

汉英机器翻译扩充词典的建造*

刘 群¹ 张 彤²

¹ 北京大学计算机语言学研究所

² 中国科学院计算技术研究所

E-mail: liuqun@ict.ac.cn

摘 要: 本文首先介绍了我们在“面向新闻领域的汉英机器翻译系统”中语言资源建设的总体框架,然后着重介绍了其中扩充词典的建设方法。扩充词典是相对于核心词典而言的,特点是词汇量大,每个词的信息相对较少。扩充词典的建设包括以下几个阶段:收集;格式整理;词条拆分;词性标记归一化;词性标记补齐;词条合并。实验结果表明,采用这种方法,利用较少的人工,就可以得到规模很大、并初步可用的双语机器翻译词典,并且这部词典可以大大减少翻译中未定义词的数量,提高翻译的质量。

关键词: 机器翻译 双语词典获取

一、引言

现在的机器翻译研究,从一种劳动密集型的研究方式逐步过渡到了一种资源密集型的研究方式。这二者的区别在于,在劳动密集型的研究方式中,语言学家的工作是为某个具体的机器翻译系统开发词典、规则库等语言知识库,其工作依附于某个具体的机器翻译系统,不具有独立性。而在资源密集型的研究方式中,语言学家和计算机工作者的分工更加明确。语言学家的研究成果以语言资源的形式呈现出来,而不是仅仅为某一个机器翻译系统服务,其工作具有一定的独立性。同时,这种分工导致了语言资源的共享,也使得计算机工作者在算法的研究中对于语言资源的使用有了更广泛的选择余地。

本文简要介绍了我们在“面向新闻领域的汉英机器翻译系统”开发过程中语言资源建设的总体规划和实施情况,然后详细介绍了其中一个子任务——“扩充词典”建设的工作流程、算法设计、实施过程和结果评价。

二、语言资源建设的总体规划

我们在“面向新闻领域的汉英机器翻译系统”(Liu et. al. 2001)的开发过程中,非常重视语言资源的建设。我们建立了一套完整的,覆盖语言知识各个层面的机器翻译语言资源建设规范,并启动了一系列的语言资源建设子任务。这些子任务包括:

1. 汉语语义词典子任务

在原有语义词典基础上,完善规范、修改错误、补充词语,构造一部完整的机器翻译用汉语语义词典。目前此项工作正在进行,目标是将北京大学语法信息词典中的七万多汉语词

* 本项目接受国家重点基础研究项目(973)资助,课题号为:G1998030507-4。

语加入到语义词典中。

2. 汉英扩充词典子任务

在原有核心词典的基础上构造一部机器翻译用扩充词典，本文后面将详细介绍。

3. 汉英短语库子任务

收集汉英对照的短语，标记出短语结构信息，此项工作已告一段落，目前收入短语 5 万余条。

4. 汉英双语语料库收集整理子任务

收集、整理双语语料库、进行格式整理、篇章标注、篇章对齐，此项工作第一阶段已经完成，收录大约 100 万汉字的汉英对照新闻语料以及大量的其他类型语料，目前正在进行第二阶段工作。

5. 汉英双语语料库句子对齐子任务

在篇章对齐语料库基础上，完成段落对齐和句子对齐，与上一个子任务一样，第一阶段已完成，完成了 100 万汉字的汉英对照新闻语料的句子对齐（约 4000 句对），并整理其他类型句子对齐语料约 18 万句对。目前正在进行第二阶段工作。

6. 汉英双语语料库短语对齐子任务

在句子对齐语料库基础上，完成子句、最长名词短语、基本名词短语以及其他一些特定类型短语的对齐，此项工作第一阶段已经完成，大约完成了 3000 汉英句对的短语对齐工作。第二阶段还没有开始。

本文主要介绍其中的第二项子任务：汉英扩充词典子任务。

三、机器翻译的词典建设

双语词典建设是机器翻译系统开发中的一个重要环节。

常见的机器翻译词典开发有以下几种方法：

1) 人工开发：组织人力物力，从头开始开发。当然，在实际的开发中，都要借鉴已有的词典，甚至以已有的某部词典为基础。这种方法费时费力，但可以根据自己的需要任意裁减所需信息，易于和翻译算法紧密结合。传统的机器翻译词典开发一般都采用这种方法。《知网》（董振东）也可以看作一部采用这种方法开发的可用于机器翻译的双语词典。

