摘要

本研究主要探讨汉语限定词和关系从句共现句式的位序选择问题，考察人们在实际言语运用中如何使用这一位序灵活的句式结构，以及影响该共现句式位序变换的支配因素何在，以期从理论和实证的角度更好地理解这一位序变换现象。

汉语的基本语序是 VO，但关系从句的核心名词却置尾，在语言类型学上较为罕见（Greenberg, 1966; Dryer, 1994）。此外，汉语较印欧语言又缺乏形态标记，而汉语量词要求与所修饰的名词在语义上匹配。但量词不能单独修饰名词，而需要跟指示代词/和数词一起，组成指量词/数量词（以下统称限定词），共同修饰名词短语。当汉语指/数量词与关系从句共同修饰名词时，主要有两种位序：一种是"限定词前置（DetP-first）"，即限定词+关系从句+中心名词；另一种是"限定词后置（DetP-second）"，即关系从句+限定词+中心名词。近年来，已有一些研究者考察这一共现结构在自然语言（主要是语料库）中的位序分布态势，结果得到较为一致的不对称分布态势，即：在主语提取的关系从句中，限定词倾向于前置；而宾语提取的关系从句中，限定词倾向于后置 （Pu, 2007; Wu, 2009; Tao & Ming, 2010; Sheng & Wu, 2013）。

但已有研究主要还存在以下两点不足： 1）在语料库分析方面，仅从单一变量视角进行因果分析（如，关系从句的提取类型，中心名词在主句中的语法位置等），然而在实际言语运用过程中，位序选择偏好是由多种因素共同作用的结果；2）在语言加工方面，以往研究主要集中考察前置限定词对关系从句加工的影响，而鲜有研究考察*限定词位置*和*关系从句类型*这两个变量对加工模式的影响。

鉴于此，本研究主要结合语料库分析和两个在线理解实验，从心理语言学的角度，考察影响汉语限定词在关系从句位序选择的各个因素的权重及其在实时加工的进程。在语料库分析方面， 本研究借助大型语料库（Chinese Treebank 8.0），运用自动检索软件穷尽式检索出目标句式，对可能影响位序选择的各因素（如关系从句类型、词长、中心名词在主句中的语法位置、中心名词和嵌入名词的生命性、

嵌入名词词汇形式、限定词类型等）进行编码。首先对这些因素作了较为全面的描述，在运用单因素描述统计分析的同时，采用多元回归分析的统计步骤，计算各因素对位序选择的权重。研究结果验证了"限定词根据关系从句类型呈不对称的分布态势"的这一结论。多因素回归模型结果显示：一、**关系从句类型、词长和从句嵌入名词词汇形式**为影响共现句式位序选择的主要因素；二、**中心名词在主句中的语法位置**和**从句嵌入名词的生命性**也会影响位序选择，但其作用相对较小。

为了进一步从语言加工的角度考察不对称分布态势的在线理解模式，在语料分析的基础上，本研究设计了两个在线实时言语理解实验，操控*限定词位置*和*关系从句类型*两个自变量（实验1和实验2），句子结构复杂度（实验2），因变量是各条件下不同区域的反应时间。具体考察以下三个问题：1）基于语料库分析得到的分布偏好是否具有心理现实性，即目标句式的分布频率是否与汉语母语者的在线阅读反应时数据模式一致？2）限定词在关系从句加工进程中的作用，是否存在指量词位置与关系从句类型的交互作用？3）句法结构（复杂度）是否影响关系从句在线加工？实验结果显示：1）在线阅读反应时模式与语料库中的不对称分布态势一致：主语关系从句中，限定词前置条件比限定词后置条件更容易加工；而宾语关系从句中，限定词前置条件比限定词后置条件更难加工，该研究结果支持（受频率驱动）基于经验的理论（Experience-based accounts）；2）有限定词的条件比没有限定词的条件阅读时间更快，该发现进一步证实限定词对关系从句加工的必要性（Wu & Sheng, 2016），且研究结论与基于工作记忆的理论预测不符 (实验2)；3）句法结构复杂度影响关系从句的在线加工，当限定词与关系从句共现结构嵌套在更为复杂的从句中，限定词对关系从句加工的重要性更为凸显。

