

北京日本学研究中心学术专著 (八)

汉日主题句结构

对比研究

施建军 著

——兼论主题句的计算机处理

世界知识出版社

北京日本学研究中心学术专著（八）

汉日主题句结构对比研究

——兼论主题句的计算机处理

中日主題文構文文型

の対照研究

——主題文の計算機処理をかねて

施建军 著

世界知识出版社

责任编辑: 林 昶 王 江

封面设计: 孙 敏

责任出版: 夏凤仙

责任校对: 何 莉

图书在版编目(CIP)数据

汉日主题句结构对比研究: 兼论主题句的计算机处理 / 施建军著. —北京: 世界知识出版社, 2001.5

ISBN 7-5012-1490-5

I. 汉… II. 施… III. 句对句翻译(机器翻译) —对比研究 —汉、日 IV. H365.9

中国版本图书馆 CIP 数据核字(2001)第 22991 号

汉日主题句结构对比研究

Hanri Zhutiju Jiegou Duibi Yanjiu

世界知识出版社出版发行

(北京市东城区干面胡同 51 号 邮政编码: 100010)

网址: <http://www.wapbook.com>

北京市朝阳展望印刷厂印刷 新华书店经销

850×1168 毫米 32 开本 印张: 8.5 字数: 198000

2001 年 5 月第 1 版 2001 年 5 月第 1 次印刷

ISBN 7-5012-1490-5 / G·455 定价: 13.60 元

版权所有 翻印必究

“北京日本学研究中心学术专著”

出版说明

北京日本学研究中心是中国国家教育委员会与日本国际交流基金会与 1985 年合作创办的、研究生院级的日本学教学和研究机构。设有日本语言、日本文学、日本社会和日本文化四个专业，招收硕士和博士研究生，并具有外国语言文学博士后流动站招生权。日研中心在教学上实行中日联合培养，在研究上实行个人课题和国内外合作研究项目并重。中心研究部设有日本语言、日本文学、日本社会和日本文化四个研究室，并相应地建立了客座教授和客座研究员制度，每年从国内的日本学及相关人文社会科学界聘请十几名专家学者和中青年研究人员来中心从事教学和专门研究。经过十年经营和积累，中心的教学和研究均已规模初具，收获在望。为了向国内外介绍中心的研究成果，促进学术交流，繁荣学术研究，我们特地组织编撰、联系出版“北京日本学研究中心学术专著”系列丛书。主要为中心研究人员、历届毕业生、历届客座教授和客座研究员以及中心学术协作网的专家学者提供最新研究成果的发表园地。我们希望通过这一项新的学术工程，为中国的日本学大厦不断地添砖加瓦。

北京日本学研究中心

序

本书的研究课题是《中日主题句结构对比研究——兼论主题句的计算机处理》。首先，主题句研究这一课题可以说是中日语言对比研究中的一大热门课题，也是难点所在。在日语中，除了有一般语言概念中的“主语”以外，还存在一种不同于“主语”的“主题”这样一个重要的语法范畴。这一问题在日本经过近百年来的争论，基本得到了澄清和认可。这是因为日语在语言形态上，两者有着比较明显的区别，即助词“は”与“が”的区别。当遇到像“象は鼻が長い”这样的句子时，人们不仅能从意义上辨明“象”和“鼻”在句子中起的语法作用的不同，而且可以从其后续的助词“象は”、“鼻が”的形态上，很容易证明它们在语言形态上也是不同的。另外，经过研究证明，日语中助词“は”前面出现的名词，决不仅限于助词“が”所表示的句子谓语所言性质的主体或动作的主格，而可以有多种多样的内涵。尽管直至今日，日语自身语言中助词“は”与“が”的区别问题尚未得到彻底的解决，但终归在“主语”和“主题”研究方面，已经取得了比汉语深得多的研究成果。

然而在汉语中是否存在“主题”，这还是一个很值得研究的重要问题。在“妹妹眼睛很大”这样的句子中，是把“妹妹”

看作“眼睛”的修饰语还是看作句子的“主语”？如果将其看作“主语”，那么句中的“眼睛”又是什么语法成分。在这一问题上，汉语研究领域里还存在很大的争论。但是，如果把这样的句子与日语的“象は鼻が長い”这样的句子相比较，就不难看出，二者在句子结构上有极其相似的地方。因而可以说将两种语言中存在的这样的句子从结构上进行对比研究，一定会对进一步深入研究汉语句子结构有重大的贡献。

