

人 机 通 用

现代汉语 动词大词典

中日合作MMT汉语生成组 编著

林杏光 王玲玲 孙德金 主编

北京语言学院出版社

人 机 通 用

现代汉语 动词大词典

中日合作MMT汉语生成组 编著

林杏光、玉玲玲、孙德金 主编

北京语言学院出版社



(京)新登字 157 号

图书在版编目(CIP)数据

现代汉语动词大词典/中日合作项目 MMT
汉语生成组编著;林杏光等主编,一北京:
北京语言学院出版社,1994. 11

ISBN 7-5619-0431-2

I. 现… II. ①中…②林… III. 汉语-
动词-词典 IV. H146. 2-61
中国版本图书馆 CIP 数据核字(94)第 07626 号

人 机 通 用

现代汉语动词大词典

中日合作 MMT 汉语生成组 编著
林杏光 王玲玲 孙德金主编

*

北京语言学院出版社出版发行
(北京海淀区学院路 15 号 邮政编码 100083)

全国新华书店经销

北京语言学院出版社印刷厂印刷

开本 787×1092 毫米 1/32 36.5 印张 1772 千字

1994 年 11 月第 1 版 1994 年 11 月第 1 次印刷

印数 0001—5000 册

ISBN 7-5619-0431-2

H·301 定价:46.00 元

编 著 者 名 单

顾问： 鲁 川

主编： 林杏光 王玲玲 孙德金

编委： 沈建华 张庆旭 李绘新 苗传江 祁素芳

刘士娟 鲁 川 林杏光 王玲玲 孙德金

陈满华 赵守辉

资料： 刘晓梅 王少英 潘 伟 潘 东 林 间

林 貌 林 难 许光明 郜长安 郜长生

王 南 林 滔 王小许 王造时 戴欣秋

林 翰 何耀祖 王月英 王 宁 何毛英

序

从计算机对自然语言的理解或翻译来看,述语动词和形容词是句子句法结构和语义解释的中心。因此,如果在一部电子词典中能对句子中的述语动词与其周围的名词性成分所发生的语义组合关系(即格关系)^①作出具体详尽的描写,就可以大大提高自然语言理解系统或机器翻译系统的性能。在这个意义上,《现代汉语动词大词典》(以下简称《词典》)的编辑出版对我国的语言信息处理事业作出了开拓性的贡献。

自从1968年美国语言学家菲尔墨(C. Fillmore)发表论文“格辨”(The Case for Case)以来,采用动词格框架来表达句意的做法已被越来越多的研究人员所接受,并广泛应用于各国的自然语言系统中。这说明,语义组合关系理论不依赖于特定语种,具有语言上的普遍性。但是令人困惑的是:多年来我们几乎看不到有人对一种语言的这种语义组合关系作出系统的、大规模的描写。^②相反,不少研究者对于究竟应当设置多少种格以及怎样来命名这些格津津乐道,争论不休。这种情况实际上已经延误了语义组合关系理论在大规模自然语言系统中的应用。比如,有些很著名的机器翻译系统也采用格关系作为从源语到目标语转换的中间表示,设计者虽然定义了几十种格,但由于在电子词典中没有相应给出每个动词义条的格框架信息,导致系统在源语分析和目标语生成的过程中频频出错。

笔者对《词典》的主编在学术上的洞察力和胆识深表钦佩,他们明知这是一项难度高、工作量大的语言工程,却知难而进,锲而不舍,花费四五年的时间对现代汉语的几千个常用动词的义项逐一作了详尽的描写。他们用自己的辛勤劳动证明了这是

一条正确的研究路线,国内外学术界从中可以获得不少有益的启迪。

笔者认为《词典》至少有以下三方面的特色:

一、任何系统都具有层次性,格系统也不例外。因此,编著者把由 22 个格组成的格系统划分成三个层次:第一层由“角色”和“情景”组成;“角色”下包括“主体”、“客体”、“邻体”、“系体”等四个要素,“情景”下包括“凭借”、“环境”、“根由”等三个要素,这七个要素构成了格系统的第二层,它们以述语动词为核心完整地表达了一个句子的意义;第三层是 22 个具体格,它们分属于上一层的七个要素。这种分层的格系统表述比以往几十种格不分层次的并列格局清晰多了。

