

代汉语语义词典多义词词库的校正和再修订

摘要

本文依据面向汉语信息处理的词语义区分的完备性和操作性原则，基于代汉语语义词（SKCC）和代汉语语法信息词（GKB）的词对，以语料为依托并结合代汉语词、义词词林和代汉语搭配词词等资源进行代汉语语义词的多义词词校工作。首先SKCC和GKB的词对入手，设计了候选拟修改多义词的取算法，对取出来的1605个多义词进行了义补录、合并、删除，补充释义，修改翻译和示例等工作，针对类特殊的食物+作物类多义词建立了义的树形结构以满足粒度的词义消歧任要求。方便续的修改工作，本文开发了SKCC和GKB之间的词映射，在的基础进行了多义词映射工作

关键词 义区分 代汉语语义词 多义词映射 多义词修改

New editing and checking work of the Semantic Knowledge base of Contemporary Chinese (SKCC)

Abstract:

This paper is rooted in the two principles and methods that should be followed by sense discrimination for Chinese language processing: Completeness and discreteness. Built on the comparison of Semantic Knowledge-base of Contemporary Chinese (SKCC) and Grammatical Knowledge base of Contemporary Chinese (GKB), supported by large scale corpus, we conducted our new editing and checking works. Firstly, we designed a novel multi-sense lexicon candidate abstraction algorithm based on lexicon comparison between SKCC and GKB. For all 1605 candidate multi-sense lexicon, we conducted editing work on the senses, explanation, and its translation. Then, we built a tree structure to process a special food and plant lexicon. Thirdly, a mapping platform between SKCC and GKB has been built to help us built mapping relationships between multi-sense lexical between SKCC and GKB. Finally, we finished mapping work for all multi-sense lexicon in SKCC.

Key words: Distinguish word sense; SKCC; Multi-sense word mapping; Multi-sense word editing

一、引言

代汉语语义词（Semantic Knowledge-base of Contemporary Chinese，以简称SKCC）是个面向汉英机器翻译的大规模汉语语义知识，目的是在语法分析的基础上，给自然语言处理提供更加全面、深入的语义信息。[1]作国家科技进步二等奖获得项目“综合型语言知识库（Comprehensive Language

Knowledge Base, 以 简称 CLKB) ”的 部分, SKCC 被广泛 用于 算词
汇语义学的基础研究和 用研究之中, 例如魏雪 袁 林 2014 以 SKCC 为
依托建立了 词语义 类组合模式 [2], 张仰森 (2012) 利用 SKCC 进行了词汇语
义相似度的计算[3]等 来 计算词汇学特别是词义划分理论领域取得了较大的
进展, 吴云芳、 士文 (2006) 出了 实际操作性的面向汉语信息处理的
词语义 区分原则和方法 [4], 而 SKCC 自 2003 第 版发布以来 直没 进行
大规模更 因 , 必要结合自 SKCC 发布以来语义词 编纂和词义划分理
论的 果, 对 SKCC 的多义词词 进行 修 , 使 更好地 自然语言处理服
务。

本文首先建立了 SKCC 和 GKB 的多义词映射平台, 据映射的结果 确定拟进行
修改的多义词词条, 然 结合 SKCC、《现代汉语语法信息词典》(Grammatical
Knowledge-base of Contemporary Chinese, 以 简称 GKB) [5]、 义词词林
[6]、 代汉语搭配词 词 [7]和 代汉语 词 [8]等多部词 资源, 以 2000
年 1-3 的 人民日报 语料 和 京大学 CCL 语料 语料资源进行 SKCC
多义词词 校 和 修 工作 工作包括对 1356 个 SKCC 多义词词条的校 和
1508 个多义词的映射, 经过校 和 修 工作的 SKCC 在多义词部分规模和
质量方面有了较大提高, 可以为计算语义分析和词义消歧等任务提供更好的支持