2) 利用已有的电子词典资源：目前通过各种方式可以获取的双语电子词典已有很多，通过对这些词典进行收集整理，可以得到机器翻译用的双语词典。这种方法的优点是可以充分利用已有资源，省时省力，缺点是各种不同来源的词典格式、体例、属性字段等都不一致，词典质量参差不齐，整理加工比较麻烦。

3) 从双语语料库中抽取词典：在各种双语对齐语料基础上，利用统计方法，从中抽取双语对照词语。这种方法的优点是直接从语料中抽取，可以学习到一些活的翻译方法，而这些译法很可能在词典中是找不到的，特别适用于一些专业领域的词典抽取，或新词语的学习。不过这种方法依赖于学习算法的好坏，会产生很多垃圾信息，要得到高质量的词典也需要较多的人工校对；学习到的词典信息依赖于语料库，往往信息不够全面，最好作为已有词典的补充。这方面的研究现在比较多（Melamed, 1996），（Church et. al., 1990），（Smadja et. al., 1996），（Haruno et. al., 1996），（Fung, 1995）。

在我们的汉英机器翻译系统中，已经有了一部核心机器翻译词典（王慧等，1998），这部词典含汉语词条（以词形 word type 计算）约五万，这部核心词典含有丰富的词法、句法、语义搭配信息如图 1 所示。这部词典是采用上面所述的第一种方法，通过自己组织人力物力从头开发的。词典采用文本形式存储，使用我们专门编写的词典管理工具进行管理。不过，在面对真实的文本的时候，这个词典的规模就显得有点太小了，远远不能满足实际翻译的需要。在我们的新版本机器翻译系统中使用了一个独立的分词程序，而这个分词程序所用的分词词典就有词语 8 万多条，这样就导致很多分词程序切分出的词语无法翻译。面对这种情况，我们决定采用上述的第二种方法，构造一部扩充的汉英机器翻译词典。目前我们还没有看到采用这种方法构造机器翻译词典的文献，所以本文介绍的方法完全是我们自己摸索出来的。

```

$$ 背
  **{vA} v $=[义项:"(1)驮;(2)负担",谓词性主语:NO,系词:NO,助动词:NO,趋向动词:NO,补助动
  词:NO,……,语义类:搬移,配价数:2]{主体:[语义类:人|动物],客体:[语义类:具体事物],处所:[语义类:空间|
  集体]}
  => V<bear> $=[VSUBC:VO,VMORF:IREG,VD:bore,VN:borne,OBJTYPE:NP]
  **{vB1} v $=[义项:"(1)向~;(2)避",谓词性主语:NO,系词:NO,助动词:NO,趋向动词:NO,补助动
  词:NO,……,语义类:自移自为,配价数:1]{主体:[语义类:人],客体:[语义类:人类]}
  => D<"behind one's back"> $=[DVPRE:NO,DVEND:YES,DADJV:NO,DMORF:NONE]
  => V (!V<do> P<behind> N<back>) %V=[VSUBC:VI,VMORF:IREG,VD:did,VN:done]
  **{vB2} v $=[义项:"(4)~诵",谓词性主语:NO,系词:NO,助动词:NO,趋向动词:NO,补助动词:NO,形式
  动词:NO,……,语义类:对待,配价数:2]{主体:[语义类:人],客体:[语义类:人为事物|抽象事物]}
  => V<recite> $=[VSUBC:VO,VMORF:REGU,OBJTYPE:NPICS]
  **{n} n $=[名词子类:na,前名:NO,前动:NO,后名:NO,名状语:NO,临时量词:NO,语义类:构件]
  => N<back> $=[NSUBC:NCONT,GEND:NONE]

