

# 基于集合运算的德语冠词歧义消解

冯志伟

语言文字应用研究所, 100010 北京

E-mail: [zwfengde@public.bta.net.cn](mailto:zwfengde@public.bta.net.cn)

**摘要:** 在德语冠词的自动分析中, 存在着不少功能歧义的问题。本文总结了德语冠词功能歧义类别, 并采用特征值集合运算的方法来消除这样的功能歧义。

**关键字:** 德语、冠词、功能歧义、歧义消解、集合运算。

## German Article Disambiguation Based on Set Operation

Feng Zhiwei

Applied Linguistics Institute, 100010 Beijing

E-mail: [zwfengde@public.bta.net.cn](mailto:zwfengde@public.bta.net.cn)

**Abstract:** In the automatic analysis of German articles, there are some functional ambiguities. All the ambiguities in German Articles can be resolved by the set operation of feature values.

**Keywords:** German, article, functional ambiguity, disambiguation, set operation

### 1 引言

德语是一种重要的语言, 德国是我国在欧洲主要的贸易伙伴, 中德两国在官方和民间的经济、文化、科技交流日趋频繁。德国是经济高度发展的国家, 中德之间有着长期稳定的友好关系, 随着中德交流的日益发展, 德汉机器翻译显得越来越加迫切。除了德国讲德语之外, 奥地利、瑞士、比利时、卢森堡也有很多人讲德语, 在东欧的匈牙利、波兰、罗马尼亚, 也有不少的地区讲德语。德汉机器翻译系统的研制, 对于我国与这些国家和人民的交流, 也是很有用处的。本文是作者正在研制的德汉机器翻译系统的一项基础工作, 讨论冠词的功能歧义及其消解方法。作者采用了基于集合运算的歧义消解方法, 这样的方法所根据规则, 都是从语言学研究中总结出来的, 因此是一种理性主义的方法。

作者曾经试验过采用 CLAWS 算法和 VOLSUNGA 算法等统计方法来进行德语冠词的歧义消解, 但统计数据要根据大规模的真实德语语料库来获取, 目前我们试验的德语语料库规模较小, 通过训练集所得到的统计数据的可信度较差, 试验效果并不理想。因此, 我们改用基于集合运算的方法, 完全根据语言学规则, 通过集合运算来消解德语冠词的歧义, 效果良好。这种情况说明, 在我们进行大规模真实文本的语料库研究的时候, 我们也不可忽视语言

学的规则，语言学规则是语言学家长期研究成果的结晶，是计算语言学研究的宝贵财富。在计算语言学研究中，我们应该把基于语料库的经验主义方法与基于规则的理性主义方法结合起来，取长补短，相得益彰，才能进一步推动学科的发展。

## 2. 德语的定冠词和不定冠词

每个德语名词都有一个冠词。冠词是伴随名词一起使用的。它与后面的名词在性，数，格保持一致。

例如，Der Mond geht auf. 月亮升起来了

Herr Ober, ich bekam ein Glas Bier. 服务员，我要一杯啤酒。

Der Chef ueberreicht den Kollegingen Urkunde. 上司向职员颁发证书。

“Mond”前面有个冠词“der”，它表示后面的名词是阳性，单数，第一格。“Bier”前面有一个冠词“ein”，它表示后面的名词是中性，单数，第四格。“Kolleginnen”前面的冠词是“den”，它表示后面的名词是阴性，复数，第三格。

