《现代汉语语言资料索引》第一辑序*

吕　叔　湘

武汉大学中文系和计算机科学系的同志们合作，把老舍先生的《骆驼祥子》全文存入了RD—11微型机软盘，并且利用计算机对《骆驼祥子》的语言资料作了自动加工处理。他们的软件系统可以进行自动查频、自动编索、自动检索、自动校对、自动统计标点及句长等工作。他们可以在计算机上对语言工作者提出的任意字、词、词组、短语、句子进行检索，打印含有这些字、词、词组、短语的全句原文。对我国语言研究的现代化，特别是研究手段的现代化来说，这是一件很有意义的事。

他们已经用计算机打印了《骆驼祥子》的《单字频度表》、《标点频度表》、《句长频度表》、《音序检字表》、《部首检字表》、《音序逐字索引》。现在，他们把《骆驼祥子》的《逐字索引》、《部首检字表》、《单字频度表》抽出来，作为《现代汉语语言资料索引》的第一辑印行。他们还准备逐步将老舍的其他著作和其他一些现代语言大师的著作存入计算机，分辑继续出版。这对于那些暂时还没有计算机的语言研究单位和个人将是极大的帮助，可以免除一大部分用手工搜集语言资料的劳动。对于研究文学的人，这样一套索引，也将是很有用的。

"索引"，过去又叫"通检"，有人根据英语index音译为"引得"。解放前，燕京大学有个引得编纂处，以哈佛燕京学社名义印行了引得六十三种；中法汉学研究所印过通检八种，又以巴黎大学北平汉学研究所名义印过通检六种；中华书局、商务印书馆也都在解放前和解放后印过一些索引。此外，台湾省、香港以及日本也编有多种汉籍索引。这些索引、引得、通检是为古代文献而作的，并且是人工编纂的。人工编纂索引手续繁复，容易出错，要进行大量的摘抄、校对、编号、排列、过录等枯燥劳动。在各种索引中，逐字索引的编纂又是最繁复的，所以除非十分必要，不肯轻易从事。现在有了计算机，就有可能编纂出版一整套逐字索引。

以上是就出版索引说的。如果不考虑出版，只是在机器里建立一个语言资料库以供人们自由检索，那末，研究语言的人就可以提出各种各样的要求，向计算机索取资料，索取过去纯靠手工劳动很难取得甚至不敢想象的资料。

我在《汉语语法分析问题》的《序》中曾说过有两件事要注意："由于汉语缺少发达的形态，许多语法现象就是渐变而不是顿变，在语法分析上就容易遇到各种'中间状态'。……但是这不等于说一切都是浑然一体，前后左右全然分不清，……积累多少个'大同小异'就会形成一个'大不一样'。"这是一件事。另一件事是"由于汉语缺少发达的形态，因而在做出一个决定的时候往往难于根据单一标准，而是常常要综合几方面的标准。……既然要综合几方面的标准，就有哪为主哪为次、哪个先哪个后的问题，就会得出不同的结论。"有了现代化的搜集资料的手段，能够采用统计法去分析研究，这两件难办的事就相对地比较好办一点儿了。

武汉大学的同志们要我写一个《序》，我就简单地说了上面那点意见。他们的工作在语言研究手段现代化这件事上做了一个良好的开端，我希望有更多的语言工作者和计算机专家结合起来，把这项有重大意义的工作推向前进，取得更丰硕的成果。

* 《现代汉语语言资料索引》第一辑将由武汉大学出版社出版。