二、编著者首次根据格关系将汉语动词划分为六个次类:他动词、自动词、外动词、内动词、领属动词和系属动词。他们从两条视线来考察一个动词:从动词往后看其客体,将动词分为两类,带客体的叫及物动词,不带客体的叫不及物动词;从动词向前看其主体,也可以把动词分为两类,连接施事主体的叫自主动词,连接当事主体的叫非自主动词。据此,自主而及物的是他动词(如踢、吃、研究等);自主而不及物的是自动词(如跑、蹲、飞等);非自主但及物的是外动词(如听见、看见等);非自主且不及物的是内动词(如病、死等)。再从主体和动词的领属关系、系属关系划分出领属动词(如具有、属于等)和系属动词(如是、等于等)。这种将动词的语法范畴同其格关系直接挂钩的分类思想是一个创举,对推动语义组合关系理论的研究与应用都有积极意义。

三、格语法的一条基本原理是:表层句法形式各异的一些句子,只要它们包含同一个述语动词的义条,而且拥有相同的格关系,它们的格框架表示应是唯一的。因此,对每个动词的义条来说,大都含有一个基本句式 and 若干同义句式(或变换句式)。比如,“撕”的基本句式是“我撕了那封信”,其同义句式有:“我把那封信撕了”、“我那封信撕了”、“那封信我撕了”、“那封信被我

撕了”等。不同句式仅仅反映了言谈者不同的意图,而它们共同的格框架都是:施事(我)+撕+受事(那封信)。然而,动词的次类不同,不仅其基本句式不同,而且同义句式也不同;即使动词次类相同(如他动词),同义句式也会因特定动词的不同而发生变异。《词典》对基本句式和同义句式的描写丰富了动词的用法信息,这不论对语言学习还是对计算机的语言处理都具有实用价值。

当然,这样一项庞大的语言工程在初创阶段难免有这样那样的缺点。笔者认为,《词典》对动词格框架的描写目前还只针对汉语的简单句,而对复句、流水句等语言现象的描写尚有待深化;此外,对于格角色的描写目前只限于举例,如果能进一步给出它们的语义分类码或语义特征就更好了。据悉,《词典》的编著者已制订计划在上述两方面开展深入的研究,这是值得庆幸的。

清华大学计算机科学与技术系教授

黄昌宁

1994年7月28日于盘山憩园

①由于“格”(case)在语言学中又分为表层格(如主格、宾格、所有格等)和深层格(如施事、受事、工具等),为避免术语上的混淆,近年来国际学术界将语义组合关系定名为论旨属性(thematic property),把相应的深层格称作论旨角色(theta-role 或 θ -role)。

②据笔者所知,只有小泉保等编著的《日本語基本动词用法辞典》(文修馆书店,1989)曾对728个日语动词做过类似的描写。

PREFACE

Judging from the understanding and translation of natural languages by computers, the predicate verb and adjective are both the syntactic and semantic core of a sentence. Therefore, if an electronic dictionary can make a detailed description of semantic combinations (i.e. the case relations)* existing between the predicate verb in a sentence and the nominal parts surrounding it, the performance of natural language understanding systems and machine translation systems can be greatly improved. In this sense, the compilation and publication of the *Dictionary of Verbs in Contemporary Chinese* (hereafter, the *Dictionary*) has made an innovative contribution to Chinese information processing.

Since the American linguist C. Fillmore published his paper *The Case for Case* in 1968, more and more researchers have accepted using the case frame for representing the sentence meaning, and the idea has been widely applied in natural language systems all over the world. This shows that the theory of semantic combinations does not depend on a specific language, but rather can be universally applied in all languages. What is perplexing is that for all these years no efforts have been made to work out a large-scale systematic description of the semantic combination relations in any languages. ** In contrast, many researchers have kept themselves busy with arguing how many cases on earth are needed and how to name

them. This, in fact, has delayed the application of the theory in large-scale natural language systems. For instance, some well-known machine translation systems do adopt the case relation as the medium for the conversion from the source language to the target language and the designers define several dozen of case roles; however, since the attached electronic dictionaries do not provide the relative information concerning the case frame for every verb entry, errors frequently prop up in the analysis of the source language and the generation of the target language.

