A Corpus-Based Study on the Categorization and Composition of Personal Names and Terms of Address in Chinese Classical Novels

Xiong Dan[1], Lu Qin[1], Lo Fengju[2], Shi Dingxu[3] and Chiu Tin-shing[1]

[1]Department of Computing
The Hong Kong Polytechnic University
Hong Kong
{csdxiong, csluqin, cstschiu}@comp.polyu.edu.hk

[2]Department of Chinese Linguistics & Literature
Yuan Ze University
Taiwan
gefjulo@saturn.yzu.edu.tw

[3]Department of Chinese & Bilingual Studies
The Hong Kong Polytechnic University
Hong Kong
dingxu.shi@polyu.edu.hk

Abstract. *As an important linguistic category, terms of address not only carry particular information, but also express feelings and emotions. They are therefore widely used in literature works. In natural language processing (NLP) and its applications, terms of address are one of the key elements in named entity recognition, which can affect the overall performance of an NLP system. Based on the analysis of a manually-annotated corpus of four Chinese classical novels in the Ming and Qing dynasties, this paper presents a classification and annotation system for personal names and terms of address from the perspective of named entity recognition and information extraction in NLP. Personal names and terms of address are categorized into simple types and compound types and the compound-type is further categorized into four subtypes, namely, fixed expressions, appositive constructions, subordinate constructions of affiliation, and other subordinate constructions.*
Keywords: terms of address; NLP; novels in the Ming and Qing dynasties; corpus; named entity recognition

1. **Introduction.** Terms of address, as important information carriers, are essential components of any language in social communication as well as crucial elements in the processing of natural languages. People's habits of using terms of address change over time, which reflect the cultural connotations and complexity of interpersonal relationship in a certain period. In literature works, especially novels, terms of address can not only convey information on titles, but also bring certain characteristics to designated figures to produce some expected literary effects. This paper presents a classification and annotation system for personal names and terms of address from the perspective of named entity recognition and its applications in NLP based on a computer-aided, manually-annotated corpus [1] which contains four influential novels produced in the Ming and Qing dynasties, namely, *A Dream of Red Mansions* (《紅樓夢》), *Romance of the Three Kingdoms* (《三國演義》), *Water Margin* (《水滸傳》), and *The Golden Lotus* (《金瓶梅》). To get a more complete picture of the categories and composition rules of terms of address, this paper further analyzes the components of compound terms and the typical ways in which they are structured through full data statistics of the corpus.

The rest of the paper is organized as follows. Section 2 gives a brief literature review. Section 3 introduces the overall classification system developed based on the analysis of the corpus. Sections 4 and 5 describe the simple types and compound types with examples respectively. Section 6 presents a statistical analysis of the full corpus. Section 7 is the conclusion.

2. **Literature Review.** Since ancient times, terms of address have attracted great interest of scholars in China. *Erya-Explanations of Kinship Terms* (《爾雅•釋親》) [2], which explains terms of address with respect to family and marriage, is commonly regarded as the earliest known monograph introducing kinship terms in a systematic way. *Records of Addressing* (《稱謂錄》) [3], written by Liang Zhangju in the Qing dynasty, is a comprehensive collection of terms of address used in ancient China, which describes and traces terms of address from the perspective of kinship, social status, and official title, etc. In contemporary China, substantial research results have also been produced from various dimensions. A review of the study on terms of address in modern Chinese over the last 20 years is given in [4].

In the field of NLP, the specification for modern Chinese processing commonly used in mainland China [5] regards personal name as a type of named entity tagged with "/nr". For a Chinese full name commonly formed by a surname followed by a given name, the surname and the given name are tagged individually, such as "張/nrf 仁偉/nrg" (*Zhāng/nrf Rénwěi/nrg*). However, the current researches on Chinese information processing have not set up a classification and annotation scheme for terms of address. The titles and other terms used to address people are treated as common nouns, which are tagged with "/n", for example, "李/nrf 主席/n" (*Lǐ/nrf zhǔxí/n, President Li*) and "劉/nrf 阿姨/n" (*Liú/nrf āyí/n, Aunt Liu*). The annotated corpus built by Academia Sinica, Taiwan [6-7] has included some Chinese classics in the Ming and Qing dynasties. It treats personal names as proper nouns (tagged with "Nb"). Yet terms of address are treated as common nouns (tagged with "Na"),

for instance, "政(Nb)老爹(Na)" (*Zhèng(Nb) lǎodiē (Na), Sir Zheng*).