二、 SKCC 改进的理论分析

义 划分是 SKCC 编制的 心问 , 也是词义消歧等自然语言处理任 的基础
性工作 Palmer (2001) [9]指出自然语言处理技术的瓶颈之 在于 计算机 确
辨析词义。吴云芳、 俞士文 (2006) [4]提出了作为自然语言处理任务的多义词
词义划分的 个原则

- (1) 般限定 词类 的多义词而 包括跨词类的多义 象
- (2) 将 形异义词和词语的 义 在 个 面 考察而 追求 形和多
义的 格区分
- (3) 消歧处理的对象 要是词汇义而 是在词语或固定组合中出 的语素义

基于 个原则 , 出了 “完备性”和“离散性”的要求, 并 出 基于语料 实证进
行词语义 粒度划分的准则

“完备性”的要求 据词 语的义 区分 , 操作者 以 利对语料中的每 个目标词
标注出义 。SKCC 中部分词条 符合 “完备性”的要求, 例如 SKCC 对“大小”区
分了 个义

- 〔大小〕 (1) 量 属性, 物体的 大小
(2) 身份, 大人和小孩

但 遇到 人民日报 语料 中 列词组时 , 上述 2 个义 显然无法 之对
大小领 /大小领 /没大小

就需要添 个义 : 表述关系类的“大小”, 释义 “表示关系的尊卑或长幼”。

“离散性”是指意义分析系统中 义 的内涵 重合 ，SKCC 中 部分词义 之间存在义 重合甚 蕴含的 象，例如 SKCC 中对“驾驶”区分了四个义

- 〔驾驶〕
- (1) 操纵 载工
 - (2) 操纵车 船等 载 工 行驶
 - (3) 操作 他工 作业
 - (4) 操纵航空器

四个义 中 ，义 2、3、4 仅释义 以包含在义 1 中 ，在 文 境 中 体 体论元角色和语义类 也完全

除了“完备性”和“离散性”的要求外，词 划分 粒度 是 SKCC 中的 问 ， Ide(1998) 指出，传统辞书中的义 划分过于细微， 对 这些义项进行归 并[10]，对于 并的 粒度， 基于语料 实证并 持词 内部的 性 。然 而，在 SKCC 中部分词条之间 粒度划分存在 ，例如， 语义类 使用领 域和联系的谓词来区分，单音节词“纲”在 代 汉语中 3 个义

- 〔纲〕
- (1) 物的关键部分
 - (2) 生物学分类的 种 类别
 - (3) 统治者认 维持 常秩序的必 少的行 规范

义 1 的语义类是“ 件 ”，常用于 府文告 例如：“学好文件抓住 ”);义 2 的语义类是自然物，常用于科学领域，和生物类别相搭配；义 3 语义类是人工 物，常和“ 件 ”语义类的 词搭配 但是 SKCC 仅 个义 人工物

外 ，SKCC 多义词词库 部分条目 存在着翻译 、用例、释义 者 对 的 象，本文将在义 过程中 并进行修改 。

三、 候选修改多义词的 取 和子 集成

经统 ，SKCC 中共 多义词 3052 个，分布在 21 个子 之中 ，用逐 阅对 的方法会耗费大量的人力物力，而 在总 中的修改难以 映到各个分 中 。为 了 影响总 和子 的 性 ，并使对总 的修改 映到子 ，本文将各个子 集 到总 中，集 的总 字段是集 前各子 字段的并集，包括 21 个字 段 表)。集 工作完 所 的对 修改工作都 以在总 中进行