最后，本研究结合语料库分析和在线句子理解实验得到的结果，试图运用并拓展"语言加工决定语法形式"（grammars are shaped by processing systems, MacDonald, 2013; Hawkins, 2014; Kurumada & Jaeger, 2015）这一假说，并将所有可能支配位序选择的限制条件都整合到 "交际效度"（Communicative Efficiency）这一个语言加工框架之中加以解释。

关键词：限定词短语；关系从句；语料库；在线理解

Abstract

Focusing on the relative ordering of determiner phrase (DetP) in Mandarin relative clause (RC), this dissertation aims to identify factors affecting the positioning of the DetP in RCs, thus gaining better theoretical and empirical insights into how two alternative forms of syntactic structure are chosen in language use.

RC construction in Chinese is typologically unique, because it combines the (S)VO word order and noun-final properties (Greenberg, 1966; Dryer, 1994). Chinese is also known as an isolating language with little morphological inflections. One exception is its numeral classifier system, in which a classifier must agree with its modified noun phrase. When a DetP sequence (consisted of a demonstrative/numeral and a classifier) co-occurs with an RC, the DetP can either precede or follow the RC, forming two flexible word orderings, namely, DetP-first and DetP-second constructions. In recent years, some researchers have begun to investigate the distribution pattern of DetP-RC co-occurrences, and found a relatively consistent asymmetrical distribution pattern in corpora, that is, SRCs prefer DetP-first constructions, while ORCs prefer DetP-second constructions (Pu, 2007; Wu, 2009; Tao & Ming, 2010; Sheng & Wu, 2013).

These previous studies are certainly informative, but thus far there has been lacking even a descriptively adequate account of determiner positioning for two different reasons. First, almost all of these studies examined the impact of each confounding variable (i.e. RC extraction type, the grammatical role of head noun) in isolation, by applying mono-causal analyses. However, given that the mind of the speaker engaged in a communicative situation, there are neither variables in isolation nor will the choice for a structure take place as a purely sequential comparison strategy. Second, little work has investigated the psychological reality of the observed asymmetrical distribution pattern by cross-manipulating the factor of *determiner position* and *RC type* in online comprehension tasks. To fill these gaps, this dissertation endevors to find out, through a corpus study and two self-paced reading experiments, the factors affecting the distribution of DetP in RCs, their respective weight in accounting for its positioning bias shown in the corpus, and how DetP-first vs. DetP-second influence real time processing of different types of RCs.

For the corpus investigation, I present a multifactorial analysis of 1462 cases of determiner and RC co-occurrence construction retrieved from the Chinese Penn Treebank 8.0. The data were firstly coded for eight predictors of determiner positioning, including relative clause extraction type, RC length, grammatical role of head noun in the matrix, animacy of head noun and embedded noun, determiner phrase type, and lexical form of the embedded noun. Other than conducting univariate analysis, I use a multiple logistic regression model to analyze the effects of these various factors. The modelling results show that overall, the choice of two different determiner positionings is mainly predicted by RC extraction type, RC length, and lexical form of the embedded noun. However, the grammatical role of head noun in the matrix clause, and the animacy of embedded noun play a secondary role in the choice of determiner positioning.

To further examine the interaction of determiner phrase position and RC extraction type on real-time processing of DetP-RC co-occurrence constructions, I also conducted two self-paced reading experiments, in which I manipulated *RC Type* and *DetP Position* (demonstrative classifier phrase) (Experiment 1 & Experiment 2), as well as the complexity of sentence structure (Experiment 2). Results show that i) the online reading comprehsension pattern exhibit an asymmmetrical pattern with respect to the ease of processing, which is consistent with the pattern observed from the corpus investigation, supporting the experience-based accounts; ii) the necessity of adding a determiner phrase for the RC-modifying head noun is confirmed; iii) the complexity of sentence struture where the relative clause is embedded do affect the ease of processing.

Taking together the corpus findings and comprehension pattern observed, this dissertation argues for an extension of the proposal that grammars are in part shaped by processing systems (MacDonald, 2013; Hawkins, 2014; Kurumada & Jaeger, 2015), specifically with regard to multifactorially determined phenomena such as the relative position of determiner and relative clause.

Key Words: determiner phrase; relative clause; corpus investigation; online comprehension