The function of terms of address as a reflection of the cultural and linguistic features of a certain age becomes more prominent when they are used in novels. Therefore, this paper regards them as a major category of named entity and further classifies them for the purpose of information processing. Different from the previous work on Chinese language processing and other researches conducted in the fields of literature and history, this paper studies the constituent parts of terms of address and their composition rules to set up a classification and annotation system for them. In practical work, this system was applied to the building of the corpus consisting of the four Chinese classics in an iteratively refined process.

3. **Overall Categorization System.**

3.1. **Definition and Scope.** There have been rather diverse opinions about the exact definition and scope of terms of address, and some represent several mainstream views in China [4]. As picking a specific definition is not the focus of this work, we look at terms of address in a rather broad sense. We investigate the terms used for direct addressing in face-to-face situations as well as those used to indirectly refer to a particular person. In this sense, all kinds of titles indicating a person's social status, rank, and occupation are included. With regard to word classes, the terms of address included in this work comprise only nouns and nominal phrases while exclude pronouns such as "you" which has no ambiguity. In addition, only the terms of address with a specific referent are treated as named entities in consideration of future application in the area of NLP, and others are treated as common nouns. For example, the word "姑娘" (*gūniáng, girl*) in *A Dream of Red Mansions* is treated differently depending on situations. If the girl refers to a specific person from the context, it is tagged as a term of address, for example, "姑娘/na2 到/了" (*gūniáng/na2 dào/le, Miss Lin is here*). Otherwise, it is treated as a common noun as in "一/個/姑娘/領著/他" (*yī/gè/gūniáng/lǐngzhe/tā, a girl led him*) and "姑娘/們" (*gūniáng/men, the girls*).

It should be pointed out that the terms of address consisting of several units are segmented and tagged in nested fashion in accordance with the unified specification applicable to classical novels [1] so as to ensure the consistency of segmentation and enhance the efficiency of system training. In general, the annotation system follows the basic approach developed by Peking University for Part of Speech (PoS) tagging [5]. That is, square brackets ([]) are used to enclose compound terms, in which the units are tagged accordingly. If a compound term contains other types of named entities, such as names of place and organization, they are also explicitly tagged in order to show the roles of other categories of entities in serving as terms of address.

3.2. **Purpose of Categorization.** Due to the difference in cultural conventions, terms of address change over different historical periods, geographic regions, and social communities. To make the story more attractive, the way in which a certain person is addressed changes along with the status of the addressee, the situation, the interlocutors,

and even the emotions the author intended to express. For example, Cai Jing (蔡京), although not a major character in *The Golden Lotus*, is addressed in numerous forms. A neutral way is to address him as his surname or full name. In formal circumstances, official titles are mentioned, such as "[左/丞相]/nu1" (*[zuǒ/chéngxiàng]/nu1, the Left Prime Minister*), "太師/nu1" (*tàishī/nu1, Imperial Tutor*), and "[吏部/nt 尚書]/nu1" (*[lìbù/nt shàngshū]/nu1, Minister of the Board of Civil Service*). However, the most common form for addressing in the novel is to combine different elements into a compound one, such as "[蔡/nr1 太師/nu1]/na1#[1]" (*[Cài/nr1 tàishī/nu1]/na1#, Imperial Tutor Cai*) and "[蔡/nr1 老先生]/na1#" (*[Cài/nr1 lǎoxiānshēng]/na1#, Old Gentleman Cai*). With high status, he is mostly addressed in certain respectful and polite forms, while in some private conversations, unfriendly or casual forms are also used, such as "老賊/na2#" (*lǎozéi/na2#, a disparaging term for addressing an old person*). In addition, the four classical novels involve different dynasties, historical backgrounds, and social groups, which result in a rich system of address. Therefore, it is necessary to classify them in a systematic way to establish an annotation system for the building of the classical Chinese corpus. The results are also expected to lay a foundation for future work, such as intra- and inter-textual entity linking throughout the whole diachronic language knowledge base.

