Chinese Kinship Semantic Structure and Annotation Scheme

Yufang Qian[1] and Scott Songlin Piao[2]

1. Introduction

One of the most fruitful approaches to the study of meaning is to regard it as conceptual in nature (Cruse, 2004: Chapter 7). Cruse argues that "concepts are vital to the efficient functioning of human cognition". We need to organize the words carrying complex meanings into certain conceptual structure in order to efficiently access it. Such an approach also provides a practical way of constructing semantic lexicons, which in turn, can be used for annotating semantic information in corpora.

One of the major semantic lexicon annotation schemes in existence is the Lancaster USAS Semantic taxonomy, which has been used for constructing semantic lexicons for semantic taggers for English, Finnish and Russian (Wilson and Rayson, 1993, Rayson et al., 2004; Piao et al., 2005; Lofburg et al, 2005; Serge et al., 2006). Since the semantic scheme has been successfully ported from English to the other two languages, it can possibly be ported to other languages. However, it is not straightforward to port a semantic annotation scheme between different languages. In two previous projects (BENEDICT and ASSIST carried out at Lancaster), where the annotation scheme was applied to Finnish and Russian, it was found that different languages may have discrepancies between their semantic conceptual structures due to the diversity of cultures. Such cultural difference occurs between Chinese and English as well. A typical example of such differences is the kinship semantic structure between Chinese and English languages.

Over the entire Chinese history, family has been in the central place in people's lives. It is linguistically manifested by the finely grained semantic structure of kinship terms, a distinct feature compared to English counterparts. In Chinese language, on the one hand, there are two sets of titles for addressing corresponding relatives derived from the mother and father sides, or **affinal** relationship, which reflect relatives' precise positions in a family structure, e.g. *tang ge vs. biao ge*. Here *tang ge* refers to an elder male cousin of paternal side while *biao ge* means an elder male cousin of maternal side. In English, on the other hand, these two sets of titles are generally merged together into a single set, only reflecting the hierarchy of generation and gender. Sometimes even the gender is ignored. For example, the *(first) cousin* corresponding to both *tang ge* and *biao jie* (elder female cousin of maternal side) in

[1] Department of Linguistics and English language, Lancaster University
   e-mail: y.qian@lancaster.ac.uk
[2] Text Mining Group, School of Computer Science, University of Manchester
   e-mail: scott.piao@manchester.ac.uk

Chinese only reflects generation hierarchy. As this example shows, the semantic conceptual structure of kinship in different languages is an interesting issue for comparative studies.

In this paper, we discuss the issue of kinship studies from the corpus linguistics perspective. We identify the need and importance of encoding kinship semantic information in Chinese corpus resources in order to support such studies in corpus linguistics. We propose a semantic field framework of Chinese kinship and design an annotation scheme for the Chinese kinship annotation in corpora. We also present a case study regarding the annotation in Chinese corpora.

2. Previous related Work

Traditionally, the study of kinship has been an important topic for both the Western and Chinese cultural studies. Methodologically, studies on Chinese kinship system tend to be descriptive and analytic. The literature in this area can be traced back to Morgan's work (1871). Improving this work, Chen and Shryock (1932) used a variety of dictionary materials to study Chinese kinship system. Feng (1937) augmented the historical material used by Chen and Shryock. Kroeber (1933) adopted a more dynamic approach to associate the development of nomenclature with cultural transition. Chao (1956) devised a multiple chart as an excellent guide for componential analysis.

Different from the earlier descriptive approaches applied in the studies of Chinese kinship system, McCOY (1970) undertook a formal semantic analysis of how people are addressed and referenced in the Chinese culture. McCOY found that much of tradition still remains evidently in contemporary Chinese society, especially in the overseas Chinese communities. Baker (1979: preface) tried to access the kinship in action and he was able to indulge unrepentantly in 'chronological sleight-of-hand'. He successfully unveiled the features of a traditional scene through twentieth-century fieldwork studies, while using the historical facts to back up his analysis of modern social phenomena.

Since the 1980s, extending the previous studies on Chinese kinship terms per se, some scholars pursued the issue from a wider angle of perspectives. Watson (1982) approached the Chinese kinship as a social institution or organization, while Hirschfeld (1986) examined the issue "how and what kinship terms mean" (p. 236) from the perspective of individualized cognitive psychology. Huang and Jia (2000) classified the Chinese kinship terms by affinal, consanguineal and siblings relationship in order to make a semantic comparison between Chinese and Western kinship terms. They criticized that the past studies on Chinese kinship were not conducted from the communication perspective, and they pushed forward the study by investigating the cultural connotations and communicative functions of Chinese Kinship Terms. Qiu (2003) used the affinal and consanguineal kinship system to contrast Chinese and English languages and addressing of kinship terms. She argued that the cultural differences lie in the family structure, consanguinity, patriarch

concept, cultural adoption and courtesy principle.

From the above review, we observe that the Chinese kinship terms, embodying to a large extent the feature of Chinese culture and language, deserve much more attention and study, particularly in the corpus linguistics community. Sapir (1949: 162; cf Saeed, 2001: 42) argues: "we may think of language as the symbolic guide to culture". Following his argument, we contend that Chinese kinship terms are an important guide to Chinese culture. Indeed, they have had an important role in constituting and maintaining Chinese culture throughout Chinese history, and still enjoy the similar importance in people's daily life in Chinese society today. They form a part of the unique feature which distinguishes Chinese language and culture from others.

Despite the earlier studies from various perspectives mentioned above, as far as we know, there is no published corpus-based study on Chinese kinship. Based on Lancaster USAS Semantic taxonomy, we investigated the feasibility of porting the English kinship taxonomy into Chinese language. In the following sections, we discuss a Chinese kinship semantic structure.

3. Chinese Kinship Semantic Structure

In this section, we discuss Chinese kinship structure in comparison with English equivalents and propose a Chinese kinship annotation scheme.

