

# 介连兼类词“以”的句法语义区别特征及消歧策略

方向红 宋春阳

上海师范大学语言研究所 上海 200234

E-mail: [fangxb@163.com](mailto:fangxb@163.com)

**摘要:** 介连兼类词的词性消歧是中文信息处理难题之一。本文借鉴了汉语史研究的成果, 指出介词“以”从动词虚化而来, 因而继承了动词的语法特征, 如介宾结构前面可有副词、助动词、连词等附加成分, 并分析了“以…为…”的两种格式义及相关的名词、形容词的语义特征。分析了连词“以”连接的分句 S1、S2 中动词的语义特征。据此提出一些具体的消歧策略, 目的在于为自动句法分析、话语分析、机器翻译等提供语言学的支持。

**关键词:** 以, 介词, 连词, 虚化, 句法语义区别特征, 兼类消歧

## Studies on Syntactic and Semantic Features of “Yi” and on Some Resolutions of Ambiguity

Fang Xianghong Song Chunyang

Linguistics Institution of Shanghai Normal University, Shanghai 200234

E-mail: [fangxb@163.com](mailto:fangxb@163.com)

**Abstract:** Lexical ambiguity resolution is one of the hard problems to handle in Chinese information processing. This text draws on the study of Chinese language history, and points out that the preposition “Yi” inherits the verb “Yi”’s syntactic characteristics. This text also analyzes the semantic features of nouns or adjectives related to the construction “Yi ...Wei ...”. On these grounds, this essay brings forth some distinguishing features and disambiguating strategies based on rules and provides linguistic support for natural language parsing, discourse analysis, and machine translation

**Keywords:** Yi (“以”), prepositions, conjunctions, syntactic and semantic distinguishing features, resolution of ambiguity

### 1 自动标注中介连兼类识别的困难及本文的研究策略

1.1 兼类消歧是任何自然语言理解和处理都无法回避的问题, 但是语言类型不同, 兼类词的词性消歧的重点是不同的。目前中文信息处理技术词性自动标注中遇到的难题中与连词有

关的主要就是介连兼类词、副连兼类词的识别问题。王力先生在其《汉语史稿》中把介词连词总名为“联结词”，张世禄先生在其《古代汉语》中把介词连词合称为“关系词”。介词与连词在语法功能及语法意义既有联系又有区别。以兼类词“以”为例，下列句子中的“以”如果单从语言线性序列上来看就很难辨别是介词还是连词。

(1) 1904年，摩尔根应聘到哥伦比亚大学任实验动物学教授。他在教学之余，继续以a实验方法进行着进化论和遗传学的研究，通过几年的辛勤研究，他背叛了自己前阶段对孟德尔、达尔文理论的怀疑，以b大量事实证实了达尔文理论和孟德尔学说的正确性。

(2) 作为机械论哲学的忠实信徒，马让迪蔑视理论思维，始终坚持以实验作为出发点，认为生理学知识就是个别实验事实的逐步积累。

(3) 恩格斯指出：“群婚制是与蒙昧时代相适应的，对偶婚制是与野蛮时代相适应的，以通奸与卖淫为补充的一夫一妻制是与文明时代相适应的。”

(4) 外方希望大量裁员以提高劳动生产率，而中方却不愿看到职工砸了饭碗。

(5) 但他们都忍受了巨大的悲痛，以顽强的毅力控制了自己，始终未偏离他们所追求的目标，最终在科学上作出了重要贡献。

(6) 药店为了增大营业额，多提成私分，扩大经营了毛纺织品、日用百货和保健食品，在出售这些商品时给顾客开成药品的发票以“方便”顾客凭票报销。

(7) 因此，重要的是，应将生物体视为一个整体——每部分都有其自己的功能；但要通过各种控制过程与其他部分进行整合，以维持内环境的恒定。

从以上几个例子可以看出，介词“以”与连词“以”两者区分之难具体表现在：

从表层句法结构看，介词“以”与连词“以”很难区分。两者都可位于前后无停顿的句法环境中，如例(2)与(4)。两者前面又都可以有停顿，如例(3)与例(7)。

从句法分析的结果看，两者后面都可以出现动词性短语结构。介词“以”可带动词性短语作宾语，连词“以”引导的分句由动词性短语构成，如例(3)与(6)。

以上的各组比较中，前者(例(2)、(3))都是介词，后者(例(4)、(6)、(7))都是连词。

1.2 造成词性自动标注中介连兼类词识别困难的原因主要在于，从语言发展的历时角度看，兼类的实质是语法化的不同层次在共时平面上共存的体现。汉语中的介词“以”是从动词“以”虚化而来的，而连词“以”则是从介词“以”进一步虚化而来的。一方面，它们之间有语法功能上的区分，另一方面又有语义上的联系，即共有一个语义基核。