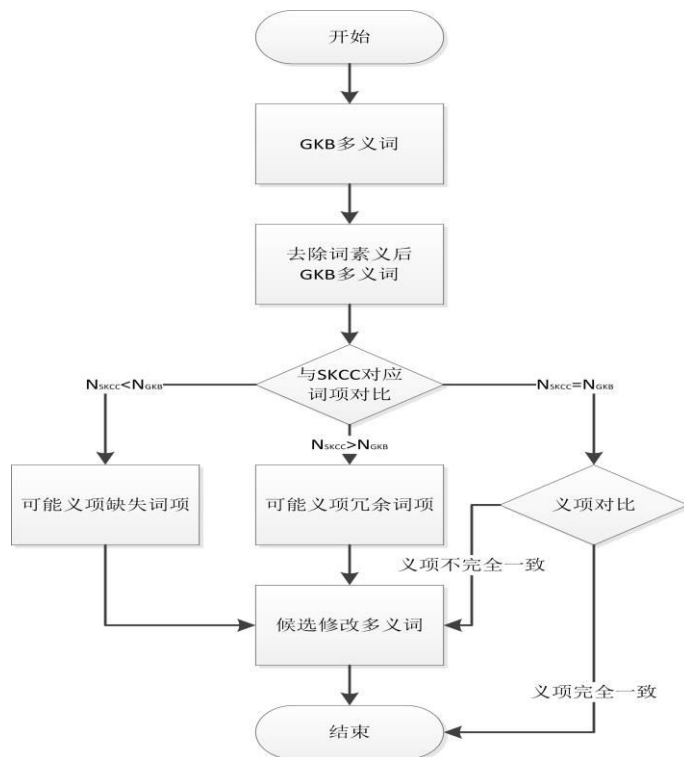
表一：集成后总库字段

序号	字段	长度	备注
1	词语	8	收录 1~4 个字的词语
2	拼音	24	每个词语的汉语拼音
3	词类	2	词语所属词类的代
4	子类	2	词语所属词类的子类代码
5	同形	2	词语的同形词
6	兼类	4	该词语兼属的词类代码
7	义项	2	对“同形”字段相同的词条进一步加以区分
8	释义	10	该词语的简明释义

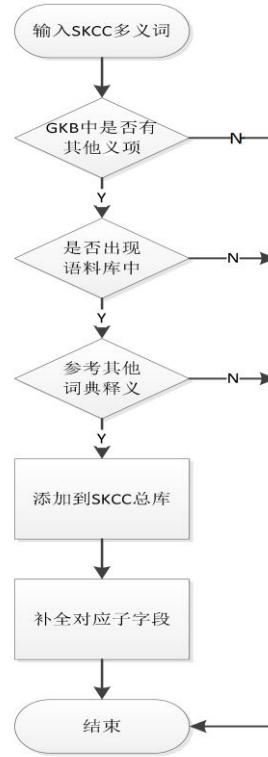
9	语义类	20	该词语的语义类别名称
10	WORD	40	该词语对应的英语译词或短语
11	Ecat	40	该词语的英语译词的词性代码，或短语组成结构
12	备注	20	填写词语用法的简明说明(仅用于与古汉语标注填写‘G’,仅用于方言标注填写‘F’)
13	配价数	2	(动词)说明一个动词能支配多少名词性成分/(名词)要求与之共现的从属名词个数/ (形容词)要求与之共现的名词性成分个数
14	例句	20	简要提出几个该义项的示例
15	主体	20	动作主体所属的语义类名称
16	客体	20	二价和三价动词的客体语义类名称
17	与事	20	三价动词的与事所属的语义类名称
18	参照体	20	(名词)写一价和二价名词参照体(配项成分)的语义类名称
19	对象	20	(二价名词)二价名词的对象的语义类名称/(二价形容词)关涉对象的语义类名称
20	直接上位	20	(名词)直接上位的概念
21	主体	20	(形容词)主体的语义类名称

由于现代汉语语法信息词典（GKB）自发布以来经过了反复的修改和校正，其多义词部分的义划分已较合理，本文使用开发较完备的代汉语语法信息词典（GKB）为参照词典，通过两部词典多义词对比取需要进行审订和修改的候选多义词，多义词的较在集的 SKCC 总 和 GKB 总 之间进行，通过词对 取候选多义词的算法流程如图一：