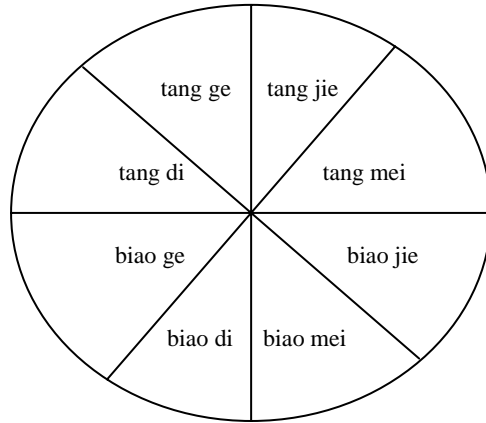## 3.1. Features of Chinese kinship conceptual structure

According to the linguistic relativity hypothesis of Sapir – Wharf (1929), every language has its own way of encoding the real world concepts, which bears a deep relation with national culture (Sidorov *et al.* 1999). Although every language is capable of denoting every kinship relationship conceivable, sometimes we can not find direct equivalent kin terms across languages. Mirroring the cultural importance of family relationships in Chinese society, Chinese language contains numerous specific kinship terms which facilitate a fine classification of the relationship among family members. Indeed, Chinese kinship system has its roots deep in Chinese culture. It is a product of centuries of an agricultural society, a family-centered economy, and a strictly hierarchical family system (Huang and Jia, 2000). Many Chinese kinship terms do not have their matching terms in English; they have to be paraphrased in translation. Below we discuss some prominent features of Chinese kinship structure in comparison to English counterpart from four perspectives (refer to Collins English Dictionary 5th edition 2000):

- Consanguineal relation: Relationship by blood,
- Affinal relation: Relationship by marriage,

- Lineal relation: in a direct line of descent from an ancestor,
- Collateral relation: Descended from a common ancestor but through different lines.

With respect to the consanguineal line, Chinese has a set of kin terms for an extensive range of hierarchical order. For example, for the male ancestors and offspring, there are terms such as *zeng zu fu* (great grandfather), *zu fu* (grandfather), *er zi* (son), *sun zi* (grandson), *zeng sun* (great grandson) etc. A rich vocabulary denoting various affinal relations is another prominent feature of Chinese language. For example, the relations subsisting between a husband and his wife's relatives are marked with affixes *yue*, *jiu* or *yi*, such as yue-fu (wife's father), qi-jiu (wife's brother), qi yi (wife's sister) etc. Relatively, the relationship derived from husband's relatives is less marked. An interesting finding of Huang and Jia (1998) is about the Chinese kinship terms for maternal grandfather and grandmother, *wai zu fu* (mother's father) and *wai zu mu* (mother's mother). The maternal marker *wai* in Chinese literally means "outside," which implies a marginal social status ascribed to the maternal relations in the Chinese tradition. On the other side, English lacks such a finely-grained terminological mechanism for distinguishing the complex relations derived from a marriage. This may reflect the English culture in which the differences between consanguineal and affinal relations are of less significance. Consequently, these Chinese terms have to be paraphrased in English in translations.

Furthermore, in contrast to English kinship terms, which focus on direct lineal relations such as father, mother, etc., Chinese has a rich set of kinship terms in this regard, precisely locating an extensive range of collateral relations. Each of these terms has rather narrow sense, pinpointing a small area of the kinship semantic space. For instance, an "uncle" in English refers to a brother or brother-in-law of either parent; Chinese has as many as five terms for the different aspects of *uncle* sense: *bo fu* (father's elder brother), *shu fu* (father's younger brother), *gu fu* (husband of father's sister), *jiu fu* (mother's brother) and *yi fu* (husband of mother's sister), each of which denotes different relation with ego. Similarly, the concept of *first cousin* in English is broken down into eight sub-concepts in Chinese, as shown in Fig. 1. Each of the sub-concepts accurately represents a place in the collateral hierarchical structure with respect to age, gender and the maternal/paternal origin. Here the consanguineal and affinal relations are labeled with the prefixes *tang* and *biao* respectively. Such a complex kinship terminology reflects the importance of extensive network of relations among people in the Chinese society, which entails maintaining close relationship even among those who are remotely related in a family tree.

Figure 1: Chinese conceptual structure corresponding to "first cousin"

Moreover, there are often several propositional synonyms in Chinese language referring to a single kinship concept. This observation has an important implication in a number of areas in which text analysis is undertaken. For instance, there are more than five Chinese kinship terms referring to *father* (e.g. 爸爸/*ba ba, /父亲 fu qin, 老爸/lao ba*，爹/*die* and 老子/*lao zi*). The differences between them mainly lie in the different non-propositional meanings, such as implying the degree of intimation between a father and his offspring, colloquial usage etc. Although similar kinship synonyms exist in English, too, e.g. *mother*, *mum* and *mummy*, Chinese language contains significantly richer vocabulary in this regard.

It should be noted that the Chinese kinship system and terminology keep evolving today. Traditionally, the Chinese kinship can be distinguished by gender (e.g. difference between a brother and a sister), between generations (e.g. difference between a child and a parent), as well as between relatives by blood and marriage. With the evolution of composition and conception of families in contemporary China, people have started paying more and more attention to the single child family, single parent families, and the families headed by same-sex couples or by heterosexual partners. For example, since the implementation of China's 'Single Child Policy' in 1980s', most of the children grown up since then in China have no siblings. Only when they are involved in a remarried family, they may have *half-brother*, *half-sister*, *step-brother* or *step-sister*.

In sum, Chinese has a rich, fine-grained and descriptive kinship terminology. Each term reflects a dimension of relations, gender and generation. In comparison, English kinship terms are more coarsely-grained. The more aliener the relation is, the vaguer the kinship terms are (Qiu 2003). As mentioned previously, quite often we can not find straightforward equivalent terms between these two languages. In particular, some English kinship terms has to be disambiguated in terms of age, gender etc. in order to determine their correct Chinese translations. Such a wide semantic discrepancy between Chinese and English kinship vocabularies presents a tough challenge in porting the existing English kinship annotation scheme to Chinese language. In the following section, we propose a Chinese kinship semantic model, which organises the Chinese kinship terms in a logical structure. Such a model would

help us to bridge between Chinese and kinship terms and their counterpart in English.