1.3 兼类词的词性自动消歧有基于规则的方法和基于统计的方法。具体到用基于统计的方法研究与连词有关的兼类消歧的并不多见，高维君等人讨论了如何将机器学习应用于汉语关联词的歧义辨别——原因，方法和效果。本文拟借鉴汉语史研究的有关成果，对介连兼类词“以”的句法语义特征作一些探讨，并提出一些相关的介连兼类消歧策略，目的在于为自动句法分析、话语分析、机器翻译等提供语言学的支持。

## 2 现代汉语中介词“以”的句法语义特征

2.1 从历时的角度看，介词“以”是由动词“以”虚化而来的，表使用义的动词“以”在连

动结构中处于次要地位，是它虚化为介词的语法条件。从句法位置上看，虚化后的介词“以”保留了动词的语法特征，一是介词“以”及其宾语所处的句法位置仍然是在主语之后，行为动词之前；二是介词“以”及其宾语组成的介宾结构前面可有副词、助动词、连词等附加成分。根据我们对一定数量的语料进行统计分析的结果显示，可出现在介词“以”前的副词有“不、就、都、又、经常、再、竟、却、几乎、还……”等；助动词如“要、能、应……”等，（助动词“可”位于介词“以”之前会构成组合型歧义字段）；连词如“并、但是、若、及、而是……”等。这些词在介词“以”前面出现时，按其语法范畴的不同，排列顺序依次为：连词>副词>助动词>“以”（注：“A>B”表示在语言的线性序列上A句法位置在前，B句法位置在后）。如下例中“以”前有连词“并”，据此我们可识别该句中的“以”是介词。

（8）资本主义从一开始就要把人对自然的统治作为它的前提，在它那里，自然界只不过是社会的人的改造对象，“不过是有用物”，“其目的是使自然界……服从于人的需要”，并以惊人速度发展生产。

我们也可以把这作为一条标准识别上文例（1）中的在副词“继续”后面的“以a”是介词。由于能够出现在介词“以”前面的这些附加成分，都不能出现在连词“以”的前面。据此可作为识别介词“以”与连词“以”的一个语法标记。

2.2 从语义特征来看，表实义的行为动词“以”虚化为介词“以”，其动作性减弱，语法意义增加。介词“以”的基本语法意义有二：表行动作赖以完成的凭借和表行动得以实现的原因。

（一）表凭借，引进动作行为赖以进行的工具、手段、前提等，即广义的工具。整个介词结构表示动作的凭籍。

A类：介词“以”的宾语为名词性短语且中心词是抽象名词，可充当介词“以”的宾语的抽象名词有很多，例如：“事实、实验、成果……”，语义特征动作凭籍的依据，记为[+客观事实]；“面目、身份、名义……”；“理想、思想、意识、自尊……”；“全力、毅力、思维力、气势……”等类，语义特征为实现某一行为凭籍的个人素质，语义特征为[+素质]。

B类：介词“以”的宾语为动词性短语且其中心词是名物化的过程动词，例如：实验、计算、结合……。这类动词性短语的语法特征是过程动词前有时有“的”。语义特征则是[+方式]

C类：介词“以”的宾语为形容词性短语且其中心词是表示品质的形容词，例如：沉稳、刻苦……。

（二）表方式，方式其实就是凭籍的扩展，只是意思抽象一些。

A类：介词“以”的宾语为表方式、方法的名词性短语，其中心语为“方式、形式、方法……”一类的名词，共同的语义特征为[+方式]。

B类：介词“以”的宾语为具有可度量性质的名词性短语，度量义体现在“（数词+量词）+名词”的结构上。其中的名词如“租金、成本、价格、价值、邮资、罚金……”，语义特征为[+价值]，或“数量、速度、精确度、深度、广度、坡度、规模、机制……”，语义特征为[+可度量]等。

C类：介词“以”的宾语为表时间的名词。时间也是动作进行方式的诸要素之一，是

由表示动作进行方式扩展而来的，语义特征为 [+时间]。这种用法并不多见。如：

(9) 鲁迅以1881年于浙江绍兴城内姓周的一个大家族里。

(三) 表原因，原因也是一种凭籍，它也是表示动作凭籍的引中。在这类用法中介词“以”的宾语大多为表性质的形容词，语义特征为 [+品质]；或表事件的主谓短语。并且“以”的后面可有一个共现词语“而”，构成非连续结构格式“以…(而)…”。例如：