图一：提取问题候选词流程



图二：补充未收录义项流程



对 取算法旨在将 SKCC 和 GKB 多义词进行较 首先去除掉 GKB 中收录

的词素义，去除词素义的 GKB 多义词和 SKCC 中的对 词汇进行 义 对 ，如果 SKCC 对 词 义 数少于 GKB 的，认 能存在义 缺失的情况 ；SKCC 对 词 义 数多于 GKB 的，则 能存在义 划分过细的情况 如果 部词 义 数相 但义 的 ，则 能义 遗漏 或划分过细的问 兼而 之 ，以 部分构 了候选修改多义词词表 共 取出 97 个 SKCC 和 GKB 义 数相 但义 的词 ，SKCC 义 数多于 GKB 对 义 的词 1448 个，SKCC 义 数目 少于 GKB 的词 18 个和 SKCC 未收录的 GKB 多义词 42 个，共 1605 个校对修 改候选多义词词

四 对 SKCC 的修改工作介绍

(一) 补录未收录词和未收录义

部分工作旨在使 SKCC 中多义词的义 划分接 “完备性”的要求，义 的补充 建立在语料 实证基础之 ，要以 2000 年 1-3 个 已标注的人民日报语 料 和 京大学 代汉语 CCL 语料 语料 资源 对于 GKB 收录但 SKCC 未 收录的义 ，本文参考了 代汉语词 第 版 》和 代汉语搭配词 ， 以补录 对于 GKB 以“规范” 例， GKB 中收录了 个义

〔规范〕 (1) 动词，谓使合乎模式

(2) 词，约定俗 或 明文规定的标准

(3) 形容词，符合标准的

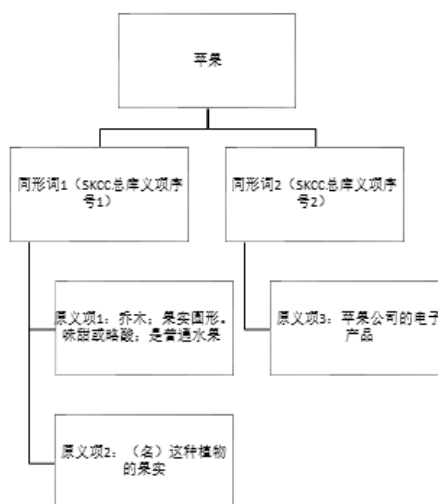
SKCC 中仅收录了义 (1)和(2)，在决定是否将义 (3)补录到 SKCC 中前，本文 在 个 已标注人民日报语料 中查找得知“规范”共出 493 例，其中符合义 (3)的“规范”词条共 123 例，占 24.9%。时 ， 代汉语词 第 版 》也收 录了“规范”的形容词义 本文将 GKB 中第 个义 入 SKCC 总 中，整个 补录未收录义 工作 2。

在补录未收录义 ，本文共对 264 个多义词进行了补充义 ，补充义 的词类 分布 表 2， 个工作是补录 SKCC 未收录的多义词，由于 SKCC 编辑的时候 难免 遗漏的词条 ， 们选取了 断进行修 增改的 GKB 作 补录 多义词的来 源，通过对 共补录 42 个 SKCC 第 版未收 录的多义词

表二：补充义项的词类分布

	词类		词条数
	词类	词条数	
实词	名词	48	
	动词	33	
	形容词	23	
	量词	42	
	代词	2	
	区别词	40	
	介词	14	
虚词	连词	16	
	副词	46	
总计			264

图三：‘苹果’的树形结构图



(二) 原义的并和备注标

对 SKCC 中的多义词义项进行合并和备注标要基于两个原则，第一是虽然“完备性”是义项收录的重要要求之一，但是目前的自然语言处理现状决定义项“完备性”要限于共时和书面的现代汉语普通话语料。因此，对于词语的一些历时意义予以标明，根据自然语言处理的体任决定是否使用。第二，面向自然语言处理任务的词语义项区分是“实证型”的，也就是说计算机依据词所在的文章关系找出清辨析词的，进行形式描述的法行。就要求面向自然语言处理的词在义项选取时依据形式的标准。例如论元结构、语法框架和选择限制。而能基于世界知识区分意义，王惠(2004)也指出[12]，词语组合对词语义项划分重要的影响。因此，严格地按照词语的法行并基于语料的实证来区分义项更容易取得较好的结果。

