

# 仅供参考

## 冯志伟序

“依存关系”和“依存语法”一直是刘海涛博士多年来关心的问题，他广泛收集了国内外有关依存语法的著作细心研读。这些著作涉及多种不同的语言，除了汉语和英语的著作之外，还有德语、法语、俄语、荷兰语、日语等外语写成的著作。为了读懂这些著作，他还学习了德语、法语等外语，具备了一定的阅读多种语言原著的能力，获取了第一手的资料。由于他尽量从原著中获取资料，他关于依存语法的这些论述，自然就具有很强的说服力。

海涛并不满足于依存语法的理论研究，他还勇于实践，建立了一定规模的依存树库，使用计算机来验证他提出的基于配价模式的句法分析理论，这就使他的研究，不仅在理论上翔实可靠，在实践上也经得住检验。

我认为海涛的这本专著具有以下三方面的特色。

第一，系统地梳理了国内外依存语法理论研究的历史，提供了内容丰富、翔实可靠的大量资料，有助于我国学术界全面地理解依存语法这一重要的理论。

20世纪70年代末期，我在法国格勒诺布尔理科医科大学应用数学研究所研制汉—法/英/日/俄/德多语言机器翻译系统 FAJRA 时，为了克服乔姆斯基(N. Chomsky)的短语结构语法在汉语自动分析中的困难，我的导师沃古瓦(B. Vauquois)教授和巴黎第七大学的法国朋友格罗斯(M. Gross)教授都建议我读一读泰尼埃(L. Tesnière)的《结构句法基础》(*Éléments de Syntaxe Structurale*)，这是我第一次接触“依存语法”。我用两个多月的时间认真阅读了该书的法文版，眼界大开。在我提出的中文信息 MMT 模型(多叉多标记树模型)中，我根据依存语法，确定了汉

语句法—语义分析的“行动元”(actants)和“状态元”(circonstants),这些表示依存关系的标记成为 MMT 模型的多标记中的最重要的标记。我还根据“价”(valence)的概念,赋予汉语动词不同的“价”标记,通过动词的“价”来控制句子的句法—语义结构,并吸收依存语法中“支配”和“被支配”的理念,给每一个短语都标注出“中心词”(标记为 gov),给每一个句子都标注出“轴心”(标记为 pivot),明确了短语和句子中的“支配”和“被支配”关系,大大地提高了 MMT 模型的性能。从法国回来之后,我马上就在当时的《国外语言学》杂志上写文章介绍了泰尼埃的语法理论(冯志伟 1983)<sup>1</sup>,这是国内最早系统介绍依存语法的文章。

当时我只注意到动词的“价”,没有注意到名词和形容词的“价”。上世纪 80 年代我有机会到了德国,访问了曼海姆的德语研究所(IDS),才知道德国语言学界除了研究动词的“价”之外,还研究了名词的“价”和形容词的“价”。德国语言学家早就出版了《德语动词配价词典》、《德语名词配价词典》和《德语形容词配价词典》,德语研究所的托依拜特(W. Teubert)在 1979 年出版的《名词的价》,就是他在海德堡大学的博士论文。这时,我才认识到不仅仅动词有“价”,名词和形容词也有“价”,我开始把依存关系看成是普遍存在于语言中的一种支配与被支配的关系,并且深深感到自己过去对于依存语法的学术发展的情况实在是了解得太少了。

近年来,我国出现了研究配价理论和依存语法的热潮,在具体的细节研究方面取得了很好的成果,但是,在系统性和整体性方面还显得很不足。

海涛在这本书中,根据大量的历史资料,为我们勾画出了依存语法和依存关系的观念发展的历史脉络,告诉了我们如下鲜为人知的重要历史事实:

---

1 当时我把 Tesnière 翻译成特斯尼耶尔,不妥,按照法语人名的规范读音,最好翻译为泰尼埃。

早在12世纪,语言学家赫利亚斯(Petrus Helias)在他的著作中就提出了“动词中心说”,他认为,动词要求的句子成分的数量是不同的,动词的必有成分一般是指名词性的,这些成分是构造一个 perfectio constructionis 所必需的。这种“动词中心说”指出了动词对于句子成分的要求,已经隐含了“配价”的理念。

德国普通语言学家梅讷尔(Johann Werner Meiner)在1781年的著作里就明确将谓语(动词)分为:一价动词,二价动词和三价动词,只不过他没有直接使用“价”这个词,而是用了一个德语词“seitig-unselbständig”,但是其实质已基本无异于现代人定义动词“配价”的说法了。

1934年,奥地利语言学家卡尔·比勒(Karl Bühler)在其《语言理论》中说,“每种语言中都存在着选择亲缘性;副词寻找自己的动词,别的词也是如此。换言之,某一词类中的词在自己周围开辟一个或几个空位,这些空位必须由其他类型的词来填补。”卡尔·比勒关于“空位”的见解,揭示了“配价”的本质。虽然他没有使用过“配价”这个词,但是国外研究配价理论的学者普遍将卡尔·比勒看做是配价理论研究的先驱。

1948年,苏联语言学家科茨年松(Kacnel'son)首次提出“配价”这个术语。他说,“在每一种语言中,完整有效的具体化的词不是简单的词,而是带有具体句法潜力的词,这种潜力使得词只能在严格限定的方式下应用,语言中语法关系的发展阶段预定了这种方式。词在句中以一定的方式出现以及与其他词组合的这种特性,我们可以称之为句法配价。”

科茨年松特别强调“配价”的“潜在性”。他认为,明显的语法范畴、功能和关系是“通过句法形态来表现的”,而在词的句法组配和语义中隐含了潜在的语法范畴、功能和关系。他还说,“语法如同一座冰山,绝大部分是在水下的。”

1949年,荷兰语言学家格罗特(A. W. de Groot)在他用荷兰语出版的《结构句法》(Structurale Syntaxis)一书中也使用了“配价”这一概念。格罗特在他的书中写道,“与其他词类相比,某些词类的运用可能性受到

限制，即词类具有不同的句法配价。配价是被其他词所限定或限定其他词的可能性或不可能性。”他在句法研究中使用了“valentie”和“syntactische valentie”这两个术语。

这些鲜为人知的历史资料零星地隐藏在国外文献的汪洋大海之中，互联网上也不容易搜索到，获取它们犹如大海捞针，而且，这些文献涉及德文、俄文和荷兰文等不同的外文，要读懂它们需要丰富的外语知识，特别是格罗特的《结构句法》(*Structurele Syntaxis*)一书，尽管过去我曾有所耳闻，但由于是荷兰语写的，一直不敢问津，海涛在外国朋友的帮助下，弄清楚了荷兰文的原意，使得我们有机会了解到这本重要文献的内容。海涛做的这些钩沉探源的工作是非常有意义的。

我一再对我的博士研究生们说，“治史须读原著”，鼓励他们尽量阅读外文原著。我们绝对不能仅靠翻译成中文的材料来研究国外语言学的历史，因为可能会出现翻译错误，因错就错；我们更不能仅靠道听途说的资料来研究国外语言学的历史，因为可能会以讹传讹，谬种流传。因此，阅读原汁原味的外文原著是非常重要的，特别是对于研究语言学的博士生来说，我认为是一必不可少的。既然是博士，就不能像其他人那样只依靠中文译文来进行研究，既然是语言学的博士，就应当在外文水平方面胜人一筹。海涛坚持了“治史须读原著”这个原则，严格要求自己，进行了不懈的探索，终于揭开了语言学历史上这些鲜为人知的神秘面纱，这是令人高兴的。

第二，全面地讨论了依存语法的形式化方法和句法分析算法，提出了“概率配价模式”，这个模式不仅可以定性地描述依存树中的支配和被支配关系，而且，还可以定量地计算依存关系的强度。