本课题的研究者施建军于1995年9月考入北京日本语研究中心博士课程，由严安生教授指导。由于他的专业是语言专业，我也就参与了他的论文指导。施建军在学期间，学习非常刻苦，对遇到的问题一丝不苟，表现出了年轻学者的一股韧劲儿。对于我们这些专门从事语言研究的人来说，施建军还有着自己的一技之长，即他大学本科学习期间，曾经受过全面系统的计算机方面的专业教育，硕士课程学习的是日语语言文学专业。语言与计算机的结合，使他如虎添翼，给他的研究开拓了更加广阔的天地。90年代中期，又正好是计算机行业在中国迅速普及的时代。在这样一个时代，选择了这样一个课题，可以说是充分发挥出了施建军的专业特长，也为计算机研究与我国日语研究的相结合闯出了一条新路。

本书正是在施建军博士论文的基础上改写而成的。他在本书中全面探讨了中日主题句研究的问题。首先回顾了日语主题句研究的百年历史，分析了汉语主题句研究的现状和问题，通过与日语的对比研究，明确提出了汉语语法中“主题”的概念。并对中日语言中主题句的相同、相似之处与不同之处分别进行了分析研究。这种比较研究不仅对汉语句子研究是一种新的尝试，反之对日语主题句的深入研究也有积极的意义。最后，施

建军有提出了由计算机自动抽取日语主题句的模式。这种模式的提出，对日语自然语言处理的研究，以及对中日语言机器翻译的研究都是非常有意义的尝试。

施建军除完成了博士论文以外，在北京日本学研究中心还积极参加了我中心申报的《中日对译语料库的研制和应用研究》国家社会科学基金项目，在项目执行中发挥着重大的作用。我相信，此次本书作为“北京日本学研究中心学术专著”出版，不仅为我中心增加了新的研究成果，也一定会为中国的日语研究乃至中日语言对比研究做出新的贡献。

北京日本学研究中心主任

北京外国语大学教授、博士生导师

徐一平

2001年3月20日于北京

前 言

汉语句子的成立一直是一个悬而未决的问题。一般认为，句子是由表示事情的客观成分和表示态度的主观成分两部分构成的。汉语由于缺乏形态特征，因此，很难从形态上找到句子成立的证据。而日语却有丰富的形态特征可以把握，相对汉语而言人们对日语句子的认识比较深入。汉语和日语中存在着许多极为相似的现象，作为句子成立条件的主题问题就是其中之一。在解决这些问题时相互借鉴对方的研究方法和成果，有时可以起到“他山之石”之功效。

日语有“象は鼻がながい”、汉语有“她眼睛很大”一类的句子，这些句子的结构汉语和日语有相似之处。日语由于存在着“は”“が”这种形态上的差别，因此主题和主语的问题已基本得到了解决，而且有关主题句句型系统的研究也相当深入。而汉语由于缺乏形态上的差别，“主语”问题一直是汉语学界争论的焦点。

汉语中处于句首的名词性成分和主谓结构中动词或形容词前面的名词性成分过去一直被认为是主语。在对汉语的主谓谓语句和定语从句进行考察时，我们可以发现处于主谓结构中的动词前面的名词性成分和句首的名词性成分的句法功能是有差别的。汉语句首的名词性成分和处于句子叙述部中

的谓语动词等有着各种各样的格关系,它的势力可以贯穿整个句子,将多个叙述部统一在一个句子中,而主谓结构中动词或形容词前面的名词性成分和动词或形容词之间的关系比较单一,一般只能是动作的执行者或者是状态、性质的主体,而且这些名词性成分的势力只局限于主谓结构当中,而不能超越主谓结构。

日语的主题句和汉语的主题句在结构上有相似之处。这主要表现为两种语言的主题都可以由格成分形成,也可以是格成分以外的成分形成的,主题句的类型也基本上一样。但是,两种语言主题句的主题形成过程有不同的地方。在格成分转化成主题时,汉语的凭借格中只有表示方式的格成分不能转化成主题,而日语中表示凭借的“で格”均不能转化成主题。汉语和日语的领格虽然都可以转化成主题,但是转化过程不一样。格成分以外的名词谓语的定语从句和动词结构转化成主题时,日语为了使句子简洁明了,经常用这样的句子,但是,汉语的这种转化是有条件的,通常只有这些成分很简单时才可以转化成主题,或者是原来的主格(主语)成分很复杂时谓语节才可以转化成主题。这是由于中国人尽量避免“头重脚轻”的语言意识引起的。这是中日语言差别的根本之所在。