德语冠词有两种：

### -- 定冠词 (der bestimmte Artikel)

定冠词有三种形式：der, die, das，它们表示后面的名词是读者或交谈者熟悉的，或是提到过的。

例如，

Der neue Kamera ist sehr schoen. 这台新的照相机很好看。

Die Sonne geht auf. 太阳升起来了。

Das Klima in Deutschland ist mild. 德国的气候温和。

### -- 不定冠词 (der unbestimmte Artikel)

不定冠词也有三种形式：ein, eine, ein，它们表示后面名词所表示的人或物是某一类别中的任意一个个体。

例如，

Die Kinder singen ein deutsches Lied. 孩子们在唱一首德国歌。

Ich moechte eine neue Armbanduhr kaufen. 我想买一块新錶。

Vor der Haustuer sass eine Katze. 门前坐着一只猫。

## 3. 德语冠词的功能歧义

冠词和名词在句子中一起使用时要变格(Konjugation)。定冠词和不定冠词的变格形式是不同的，其功能歧义的表现也有所区别，但基本上是大同而小异。

### 定冠词变格

阳性单数第一格	der
阳性单数第二格	des
阳性单数第三格	dem
阳性单数第四格	den

阴性单数第一格	die
阴性单数第二格	der
阴性单数第三格	der
阴性单数第四格	die

中性单数第一格	das
中性单数第二格	des
中性单数第三格	dem
中性单数第四格	das

阳性复数第一格	die
阳性复数第二格	der
阳性复数第三格	den
阳性复数第四格	die

阴性复数第一格	die
阴性复数第二格	der
阴性复数第三格	den
阴性复数第四格	die

中性复数第一格	die
中性复数第二格	der
中性复数第三格	den
中性复数第四格	die

由此可以看出：

- 阳性和中性定冠词单数二格均为 des，单数三格均为 dem。
- 阴性定冠词单数一、四格均为 die；单数二、三格均为 der。
- 中性定冠词单数一、四格均为 das。
- 复数时，阳性，阴性，中性定冠词各格的变格形式相同。

这种情形说明，德语的冠词的语法属性是有歧义的，同样的冠词形式可以表示不同的语法意义。如何区分同形歧义，是德语冠词计算机处理的关键问题，为此我们必须研究德语冠词的功能歧义。

首先我们来看德语定冠词的功能歧义。

从形式分析的角度来观察，德语定冠词共有 der, des, dem, den, die, das 六种形式。每种形式都是多功能，在机器词典中，应标出它们的多种功能，多种功能也就是功能歧义。

六种形式的功能歧义归纳如下：

der: ms1, fs2, fs3, mp2, fp2, np2.

des: ms2, ns2.

dem: ms3, ns3.

den: ms4, mp3, fp3, np3.

die: fs1, fs4, mp1, mp4, fp1, fp4, np1, np4.

das: ns1, ns4.

符号说明：m:阳性，f: 阴性，n:中性，s:单数，p:复数，1,2,3,4 分别表示 1 格，2 格，3 格，4 格。例如，ms1 表示阳性、单数、1 格；fs2 表示阴性、单数、2 格；np2 表示中性、复数、2 格。

德语不定冠词变格如下：

阳性单数第一格	ein
阳性单数第二格	eines
阳性单数第三格	einem
阳性单数第四格	einen

阴性单数第一格	eine
阴性单数第二格	einer
阴性单数第三格	einer
阴性单数第四格	eine

中性单数第一格	ein
中性单数第二格	eines
中性单数第三格	einem
中性单数第四格	ein

不定冠词没有复数。

例如， Im Garten steht ein Baum. 花园里有棵树。

Im Garten stehen Baeume. 花园里有些树。

由此可以看出：

- a)不定冠单数变格的词尾与定冠词的变格词尾极为类似。
- b)不定冠词没有复数形式。

据此我们对于德语不定冠词进行功能分类。

从形式分析的角度来观察，德语不定冠词共有 ein, einer, einem, einen, eine, eines 六种形式。其中，einen 只表示 ms4，没有功能歧义，其他五种形式都是多功能，有功能歧义，在机器词典中，应标出它们的多种功能歧义。

六种形式归纳如下：

-- 无歧义的形式（一种）：

einen: ms4.

-- 有歧义的形式（五种）：

ein: ms1, ns1, ns4.

eines: ms2, ns2.

einer: ms2, fs2, fs3.

einem: ms3, ns3.

eine: fs1, fs4.