海涛并不满足于对历史事实的考察，他考察历史的目的是吸取前人的学术成果，“古为今用”，建立现代汉语的依存句法，因此，他更多的工作是研究依存语法的理论。

他讨论了依存语法的形式化方法和句法分析算法，并根据依存语法

的基本原则，深入地研究了汉语依存树库中众多树形图节点之间的支配和被支配的依存关系，构建了现代汉语配价模式；为了表示这种依存关系的不同强度，他在模型中引入了概率成分，提出了“概率配价模式”，在配价模式图中，用粗细不同的线条来直观地表示依存关系的强度，这样，就可以从定性和定量两个方面来描述树形图中的依存关系。在这本书里，海涛也简单介绍了如何通过依存树库来提取配价模式的方法。

在国内外依存语法研究中，大多数都只关注依存语法的定性研究，还没有明确地提出“概率”的概念，“概率配价模式”是海涛的创新，这个模式使我们有可能对依存关系进行定量的研究，这是海涛对于依存语法研究的新贡献。关于“概率配价模式”，我和海涛在2007年写了一篇论文，发表在《语言科学》上，有兴趣的读者可以阅读，以便加深对于本书的理解。

第三，通过一定规模的依存树库，检验了作者提出的基于配价模式的句法分析理论，并对汉语进行了初步的定量分析，发现汉语的依存距离远远大于英语、德语和日语的依存距离，这一发现有助于推动语言复杂网络研究的进展。

海涛构建了包含13个词类和34种依存关系的现代汉语依存句法，提出了汉语的树库格式，标注了一个包含两万词次的、实验性的汉语依存树库，同时还使用了哈尔滨工业大学的依存树库，在这样的基础上来检验他的理论。他采用了XDK和MaltParser等软件对汉语进行自动句法分析实验，并且在实验中适当地调整对某些语言现象的处理方法和标注的精细程度，把计算机的自动句法分析与语言学家的语言知识结合起来，有效地改善了句法分析的效果。

海涛的研究还发现，汉语的依存距离为2.81，远远大于英语和日语的依存距离（英语为1.386，日语为1.43）。这是海涛的一个重要发现，而这样的发现是使用其他的语言研究方法难以做到的，这从另一个侧面说明了依存语法的长处。

依存距离的研究有助于把语言作为一种复杂网络来进行研究。目前关于复杂网络的研究结果表明,语言网络是一个无标度的复杂网络(scale free complex network),是一个“小世界”(small world)的复杂网络。采用复杂网络来研究语言,有助于把语言与其他的复杂网络相比较,可以提高语言研究的普适程度,这方面的研究的价值是不言自明的。在这一方面,本书也有一定的探索。希望海涛能继续进行这种研究,把语言研究与复杂网络研究结合起来,做出更多的创新。

我和海涛认识已经将近30年了,他原来是学习自动化的,他的本职工作是从事企业信息化的高级工程师,是一个语言学的业余爱好者,与我常有书信往来,利用业余时间探讨语言学的各种问题。进入21世纪以后,他出于对语言研究的热爱,毅然改变了原来的专业方向,潜心投入清苦而艰巨的语言研究中,成为一个专业的(计算)语言学家。

在当今经济大潮下,很多人都忙于赚钱,并以此为乐,海涛凭着他的技术水平和外文功底,在这一方面也不乏机会。但他却选择语言学作为他的努力方向,宁愿与我等这样收入菲薄的语言学家为伍,下决心坐冷板凳,以探索学问作为自己的乐趣。这种品德,是值得我们学习的。

海涛有很好的自然科学和工程技术的基础,又有广博的语言学知识,对于语言研究充满了热情,这正是从事计算语言学研究的极好条件。本书是他出版的第一本专著,我对他表示热烈的祝贺,特作此序,算是我阅读此书的粗浅体会。希望他不断努力,今后做出更多的创新,出版更多的著作。

冯志伟

2009年5月7日于德国海德堡