以上所有差别通过两种语言的对比研究可以看得十分清楚。这也许是对比语言学研究方兴未艾的重要原因。

近年来,随着信息处理技术迅猛发展,自然语言处理也对语言学研究提出了许多新的课题,语言学研究的滞后已成了制约语言信息处理发展的瓶颈。我在对中日语主题句系统进行对比研究的同时,还对主题句的计算机自动抽取作了一些尝试,提出了以句节为单位、形态和词典相结合的自动分

词和标注算法，并且开发出了主题句自动抽取、自动分词、自动标注的实验系统。

本书顺利完稿，是与许多前辈、朋友的鼓励、支持分不开的。该课题的研究是在恩师严安生教授、徐一平教授悉心指导下完成的，可以说本书也倾注着他们的心血。北京日本学研究中心曹大峰教授、上海外国语大学戴宝玉先生、北京外国语大学于日平先生、东京外国语大学平井和之先生都给予了很多有益的启示。在本书出版过程中得到了中国社会科学院日本研究所林昶先生、世界知识出版社王江先生、洛阳外国语学院徐明淮教授的大力支持。本书付梓之际一并表示感谢。

另外，已故恩师日本国立国语研究所日本语教育中心前主任中野洋先生生前对该研究寄予厚望，并在资料和技术上给予了大力帮助，在此对中野洋先生表示深切怀念。

施建军

2001年1月于古都洛阳



施建军 1966年

3月生，江苏南通人。

1984年入洛阳外国语学院科技日语专业学习，先后获理学士、文学硕士学位。1995年入北京外国语大学日语语言文学专业，1999年3月获文学博士学位。现为洛阳外国语学院日语副教授。主要从事计算语言学、汉日对比语言学研究，发表论文20余篇。

目 次

第一章 文型の対照研究	1
第一節 緒 論	1
第二節 文型対照研究の意義	6
第二章 主題文に関する先行研究	20
第一節 日本語の主題をめぐる論争	20
第二節 中国における主題についての論争	39
第三節 中日両国における主題論争の相違と 本研究の位置づけ	56
第三章 中国語の新しい文法カテゴリー——主題	66
第一節 主語と対立する主題	66
第二節 主題文とは何か	91
第三節 主題と主語の取り方	96
第四節 従来の主題説との相違	102
第四章 中日主題文の対照研究	118
第一節 格成分と文の直接の成分	118
第二節 中国語の格成分	123
第三節 中国語の格成分の主題化	130
第四節 格成分以外のものの主題化	158

第五章 コンピュータによる日本語の	
主題文の抽出	187
第一節 単語の分割と品詞の認定	189
第二節 文型抽出の試み	203
第三節 文型抽出の難しさ	208
第四節 文型抽出のステップ	214
第五節 主題文の抽出	216
第六章 中日主題文の特徴	220
 付録1 MCL結果の整理プログラム	227
付録2 辞書のソートプログラム	230
付録3 実験用テキスト	233
付録4 実験用辞書1 (実詞)	235
付録5 実験用辞書2 (虚辞)	238
付録6 表記解析プログラムの解析結果	248
付録7 形態素解析の結果	250
付録8 抽出された文型	253
 参考文献	257

第一章 文型の対照研究

第一節 緒 論

日本語は難しいと言われる。外国人にとって習得困難な言語の一つと言える。だが、その難しさはいったいどこからくるのかを考えて見ると、さまざまな要因があるように思われる。日本語の構造は学習者の母語にかなり近似しているところもあるだろうが、かなりずれているところもあるだろう。それも音節、語彙、文法など広い分野に渡って存在しているはずである。と言っても、そのすべてを研究の土俵に載せることはもとより望むべくもないので、ここでは、主として、文法に焦点を当てて、日本語の構文について掘り下げて考えてみたい。

