicit causality on pronoun resolution in Mandarin Chinese: Evidence from sentence production and comprehension. *Modern Foreign Languages*, (4) : 331-339.
- Xu, X.D., Chen L.J. & Ni C.B. (2017). How is pronoun resolution modulated by topic structures and verb-based implicit causality in Mandarin Chinese? An ERP investigation. *Foreign Language Teaching and Research*, (5):323-334.
- Xu, Y.L. (1995). *Resolving third-person anaphora in Chinese texts: Towards a functional-pragmatic model*. Doctoral dissertation. The Hong Kong Polytechnic University.
- (2003). Inter-clausal anaphora in Chinese complex sentences. *Contemporary Linguistics*, (2):97-107.
- (2004). *Towards a Functional-pragmatic Model of Discourse Anaphora Resolution: A Study Based on a Data-base Driven Analysis of Chinese Folk Stories and Newspaper Articles*. Shanghai : Shanghai Foreign Language Education Press.