```

图 1 核心词典示例

四、扩充词典的收集与整理

1. 扩充词典的目标

扩充词典（下面我们又称目标词典）应该满足以下三方面要求：

- 1) 面向通用领域：扩充词典不是专业词典，不应该收入专业性太强的词语；
- 2) 词汇量足够大：扩充词典的词汇量不应太少，按照我们事先的估计，至少应达到 20~30 万词对的数量级；
- 3) 每个词对的信息虽然不必像核心词典那么丰富，至少应该含有以下几个字段：汉语词语、汉语词性、英语词语、英语词性、置信度而且汉语词性和英语词性的标记集应该与核心词典一致。我们称之为一个五元组。
- 4) 通过在中多部词典中出现的频次的统计，也可以得到双语词对的置信度信息。这种置信度信息虽然不是直接从双语语料库统计得到的，但也有一定的统计意义。

词典在收集过程中采用数据库进行管理，这样便于利用成熟的数据库管理工具进行操作。最后得到的词典再通过文本形式提交给翻译程序，由翻译程序建立索引、进行查询。

2. 扩充词典的收集

确定了目标以后, 我们开始从各种途径收集一些英汉和汉英词典, 加上我们原先的一些积累, 总共得到6部电子版的词典, 本文把这六部词典称为原始词典。这些词典的总体情况如下表所示:

表1 扩充词典的来源

词典	方向	规模	类型
A	英汉	12万词对	背单词软件(各类大中学学英语课本、英语学习类书籍共30余本)
B	汉英	18万词对	书面词典
C	汉英	17万词对	书面词典
D	汉英	12万词对	网络词典
E	汉英	8万词对	网络词典
F	汉英	693词对	一些常用新词语

值得注意的是, 这些词典里既有英汉词典, 也有汉英词典。我们的实验表明, 即使在汉英机器翻译中, 加入英汉词典中获得的信息也是有效的。例如, 我们的汉英机器翻译系统曾经遇到一个汉语词语“中西部”, 在汉英词典中都查不到, 而在英汉词典中确可以找到其对应词语“Midwest”。

3. 扩充词典的整理

利用原始词典并不能通过简单合并得到所需的扩充词典。由于词典来源不一, 导致了很多问题。其中的主要问题包括:

- 格式不一致: 由于词典来源不一, 格式、体例都不一致, 有的格式比较规整, 有的就比较混乱;
- 录入排版错误和格式错误: 录入、排版、格式方面的错误很多, 需要认真清理;
- 词性标记不一致: 每部词典的词性标记都不一致, 而且都与核心词典的词性标记不一致;
- 词性标记不完整: 很多词典甚至没有词性标记。

可以看到原始词典面临的问题是非常杂乱的, 由于数量巨大, 如果都采用人工校对的方法, 人力、物力的消耗都是难以承受的。为此我们主要采用了人机互助的办法, 以自动处理为主, 对于一些自动处理无法解决的问题, 辅以人工校对。实践证明, 这种方法是成功的, 我们只采用了较少的人工校对, 就得到了一部规模较大、初步可用的机器翻译扩充词典。

扩充词典的整理按照以下几个步骤进行:

1) 词条拆分: 由于我们的目标词典的每个词条是一个五元组(汉语词语, 汉语词性, 英语词语, 英语词性, 置信度), 而原始词典中, 往往是将一个源语言词语解释成词性相同或不同的多个目标语词语, 因此需要将一个原始词典的词条拆分成一个或多个目标词典的词条。通过词条拆分, 词典就可以装入数据库中, 后续的步骤都是在数据库上面进行操作。

2) 格式整理: 原始词典中存在大量的格式错误, 需要校正, 还有很多的附加信息(如领域信息、解释性信息)需要滤掉。

3) 词性标记归一化: 由于每部原始词典的词性标记信息都不一致, 而且与核心词典的词性标记也不一致, 因此要将原始词典的词性标记映射到核心词典的词性标记;

4) 词性标记补齐: 由于原始词典中很多词没有词性标记, 而目标词典中所有词条(五

元组)都必须有完整的词性标记,因此对于原始词典中缺少的词性标记要补上,并给出置信度信息:

5) 词条合并:将来自不同原始词典的相同词条合并为一个词条,同时将置信度求和。

在词典整理过程中,必然设计一个人工校对的问题,如何实现人机互助的校对?如何用最少的人力、快速有效并且高质量的完成以上工作,是这项工作一个关键。我们把握的原则是:能够用机器做的事尽量不用人做;通过计算机程序,将无法由人工完成的复杂的校对工作分解成很多非常单一的校对任务,以尽可能地提高校对的质量和效率。