解决第一个问题，我们对 SKCC 中多义词词数大于 GKB 的 1448 个词收录的义项进行逐条校对。在校过程中对被认定是仅存在现代汉语或者部分方言地区的词汇义项，本文在人民日报语料和京大 CCL 现代汉语语料中进行查找，查得义项确实未出现在上述语料中或虽然出现但基本引用代文献用法的，本文在 SKCC 中的“备注”字段添备注说明。如 SKCC 中，“山陵”共以 3 个义项

- 【山陵】 (1) 词，山川丘陵等地貌
 (2) 词，对帝王的代称
 (3) 词，指帝王的陵墓

GKB 中对“山陵”只收录了第一个义项，在人民日报语料中三个义项均未出现，在 CCL 现代汉语语料中“山陵”三个义项共出现 32 次，中义 1 出现 5 次，义 2 和 3 共出现 27 次，义 2 和 3 全部代人在谈话或文章中引用。籍因，我们在修改 SKCC 中对原的第一个义项予以保留，在义项 2、3 添了表示义项仅用于现代汉语的备注。对于部分常用于现代汉语，

在相义在代汉语中然留的，们将义合并到代汉语然留的义中，例如 SKCC 中“参”词以 3 个义：

- 【参】 (1) 词，参的人
(2) 词，府官职
(3) 动词，参

义 (2) 是古代中国、日本常用的官职名，在现代汉语中已经不再使用，但是义项 (1) 的用法在些企业和社团中经常使用，在 CCL 语料机抽取的 500 条含“参”的录中，有 18 条是义 1 的用法，因们将义 1 和 2 进行合并

解决第个问，们依据词语在文语境中的法行，对于义意义虽然在世界知识细微差异，但没明确的法的差异的，们将进行义合并 合并的依据要据 GKB 各子中列出的法差异例如，前文中到过的“驾驶”，们将 4 个义均尝试填入 GKB 中的语法述框架之发，4 个义都以接词都以做动趋式都以搭配“着了过”和“在”，共的接分，所以将 4 个义合并个义部分共对 247 个多义词的义进行了备注添和合并

(三) “食物+作物”类词汇的树形结构处理

外类特殊的词部分植物和果实在 SKCC 中的释义和语义类，在代汉语语法信息词和代汉语词典》中也将作物和果实食物区分个的义。但是在词义消歧过程中，由于个义的内涵重叠部分，区分作物和果实较困难，时个义在文中的法和搭配境也高度相似以“苹果”例，在 CCL 语料中要 3 个义项：

- 【苹果】 (1) 词，乔木果实圆形味甜或略酸是通水果
(2) 词，种植物的果实
(3) 词，苹果的电子产品的总称

中义 1 和义 2 是作物和果实食物区分，义 2 的外延蕴含在义 1 的外延之中对于这类词语，我们在校订过程中建立了该类词的树形义项结构：在总中将义 1 和义 2 进行合并，然将原的义 1 和义 2 存储在建的子中，用“词语”、“形”、“词类”进行链接，构上图三所示的树形结构

针对粗粒度的词义消歧任，们以直接调用总中的个形词，区分作植物或食物的苹果和作电子产品的“苹果”，对于进一步要求区分作为植物的“苹果”和作果实的“苹果”，们以使用分中的个义进行进区分 本文共对 152 个作物/食物果实类多义词建立了如的树形结构

四 修改释义示例和翻译

在基于语料的词义辨析模型中，依据语料中的词语列进行义是重点和

心 Kilgarriff(1997) [12])，而词语义的释义较少的被及。们在利用词语的语法特征进行词语义的增删改并工作之，也以利用语法特征来完善词的释义。因，们在释义修改的基础对释义空缺的多义词词进行填充。释义的填充以语义类和子表中的相关字段作选择对释义的判断依据，参考代汉语词中对释义，并结合义在文搭配中的语法特征填写的释义。例如：多义词“地道”在原SKCC中3个义，释义均未填写，但是SKCC中给出了每个义的语义类，们据语义类判断出义3语义类建筑物在代汉语词中解释“地的道路或坑道”。义1和义2的语义类都是“性质”，但是们可以通过英文翻译(WORD字段得知，义1对“技能、工作或材料的质量够标准以重叠，语是技能或物体”，义2解释“为人合乎定的道德规范重叠，语是人或行等”。