3.2. Chinese Kinship scheme

In order to accommodate the unique features of Chinese kinship, we propose an annotation scheme with a hierarchical structure. We extended and refined a traditional kinship scheme (Huang and Jia 2000) in order to obtain a wider coverage of the kinship semantic fields. At the top level, the scheme has five semantic categories: Parents, Siblings, Marital relations, Offspring and Partnership (see Appendix 1). This semantic tree is a projection of the Chinese kinship relationships from the perspective 'ego'.

A tagset is used to denote the semantic categories. In addition, some auxiliary codes are used to denote secondary sub-categories – a borrowed idea from the USAS encoding scheme. For example, letters *m* and *f* denote male and female respectively (husband (m) and wife (f)), *e* and *y* denote the elder and younger relatives respectively (elder brother (e), younger brother (y)), etc. The gender differences between the similar positions in the taxonomy tree, such as *nephew* versus *niece*, are denoted with *m*, *f*, *e* and *y*, or with their combinations (see tags *S4.2.1.3mfey* and *S4.2.2.3mfey*). Other similar auxiliary codes include:

h ~ half brother/sister,
o ~ orphan/widow/widower,
s ~ step father/mother,
t ~ foster son/daughter,
x ~ ex wife/husband,
- (minus) ~ engaged,
g ~ god father.

As mentioned earlier, the division of the five categories is based on the egocentric network of family relationship. We use the code S4 for ego, S4.1 for Parents, S4.2 for Sibling, S4.3 for Marital relatives, S4.4 for Offspring and S4.5 for Partnership. Each category has its own sub-divisions. As can be seen in the category of *Parents*, there are two sub-categories, *Paternal side* and *Maternal side*, which show the relationship between 'ego' and paternal and maternal sides respectively (see Table 1 and 2). In each column of the tables, three generations of kinship terms are put hierarchically with each couple in the same hierarchy. *Step parents*, *god parents* and *foster parents* share the same position as biological parents in our kinship scheme (see Table 1).

In the category *Siblings*, we put *half brother*, *half sister* under the sibling category (see appendix 1: S4.2.1.1, S4.2.1.2, S4.2.2.1, S4.2.2.2). Meanwhile, we add *ex-wife*, *ex-husband*, *widow* and *widower* to the conjugal relation part (See appendix 1: S4.3.1.1, S4.3.2.1). *Orphan* is put under the Offspring category (see S4.4.1.1, S4.4.2.1).

Paternal Side	
fu qin (father), ji fu (step father), jiao fu (god father), yang fu (foster father)	
zu fu (paternal grandfather)	zu mu (paternal grandmother)

zeng zu fu(father' father's father)	zeng zu mu(father' father's mother)
bo fu (father's older brother)	bo mu (father's older brother's wife)
gu mu (paternal aunt)	gu mu (husband of paternal aunt)
shu fu (father's younger brother)	shen mu (father's younger brother's wife)
tang ge (paternal elder male cousin)	tang sao (paternal elder male cousin's wife)
tang di (paternal younger male cousin)	tang di xi (paternal younger male cousin's wife)
tang jie (paternal older female cousin)	tang jie fu (paternal older female cousin's husband)
tang mei (paternal younger female cousin)	tang mei fu (paternal younger female cousin's husband)

Table 1: Paternal side

Maternal Side	
mu qin (mother), ji mu step mother), jiao mu(god mother), yang mu (foster mother)	
wai zu fu (maternal grandfather)	wai zu fu (maternal grandmother)
zeng wai zu fu (mother' father's father)	zeng wai zu fu (mother' father's mother)
jiu fu (mother's brother)	jiu mu (mother's brother's wife)
yi mu (maternal aunt)	yi fu (husband of mother's sister)
biao ge (maternal elder male cousin)	biao sao (maternal elder male cousin's wife)
biao di (maternal younger male cousin)	biao di xi (paternal younger male cousin's wife)
biao jie (maternal older female cousin)	biao jie fu (maternal older female cousin's husband)
biao mei (maternal younger female cousin)	biao mei fu (maternal younger female cousin's husband)

Table 2: Maternal side

One of the unique features of this scheme is that we consider both the heterosexual and homosexual cohabit partnerships as family forms (see S4.5.1.1, S4.5.2.1). Such family forms are still rare in contemporary Chinese society. However, with the rapid economic development and the advance of the women's liberty, both in terms of social status and economic independence, more and more people choose the non-matrimony cohabits relationships (Li, 2005). Similarly, the homosexual partnerships also start to draw attention in China. Departing from the traditional scheme of family, we included the cohabit partnership in our Chinese Kinship Semantic Structure.

In addition to the Chinese kinship scheme, we also collected regularly-used Chinese kinship terms for each of the categories in the scheme to form a mini-lexicon of this domain, as described below.

3.3. Chinese kinship semantic lexicon

As already mentioned, Chinese language has a rich kinship vocabulary. For a single kinship title, there often exist several terms denoting it. For example, we found eleven commonly used terms for the kinship category of *father* and nine terms for *mother*. Obviously it is difficult to exhaustively collect every Chinese kinship term occurred

over the Chinese history, however, we made use of various resources including online resources and corpora in an effort to collect as many as possible, particularly those frequently used in daily life in modern Chinese society.