在我们的实际工作中,由于时间的限制,我们只在上述的第一步和第二步(词条拆分和格式整理)完成了人工校对,后面几步都是全自动进行的。虽然这免不了有很多错误,不过总体上词典还是可用的,而且概率信息的引入,也使得这些错误的影响降低到很低的程度。

下面我们分别就以上五个步骤介绍我们的具体做法。

3.1 词条拆分

原始词典的词条可能是以下形式:

complex=a. 复杂的, n. 复合体; 综合建筑

通过词条拆分,要变成以下形式:

汉语词条	汉语词性	英语词条	英语词性
复杂的		complex	a.
复合体		complex	n.
综合建筑		complex	n.

当然,实际操作并不这么简单,会遇到很多复杂的格式问题,有时还需要引入人工校对。通过这种方式,将原始词典从文本形式转换成了数据库形式,便于后面的进一步处理。

3.2 格式整理

每一部原始词典的格式都不相同,所以处理的方法也不尽相同。图2是其中一部词典的格式处理过程。

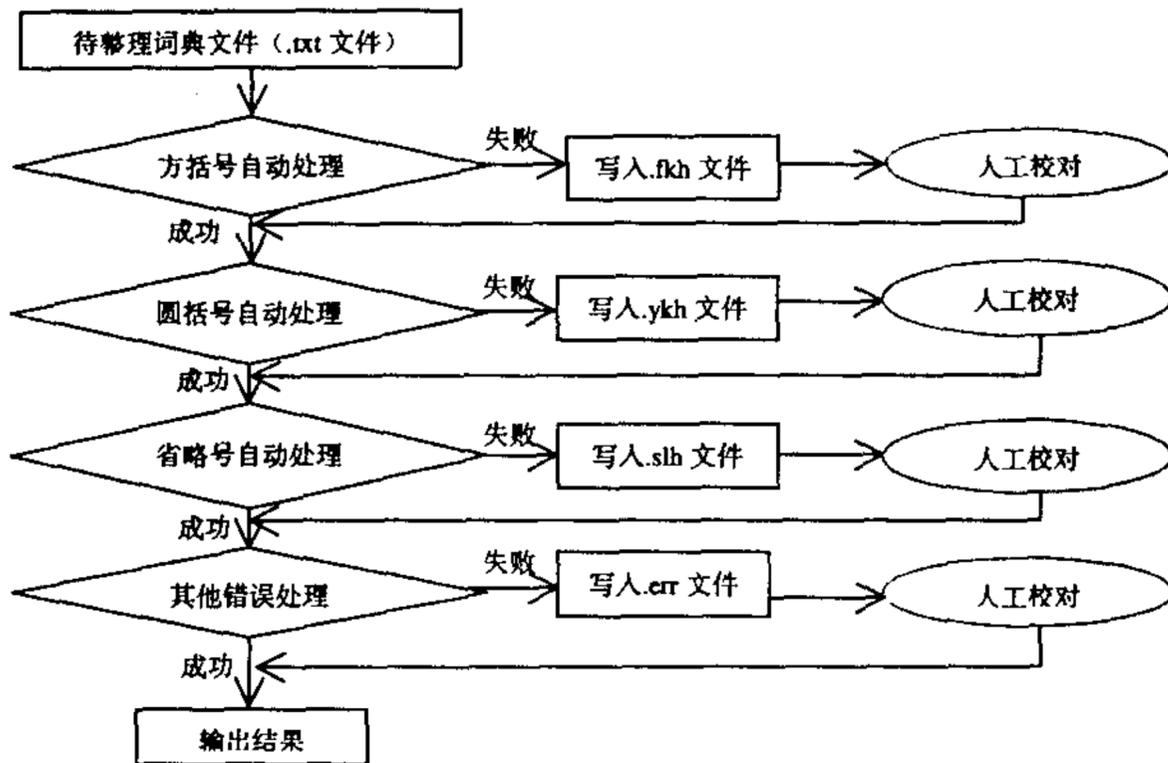


图2 词典格式整理过程示例

这种做法的好处有两点：1.自动处理阶段处理了大部分的格式问题。比如方括号的处理，很大一部分的方括号都是领域信息，而领域信息的种类是有限的，只要把表示领域信息的方括号处理好，剩下需要人工校对的工作量就很小了；2.通过这种逐级处理的方式，每一次人工校对的工作都变得非常单一，这一方面减轻了校对者在工作中的大脑思维活动的负担，使得校对国内工作的速度和质量都有所提高，同时也使得校对的工作量计算更为合理，有利于控制成本。