符号含义与上同。

#### 4. 基于集合运算的歧义消解方法

在词典中标出定冠词和不定冠词的功能歧义之后，计算冠词的语法特征同它所限定的名词的语法特征的交，进行复杂特征的集合运算，所计算出来的交的值就应该是冠词的正确语法特征的值，从而根据上下文来消除冠词的功能歧义。

下面我们举例说明如何根据上下文来消除定冠词和不定冠词的功能歧义。

例 1. Stephan hat eine neue Brille. 斯蒂芬有一付新的眼镜。

[歧义消解] 不定冠词 eine 有 fs1, fs4 的歧义，它修饰的名词 Brille 是 sf1 或 fs4，但因作为动词 hat 的直接宾语，应为 fs4，与 eine 的值 fs1, fs4 求交，其交为 fs4，故可以判定 eine 为 fs4。

例 2. Die Brille ist schoen, aber teuer. 这付眼镜好是好看，但很贵。

[歧义消解] 定冠词 die 有 fs1, fs4, mp1, mp4, fp1, fp4, np1, np4 的歧义，它修饰的名词 Brille 做主语，为 fs1，与 die 的值求交，其交为 fs1，故可以判定 die 为 fs1。

例 3. Vor der Haustuer stehen unbekannte Leute. 门前站着不认识的人。

[歧义消解] 定冠词 der 有 ms1, fs2, fs3, mp2, mp3, np2 的歧义, 它修饰的名词 Haustuer 在介词 vor 之后, 应为 fs3, 与 der 的值求交, 其交为 fs3, 故可以判定 der 为 fs3.

例 4. Die Leute kenne ich auch nicht. 这些人我也不认识。

[歧义消解] 定冠词 die 有 fs1, fs4, mp1, mp4, fp1, fp4, np1, np4 的歧义, 它修饰的名词 Leute 在这里是动词 kenne 的直接宾语, 为 p4, 与 die 的值求交, 其交为 p4, 故可以判定 die 为 p4.

例 5. Ich habe mir ein neues Auto gekauft. 我买了一辆新车。

[歧义消解] 不定冠词 ein 有 ms1, ns1, ns4 的歧义, 它修饰的名词 Auto 做动词 gekauft 的直接宾语, 为 ns4, 与 ein 的值求交, 其交为 ns4, 故可以判定 ein 为 ns4.

例 6. Das Auto faehrt sich gut. 这辆车很好开。

[歧义消解] 定冠词 das 有 ns1, ns4 歧义, 它修饰的名词 Auto 做主语, 为 ns1, 与 das 的值求交, 其交为 ns1, 故可以判定 das 为 ns1.

由此可以看出, 特征值的集合运算(主要是求交运算)是我们消除德语冠词歧义的基本运算。在德汉机器翻译中, 我们采用这样的基于集合运算的歧义消解方法, 效果良好。这种情况说明, 我们在计算语言学研究中, 切不可忽视传统语言学所积累起来的丰富的规则, 这些规则, 不仅是传统语言学的宝贵财富, 同样也是计算语言学极为重要的知识资源。

## 参考文献

- [1] 冯志伟: 自然语言机器翻译新论, 语文出版社, 1995年, 北京。
- [2] 冯志伟: 自然语言的计算机处理, 1996年, 上海。
- [3] 冯志伟: 应用语言学综论, 广东教育出版社, 1999年, 广州。
- [4] 冯志伟: 计算语言学基础, 商务印书馆, 2001, 北京。
- [5] 刘开瑛: 中文文本自动分词和标注, 商务印书馆, 2000年, 北京。
- [6] 张才尧: 实用德语语法, 外语教学与研究出版社, 1997年, 北京。
- [7] Goodman. K, Nirenburg. S.: The KBMT Project: A case study in knowledge-based Machine Translation, Morgan Kaufmann Publishes, 1991, San Mateom California.