本着更好地配合释义、体用法以及扩大词汇信息量的原则，对例词、例做了相的修改，特别是对释义和示例能对的情况，本文结合代汉语词和代汉语搭配词词，删除了部分重复示例，并对示例进行修改，示例选择述部参考词中相义列出的示例以进确示例的准确性。在校正过程中，们共对449个多义词进行了义释义添，在义释义添的基础，对100个多义词的错误示例进行了修改

在对义释义补全修改释义之，们对468个多义词的英文翻译进行了修订。修时依照《柯林高阶英汉解词[13]，查找得出对的多义词义的翻译，和原的人工翻译进行对。对部分多个英文翻译的单音节词，依照修改的SKCC示例，查找在线英语语料中对SKCC示例的翻译，经人工对给出多义词的翻译。选取英文翻译时，本文尽量选择单词而是短语，对人、地等词的翻译据世界人翻译大辞[14]和《外地译手[15]。对喻意义的多义词，本文在查阅相关资料发掘出喻意义后，在柯林高阶英汉解词[16]中找多义词义对的翻译。例如多义词‘清流’以两个义

【清流】 (1) 词，清澈的流水

(2) 词，喻指德行高洁负望的士大夫

第二个义项是‘清流’的喻义，在柯林高阶英汉解词》中以下平行例

中文例 他们都是明朝著的清流。

英文例 They are all famous virtuous scholars of the Ming Dynasty.

据例的翻译，‘清流’的第一个义对‘virtuous scholars’，字面翻译‘有德行的士大夫’，释义基本对，被选对翻译

五 SKCC和GKB映射的开发和多义词映射工作

了对所修改结果统管理，并满足日多人时操作修改词的需要，本文在卞峰(2015)[17]的基础建立了SKCC和GKB的映射能能够自动查找待映射词语在部词中的词义解释，需要人工逐个查找；同时，

能够快速建立映射关系，将建立的映射关系存储到映射关系数据库中，减轻了词映射的工作量。映射要分个模块：词语查询，建立、存储映射关系，查询映射结果，映射工作流程如流程4。

词语查询 浏览器输入待映射词语并发送给服务器端，服务器端通过 SQL 查询语词数据中查找，返回查询结果，显示到浏览器中。对于 GKB 的查询由于分的存在需要入额外的查询步骤：服务器接收到查询词语时，首先需要先在 GKB 总表中找出所查词语的所词性，然后根据词类分别对分中查找相词语信息，将各分结果汇总返回给浏览器端进行显示。

建立存储映射关系 SKCC 数据中“词语”、“词类”、“义”、“形”字段提供了词的唯一标识，在 GKB 数据中，“词语”、“词类”、“形”等字段也能够作唯一标识。由于每次查询结果的序都是变的，因在存储映射结果时，本文选择部词的唯一标识进行存储。在建立映射关系时，为了减轻人工映射的工作量，将查询结果进行编号，映射时只需要进行编号的填写。填写完后，向服务器端交映射结果，服务器端根据编号将对标识存储到数据库中，存储结果如表四所示。