3.3 词性标记归一化

要为每一步原始词典的词性标记建立到核心词典的词性标记之间的映射关系。这样，绝大部分词性标记归一化工作可以自动完成。个别情况需要人工校对（暂时没做）。

3.4 词性标记补齐

原始词典中，很多词语都没有标记词性。英汉词典中，汉语词语没有标注词性。汉英词典中，几乎都没有任何词性信息。所以，词性标记补齐是一件非常重要的工作。

词性补齐需要用到两部（源语言和目标语言）带词性标注的单语词典。这两部单语词典的来源除了已经整理的原始词典以外，还利用了机器翻译系统的核心词典。

另外，需要用到一个双语词性对照表，也就是说，对于源语或译语的每一种词性，要给出其最可能的对译词性。这个表根据语言知识很容易构造。

在得到两部单语词典以后，对于每一个五元组，如果没有词性缺失，那么设置其置信度为1。如果有词性缺失，那么利用以下算法进行词性补齐并计算置信度：

1. 如果一种语言词语A词性已知，另一种语言词语B词性未知：
 - a) 按照算法2或者算法3，判断词语B的词性；
 - b) 如果词语B有多个词性
 - i. 如果词语B的词性中有与词语A的词性对应的词性，那么只保留这种词性，删除词语B的其他词性，生成一个五元组，置信度为1；
 - ii. 否则保留词语B的所有词性，将该五元组拆分成多个五元组，这些五元组的置信度都是 $1/n$ ， n 为B的词性个数；
 - c) 如果词语B只有一个词性，那么生成一个五元组，置信度为1；
 - d) 如果词语B的词性没有判断出来，将词语B的词性设置成与词语A的词性相对应的词性，生成一个五元组，置信度为0.5；
2. 如果双语词性都未知：
 - a) 首先分别使用算法2和算法3，判断两个词的词性；
 - b) 如果两个词语的词性都已判断出来
 - i. 如果两个词的词性中有些词性组合满足对应关系，那么只保留这些词性组合，生成相应的五元组，删除其他词性，生成的五元组置信度为1；
 - ii. 如果两个词的词性中没有词性组合满足对应关系，那么都保留，两两对应生成五元组，这些五元组置信度为 $1/n$ ，其中 n 为五元组的个数；
 - c) 如果只判断出一个词的词性，那么将另一个词的词性设置为已知词性的对应词性，生成相应的五元组，置信度均为0.5；

3. 如果两个词语都无法判断出词性，一律标为名词，生成五元组置信度为0.1。

算法1 词性补齐算法

1. 查单语词典，如果查到则返回成功；
2. 按照以下规则判断汉语词性：
 - a) 尾字为“的”=>形容词
 - b) ……
3. 如果判断出汉语词性，返回成功，否则返回失败；

算法2 汉语词性判断算法

1. 查单语词典，如果查到则返回成功；
2. 按照以下规则判断英语词性：
 - a) 首字母大写=>专有名词
 - b) 如果是单词=>根据词缀判断
 - c) 如果是短语=>根据首单词判断
 - i. 首单词是冠词=>名词短语
 - ii. 首单词是不定式 to=>动词短语
 - iii. 首单词是动词=>动词短语
 - iv. 首单词是介词=>介词短语
3. 如果判断出英语词性，返回成功，否则返回失败；

算法3 英语词性判断算法

3.5 词条合并

最后一个步骤是词条合并。词条合并就是将汉语词语、汉语词性、英语词语、英语词性相同的五元组合成一个五元组，其置信度是被合并的所有五元组的置信度之和。

4. 输出结果

词条合并之前，共有 681,676 个词条（五元组）。词条合并之后，还有 460,455 个词条。其中，不考虑词性，汉英词对大约有约 41 万，单独计算汉语词形和英语词形，分别有大约 21 万。