(10) 以吃苦耐劳著称于世。

(11) 我们以祖国有这样的英雄而自豪。

我们认为，“以”作介词的语义特征与介词宾语(NP)的语义特征一致。由此，宾语NP的中心语N(head)可以作为词汇标志识别介词“以”的不同语义，并可进一步作为介词“以”与连词“以”识别的词汇标记。据此，我们可判定上文中的例(1)、(2)、(3)、(4)中的“以”都是介词。

2.3 “以”作介词构成的语法格式是“以+NP+VP”，能够进入这一格式的使用频率最高的典型的动词就是“为”，在现代汉语中还有一个同义的双音节词“作为”。我们认为，介词“以”与共现动词“为/作为”组成的这一非连续结构“以…为/作为…”可用来作为区分“以”作介词还是连词的一个词汇标记，如例(2)、(3)中的“以”都是介词。下面我们讨论“以…为/作为…”格式，将“为/作为”代入语法格式“以+NP+VP”，并结合语料统计分析的结果，得到以下两种具体格式：

A式：“以+NP1+为/作为+NP2”，表示“把……当作……”的意义。常出现在NP2位置的名词有：宗旨、目标、目的；内容、中心、核心、原点、出发点、重点、焦点、突破口、龙头、线索；前提、基础、背景、补充、依据、标准、原型、例、名；代表、领导、先导；工具、载体；任务、方针；这些名词共同的语义特征是实施某一行为时所参照的标准，记为 [+标准]。例如：

(12) 我们以贝尔纳(C. Bernard, 1813-1878)“内环境”学说的创立及“生命体是一部热机”这一思想的确立作为两条主要线索，来探讨十九世纪下半叶生理学中思想及方法的演变特色。

B式：“以+NP+为+AP”表示的意义是“认为……是……”，常出现在A位置上的性质形容词有：主、重、时髦、大……等，语义特征为 [+性质]，如：

(13) 科技史表明，优秀的自然科学家都是以事业为重的人。

### 3 连词“以”的句法语义特征

3.1 连词“以”由介词“以”虚化而来。虚化的句法条件是，由于古汉语中介词“以”的宾语既可以前置，又可以省略，加上也可以用动词性短语充当，因而使它具有了进一步虚化为连词的条件。连词“以”是从表示“凭籍”的介词虚化而来的。当介词“以”的动词性宾语发生移位至介词“以”前，语法地位得到提升，从外部形式上看，“以”处在两个动词性结构之间，这已经不是介词能够出现的典型语境，意义进一步虚化，由介词虚化为连词，虚化后的连词又在句法功能、语义方面与介词相区别。从语言的共时平面看，在汉语语法中，介词与连词都具有连接功能，但介词与连词属于不同的语言结构层次。介词及其宾语

组成的介宾结构在句内可以充当各种句子成分，如状语、补语、定语等，属于句法层次。连词不作句子成分，用于连接小句或句子，连词既属于句法层面，又属于篇章层面。介词表示实词与实词之间的关系，用于给动词引进时间、地点、方式等。连词表示小句或句子之间的逻辑语义关系，连词的意义比介词更虚。

3.2 连词“以”的基本功能是连接两个语言成分，以显示前后两项之间的逻辑语义关系。古汉语中的“以”可以连接词或短语，多为显示其间的目的、结果关系，在语言表层结构上是连续的。连词“以”连接两分句，单在表层结构上两分句之间连续的，这种结构在现代汉语中仍然存在，如上文中的例（4）、（6），给介词、连词的区分带来一定的困难。另外一种结构式，也就是连接两小句，显示其间的因果关系，即现代汉语中连词“以”的基本连接功能：“S1，以 S2”。这个表达式有两方面的特点：

①“以”虚化后是一个典型的连词，其所处的句法位置是在第二个分句的句首。但是在如我们在上文中已经指出的那样，这一句法位置不足以将连词“以”与介词“以”区分开来。

②“以”是一个句内连词，只能连接分句，不能连接句子。所以，形式标记“。”可作为区分介词“以”与连词“以”的一个标志。即，如果有句中的“以”其前面的邻接字符为“。”，那么“以”的语法功能一定为介词而不可能是连词。如下例：

（14）但是，那些另立名目的收费单位，都有实权，你不交纳，木材就动不得。以火车站联营费为例，木材卖主只有与火车站联营才有可能得到车皮，否则纵有车皮指标也拿不到车皮。

3.3 能够据以识别连词“以”的词汇标志是其后面的邻接成分“使”、“区别于/别于”。“以”在这些句子结构中表达的逻辑语义为：前一分句中的行为是后一分句中行为的目的。“以”引导目的分句。如下面的例子：