表四：映射关系在数据库中存储结果

id	word	sem_tag	sem_no	sem_tongxing	gkb_tag	gkb_tongxing
4	南瓜	n	1	1	n	1
5	南瓜	n	2	2	n	2
6	照会	v	1	None	v	None
7	照会	n	1	None	n	None
8	清爽	a	1	None	a	None
9	清爽	a	2	None	a	None
10	预见	v	1	None	v	None
11	预见	n	1	None	n	None
12	内线	n	1	None	n	None
13	内线	n	2	None	n	None
14	血脉	n	1	None	n	None
15	血脉	n	2	None	n	None
16	成就	v	1	None	v	None
17	成就	n	1	None	n	None
18	冶炼	v	1	None	v	None
19	冶炼	v	2	None	v	None
20	收敛	v	1	None	v	None
21	收敛	v	2	None	v	None
22	年收入	n	1	None	n	None
23	年收入	n	2	None	n	None

在表四中，“id”属性为数据库自动创建，“word”表示多义词语，“sem_tag”、“sem_no”和“sem_tongxing”分别表示 SKCC 中词语的词类、义、形标，这个属性信息“word”能够在 SKCC 中唯一标识一个词，而“gkb_tag”、“gkb_tongxing”表示 GKB 中词语的词类、形标，在 GKB 中一个属性“word”能够唯一标识一个词，便组了条映射关系。

查询映射结果：映射结果的查询词语的查询是时进行的，若查询的词语已经进行了映射，在结果显示时就会将词语的映射关系时显示出来。由于存储的映射关系是以部词的唯一标识进行存储的，因在显示映射关系前需要将唯一标识转换词语查询结果的编号。对于映射过程中出映射错误的情况，平台可以对问的映射关系进行增删改操作。

在映射的基础，本文完了修改 SKCC 全部 3282 个多义词和 GKB 中对词条的义映射工作。映射工作要依据释义和示例，如表五的“翻译”，在经过第四章的修改操作 SKCC 中收录的个义。

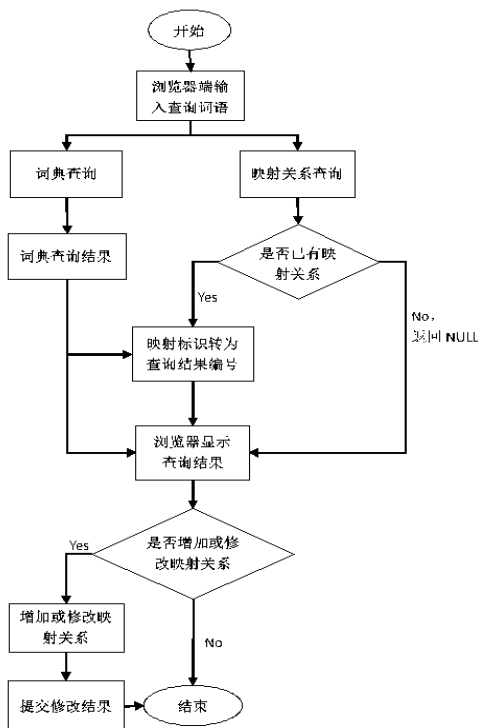
〔翻译〕 (1) 动词，用种语言文字来表达种语言文字。

- (2) 词， 翻译的 人
- (3) 词，把 种语言 译 种语言 的文本

GKB 中“翻译”有 3 个义 义 释义空缺 :

- 〔翻译〕 (1) 动词
(2) 词，职业

中， GKB 中“翻译”的第 个义 和 SKCC 中义 1 词性对 ， 据词性 建立 起 SKCC 中义 1 和 GKB 中义 1 的映射， 据词性和示例 SKCC 中义 2 则 对 GKB 中义 2，SKCC 中义 3 在 GKB 中没 对 义 ， 映射结果如表五 所示



图四：映射平台工作流程图

表五：多义词“翻译”的映射结果

《语义词典》

序号	词	拼音	词性	义项标号	同形	释义	语义类	示例	WORD
1	翻译	fanlyi4	v	1	None	用一种语言文字来表达另一种语言文字	创造	翻译小说/翻译名著/现场翻译	translate
2	翻译	fanlyi4	n	1	None	从事翻译的人	职业	随行翻译/两位翻译/俄语翻译	translator
3	翻译	fanlyi4	n	2	None	把一种语言译成另一种语言后的文本	创作物	一篇翻译/一种翻译	translation