So far, we have collected 212 related terms and classified them according to our annotation scheme. Each term is manually assigned with a kinship tag from the scheme. Table 3 shows the distribution of these terms for the main categories. A prominent feature of this distribution table is the widely different numbers of the terms under different categories. As shown in the table, the biggest category is *Parent →Father's line*, tagged with *S4.1.1*, containing fifty-five terms. It reflects the prominent status of the paternity line in the kinship relationship in the Chinese social life. The second category in the rank, as can be expected, is the *Parent → Mother line*, corresponding to the importance of mother's status in family life. If we rank all the categories by the size of terms, the remaining categories can be ordered as follows: *wife's line*, *Homosexual line*, *Husband's line*, *Sister's line, Brother's line*, *Son's line*, *Daughter's line*, *Heterosexual line*. We are well aware that our collection of the terms is by no means complete, and the distribution of the terms may not precisely reflect the whole picture of the usage of Chinese kinship vocabulary. However, we contend that such information can provide important information for the studies on Chinese kinship structure and vocabulary.

Main Category	Sub-categories	Number of terms
S4.1: Parent line	S4.1.1: Father's line	55
	S4.1.2: Mother's Line	39
S4.2: Brother and Sister's Line	S4.2.1: Brother's Line	16
	S4.2.2: Sister's Line	16
S4.3: Marriage line	S4.3.1: Husband's Line	17
	S4.3.2: Wife's Line	30
S4.4 : Children line	S4.4.1: Son's Line	11
	S4.4.2: Daughter's Line	8
S4.5: Partnership	S4.5.1: Heterosexual Line	2
	S4.5.2: Homosexual Line	18
Total		212

Table 3: Terms in main Chinese kinship categories

An interesting finding from this lexicon is the symmetrical distribution of the terms under the *Brother's* and S*ister's Lin*e*s*. We have collected sixteen terms for each of them, and they are almost perfectly parallel (refer to the lexicon and scheme). For every kin term in the *Brother's Line*, we can find a corresponding term in the other line, perfectly distinguishing between the relatives derived from these two relationships.

Another interesting finding from the terms' distribution concerns the cohabit partnership categories. So far, we have found only two commonly used terms denoting heterosexual cohabit partnership, while we found eighteen terms for homosexual partnership. The lack of the formal type of terms may reflect the social situation of China where the majority of Heterosexual relations are constrained within

marriages whereas cohabit partnerships outside marriages are still relatively rare. On the other hand, the considerable size of the homosexual-related terms can possibly be explained by the marginal social status of such relationships in China today. In such a situation, people in the homosexual community tend to use metaphorical terms, such as *tong zhi* (meaning comrade), *tu zi* (meaning *rabbit*), *bo li* (meaning *glass*), etc. Such terms are used for communicating among the members of this community exclusively as a slang. In such a circumstance, the terms are unlikely standardised across different regional homosexual communities, resulting in a wide diversity of the terms.

Although there is still much room for improvement, our current Chinese kinship lexicon, as it stands, already provides a wide coverage of the Chinese kinship vocabulary being used in modern China, and it can provide a basis for kinship annotation in corpora. We will make continuous efforts to increase the coverage of this lexicon.

4. Case study

In this section, we present our case study based on the Lancaster Corpus of Mandarin Chinese (LCMC) (Xiao *et al.* 2004), a Chinese balanced corpus widely used in corpus linguistics community. Through this case study, we will exam the applicapability of our Chinese kinship classification scheme for the semantic corpus annotation. We examine the usage of some kin terms in order to investigate the potential problems anticipated for the practical semantic annotation.

In practice, we chose four kin terms, *zhang fu, qi zi*, *ai ren* and *shu shu*, and extracted concordances of them using LCMC. As a result, we collected 258 occurrences of *zhang fu* in LCMC. Our analysis revealed that, in most of the concordances, the term *zhang fu* refers to *husband*. There were only six exceptional cases, where the word *da* (big) precedes the key term and the combined *Da zhang fu* form a compound term which means "a man of character", as shown below.

Example 1）不错，	大丈夫	n 做事 敢 做 敢 当

Translation 1: Right, Da zhangfu (a man of character) should be responsible for what he has done.

Similarly, the concordance line of *qi zi* shows that, almost exclusively, it denotes the concept of *wife*. However, some kin terms denote more than one kinship concept. For example, the term a*i ren* can refer to either male spouse or female spouse, as shown by the concordances below. For such terms, we can only determine their true senses from their contexts.

Example 2）她 的	爱人	n ， 南浦 大桥 主桥 设计师 —— 张 介 望 因 操劳 过度 ， 在 大桥 建成 的 前 一个 月 倒下 了 ， 永远 地 走 了 。
Example 3）" 你 是 他	爱人	n 吗 ？ "

Example 4) "	爱人	n 之间 不能 谈 真理 ， 因为 真理 太 冷酷 了 。 "

Translation 2: Her husband, Zhang Jiewang, the designer of Nanpu Bridge, who was over exhausted, fell sick and passed away one month before the bridge was completed.

Translation 3: "Are you his wife?"

Translation 4: "There is no truth between husband and wife, because truth is always too cold."

There are other groups of kin terms which are also used for social addressing. For instance, in Chinese community, a man who is older than one's father and mother is generally addressed as *bo bo*, and a man who is younger than one's father and mother is generally addressed as *shu shu*. Similarly, a young female is generally called *a yi* by children. In northern China, she is also called *da ma*. The elderly people of grandfather's and grandmother's generation can also be called *ye ye* and *nai nai*. Let's see what the corpus says about *shu shu*:

Example 5) 在 所有	叔叔	n 中 ， 五 叔 和 我家 走 的 最 勤 。
Example 6) 可是 ， 文化大革命 前 ， 王	叔叔	n 一家 就 从 中南海 迁 出去 了 ， 那个 院子 一时 冷清 了 下来 。
Example 7) 孙子 上 了 汽车 ， 说 ： " 爷爷 ， 你 看 那位	叔叔	n 坐 在 那儿 睡着 了 ， 叫醒 他 吧 ， 别 感冒 了 。 "
Example 8) 还有 一 次 在 马路 上 拾 到 一块 手表 ， 交给 了 警察	叔叔	n 。

Translation 5: Among all uncles, the fifth uncle is the closest to my family.