I admire the academic insight and courage of the compilers of the *Dictionary* in that they perserved in the vast and difficult language project and spent more than four years working on a detailed description of the sense entries of several thousand frequently-used Chinese verbs. They have proved with their conscientious work that they are heading in the right direction and scholars both at home and abroad can learn a lot from their work.

I attribute to the *Dictionary* at least the following three characteristics:

1. All systems are hierarchical and so is the case system. The compilers arrange the case system of 22 cases in three levels, the first one made up of "role" and "situation". The "role" includes four elements: "subject", "object", "company" and "link"; the "situation" includes three elements: "means", "circumstance" and "reason". The seven elements together constitute the second layer of the case system and they can, with the predicate verb as the core, express the

integrate meaning of a sentence. The third layer includes 22 specific cases, each belonging to a certain element in the layer next above. Such a hierarchical case system is much clearer than an indiscriminate listing of several dozen of cases.

2. According to the case relation, the compilers divide verbs into six subcategories: *ta*-verbs, *zi*-verbs, *wai*-verbs, *nei*-verbs, possessive verbs, and linking verbs. They consider a verb from two points of view. On the one hand, if a verb is followed by an object, it is a transitive verb, and if it is not, it is an intransitive verb. On the other hand, if the verb follows an agent-subject, it is an intentional verb, and if it is not, it is an unintentional verb. The two ways of differentiating taken together, an intentional transitive verb is a *ta*-verb (e. g. 踢 *tī* —kick, 吃 *chī* —eat, 研究 *yánjiū* —study), an intentional intransitive verb is a *zi*-verb (e. g. 跑 *pǎo* —run, 蹲 *dūn* —squat, 飞 *fēi* —fly), an unintentional transitive verb is a *wai*-verb (e. g. 听见 *tīngjian* —hear, 看见 *kànjian* —see), and an unintentional intransitive verb is a *nei*-verb (e. g. 病 *bìng* —fall ill, 死 *sǐ* —die). Meanwhile, the possessive relation and linking relation between the verb and the subject enables the differentiation of possessive verbs (e. g. 具有 *jùyǒu* —possess, 属于 *shǔyú* —belong) and linking verbs (e. g. 是 *shì* —to be, 等于 *děngyú* —equal) from other verbs. It is original to relate the grammatical category of a verb to its case relations directly and it will help promote the research on, and application of, the theory of semantic combinations.

3. A fundamental principle of the case grammar is that sentences with different surface expressions but including the

same sense entry of a predicate verb with the same case relation should have only one case frame. Therefore, almost every sense entry of a verb has one basic sentence pattern and several synonymous sentence patterns (or alternative patterns). For instance, the basic sentence pattern for “撕 *sī*—tear” is: “我撕了那封信。Wǒ *sī* le nà fēng xìn. I tore the letter” and its synonymous sentence patterns are: “我把那封信撕了。Wǒ bǎ nà fēng xìn *sī* le. I[auxiliary] the letter tore”, “我那封信撕了。Wǒ nà fēng xìn *sī* le. —I the letter tore”, “那封信我撕了。Nà fēng xìn wǒ *sī* le. —The letter I tore”, and “那封信被我撕了。Nà fēng xìn bèi wǒ *sī* le. —The letter was torn by me”. Different sentence patterns reveal the different purposes of the speaker but they have in common the case frame: agent (“I”) + “撕 *sī* tore” + patient (“信 *xìn* the letter”). In contrast, different subcategories of a verb not only have different basic sentence patterns but also have different synonymous sentence patterns; even if two sentences have verbs of the same subcategory, their synonymous sentence patterns may also be different with different verbs. The description in the *Dictionary* of the basic sentence patterns and synonymous sentence patterns can enrich the information on the usage of verbs, which is of practical value both to language study and to language processing by computers.