3.3. **Overall Categorization System.** Having a close connection with the culture of the related times, the personal names and terms of address used in the novels show distinctive features different from those in modern Chinese, which should be taken into consideration. For instance, besides given names, other forms of names can be found in the novels, such as courtesy name (*字, zì another form of a given name as a symbol of adulthood and respect*), art name (*號, hào, a self-selected alternative name*), and posthumous name (*諡號, shìhào, a name awarded to an honorable person after death*).

Based on the statistics and analysis on the four classics, this paper classifies personal names and terms of address from the perspective of their composition for the purpose of more effective processing in NLP. At the first level, they are categorized into simple types and compound types. As the names imply, the former includes those consisting of only one unit which can be used for addressing or reference independently, and the latter is constructed with several units in various ways. For instance, some compound terms are constructed with several simple-type terms, all of which can be used for addressing, such as "[令郎/先生]/na2" (*[lìngláng/xiānshēng]/na2, your son/Mr., the two are used together to show respect*). Some compound terms are composed of a personal name and some descriptive words or phrases that cannot be used to address people independently, such as "[周瑞/nr3 家/的]/na1" (*[Zhōu Ruì/nr3 jiā/de]/na1, the wife of Zhou Rui*), in which "家/的" (*jiā/de, of the family*) cannot be used for addressing. Based on their structure and the relations among the components, terms of address are classified into several subtypes as illustrated in Figure 1.

---

[1] To facilitate future researches, real named entities are labeled with "#" to distinguish them from fictional ones.
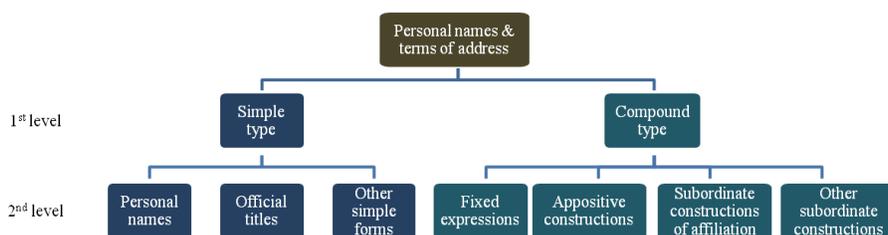
FIG. 1. OVERALL CATEGORIZATION

4. **Categorization of Simple Types.** As Figure 1 shows, the simple types in the classical novels are classified into three subtypes: personal names, official titles, and other simple forms. This section gives a detailed description of all the subtypes. To have a better understanding of the annotation system, the tags used for the modern Chinese corpus of Peking University (PKU) [5] are also provided for reference.

4.1. **Personal Names.** Personal names fall into six categories: surname (S), given name (GN), S + GN, courtesy name (CN), S + CN, and all other alternative names (AN). The last category includes all forms of names not included in the first five types, such as milk name, nickname, pen name, art name, posthumous name, temple name (*廟號, miàohào, given to an emperor after death*), and the names from foreign countries and minority ethnic groups in which the surname and given name cannot be identified since they are translated or transliterated from foreign languages. Table 1 explains each of these categories with examples.

TABLE 1. CATEGORIZATION OF PERSONAL NAMES

2nd-level	3rd-level	Tags used by us	Tags used by PKU	Examples and description
Personal names	S	/nr1	/nrf	A surname may contain one, two, or multiple characters, such as 林/nr1 (*Lín/nr1*), 諸葛/nr1# (*Zhūgě/nr1#*)
	GN	/nr2	/nrg	黛玉/nr2 (*Dàiyù/nr2*)
	S+GN	/nr3	/nrf /nrg	林黛玉/nr3 (*Lín Dàiyù/nr3*)
	CN	/nr4	/nrg	士隱/nr4 (*Shìyǐn/nr4*)
	S+CN	/nr5	/nrf /nrg	甄士隱/nr5 (*Zhēn Shìyǐn/nr5*)
	AN	/nr6	/nr	顰兒/nr6 (*Pín ér/nr6, nickname*), 阿會喃 (*Āhuìnán/nr6, the name of a tribesman*)

4.2. **Official Titles.** Official titles in this paper include only names of official positions (NOP, *官職, guānzhí*) and titles of nobility or honor (TNH, *爵位 juéwèi, 封號 fēnghào*) which designate particular referents as shown in Table 2.

