A Study on the Evolution of Post-Nominal Plural Expressions in Chinese based on Tagged Corpora

1 Introduction

"Number" is one of the most important grammatical categories and its expression is usually associated with the type of a specific language. Different languages use either classifiers or number markers to indicate the category of number. Generally, in classifier languages, the extensions of nouns are collective and there are no plural markers. However, in Chinese, a classifier language that uses quantifiers, there also exists a plural marker "men" (们), which is used after common nouns and pronouns to indicate plurals. At present, the academia believes that "men", when used after nouns, can be divided into two types. One is to serve as a plural marker to indicate plurals (only indefinite plurals), when used after common nouns or common noun phrases; the other is to serve as a collective marker to indicate collections, when used after proper nouns or proper noun phrases. Plural expressions after proper nouns and common nouns are of distinct properties.

The plural expression "men" (们) did not emerge until in Pre-modern Chinese. So except for "men" (们), what were the plural expressions in Old and Middle Chinese? Were they also of two distinct properties? How did plural expressions of different properties come into being? Did these two types of plural expressions emerge at the same time or not? If one type came into being before the other, which one came first and were they related with each other in some way? These issues have not been systematically studied or fully discussed.

This article, basing its research on tagged corpora, intends to obtain an exhaustive collection of plural expressions that appeared in Old Chinese, Middle Chinese, and Pre-modern Chinese, and then according to statistics on word frequency, take the dominant plural expressions in various historical periods as the research object, conducting a quantitative survey and providing a systematic description of them in regard to their collocation with proper nouns and common nouns as well as their distribution, in order to present the trends and characteristics of the historical evolution of post-nominal plural expressions of different properties.

2 Tagged corpora and research methods

The research of this article is mainly based on the Academia Sinica Ancient Chinese Corpus (simplified as Sinica Corpus) in Taiwan, China, which has three sub-corpora, namely, the Old Chinese Tagged Corpus, the Middle Chinese Tagged Corpus, and the Pre-modern Chinese Tagged Corpus. So far the Sinica Corpus covers the majority of the most important linguistic data in Old Chinese (from Pre-Qin to Western Han), Middle Chinese (Eastern Han, Wei, Jin, Southern and Northern Dynasties), and Pre-modern Chinese (from Tang and Five Dynasties on), some of which have been tagged and are available to be retrieved online.

In the tagged corpora, common nouns are tagged as Na and proper nouns as Nb; post-nominal plural expressions are tagged as T (for auxiliary words) in the Old Chinese Tagged Corpus and T6 (for dependent words indicating plurals) in the Middle and Pre-modern Chinese Tagged Corpora. This article conducts a statistical survey of all the plural expressions that occurred in the three corpora, which are listed in Table 1 along with their frequencies.

<table>
<thead>
<tr>
<th>Tagged corpus</th>
<th>Post-nominal plural expressions and their frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Chinese Tagged Corpus</td>
<td>“deng”(等)/484</td>
</tr>
</tbody>
</table>

3 Trends of evolution of the post-nominal plural expressions based on tagged corpora

We calculate the average frequencies per ten thousand characters. In order to present the distribution of the post-nominal plural expressions across different historical periods in a more clear manner, we have transformed the data into bar graphs in Fig. 1.

![Fig. 1. Trends of evolution of the post-nominal plural expressions in different historical periods.](image)

4 Conclusion

This article, based on tagged corpora, conducts an exhaustive survey and analysis of plural expressions that occurred in the history of the Chinese language. Statistics show that the only plural expression in Old Chinese was the collective marker “deng”(等), which was only used after proper nouns. In Middle Chinese, especially in translations of Buddhist scriptures, “deng”(等) began to occur in large numbers after common nouns, acquiring the usage of plural marking. In this period, a new plural expression “bei”(辈) emerged, with a sole function of plural marking after common nouns. When it came to Pre-modern Chinese, “deng”(等), inheriting its usage in Old and Middle Chinese, was mainly used as a collective marker, while the function of plural marking was mostly taken by the newly emerged plural expression “men”(们).

Bing Qiu
Tsinghua Univ.
qiubing@mail.tsinghua.edu.cn