语言习得中的发现程序

赵 亮 著

Discovery Procedures in Language Acquisition
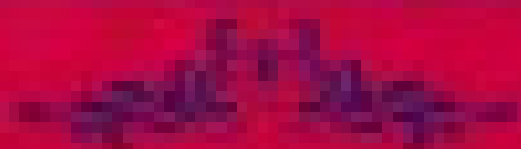
语言习得中的发现程序

赵 亮 著

Discovery Procedures in Language Acquisition

前　　言

本书从发现程序的角度探讨语言习得的机制。

"发现程序"由美国结构主义语言学家的代表人物之一Harris提出。贯穿Harris整个学术生涯的是以分布为标准的分析原则。他致力于设计一套明确的"发现程序"，用高度形式化的方法确定语言的基本单位：音位和语素，从复杂的语义中提取出语言形式中的规律和模式，减少语法的冗余。Harris曾经数次提到，语言学家们在直觉中知道语法规则，发现程序是用科学的、形式化的方法来代替简单的直觉。Harris的发现程序包括两个主要步骤，首先确定语言单位，然后综述这些单位的分布规律。

虽然Harris没有用发现程序来解释语言习得的机制，但是，他[1]也承认这方面可以作进一步的研究。发现程序的两个主要步骤看上去就是语言习得的机制。在语言习得中，儿童首先从言语输入中发现并存储重复出现的模式，然后把它们归类为有效的语言单位，接着合并为更高一级的单位。这样，每一个上一层次的语言单位都由下一级单位的分布规律来决定。用这种与Harris发现程序

---

[1] Harris, Zellig. 1986. *Structural Linguistics*. Chicago & London: The University of Chicago Press.

相似的方法，儿童能够完成自然语言的习得。

作为 Harris 的学生，Chomsky 发现，以分布和替换的发现程序作为对话语进行切分和分类的方法，只能对语言结构的表面现象作一定的描写，却不能解释语言的创造性、语言习得、语言知识和能力等重要问题。Chomsky 认为，语法是复杂的、未知的，不可能用一套机械的程序获得。试想，如果语法是未知的，怎么证实用来发现语法的发现程序本身是正确而且有效的呢？在此基础上，Chomsky 提出了评估程序，后来发展为普遍语法（Universal Grammar）。他[①②]举了几组例句，认为儿童掌握的语法知识不是学来的，用概括、类推等方法解释不了。因此，Chomsky 认为，儿童生来就具有一整套先天普遍的语法，儿童习得语言的过程是为普遍语法的"参数"赋值的过程。

这一生成语言学的理论基础遭到了一些语言学家的质疑，他们指出，Chomsky 的例子可以用一般的学习机制来解释，以此作为先天语法的证据是不充分的。

林允清[③④⑤]详细追溯并分析了 Chomsky 理论形成、发展的全过程。他指出 Chomsky 理论中的一些其他错误之处。最重要的一点是，Chomsky 和结构主义语言学家们一样，都混淆了"语法知识"（knowledge of grammar）和"语法知识是怎样习得的"（knowledge of how knowledge of grammar is acquired）这两个概念。林允清强调，每一个心智正常的成年人都清楚地、清醒地知

---

① Chomsky, Noam. 1975. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.

② Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. London: Praeger.

③ Lin, F. Y. 1999. "Chomsky on the 'Ordinary Language' View of Language". *Syntheses*, 120, pp. 151-191.

④ Lin, F. Y. 2000. "The Transformations of Transformations". *Language & Communication*, 20, pp. 197-253.

⑤ Lin, F. Y. 2002. "On Discovery Procedures". In B. Nevin (ed.), *The Legacy of Zellig Harris: Language and Information into the 21st Century*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

道自己本族语的语法，即结构主义语言学家们所描述的句型及转换规则。他认为，这些句型和转换规则是约定俗成的惯例，是人们讲述经验、表达思想的习惯方式。结构主义语言学家们的研究体会也证实了这一点。

在重新审视 Harris 研究的过程中，林允清发现，结构主义语言学家在自己身上强加了双重任务：提出发现程序和利用这些程序发现语法，正是这一点导致了这个学派的衰落。而在语法是已知的情况下，发现程序正是解释儿童语法习得的钥匙。林允清提出假设：儿童并不是生来就有一套先天的普遍语法，而是具有一种受遗传因素决定的先天的语言习得机制——发现程序，这些程序使得儿童能够在输入的语料中识别规则和模式，从而习得语法。而语言学家的任务则是在 Harris 研究的基础上发展发现程序。在这一新的语言学研究中，目标是明确的：发现儿童天生的语言习得机制——发现程序；方法也是明确的：用已知的语法知识去测试设计的程序正确与否。"发现程序"学说没有任何哲学基础或方法论上的缺陷。