Translation 6: But, before Cultural Revolution, Uncle Wang's family moved out of Zhong Nanhai. All at once that yard became desolate down.

Translation 7: The grandson boarded on the bus and said: "Granny, look at that uncle! He is sleeping. Wake him up so that he will not catch cold."

Translation 8: Another time, he picked up a watch on the road and handed it to Uncle Police.

In the above examples, the term *shu shu* in example 5) is used as a real kin term of *father's fifth younger brother*, while in examples 6) and 7) it is used as a non-relative addressing. In example 8), it is used as a social addressing term for policeman by youngsters, as the stereotype of a police is always a strong young man.

From the above examples, we can see that our proposed Chinese kinship scheme is applicable to the corpus data. It generally reflects the true Chinese kinship structure in reality. However, considering the numerous polysemeous Chinese kin terms, i.e. different usages of kinship terms in different speech community and with different communicative function, we can not apply the kinship scheme and lexicon directly for semantic annotation. We anticipate that word sense disambiguation will be a critical work in identifying the true meanings of the kin terms in various contexts.
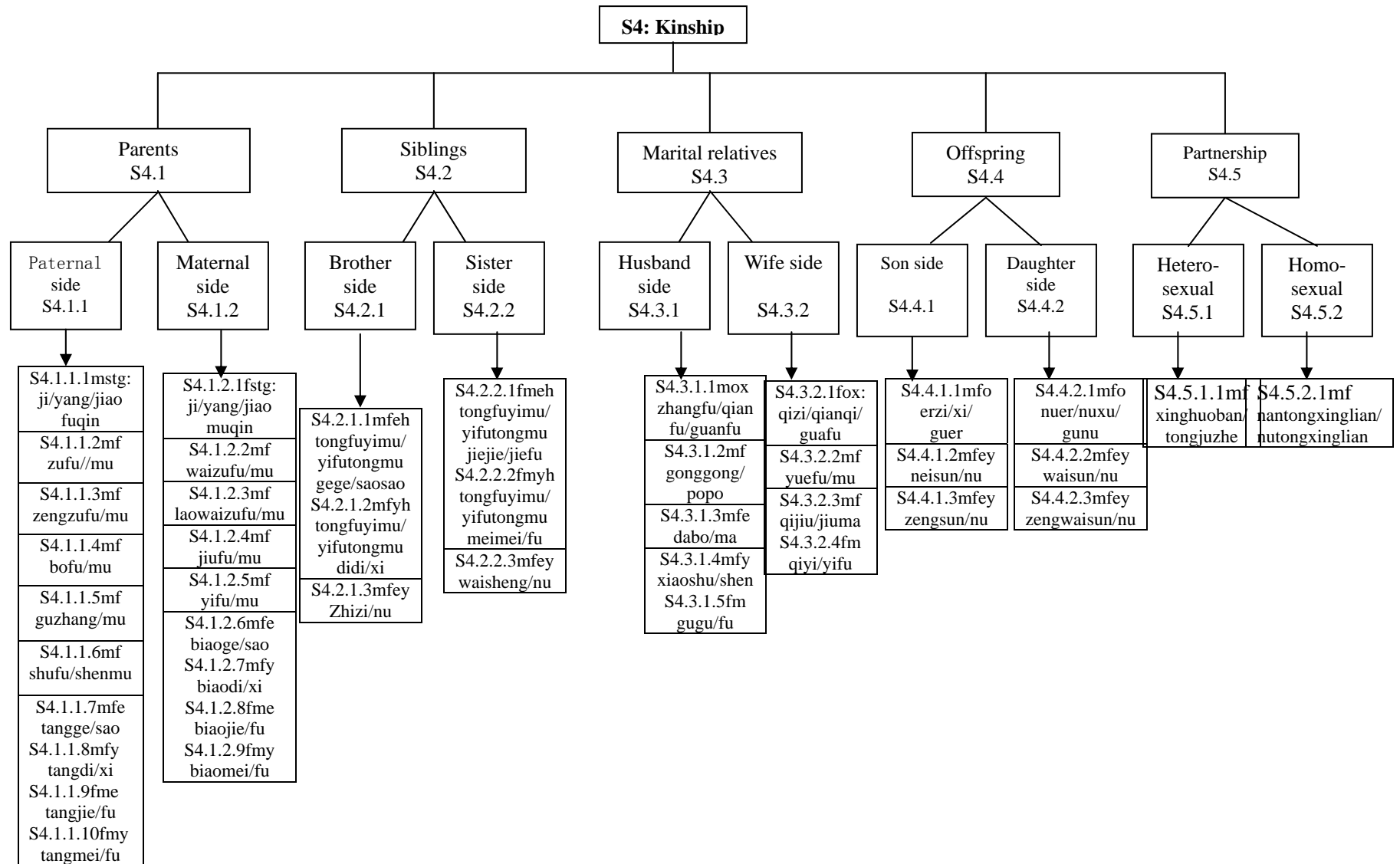
5. Conclusion

In this study, we have explored the possibility of applying existing kinship annotation, particularly the Lancaster USAS semantic taxonomy, to Chinese language. Our study demonstrates that there is a wide gap between the structures of the English and Chinese kinship terms. To address this problem, we proposed a Chinese kinship scheme which is based on the Lancaster scheme but is an extended model of it. We also presented a Chinese kinship semantic lexicon which is classified using our scheme. Through a case study, we examined the applicability of the corpus annotation. We found that our scheme generally reflects Chinese kinship structure of Chinese language. We foresee that, with continuous improvement, our scheme and lexicon will find various applications such as Chinese kinship annotation in corpora and information extraction.

References

Baker, H. D. R. (1979) *Chinese Family and Kinship*. London and Basingstoke, The MacMillan press LTD.

Chao, Y. R. (1956) "Chinese Terms of Address." *Language* vol. 32.

Chen, T. S. and J. K. Shryock (1932) "Chinese Relationship Terms." *American Anthropologist* vol. 34.