TABLE 2. CATEGORIZATION OF OFFICIAL TITLES

2nd-level	3rd-level	Tags used by us	Tags used by PKU	Description	Examples
Official titles	NOP	/nu1	/n	The names of official positions	太 師 /nu1 (*tàishī/nu1, Imperial Tutor*)
	TNH	/nu2	/n	The titles granted to members of the aristocracy or bestowed on outstanding persons	郡 王 /nu2 (*jùnwáng/nu2, Prince*) 貴 妃 /nu2 (*guìfēi/nu2, Imperial Consort*)

4.3. **Other Simple Forms.** This category consists of all the other single independent units used for addressing or reference, excluding pronouns and self-address terms. This type, tagged with "/na2", usually reflects kinship or social relations among the interlocutors. For example, "[賈/nr1 母]/na1" (*[Jiǎ/nr1 mǔ]/na1, Old Lady Jia*) in *A Dream of Red Mansions* is addressed in different respectful forms, including "老祖宗/na2" (*lǎozǔzōng/na2, Old Ancestress*), "祖母/na2" (*zǔmǔ/na2, grandmother*), and "老太太/na2" (*lǎotàitai/na2, the Old Lady*). The first one is usually used in face-to-face conversations, the second one in reference, and the third one in both situations. All of these kinds of simple forms are put into this category.

5. **Categorization of Compound Types.** Simple-type terms of address can be used either independently or as components of compound terms. As mentioned before, a compound term is enclosed in square brackets. If its components belong to a different type of named entity, they are tagged independently inside the nested structures. Based on the ways in which the compound terms are constructed, they are further classified into four categories: fixed expressions, appositive constructions, subordinate constructions of affiliation, and other subordinate constructions. This section describes the four categories in details.

5.1. **Fixed Expressions.** A compound term of this type is used as a fixed expression, in which the elements are integrated into a proper noun. The following are two typical kinds of fixed expressions.

   Titles of Nobility or Honor. Some titles conferred by the emperor on individuals contain special honorary terms, which are attached to the noble titles, such as "[北靜/郡王]/nu2" (*[Běijìng/jùnwáng]/nu2, Prince of Beijing*) and "[賢德/妃]/nu2" (*[xiándé/fēi]/nu2, Consort of Virtue*). "北靜" (*Běijìng, Beijing*) and "賢德" (*xiándé, Virtue*), being of the same type as the whole compound term, are not tagged since the system will recognize them as honorary titles by default.

   Names of Official Positions. Some official titles also consist of honorary terms which

cannot be used separately, such as "[奮威/將軍]/nu1" (*[fènwēi/jiāngjūn]/nu1, General of Valor*).

5.2. **Appositive Constructions.** A compound term of this kind comprises a simple-type one followed by an appositive to show respect and politeness or to give further explanation, confirmation, or emphasis. Usually, both the simple-type term and its appositive can be used for addressing independently, such as "[父親/大人]/na2" (*[fùqīn/dàrén]/na2, addressing father in a respectful way*) and "[都太尉/統制]/nu1" (*[dūtàiwèi/tǒngzhì]/nu1, the High Marshal*).

5.3. **Subordinate Constructions of Affiliation.** A compound term of this type usually consists of two independent named entities, in which one is affiliated with or closely connected to the other. Although the elements in these constructions are different named entities, they are enclosed to form compound terms for the purpose of information linking. This category can be further classified into the following subtypes.

Name of Place + Title of Nobility or Honor. When a person is addressed with a title of nobility or honor, the name of the land inherited by or granted to the person, if any, is usually attached to the term of address, such as "[烏程/ns2# 侯]/nu2" (*[Wūchéng/ns2# Háu]/nu2, Marquis of Wucheng*). Since Wucheng is a place name, it is tagged with "/ns2" independently.

Name of Place + Official Position. The official position of a person is sometimes mentioned together with the name of the territory, such as "[揚州/ns2# 刺史]/nu1[2]" (*[Yángzhōu/ns2# cìshǐ]/nu1, Prefect of Yangzhou*).

Name of Organization + Official Position. The names of some official positions are used for addressing in combination with the related organizations, such as "[吏部/nt 尚書]/nu1" (*[lìbù/nt shàngshū]/nu1, Minister of Personnel Affairs*).