本书的理论基础正是林允清提出的"发现程序作为儿童语言习得机制"设想。我们讨论了语言和发现程序的本质。语言是约定俗成的惯例，是人们讲述经验、表达思想的习惯方式，每个人都清楚地知道本族语的语音和语法；而发现程序被设想为先天的语言习得机制。本书的目的是验证和发展这一设想。

首先，儿童语言发展中的研究成果被用来对比和补充发现程序发现语音和语法的过程。我们设想儿童习得语言的过程跟早期的语言学家找到语言中的音位和语素类似。由任意把言语切分成单位开始，再根据更多的言语输入来调整，发现并存储重复出现的模式，把它们归类为有效的语言单位，接着合并为更高一级的单位。通过反复试验，儿童逐渐习得母语的语音和语法结构。最小对比对不仅对语言学家发现语音和语法非常有用，对儿童习得语言也同样重要。当儿童能够在日常交流中判断两个看似相近的

言语是否相同的时候，他们已经开始习得语言了。我们进一步设想，在两词话语短语结构的基础上，儿童通过替换、扩充、移位、同现、结合、嵌入、错误纠正等机制，习得基础句法结构，并不断叠加复杂化。

其次，为了证实和发展这一理论假设，Harris 的发现程序将被应用于语料分析。在一个英语语料的背景下分析一篇英语短文，生成不同等级的语法单位，并与我们已知的英语语法相对比。我们的分析从最容易识别的单词入手。一个单词包括一个或多个语素。我们分析短文中的单词在语料中的分布并进行对比，查找规律。本书用到了 Harris 的三个程序："独立和模式化组合"程序找到可能切分的语素段，"平衡独特的语素段"程序把互补的一系列语素段视为一个语素，"大体相似的环境"或"在环境中的语素类"程序将分布规律相似的语素归类。通过与已知的语法相比对，发现程序被进行进一步的补充和调整。

最后，本书设计出两个计算机程序来模拟发现程序发现语言规律和模式，用严谨的方式来验证发现程序作为语言习得机制的可行性。第一个计算机程序从语料中读出不同的字母、数字和标点符号，并计算它们的出现次数。第二个计算机程序能根据我们的设定找出语料中含有一个或多个后缀的一组组单词，并判断词根是否在语料中作为单词单独出现。例如，如果我们输入后缀：-ent, -ently, -ence，在语料中我们能够得到类似 confident, confidently, confidence; different, differently, difference; present, presently, presence 这样一组组的单词，并判断字母组合 confid，differ，pres 在语料中是否是一个独立的单词。这一程序帮助我们识别语素的分布规律，并把有相似规律的语素进行归类。这两个程序只是一个前期的工作，我们的目的是设计出一个计算机模型，模拟语言习得的全过程，使得计算机能够在大量的语言输入的基础上自动生成语法。这可以进一步证实发现程序作为语言习得机制理论的可行性。

　　发现程序作为语言习得机制这一设想提供了另一个解释人类语言秘密的途径，而这一新的理论仍处于起始阶段，值得我们进一步研究。

<div align="right">

赵亮

2013 年 7 月

</div>

List of Abbreviations

A	Adjective
Dloc	Locational Adverb
MP	Minimalist Program
N	Noun
NCE	*New Concept English*
P	Preposition
T	Article
TG	Transformational Grammar
UG	Universal Grammar
V	Verb
Vit	Intransitive Verb
Vth	Copula

Contents

Chapter 1

Introduction

Language is immensely complicated. It is seen as a system of signs from different angles, which is affected or controlled by certain external factors. General Linguistics ranges from phonology, morphology, syntax, semantics, to pragmatics. And it has close relationship with physiology, neuroscience, psychology, cognitive science, sociology, culture, etc. However, explaining how children acquire during the course of language development is even much more difficult than explaining what children acquire.

Language acquisition is a fascinating phenomenon. It could be "magic" (Bloom, 1983), "mysterious" (Gleitman & Wanner, 1982), and "emergent" (MacWhinney, 1999). Many challenging theoretical and practical questions in this field have tortured researchers for years. Thereinto, children's unique ability to acquire phonology and grammar remains at the heart of linguistic inquiry. How do young children recognize and produce the sounds of their language? How do they

manage to break up the steady stream of sounds they hear into the basic units like morphemes and words? How do they learn to figure out the basic grammatical categories of their language such as nouns, verbs, and adverbs? How do they learn to combine basic units into constructions and sentences? What is the nature of syntactic structures?