外国語として勉強するとき日本語の難しさは、学習者が母語と比べて初めて気づくものだと思われる。外国語はその構文と学習者の母語のそれとの差異が大きいほど難しいと思われる。では、なぜ、日本語が中国人の学習者にこれほど困難に感じられるのか。その理由を明らかにしようとするのがこの課題に取り組んだ動機の一つである。もう一つの動機は大学時代に日本語と並んでコンピューターにつ

いて学び、コンピューターに深い関心を持つ筆者が中国語と日本語の構文的な差異を明らかにすることによって、中日機械翻訳のシステムの開発にある程度寄与することができのではないかと考えたからである。このような知識がなければ、中日機械翻訳の開発は荒唐無稽な話に過ぎない。この意味で、本研究は単に言語学の立場から中国語と日本語の構文を明らかにしようとするものではなく、コンピューターによる構文解析から、アプローチしようとするものでもある。

タイトルを「中日主題文構文文型の対照研究」としたが、「主題文文型の対照研究」としたら、より簡潔明瞭になるのではないかと指摘されるかもしれない。しかしあえてタイトルを上述のものにしたのは次のような考えによる。文型については、いろいろな分類が試みられてきた。分類基準によって文型にもいろいろなものがある。よく見られるものとして意味的な文型と統辞論的な文型^①が挙げられる。意味的な文型とは文の表現の意図による分類だともわれる。例としては「疑問文」、「否定文」、「命令文」などが挙げられる。統辞論的な文型とは林四郎が『国語学大辞典』において「文」というまとまりがあることを前提として、その文が、語や語群のどのような配置によって成立するか^②と述べたものをいう。例えば、「…は…だ」、「…を…する」などである。本稿で取り上げる構文文型は

① 山口 光「文型分類の原理」、『日本語学』、1984年12月号。

② 国語学会『国語学大辞典』、東京堂出版、1980年、795頁。

後者の考え方に準じるものとする。なぜかと言えば、これは、やはり、言語そのものの構造の解明と機械翻訳への利用と密接な関係がある。周知のように、コンピューターによる言語認知はいまのところでは、客観的な言語構造の解明のところにとどまり、人間のように、表現意図まで理解できる所までにはいっていないのである。したがって、表現意図によって分類し、文型を研究することはコンピューター言語学の研究にとってまだこれからの課題であり、現状では意味よりも形式による文型研究の方がよりふさわしいテーマだと考え、本論のテーマを「中日主題文の構文文型の対照研究」にしたわけである。

では、数多くの文型のうち、なぜ、研究の範囲を主題文に絞ったかと言うと、中国語と日本語には次のような例が至る所で見られるからである。

(1) 東京の都は、面積広く、人口多し。

(2) 中国地大物博。

日本語の特徴となる助詞を単独に扱わず、それとその前の実詞を一単位とみて、これらの文を考察してみよう。

	(1)	東京の都は、	面積	広	く、	人口	多	し
品詞性		名	名	形	名	形		
構文		東京の都は	面積	広	く	人口	多	し
		→	←	←	←	←	←	←
	(2)	中国	地	大	物	博		
品詞性		名	名	形	名	形		

構 文 中 国 地 大 物 博
 → ← → ←

(3) この部屋は 窓が 多い
 → ← → ←

品詞性 名 名 形
 (4) 这间房子 窗户 很多
 → ← → ←

品詞性 名 名 形

上の分析からもわかるようにこの場合の中国語と日本語の文の構造は全く同じである。この場合の文(例2)は中国の言語学界では今の所では「主謂謂語文」などと呼ばれているがまだ定説がないようである。これにたいして、日本語の場合は、これらの文(例1)を主題文と呼ぶことにはまず異論はないだろう。そこで、とりあえず日本語の術語を借りて、中国語の例2のような文をも主題文と称したい。上の図から分かるように、今まで、形態上、中国語は孤立語、日本語は膠着語というように異なるものとして扱われてきたこの両者においてなぜこのような文の構造が似ているのか、そして、このような現象は例(1)(2)のような文に限って見られるものかどうか、大変、興味深い課題である。

普通、言語の記述はまず文の構造の分析から始めなければならない。日本語の主題文は膨大な数を占めており、中国語でも前述の例は少なくない。したがって、このような文の対照研究を通して主題文における中国語と日本語との違いがいったいどこにあるのかという問題もおのずから見えてくるであろう。