There is no denying that such a vast language project at its initial stage can not be perfect. For one thing, the description of case frames of verbs in the *Dictionary* is limited only to simple sentences in Chinese and further work will be needed for the description of such language phenomena as compound

sentences and running-on sentences. For the other, the description of case roles is now given only in the form of examples and it could be further improved by giving the semantic code or semantic feature of each case role. What is encouraging is that the compilers of the *Dictionary* are said to have drawn up a plan to further their research in the above two respects.

Panshan Qiyuan
28th July 1994

Prof. Huang Changning
The Department of Computer
Science and Technology
Tsinghua University

Notes:

* Since "case" in linguistics is divided into surface case (e. g. subject, object, possessive) and deep case (e. g. agent, patient, instrument), to avoid ambiguity, many scholars have agreed to define the semantic combination as thematic property and the corresponding deep case as theta-role or θ -role.

* * According to what I am informed, only the *Dictionary of the Usage for Basic Verbs in Japanese* compiled by KOIZUMI TAMOTSU (小泉保) et al. made a similar description of 728 Japanese verbs.

前書き

計算機の自然言語の理解或は翻訳からみて、述語動詞と形容詞は句文法構造、語義解釋の中心です。もし、電子辞典が文の中の述語動詞及びその周辺の名詞性の成分による語義組み合わせ関係(即ち、格関係)⁽¹⁾を具体的に詳しく、描ければ、自然言語理解システムの性能を大きく高められます。この意味で、《現代漢語動詞大詞典》(以下《辞典》と簡稱)の出版は我国の言語情報処理領域に多大な貢献をささげました。

1968年、美国の言語学者 C. Fillmore が “The Case for Case” という論文を発表されて以来、動詞フレームを用い、文の意味を表わす方法は、多くの研究者に受け入れられ、各国の自然言語システムの中に応用されています。このことは、語義組み合わせ関係理論が特定の言語によらず、言語上に汎用性があると表明しました。ただし、長い間、ことばの語義組み合わせ関係を大規模的に述べることが見られず⁽²⁾、逆に大勢の研究者は、何種類の格を置き、どう言うふうに命じるなどのことについて、くりかえし論じています。実際のところ、語義組み合わせ関係理論の大規模な自然言語システムへの応用が延期させられています。例えば、いくつかの有名な機械翻訳システムでも原文から翻訳文の中間表示として、格関係を使用しています。数十種類の格を定義しても、電子辞典に対応的な動詞義条(sense item)の格フレーム情報がないので、システムは原語の分析と目標語の生成をするとき、よく誤りを起してしまう

と言わざるをえませ人。

筆者は、《辞典》主編の学術的な洞察力と胆略に深く感服させられました。これはかなり困難で多大な労力が必る仕事だと知りながら、四、五年かかり、現代漢文の数千個の常用動詞を一つずつ詳細に記述しました。その結果は、国内国外から大きな反響を得ました。

本《辞典》は以下のような特色があります。

一、全てのシステムと同様に、格システムにも自分の層構造を持っています。編集者は22個の格からなる格システムを3層に分けます。第一層は“役割”と“情景”からなります。“役割”は“主体”、“客体”“隣体”、“系体”など四つの要素，“情景”は“凭借”、“環境”、“原因”など三つの要素を含めています。この七つの要素は格システムの第二層を構成しています。これらは述動詞を核にして、完全に文の意味を表すことができます。第三層は22個の具体的な格からなっています。それぞれ上層の要素に属しています。この様な層状格システムは従来の数十種類で層なし并列格システムよりず、と明晰です。

二、編集者らにより、漢文動詞は初めて、他動詞、自動詞、外動詞、内動詞、領屬動詞と系屬動詞、六つの種類に分けられます。動詞を二つの視線から考察すれば、次のようになります。即ち、動詞の後にある客体により、客体のある及物動詞と客体のない不及物動詞に分けられます。動詞の前にある主体により、施事主体とつながる自主動詞と当事主体とつながる非自動詞に分けられます。従って、自主・及物の形は他動詞(例、踢、吃、研究など)、自主・不及物の形は自動詞(例、跑、蹲、飛など)、非自主・及物の形は外動詞(例、聽見、看見など)、非自

主・不及物の形は内動詞(例、具有、属于など)と系属動詞(例、是、等于など)になります。この様な動詞の文法範疇とその格関係を直接につなげる考え方は、分類学の初めての試みです。語義組み合わせ関係理論の研究と応用の促進に積極的な意味があります。