1.1   The Research Topic

This book is a theoretical study of language acquisition mechanism from the perspective of discovery procedures. The author adopts Harris's discovery procedures to interpret how the child acquires distributional features of phonology and grammar. Formal features of language are the emphasis of this study.

There is a view that the understanding of language requires both an understanding of the formal composition of utterances and of their relations with the rest of the world outside language, i.e. form and meaning. Harris (1986) suggested that many aspects of meaning actually correlate with certain distributional regularities. Thus, in many cases what can be obtained through meaning can also be obtained through formal procedures based on distributional analysis. As for the intervention of semantics, it is desirable to connect utterances and elements with social situations. It seems sufficient to define meaning in such a way that utterance differences in morphemic constituency can be considered as differences in meaning, and that the differences in meaning are assumed to indicate differences in social situations in

which these utterances occur. However, the correlation of morphemic segments with social situations could not be used in establishing the segments. There is at present no way of determining meaning differences as exactly as one can measure sound differences, and there are no morphological tests of hearer's response to meaning comparable to the phonological test of hearer's response to sound.

The formal method is unambiguous, consistent, and subject to check because there is no opportunity for uncontrolled interpreting of the data or for forcing of the meaning. It has also been noticed that form of sentence patterns is independent of the meanings of their components. According to Bolinger (1968: 127), "a difference in syntactic form always spells a difference in meaning". Moreover, many studies and facts (e.g. Curtiss, Yamada & Fromkin, 1979; Curtiss, 1981; Wang, Wu & Wang, 2000; Beck, 2002) showing that the acquisition of syntax and morphology must be somewhat independent of other cognitive developments and the representation of syntax and morphology in human brain may be different from other domains.

We assume that no matter how many kinds of external factors that affect or control the language and how the internal factors interrelate with each other, its grammatical structure, which is viewed as distribution regularities and patterns of language units in this book, could be described independently.

Language use itself is a unitary activity. Grammar and pronunciation are the abstractions from language use that are employed in the description and analysis of linguistic form. There are three dominating viewpoints on the nature of the abstractions: realism, conceptualism, and nominalism. This book agrees that the nature of linguistic abstractions could not be detached from an innate biological

human cognition basis; language acquisition could not be detached from the cognitive intention to communicate with others. Based on these, we also hold a so-called "ordinary language" view of language (Lin, 1999), which is held by Wittgenstein, Strawson, Dummett, Searle, Putnam, Lewis, Wiggins, and others. According to this view, a language consists of conventions; it is rule-governed, language rules are conventionalized; a language is learnt; there are general learning mechanisms in the brain.

Some ideas from linguistic philosophy of Wittgenstein in his late period are the philosophical basis of this study. Language does not refer to real entities or mental activities. Language meaning and functions depend on its occurring environments; the same words are different in different usages. Following language rules is a behavior rather than interpretation. There is no hidden essence behind language. Family resemblances are used to characterize similarities of different things which have common names. Language rules are uncertain, unclear, and variational according to specific environments. There is no common or stable essence in language games.

According to Wittgenstein (1958: 46e), "what we call 'sentence' and 'language' has not the formal unity that I imagined, but is the family of structures more or less related to one another". The notion of family resemblance is calling for a notion of conceptual distance, which is closely related to the idea of graded sets. This notion is always related to the recent prototype theory, which will be explained in Section 3.1.2.

Since language rules in language games are uncertain, unclear and variational, the discovering process is persistent during one's life in speech perception. However, grammar analysis is possible and necessary since grammar or the formal features of language are

comparatively stable in one period of time. Innate biological and mental base of human being is even more stable than the formal features of language, which is the premise of language acquisition mechanism research.

Different from Harris's grammar, we assume that there are no transformations in grammar. The discovering results of discovery procedures are formally different sentence types. They may have some relations in meaning, which are not considered in this study. In fact, phonological and grammatical regularities are interpreted using family resemblances and prototype theory in this book, which will be discussed in Section 3.1.

The aim of this book is to testify and develop this new theory—discovery procedures as language acquisition mechanism. First, discovery procedures as innate language acquisition mechanism will be verified and modified by characteristics in language acquisition course. Then the designed discovery procedures will be used to analyze an article based on a corpus to discover grammatical elements, which will be compared with the existent English grammar. Finally, this process of discovering will be simulated using computer programs. All these steps will verify and develop that "discovery procedures" is one among many plausible ways to interpret language acquisition mechanism.