（15）对此，贝尔纳认为，在分别地研究了生命现象之后，必须进行综合，以使所有分开的部分重新组合，使得整体的性质呈现在人们面前。

（16）我们简称之为“物的依赖关系”，以区别于它所取代的“人的依赖关系”。

（17）二十世纪提出的星云说以别于康德-拉普拉斯星云说统称为新星学说。

3.4 介词的语义是连词虚化的语义基础。表结果的连词“以”是由表凭借的介词“以”虚化而来。当介词“以”的动词性宾语发生位移后，“以”的组合作用丧失，“凭借”的语义失去实际的依托。位于“以”前后的两个动词性短语，在时间上往往有顺接关系，连动结构前后项所表的行为不仅在时间上先后相继，而且在事理上以义相应，因此后一部分可看成是实施前一动作行为的后造成的客观事实，也就是结果。

另外表目的的“以”是由表原因的介词“以”虚化而来，意义相当于“用来”。表示后项是前项的目的，句义重在强调后项是施行前项行为的主观意欲。

通过语料统计分析，不难发现在连词的典型格式“S1，以 S2。”中，出现于 S1 中的频率较高的动词主要是“进行”类动词，即“进行/举行+VP”。例如：

（18）根据这一理论，在手术中将与病者可能发生接触的各种物品，都用防腐药品进行消毒，以杜绝肉眼看不见的病菌的“胚”。

（19）区及乡镇的中心校对部分教师组织的考试必要时进行重点抽测，以保证考试的质量，消除“自教自考”可能带来的虚假现象。

同样,通过语料统计分析,我们也可以发现在连词的典型格式“S1,以 S2。”中,S2语义是为实施 S1 的目的。分两大类,一类是针对事物,一类是针对人。针对事物的,一种使事物向好的方面发展,常出现在连词“以”后的动词有“引导、维持、保证、方便、帮助、反映、完成、提高、加大”等,见例(4)、(6)、(7)、(19);另一种努力避免使事物出现不利状况,常出现在连词“以”后的动词有“杜绝、抵销、消除、摆脱”等,如例(18)、(19)。同样,针对人的也分两种。一种是促进人向善的,常出现在连词“以”后的动词,如“表彰、鼓励”等。一类使阻止人向善的,如“谴责”。总之,S2中的动词分两类,根据其语义特征,一类为促进类动词,语义特征为[+促进]。一类为阻止类动词,语义特征为[+阻止]。

## 4 结 语

介词“以”从动词虚化而来,继承了动词的语法特征,如介宾结构前面可有副词、助动词、连词等附加成分。包含介词“以”的典型句法结构“以…为…”分为两种格式,有不同的意义。一是“以+NP1+为/作为+NP2”,表示“把……当作……”的意义,二是“以+NP+为+AP”表示的意义是“认为……是……”。本文分析了与该结构相关的名词、形容词的语义特征。连词“以”的句法功能是连接分句,组成结构式“S1,以 S2。”。其中 S1 中 VP 的特征为进行类动词,S2 中 VP 语义特征为致使义动词,分两类:[+促进]类与[+阻止]类。

## 参 考 文 献

- [1] 杨伯峻:《古汉语虚词》,中华书局,1981年第一版。
- [2] 尹日高:“‘以为’的语义类别及其演变轨迹”,《古汉语研究》,1991年第3期。
- [3] 李英哲、卢卓群:“汉语连词发展中的若干特点”,《湖北大学学报(哲学社会科学版)》,1997年第4期。
- [4] 郭锡良:“介词‘以’的起源和发展”,《古汉语研究》,1998年第1期。
- [5] 何乐士:“‘政以治民’和‘以政治民’两种句式有何不同?”,《古汉语语法研究论文集》,商务印书馆,2000年第一版。
- [6] 张家文:“古汉语介词‘以’支配成分的移位和省略”,《古汉语研究》,2001年第4期。
- [7] 马贝加:《近代汉语介词》,中华书局,2002年第一版。
- [8] 董秀芳:《词汇化:汉语双音词的衍生和发展》,四川民族出版社,2002年第一版。
- [9] 龚嘉镇:“《论语》中的‘以’”,《汉字汉语汉文化论集》,巴蜀书社,2002年第一版。
- [10] 俞士汶等著:《现代汉语语法信息词典详解(第二版)》,清华大学出版社,2003年第二版。
- [11] Antoinette Renouf(英国利物浦大学),靳光瑾译,郑定欧审,“从词的关系中识别词汇标记”,《语言文字应用》,2002年第4期。
- [12] 高维君、姚天顺、黎邦洋、陈伟光、邹嘉彦,“机器学习在汉语关联词语识别中的应用”,《中文信息学报》第14卷第3期