5.4. **Other Subordinate Constructions.** All other compound-type terms of address not included in the above three types are put into this category. The elements in these constructions can be either simple- or compound-type terms, which are nested in quite flexible ways. These nesting constructions, however complicated, can be segmented and tagged in accordance with the unified specification. The compound terms are enclosed in square brackets and tagged with "/na1", in which the components are tagged accordingly. However, a component belonging to the category of "other simple forms" (OSF), which are tagged with "/na2" as described in section 4.3, is not required to be tagged since it can be recognized by default. For example, in "[蔡/nr1 老爺]/na1#" (*[Cài/nr1 lǎoye]/na1#, Master Cai*), "老爺" (*lǎoye, Master*) is not tagged with "/na2"; while in "[蔡/nr1 太師

---

[2] The relationship between a term of address and its referent changes with space and time. For example, once another person takes over the post, the relationship changes accordingly.

/nu1]/na1#" (*[Cài/nr1 tàishī/nu1]/na1#, Imperial Tutor Cai*), "太師/nu1" (*tàishī/nu1, Imperial Tutor*) should be tagged since it is a name of official position (NOP). On the basis of their composition, these compound constructions are further classified into seven subtypes, as summarized in Table 3.

TABLE 3. CATEGORIZATION OF OTHER SUBORDINATE CONSTRUCTIONS

3rd-level	Examples and description
PN + OSF	[蔡/nr1 老爺]/na1# (*[Cài/nr1 lǎoye]/na1#, Master Cai*), in which Cai is a surname
	[玄德/nr4# 公]/na1# (*[Xuándé/nr4# gōng]/na1#, Mr. Xuande*), in which, Xuande is a courtesy name
NOP + OSF	[太師/nu1 老爺]/na1# (*[tàishī/nu1 lǎoye]/na1#, literally Imperial Tutor, the Master*)
TNH + OSF	[[臨安 / 伯]/nu2 老太太]/na1 (*[[Lín'ān/bó]/nu2 lǎotàitai]/na1, mother of the Earl of Lin'an*)
PN + NOP	[蔡/nr1 太師/nu1]/na1# (*[Cài/nr1 tàishī/nu1]/na1#, Imperial Tutor Cai*), in which the surname and official position are tagged separately
PN + NOP + OSF	[[蔡/nr1 太師/nu1]/na1# 老爺]/na1# (*[[Cài/nr1 tàishī/nu1]/na1# lǎoye]/na1#, literally Imperial Tutor Cai, the Master*)
PN + TNH	[元/nr2 妃/nu2]/na1 (*[Yuán/nr2 fēi/nu2]/na1, Imperial Consort Yuan*), in which Yuan is a part of the given name
PN + TNH + OSF	[[元 /nr2 妃 /nu2]/na1 娘娘]/na1 (*[[Yuán/nr2 fēi/nu2]/na1 niángniang]/na1, literally Imperial Consort Yuan, Her Highness*)
Remarks: PN = personal name (the 1st category of simple types), OSF = other simple forms (the 3rd category of simple types), NOP = name of official positions, TNH = title of nobility or honor	

It should be pointed out that the components of the above subtypes are not listed in sequential order. For example, "NOP + OSF" also includes the structure of "OSF + NOP", such as "[義士/提轄/nu1]/na1" (*[yìshì/tíxiá/nu1]/na1, literally the righteous constable*), in which the OSF is placed in front of the NOP. In addition, any component may be singular or plural, such as "PN + OSF" may contain multiple OSFs, for instance, "[[晁/nr1 頭領]/na1 哥哥]/na1" (*[[Cháo/nr1 tóulǐng]/na1 gēge]/na1, literally Head Chao, the brother*).

6. **Data Analysis.**

6.1. **General Data.** Personal names and terms of address in the classical novels are roughly classified into simple types and compound types at the first level. Figure 2 and Figure 3 in this section show the distribution of the first-level types and their second-level subtypes. Figure 2 focuses on category, which means that the same term is counted only once. Figure 3 concentrates on frequency, that is, the number of occurrences. The data in this paper includes not only the fictional characters created by the authors, but also the real historical figures mentioned in the novels.