Feng, H. (1937) The Chinese Kinship System. *Harvard Journal of Asiatic Studies*, Vol. 2, No2

Hirshfeld, L. A. (1986) "Kinship and Cognition: Genealogy and the meaning of kinship terms." *Current Anthropology* Vol. 27, No. 3217-242.

Huang, S. and W. Jia (2000) "The Cultural Connotations and Communicative Functions of Chinese Kinship Terms". *American Communication Journal* Vol.3, No 3.

Kroeber, A. L. (1933) "Process in the Chinese Kinship System." *American Anthropologist* vol.35.

Löfberg, Laura, Scott Piao, Asko Nykanen, Krista Varantola, Paul Rayson, and Jukka-Pekka Juntunen. (2005) A semantic tagger for the Finnish language. In *Proceedings of the Corpus Linguistics Conference 2005*, Birmingham, UK.

McCoy, J. (1970) Chinese Kin Terms of reference and Address in M. Freedman (ed.) *Family and kinship in Chinese society*, pp.209-226. Stanford, Stanford University Press.

Morgan, H. L. (1871) *Systems of consanguinity and Affinity of the Human Family*. Washington, Smithsonian Institution.

Parkin, R. (1997) *Kinship - An Introduction to the Basic Concepts*. Oxford, Blackwell Publishers Inc.

Piao, Scott, Dawn Archer, Olga Mudraya, Paul Rayson, Roger Garside, Tony McEnery and Andrew Wilson. (2006) A large semantic lexicon for corpus

annotation. In *Proceedings from the Corpus Linguistics Conference Series on-line e-journal* 1(1), ISSN 1747-9398.

Qiu, Y. (2003) "A comparative study of cultural differences between Chinese and English kin-term systems". *Journal of Sichuan International Studies University*, Vol 19, No 3.

Rayson, P., A. Dawn, S. L. Piao, and T. McEnery (2004) The UCREL Semantic analysis system. In *Proceedings of LREC-04 Workshop: Beyond Named Entity Recognition Semantic Labeling for NLP Tasks*, pp. 7-12. Lisbon, Portugal. Sapir, E. (1927) *Language*. New York, Harcourt, Brace & Co.

Sharoff, S., B. Babych, P. Rayson, O. Mudraya and S. L. Piao (2006) ASSIST: Automated semantic assistance for translators. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*. Trento, Italy.

Wilson, A. and P. Rayson (1993) The Automatic Content Analysis of Spoken Discourse: A Report on Work in Progress. In Souter, C. and E. Atwell (Ed.), *Corpus Based Computational Linguistics*, pp. 215-226. Rodopi, Amsterdam.

Xiao, Z., A. McEnery, P. Baker, and A. Hardie (2004) "Developing Asian language corpora: standards and practice". In *Proceedings of the 4th Workshop on Asian Language Resources*, pp. 1-8. Sanya, Hainan Island, March 25, 2004.

Appendix 1: Chinese Kinship Annotation Scheme

S4: Kinship

Parents S4.1

Siblings S4.2

Marital relatives S4.3

Offspring S4.4

Partnership S4.5

Paternal side S4.1.1

Maternal side S4.1.2

Brother side S4.2.1

Sister side S4.2.2

Husband side S4.3.1

Wife side S4.3.2

Son side S4.4.1

Daughter side S4.4.2

Hetero-sexual S4.5.1

Homo-sexual S4.5.2

S4.1.1.1mstg: ji/yang/jiao fuqin

S4.1.1.2mf zufu//mu

S4.1.1.3mf zengzufu/mu

S4.1.1.4mf bofu/mu

S4.1.1.5mf guzhang/mu

S4.1.1.6mf shufu/shenmu

S4.1.1.7mfe tangge/sao
S4.1.1.8mfy tangdi/xi
S4.1.1.9fme tangjie/fu
S4.1.1.10fmy tangmei/fu

S4.1.2.1fstg: ji/yang/jiao muqin

S4.1.2.2mf waizufu/mu

S4.1.2.3mf laowaizufu/mu

S4.1.2.4mf jiufu/mu

S4.1.2.5mf yifu/mu

S4.1.2.6mfe biaoge/sao
S4.1.2.7mfy biaodi/xi
S4.1.2.8fme biaojie/fu
S4.1.2.9fmy biaomei/fu

S4.2.1.1mfeh tongfuyimu/ yifutongmu gege/saosao
S4.2.1.2mfyh tongfuyimu/ yifutongmu didi/xi
S4.2.1.3mfey Zhizi/nu

S4.2.2.1fmeh tongfuyimu/ yifutongmu jiejie/jiefu
S4.2.2.2fmyh tongfuyimu/ yifutongmu meimei/fu
S4.2.2.3mfey waisheng/nu

S4.3.1.1mox zhangfu/qian fu/guanfu
S4.3.1.2mf gonggong/ popo
S4.3.1.3mfe dabo/ma
S4.3.1.4mfy xiaoshu/shen
S4.3.1.5fm gugu/fu

S4.3.2.1fox: qizi/qianqi/ guafu
S4.3.2.2mf yuefu/mu
S4.3.2.3mf qijiu/jiuma
S4.3.2.4fm qiyi/yifu