《语法信息词典》

序号	词	拼音	词性	义项标号	释义	示例
1	翻译	fanlyi4	v	None	None	“日语/质量/对论文的”不满意
2	翻译	fanlyi4	n	None	None	随行/两位/俄语”

语义词典序号	语法信息词典序号	
1	1	删除
2	2	删除

总结 展望

本文依据 SKCC 和 GKB 的词 对 ， 以语料 依托并结合 代汉语词 、

义词词林 和 代汉语搭配词词 等词 资源 ,完 了对 1605 个 SKCC 多义词词条的校 和 3282 个多义词的映射工作 依据 “完备性”和“离散性”原则,对 SKCC 和 GKB 义 的词 进行了 和校 ,并针对 类特殊的 “食物+作物 植 物 ”类多义词建立了 层 义 的树 形结构以满足不 粒度的词义消歧任 要求 。方 便 续的修改工作 ,本文开发了 SKCC 和 GKB 映射 ,在映射 的基础 进行了多义词映射工作

本文的工作仅限于在原 语义词 的框架 进行修改 ,了完全实 义 划分的 “完备性”和“ 操作性 ”原则,们需要基于 文聚类技术 ,在大规模语料 中进行词类义 划分 。时 ,们注意到对利用统 模型概率 词语属性的方法对 GKB 的概率 改 在进行 吴林 张仰森 2011) [18]),对 SKCC 进行概率 改 ,得到并完善能广泛 用于词 义消歧、查错等领域的词汇知识 是 们的 个目标

参考文献:

- [1] 王惠,詹卫东, 士汶 . 代汉语语义词 规格说明书 [J]. 汉语语言 算学报, 2003, 13(2): 159-176.
- [2] 魏雪,袁 林 . 基于规则的汉语 组合的自动释义研究 [J]. 中文信息学报, 2014,28(3).
- [3] 张仰森,钟鼎. 基于 SKCC 统 相结合的 词语相似度 算方法 [J]. 京信息科技大学学报, 2012,27(6): 8-12
- [4] 吴云芳, 士汶 .信息处理用词语义 区分的原则和方法 [J]. 语言文字 用 , 2006,2: 126-133
- [5] 士汶 ,朱学锋. 《代汉语语法信息词 的 进展 [J]. 中文信息学报, 2001, 15(1): 59-64.
- [6] 梅家驹、竺 鸣 、高蕴琦、殷鸿翔编 1984. 义词词林 [M], 香港: 商印书馆·海辞书出版社
- [7] 梅家驹. 代汉语 搭配词 [M]. 京 :中 人民出版社,1994
- [8] 吕叔湘, 声树 ,江蓝生等. 代 汉语词 第 版 [M]. 京 :商 印书馆,2012.
- [9] Palmer, M. Consistent criteria for sense distinctions [J]. Computers and the Humanities,1997,31:91-113.
- [10] Ide, N. and Veronis, J. Introduction to the special issue on word sense disambiguation: the state of art [J]. Computational Linguistics, 1998, 24(1):1-40.
- [11]王惠. 代汉语 词义位的组合分析研究 [M]. 京 : 京大学出 版社,2004.
- [12] Kilgarriff, A. I don't believe in word senses [J]. Computers and Humanities, 1997, 31:91-113
- [13] 梅家驹. 代汉语 搭配词 [M]. 京 :中 人民出版社,1994.
- [14] 华通 社译 室 . 世界人 翻译大辞 [M]. 京 :中 对外翻 译出版 司,2007.
- [15] 周定 等 . 外地 译手 [M]. 京 :商 印书馆 ,1993.
- [16] 姚乃强. 柯林 英 汉 解词 [M]. 京 :汉语大词 出版社 ,1999.
- [17] 卞 峰 . 面向全文标注的中文消歧研究 实 [D]. 硕士学位论文,南京师范大学,2015.

[18]吴林,张仰森,王璐. 代汉语语法信息词的概率改及用[J]. 北京信息科技大学学报, 2011,26(6): 57-61.