三、格語法の一つ基本的な原理は、表層文法の異なる文でも、同じ述語動詞の義条を含み、また、同じ格関係であれば、それらの格フレームの述べ方は一つしかありません。そのため、一つの動詞の義条には 大体一つの基本句式といくつかの同義句式(或は、換句式)があります。例えば、“撕”の基本句式は“我撕了那封信”で、その同義句式は“我把那封信撕了”、“我那封信撕了”、“那封信我撕了”、“那封信被我撕了”などがあります。異なる句式は、話者の異なる意図を表すだけで、それらの共同の格フレームは、施事(我)+撕+受事(那封信)です。なお、動詞の種類が異なっていれば、基本句式も、同義句式も違います。例え動詞の種類が同じでも(例えば、他動詞)、同義句式は特定動詞の差異により、変異が発生する可能性があります。《辞典》の基本句式と同義句式の描写は動詞使用法の情報を豊かにさせ、言語勉強にも、計算機言語処理にも、実用価値があります。

当然、この様な大仕事では、初めての段階では、いろいろな欠点があります。筆者は《辞典》が今まで、動詞格框架を描写できたのは 単に漢文の単文にとどまっていますが、複合文、流れのある文章など言語現象の研究を深くすすめるべきだと思います。また、格役割に対する記述は例を挙げるだけで、若干でも格役割の語義分類コード或は語義特徴を与えられたら、

一層よくなると思います。幸い《辞典》の編集者たちは、上述の両面から更に深く研究なさる計画を立てていると聞いています。

1994 年 7 月 28 日

清華大学計算機科学及び技術系

教授 黄昌宁 盘山憩園にて

〔1〕格(case)は言語学の中で、表層格(例、主格、賓格、所屬格など)と深層格(例、施事、受事、道具など)に分けられるので、術語上の混乱をさけるため、近年、国際学术界では、言語組み合わせ関係を論旨属性、それに対して深層格を旨角色と定名しました。

〔2〕今まで 小泉保氏らが編集した《日本語基本動詞用法辞典》(文修館書店、1989)だけが729 個の日本語動詞を同じように動詞を描写した辞典であると存知しています。

编著者的话

机器翻译就源语言和目标语言的关系来分,可分为两种类型:双边型和多边型。日本国 CICC 组织开发的亚洲五国(中国、日本、泰国、印度尼西亚、马来西亚)合作研究的《多国语言机器翻译系统》(《Multi-Lingual Machine Translation System》,简称 MMT)是很有活力的多边型机器翻译系统,它的技术路线是采用中间语言方式,即源语言不直接达到目标语言,而是借助中间一个人造概念网络进行转换,也就是说,把源语言分析成中间概念网络形式,再由中间概念网络形式生成目标语言。这个项目历时 8 年(1987 年春——1995 年春),最终取得了 5 种语言互译 3000 句达到 50% 正确率的研究成果。这一尝试走在世界前列,对自然语言处理事业无疑有很大的促进。

从 1988 年春天开始,本词典主编林杏光、王玲玲先后荣幸地应邀参加 MMT 的汉语生成的研究工作,并组织有关研究人员和研究生成立中国人民大学课题小组。该小组在 CICC 的统一部署下,根据中方汉语生成组的研究计划,先后进行了以下的研究工作:1、参加汉语生成实验。2、提供常用汉语动词的句型并写出生成规则。3、翻译 MMT 中间语言的英文版第二版(张庆旭等译)。4、绘制中间语言图。5、评测汉语生成结果。6、撰写了五篇论文:《汉语生成 P 型规则及其科学编码与排序》(林杏光 贺国耀);《MMT 汉语生成简介》(林杏光 郭锐);《MMT 中间语言 and 现代汉语语义组合关系的研究》(林杏光 王玲玲);《定语和中心词之间用“的”规律》(林杏光 王玲玲);《MMT 中间语言的一般概念关系和现代汉语格关系比较》(林杏光 张庆旭)。中国人民大学课题小组在上述的研究活动中实际体会到:从计算机对自然语言的理解或翻译来说,述语动词是句子句法