摘要

概念空间/语义地图（简称 CS/SM）是语言类型学中的一个重要工具，也是一对有用的概念。概念空间体现跨语言普遍的功能采用相同形式进行编码的倾向性；而具体语言形式所特有的语义地图是对概念空间的切分，不同语言有不同的切分方法。其中基于图论的 CS/SM 是以图论作为数学模型，在经典模型上发展出来的边加权图是目前的主流方法之一。然而这种方法不无缺陷。一方面，尽管前辈学者针对各个功能范畴绘制了大量的概念空间及其在部分语言中的语义地图，但使用的数据来源是基于各个语言中对于某形式是否能表达某功能的偶值判断（能表达或不能表达），这些语料大多来源于参考语法，往往过于粗疏。另一方面，在边加权概念空间底图上表示具体语言形式的语义地图并不方便。为了解决这两个问题，本文结合单语语料库中的形义配对频数数据，通过“功能频率积”生成算法绘制更精细的概念空间。同时使用节点大小表示形义配对的频数，使得在边加权概念空间的底图上能够表示出具体语言形式的点加权语义地图。随后，以并列结构为个案研究，选取了以个体集合和命题合取为核心的 14 项与并列结构相关的主要功能，在语料库里考察了汉语普通话、吴语上海话以及英语这三种语言中的 22 个语言形式的功能分布，并在此基础上绘制了跨语言的边加权概念空间和 16 项有效形式的点加权语义地图。本研究的结果显示通过形义配对频数数据和新算法绘制的图符合语言学家对相关功能范畴的经验性认识，并且和前人的方法相比，效果更好。文章的最后部分讨论了加权图模型中的一个理论问题，即删边的必要性，指出删边并不是必要的。

关键词：概念空间；语义地图；语料库语言学；并列结构；点加权图

Abstract

Conceptual space/semantic map (CS/SM) is an important tool in linguistic typology. The conceptual space reflects the propensity for cross-linguistically universal functions to be encoded by the same form in a language, whereas construction-specific semantic maps carve the former. Graph-based CS/SM takes graph theory as its mathematical model. The edge-weighted graph model, which evolves from the classical model is one of the mainstream methods now. However, it has two shortcomings. On the one hand, although previous researchers have plotted a number of conceptual spaces and semantics maps of certain languages on it, the data sources they use are based on dichotomous judgment of whether a form can encode a function. These data mainly come from reference grammars and are too sloppy. On the other hand, it is hard to express construction-specific semantic map on an edge-weighted conceptual space. To solve these problems, this thesis takes into considerations the frequencies of form-function pairs in language-specific corpora, and plot a more delicate conceptual space with 'relative function frequency product' (RFFP) algorithm. In addition, the size of vertices is used to represent the frequencies of form-function pairs in an attempt to demonstrate construction-specific vertex-weighted semantic maps on the underlying conceptual space. Then I take the coordination constructions for case study, single out individual collection and proposition conjunction as two core functions with 12 other main functions, and investigate the function distributions of 22 linguistic forms in the corpora of 3 languages (Mandarin, Shanghainese and English). On that basis, the cross-linguistic edge-weighted conceptual space and the semantic maps of 16 effective forms are plotted. The result of this research shows the graph plotted with the data of frequencies of form-function pairs and the new generation algorithm fits the empirical studies of linguists. Besides, the new method is shown to be more delicate and effective by comparing it with the old methods. The last part of the thesis discusses a theoretical problem in weighted map, i.e., whether deleting edges is necessary. I show that it is not.

Keywords: conceptual space; semantic map; corpus linguistics; coordination constructions; vertex-weighted graph