最后得到的词典，从直观上看，质量还是可以接受的，以下是两个词的例子：

阿姨	爱	
n,aunt,N,3.0000	v,love,V,14.0000	v,be fond of,V,2.0000
n,auntie,N,4.0000	n,love,N,11.0000	v,take good care of,V,2.0000
n,aunty,N,1.0000	n,affection,N,6.0000	v,treasure,V,2.0000
n,mother's sister,N,1.0000	v,like,V,4.0000	n,treasure,N,1.0000
n,nurse,N,1.0000	v,hold dear,V,2.0000	v,be in the habit of,V,1.0000
	v,be keen on,V,2.0000	n,fond of,N,0.2500
	v,cherish,V,2.0000	v,fond of,V,0.2500
	v,be apt to,V,2.0000	

由于整理过程中，后面几个步骤都没有经过校对，词典中也有不少错误，不过这些错误

的词条大多数置信度都比较小。

五、实验及结果

我们的机器翻译系统采用的是一种微引擎流水线体系结构 (Liu, 2000)。

扩充词典在我们的机器翻译系统中被两个微引擎所使用: 一个是词典转换微引擎。对于核心词典中找不到的词语, 如果扩充词典中能找到, 那么可以利用扩充词典的信息进行翻译; 另一个是短语分析引擎, 扩充词典中有很多词在核心词典和分词词典中都没有, 实际上这些词大部分都是短语, 这些短语被一个短语分析引擎利用作为短语引擎加入到了句法分析的过程中。

我们利用一个机器翻译的自动评分工具 Bleu (Papineni et. al., 2001) 对加入扩充词典前后机器翻译系统翻译的结果进行了比较。使用的测试集主要是新华社的新闻稿, 测试集中含有 800 多个句子。我们分别计算了使用扩充词典前后无法翻译的汉语词语数目: 使用扩充词典以前, 有 405 个汉语词语 (917 词次) 无法翻译, 使用后只有 91 个汉语词语 (177 词次) 无法翻译, 可以看到, 扩充词典的使用, 大大减少了无法翻译的词语数目。

我们利用 Bleu 对这个测试集在使用扩充词典前后的翻译结果进行了自动评测: 使用扩充词典以前, 翻译结果的评分为: 4.7580, 使用后翻译结果的评分为 5.7580。实验证明, 扩充词典的使用, 对于改善翻译的质量有相当明显的效果。

整个扩充词典的建设工作还有很多可以改进的地方。例如在词性补齐过程中, 如果引入单语的概率词典以及双语词性之间的转换概率信息, 词性猜测的正确率还可以提高很多。另外, 很多中英文词条实际上都是短语, 如果采用一些浅层分析的方法 (如有限状态自动机), 词性的判断会更为准确。

这部扩充词典除了用在我们的机器翻译系统中以外, 还在我们正在进行的双语语料库对齐中得到应用。以后, 我们计划开展基于统计方法的机器翻译研究, 这部扩充词典也可以在其中发挥较大的作用。

致谢: 在词典整理工作的初期, 李丙辰同学曾经承担了其中大部分工作并贡献了很多有价值的思想, 在此向他表示感谢。

参考文献

[1] Church K.W. and Hanks. Word association norms, mutual information and lexicography. In: Computational Linguistics 16(1).1990:22~29.

[2] Fung P. Pattern Matching Method for Finding Noun and Proper Noun Translations from Noisy Parallel Corpora. In: proceedings of the 33th Annual Meeting of the Association for Computational Linguistics, Boston, USA:1995.

[3] Haruno M., Ikehara S. and Yamazaki T., Learning bilingual collocations by word-level sorting. In: COLING96.1996:525~530.

[4] Kishore Papineni, Salim Roukos, Todd Ward etc. Bleu: a Method for Automatic Evaluation of Machine Translation. IBM Technical Report, (keyword = RC22176) <http://domino.watson.ibm.com/library/CyberDig.nsf/home>.2001.

[5] Melamed I. D. Automatic Construction of Clean Broad-Coverage Translation Lexicons. In: Conference of the Association for Machine Translation in Americas, Montreal, Canada:1996.

[6] Qun Liu. A Chinese-English Machine Translation System Based on Micro-Engine Architecture. International Conference on Translation and Information Technology, Hong Kong:2000.