S4.4.1.1mfo erzi/xi/ guer
S4.4.1.2mfey neisun/nu
S4.4.1.3mfey zengsun/nu

S4.4.2.1mfo nuer/nuxu/ gunu
S4.4.2.2mfey waisun/nu
S4.4.2.3mfey zengwaisun/nu

S4.5.1.1mf xinghuoban/ tongjuzhe

S4.5.2.1mf nantongxinglian/ nutongxinglian

13

Appendix 2: Chinese kinship lexicon

S4 Kinship
S4.1mf Parents
S4.1.1 Father's line
Word   semtag
父亲(fu qin) S4.1.1m
爸爸(ba ba) S4.1.1m
翁(weng)   S4.1.1m
爹(die)   S4.1.1m
老子(lao zi) S4.1.1m
阿爸(a ba)   S4.1.1m
家父(jia fu)   S4.1.1m
老爸(lao ba) S4.1.1m
继父(ji fu)   S4.1.1ms
养父(yang fu) S4.1.1mt
教父(jiao fu) S4.1.1mg
祖父(zu fu) S4.1.1.2m
爷爷(ye ye) S4.1.1.2m
祖母(zu mu) S4.1.1.2f
奶奶(nai nai) S4.1.1.2f
曾祖父(zeng zu fu) S4.1.1.3m
太公(tai gong)   S4.1.1.3m
曾祖母(zeng zu mu) S4.1.1.3f
太婆(tai po)   S4.1.1.3f
伯父(bo fu)   S4.1.1.4m
伯伯(bo bo)   S4.1.1.4m
大伯(da bo)   S4.1.1.4m
大爷(da ye)   S4.1.1.4m
伯母(bo mu)   S4.1.1.4f
大妈(da ma)   S4.1.1.4f
大娘(da niang)   S4.1.1.4f
大婶娘(da shen niang) S4.1.1.4f
大姆妈(da mu ma) S4.1.1.4f
姑母(gu mu)   S4.1.1.5f
姑姑(gu gu)   S4.1.1.5f
姑娘(gu niang)   S4.1.1.5f
姑妈(gu ma)   S4.1.1.5f
大姑(da gu)   S4.1.1.5f
小姑(xiao gu)   S4.1.1.5f
姑丈(gu zhang)   S4.1.1.5m
姑父(gu fu)   S4.1.1.5m
姑爹(gu die)   S4.1.1.5m
叔父(shu fu)   S4.1.1.6m

叔叔(shu shu)　　S4.1.1.6m
叔　　(shu)　　　S4.1.1.6m
小叔(xiao shu)　　S4.1.1.6m
阿叔(a shu)　　　S4.1.1.6m
婶母(shen mu)　　S4.1.1.6f
婶婶(shen shen)　　S4.1.1.6f
婶娘(shen niang)　　S4.1.1.6f
叔母(shu mu)　　S4.1.1.6f
婶子(shen zi)　　　S4.1.1.6f
堂兄(tang xiong)　　S4.1.1.7me
堂嫂(tang sao)　　　S4.1.1.7fe
堂弟(tang di)　　　S4.1.1.8my
堂弟媳(tang di xi)　　S4.1.1.8fy
堂姐(tang jie)　　　S4.1.1.9fe
堂姐夫(tang jie fu)　　S4.1.1.9me
堂妹(tang mei)　　　S4.1.1.10fy
堂妹夫(tang mei fu)　S4.1.1.10my

S4.1.2 Mother's Line
母亲(mu qin)　　　S4.1.2.1f
妈妈(ma ma)　　　S4.1.2.1f
娘(niang)　　　S4.1.2.1f
阿妈(a ma)　　　S4.1.2.1f
家母(jia mu)　　　S4.1.2.1f
老妈(lao ma)　　　S4.1.2.1f
继母(ji mu)　　　S4.1.2.1fs
养母(yang mu)　　S4.1.2.1ft
教母(jiao mu)　　　S4.1.2.1fg
外祖父(wai zu fu)　S4.1.2.2m
外公(wai gong)　　S4.1.2.2m
姥爷(lao ye)　　　S4.1.2.2m
公公(gong gong)　　S4.1.2.2m
外祖母(wai zu mu)　S4.1.2.2f
外婆(wai po)　　　S4.1.2.2f
姥姥(lao lao)　　　S4.1.2.2f
老外祖父(lao wai zu fu) S4.1.2.3m
外太公(wai tai gong)　S4.1.2.3m
老外祖母(lao wai zu fu) S4.1.2.3f
外太婆(wai tai po)　　S4.1.2.3f
舅父(jiu fu)　　　S4.1.2.4m
舅舅(jiu jiu)　　　S4.1.2.4m
娘舅(niang jiu)　　S4.1.2.4m
舅母(jiu mu)　　　S4.1.2.4f
舅妈(jiu ma)　　　S4.1.2.4f
妗子(jin zi)　　　S4.1.2.4f