Contrary to intuitive speculations, the portion of the simple types is much higher than that of the compound types in terms of both category and occurrence, as shown in Figure 2 and Figure 3. In classic literature, we do see more compound terms of address as people in ancient times adhere to social hierarchies more strictly in interpersonal communication. Still, the results indicate that personal names are used more often for addressing or reference to facilitate the telling of the stories and third-person narration. Another reason is that the name system of ancient China offers more alternatives for the authors to choose, such as courtesy names and art names.
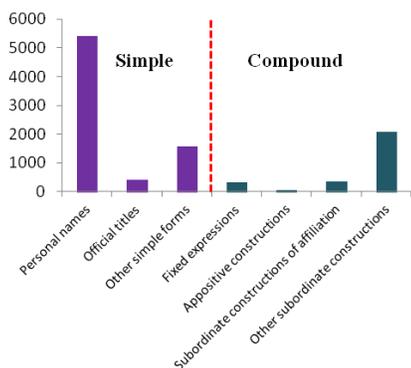
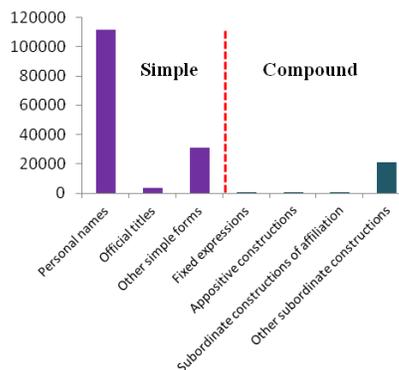**FIG.2.** GENERAL DATA (CATEGORY)   **FIG. 3.** GENERAL DATA (OCCURRENCE)

6.2. **Data of Personal Names.** In view of the high proportion of personal names, it is necessary to carry out a more in-depth analysis on their composition. In the same way as general data is displayed in section 6.1, Figure 4 illustrates the subtypes of personal names from the dimension of category, and Figure 5 from occurrence. As the data shows, the type of "surname + given name" (S + GN) has the highest percentage, which is consistent with the habits of addressing in daily communication. It is obvious that a wide range of alternative names (ANs) are used. This is because art names and other kinds of pseudonyms are popular in ancient China. And, pet names and nicknames are good choices for storytelling and casual conversations.

Figure 4 and Figure 5 also present the number of the characters contained in surname (S), given name (GN), and courtesy name (CN). As we examine the data in Figure 4, an interesting phenomenon is that the portion of single-character GNs[3] is as high as 77% of the total GNs. This is mainly because most GNs in *Romance of the Three Kingdoms* contain only one character, which reflects the characteristics of the name system in the corresponding time. As an important factor of Chinese name cultures, most CNs contain two characters, and the percentage of single-character CNs is only about 1%.

---

[3] Including the cases that only one character of a two-character given name is used, such as "迎/nr2 、/探/nr2 、/惜/nr2 三/人" (*Yíng/nr2, /Tàn/nr2, /Xī/nr2 sān/rén*) , in which Ying, Tan, and Xi are parts of the three girls' full given names
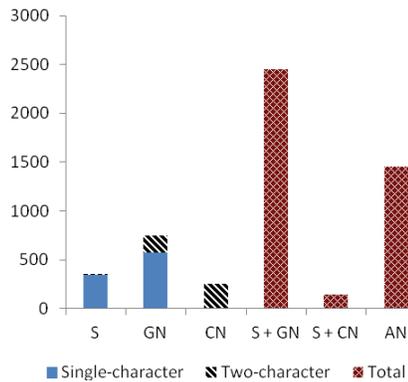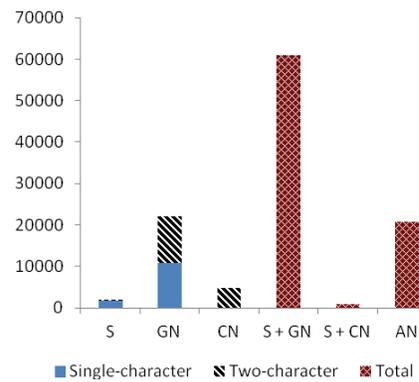
FIG. 4. DATA OF PERSONAL NAMES (CATEGORY)

FIG. 5. DATA OF PERSONAL NAMES (OCCURRENCE)