[7] Qun Liu, Baobao Chang, Weidong Zhan etc. A News-oriented Chinese-English Machine Translation System. International Conference on Chinese Computing(ICCC2001), Singapore:2001.

[8] Smadja F., McKeown K.R. and Hatzivassiloglou V. Translation collocations for bilingual lexicons: a statistical approach. In: Computational Linguistics 22(1).1996:1~38.

[9] 刘群, 詹卫东, 常宝宝等. 一个汉英机器翻译系统的计算模型与语言模型, 第三届全国智能接口与智能应用学术会议, 吴泉源, 钱跃良主编, 智能计算机接口与应用进展, 电子工业出版社, 1997:253-258

[10] 俞士汶, 朱学锋, 王惠等. 现代汉语语法信息词典详解, 清华大学出版社, 1998

[11] 董振东, 董强, 知网. <http://www.keenage.com>

[12] 王慧, 詹卫东, 刘群. 现代汉语语义词典的概要及设计, 中文信息处理国际会议论文集, 清华大学出版社, 1998

作者简介

刘群(1966-), 男, 江西萍乡人, 在职博士生, 副研究员, 主要研究领域为机器翻译, 自然语言处理与中文信息处理。

张彤(1967-), 男, 副研究员, 主要研究领域为自然语言处理。

Construction of Extended Lexicon for Chinese-English Machine Translation

Liu Qun¹ Zhang Tong²

¹ Institute of Computational Linguistics, Peking University

² Institute of Computing Technology, Chinese Academy of Sciences

E-mail: liuqun@ict.ac.cn

Abstract: This paper firstly gives a general introduction to the resource construction of "News-Oriented Chinese-English Machine Translation System", and then describes the construction of the extended lexicon in detail. Compared with the kernel lexicon, the extended lexicon has a larger vocabulary, but less information for each word entry. The construction of the extended lexicon includes the following steps: collection of original lexicons, formatting, normalization of POS-tags, filling the lacking POS-tags and merging the identical entries. The experiments show that this is an efficient way to build a large scale usable lexicon for machine translation. The extended lexicon greatly reduces the untranslated words in the output of the system and improves the translation quality.

Keywords: Machine Translation; Bilingual Lexicon Acquisition

作者: [刘群](#), [张彤](#)
作者单位: [刘群\(北京大学计算语言学研究所\)](#), [张彤\(中国科学院计算技术研究所\)](#)
被引用次数: 2次

本文读者也读过(10条)

1. [曹冬月](#), [Cao Dongyue](#) 词典与英语词汇学习[期刊论文]-河北科技大学人民警察学院学报2002, 2(3)
2. [朱海](#), [李淼](#), [张建](#), [乌达巴拉](#) 系统融合方法在汉蒙统计机器翻译上的应用[会议论文]-2010
3. [史晓东](#), [陈毅东](#) 基于语篇的机器翻译前瞻[会议论文]-2006
4. [王厚峰](#) 机器翻译中汉语词节点的识别[会议论文]-2003
5. [王海峰](#), [刘占一](#), [吴华](#) 半结构化的基于实例的机器翻译[会议论文]-2007
6. [王辰](#), [宋国龙](#), [吴宏林](#), [张俐](#), [刘绍明](#) 基于序列相交的短语译文获取[会议论文]-2008
7. [马志强](#), [周长胜](#), [丁维](#), [杨娜](#) 自扩充中文分词词典的研究与实现[期刊论文]-计算机与数字工程2007, 38(6)
8. [张步峰](#), [何丕康](#), [张冬冬](#), [李沐](#), [周明](#) 短语切分概率在统计机器翻译系统中的应用[会议论文]-2006
9. [马道山](#) 机器翻译及其翻译步骤[会议论文]-2002
10. [张振中](#), [孙乐](#), [张大鲲](#) 一种改进的统计机器翻译系统融合方法[会议论文]-2009

引证文献(2条)

1. [邓丹](#), [刘群](#), [俞鸿魁](#) 基于双语词典的汉英词语对齐算法研究[期刊论文]-计算机工程 2005(16)
2. [刘群](#) 基于微引擎流水线的机器翻译系统结构[期刊论文]-计算机学报 2004(5)

本文链接: http://d.g.wanfangdata.com.cn/Conference_4203668.aspx