姨母(yi mu)　　　S4.1.2.5f

姨妈(yi ma)　　　S4.1.2.5f

姨娘(yi niang)　　S4.1.2.5f

姨丈(yi zhang)　　S4.1.2.5m

姨父(yi fu)　　　S4.1.2.5m

表哥(biao ge)　　S4.1.2.6me

表嫂(biao sao)　　S4.1.2.6fe

表弟(biao di)　　S4.1.2.7my

表弟媳(biao di xi)　S4.1.2.7fy

表姐(biao jie)　　S4.1.2.8fe

表姐夫(biao jie fu)　S4.1.2.8me

表妹(biao mei)　　S4.1.2.9fy

表妹夫(biao mei fu)　S4.1.2.9my

S4.2 Brother and Sister's Line

S4.2.1 Brother's Line

同父异母或(tong fu yi mu huo yi fu tong mu

异父同母哥哥 ge ge) S4.2.1.1meh

哥哥(ge ge)　　　S4.2.1.1me

阿哥(a ge)　　　S4.2.1.1me

嫂子(sao zi)　　　S4.2.1.1fe

嫂嫂(sao sao)　　S4.2.1.1fe

弟弟(di di)　　　S4.2.1.2my

阿弟(a di)　　　S4.2.1.2my

幼弟(you di)　　　S4.2.1.2my

同父异母或 (tong fu yi mu huo yi fu tong mu

异父同母弟弟 di di)　　S4.2.1.2meh

弟媳(di xi)　　　S4.2.1.2fy

弟妹(di mei)　　　S4.2.1.2fy

弟妇(di fu)　　　S4.2.1.2fy

侄子(zhi zi)　　　S4.2.1.3m

侄女(zhi nu)　　　S4.2.1.3f

S4.2.2 Sister's Line

姐姐(jie jie)　　　S4.2.2.1fe

阿姐(a jie)　　　S4.2.2.1fe

长姊(zhang zi)　　S4.2.2.1fe

同父异母或(tong fu yi mu huo yi fu tong mu

异父同母姐姐 jie jie)S4.2.2.1feh

妹妹(mei mei)　　S4.2.2.2fy

阿妹(a mei)　　　S4.2.2.2fy

幼妹(you mei)　　S4.2.2.2fy

同父异母或(tong fu yi mu huo yi fu tong mu

异父同母妹妹 mei mei)　　S4.2.2.2fyh

姐夫(jie fu)　　　S4.2.2.2me

姊夫(zi fu)　　　S4.2.2.2me

姊丈(zi zhang)　　S4.2.2.2my
妹夫(mei fu)　　S4.2.2.2my
外甥(wai sheng)　　S4.2.2.3m
外甥女(wai sheng nu)　S4.2.2.3f

S4.3mf
S4.3.1 Husband's Line
丈夫(zhang fu)　　S4.3.1.1m
老公(lao gong)　　S4.3.1.1m
爱人(ai ren)　　S4.3.1.1m
先生(xian sheng)　　S4.3.1.1m
前夫(qian fu)　　S4.3.1.1mx
鳏夫(guan fu)　　S4.3.1.1mo
公公(gong gong)　　S4.3.1.2m
婆婆(po po)　　S4.3.1.2f
大伯(da bo)　　S4.3.1.3me
大妈(da ma)　　S4.3.1.3fe
小叔(xiao shu)　　S4.3.1.4my
小叔子(xiao shu zi)　S4.3.1.4my
小婶(xiao shen)　　S4.3.1.4f
姑姑(gu gu)　　S4.3.1.5f
大姑子(da gu zi)　　S4.3.1.5f
小姑子(xiao gu zi)　　S4.3.1.5f
姑夫(gu fu)　　S4.3.1.5m

S4.3.2 Wife's Line
妻子(qi zi)　　S4.3.2.1f
老婆(lao po)　　S4.3.2.1f
爱人(ai ren)　　S4.3.2.1f
内人(nei ren)　　S4.3.2.1f
前妻(qian qi)　　S4.3.2.1fx
寡妇(gua fu)　　S4.3.2.1fo
岳父(yue fu)　　S4.3.2.2m
岳丈(yue zhang)　　S4.3.2.2m
岳翁(yue weng)　　S4.3.2.2m
丈人(zhang ren)　　S4.3.2.2m
外父(wai fu)　　S4.3.2.2m
泰山(tai shan)　　S4.3.2.2m
冰翁(bing weng)　　S4.3.2.2m
岳母(yue mu)　　S4.3.2.2f
丈母娘(zhang mu niang)　S4.3.2.2f
妻舅(qi jiu)　　S4.3.2.3m
内兄(nei xiong)　　S4.3.2.3m
内弟(nei di)　　S4.3.2.3m
大舅子(da jiu zi)　　S4.3.2.3m
小舅子(xiao jiu zi)　　S4.3.2.3m

阿哥(a ge)　　　S4.3.2.3m
弟弟(di di)　　　S4.3.2.3m
舅嫂(jiu sao)　　S4.3.2.3f
舅妈(jiu ma)　　S4.3.2.3f
妻姨(qi yi)　　S4.3.2.4f
大姨(da yi)　　S4.3.2.4f
小姨(xiao yi)　　S4.3.2.4f
姨姐(yi jie)　　S4.3.2.4f
姨妹(yi mei)　　S4.3.2.4f
姨夫(yi fu)　　S4.3.2.4m

S4.4mf Son/daughter
S4.4.1 Son's Line
儿子(er zi)　　　S4.4.1.1m
儿媳(er xi)　　　S4.4.1.1f
媳妇(xi fu)　　　S4.4.1.1f
孤儿(gu er)　　　S4.4.1.1mo
内孙(nei sun)　　S4.4.1.2m
孙(sun)　　　S4.4.1.2m
孙子(sun zi)　　S4.4.1.2m
内孙女(nei sun nu)　S4.4.1.2f
孙女(sun nu)　　S4.4.1.2f
曾孙(zeng sun)　　S4.4.1.3m
曾孙女(zeng sun nu)　S4.4.1.3f

S4.4.2 Daughter's Line
女儿(nu er)　　　S4.4.2.1f
孤女(gu nu)　　　S4.4.2.1fo
女婿(nu xu)　　　S4.4.2.1m
姑爷(gu ye)　　　S4.4.2.1m
外孙(wai sun)　　S4.4.2.2m
外孙女(wai sun nu)　S4.4.2.2f
曾孙(zeng sun)　　S4.4.2.3m
曾孙女(zeng sun nu)　S4.4.2.3f

S4.5 Partnership
S4.5.1 Heterosexual
同居者(tong nu zhe)　S4.5.1.1mf
性伙伴(xing huo ban)　S4.5.1.1mf

S4.5.2 Homosexual
男同性恋(nan tong xing lian) S4.5.2.1m
同志(tong zhi)　S4.5.2.1m
妹子(mei zi)　S4.5.2.1m
金刚芭比(jin gang ba bi) S4.5.2.1m
CC　　　S4.5.2.1m
兔子(tu zi)　　S4.5.2.1m
玻璃(bo li)　　S4.5.2.1m

哥哥(ge ge)　　S4.5.2.1m
弟弟(di di)　　S4.5.2.1m
女同性恋(nu tong xing lian) S4.5.2.1f
同志(tong zhi)　　S4.5.2.1f
拉拉(la la)　　S4.5.2.1f
拉子(la zi)　　S4.5.2.1f
欧蕾(ou lei)　　S4.5.2.1f
蕾丝边(lei si bian) S4.5.2.1f
歹客(dai ke)　　S4.5.2.1f
T　　　　S4.5.2.1f
P　　　　S4.5.2.1f