6.3. **Data of Other Subordinate Constructions.** This kind of constructions is the biggest part of compound terms of address, in which components at different levels are combined in flexible ways. The components of some structures are already compound ones, for instance, "[[梁 /nr1 中書 /nu1]/na1 相公]/na1" (*[[Liáng/nr1 zhōngshū/nu1]/na1 xiànggōng]/na1, literally Minister Liang, the Master*), which is grouped by a surname, an official title, and a simple-type term of address showing respect. As Table 3 lists, these constructions are classified into seven subtypes: "personal name (PN) + other simple forms (OSF)", "name of official positions (NOP) + OSF", "title of nobility or honor (TNH) + OSF", "PN + NOP", "PN + NOP + OSF", "PN + TNH", "PN + TNH + OSF". Figure 6 illustrates the distribution of the seven subtypes by category (the same term of address counted only once). As the chart shows, "PN + OSF" is the largest slice of the pie, which is due to the complexity of the name system in ancient China. This category is therefore further analyzed, as illustrated in the secondary pie. The results show that "surname (S) + OSF", used widely in conversations, has the largest percentage.
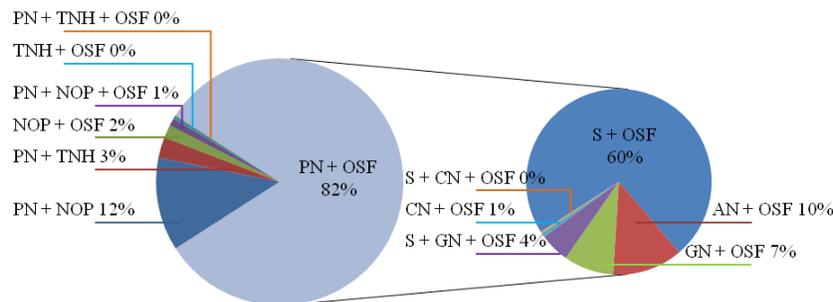

FIG. 6. DATA OF OTHER SUBORDINATE CONSTRUCTIONS

7. **Conclusion.** This paper classifies the personal names and terms of address in the classical Chinese corpus containing four famous novels at different levels from the perspective of named entity recognition and information retrieval. The classification system is established on the basis of full statistics and comprehensive analysis performed on the corpus. In the short term, the results have been employed to achieve an annotation system with feasibility, flexibility, and consistency, which is of great significance for the processing of Chinese classics. While in the long run, this study is expected to provide effective resources for the realization of intra- and inter-textual entity linking throughout the whole diachronic language knowledge base, which will consequently offer a convenient platform for future researches on literature and history.

REFERENCES

[1]   Xiong, D., Lu, Q., Lo, F.J., Shi, D.X., Chiu, T.S., Li, W.Y.: Specification for Segmentation and Named Entity Annotation of Chinese Classics in the Ming and Qing Dynasties. In: Chinese Lexical Semantics (CLSW2012 Revised Selected Papers). LNCS, vol. 7717, pp. 280-293. Springer, Heidelberg (2013)

[2]   Guo, P.: Erya Commentary (爾雅注疏). Peking University Press, Beijing (1999) (in Chinese)

[3]   Liang, Z.J.: Records of Addressing (稱謂錄). Zhonghua Book Company, Beijing (1996) (in Chinese)

[4]   Zheng, E.N.: A Review of the Study on Terms of Address in Modern Chinese over the Last 20 Years (近二十年來現代漢語稱謂語研究綜述). In: Journal of Chinese 2, pp. 120-122 (2005) (in Chinese)

[5]   Yu, S.W., Duan, H.M., Zhu, X.F., Swen, B., Chang, B.B.: Specification for Corpus Processing at Peking University: Word Segmentation, POS Tagging and Phonetic Notation (北大語料庫加工規範：切分•詞性標注•注音). In: Journal of Chinese Language and Computing 13 (2), pp. 121-158 (2003) (in Chinese)

[6]   Wei, P.C., Thompson, P.M., Liu, C.H., Huang, C.R., Sun, C.F.: Historical Corpora for Synchronic and Diachronic Linguistics Studies (建構一個以共時與歷時語言研究為導向的歷史語料庫). In: International Journal of Computational Linguistics & Chinese Language Processing 2 (1), pp. 131-145 (1997) (in Chinese)

[7]   Corpus of Academia Sinica, Taiwan, http://early_mandarin.ling.sinica.edu.tw/