

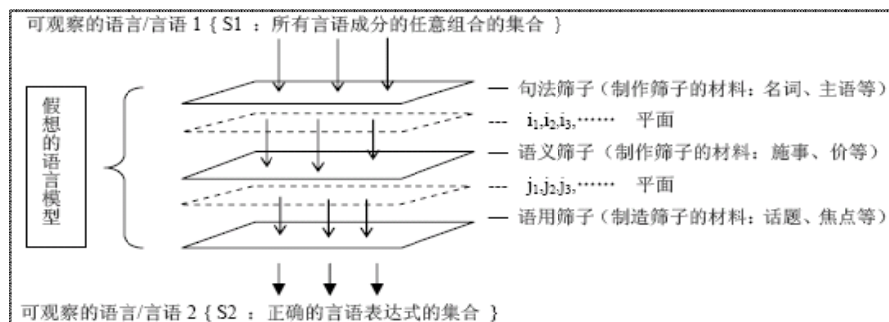
1 中国語補語動詞の記述方法

中国語補語動詞（以後 DCVC [Dictionary of Chinese Verb Complement] と略）は、その融通無碍な結合特性と、語義結合の複雑性から、統語ルールだけで言語運用の実態を把握することは難しく、定性研究の暗礁地帯とされる。従来中国語補語研究は、補語の定義や分類自体に異説が多く、辞典やテキストも個別の記述に留まらざるを得ず、習得にもネックとなっている。

本辞書は、大規模自然言語資源に基づき、実生活で常用される生きた中国語補語の結合頻度を科学的統計処理によって抽出後、中国語動詞(形容詞を含む)の語義特性(Semantic Role)について、個別に丹念に記述した Web 辞書である¹。各見出し項目には、大規模コーパスやインターネット言語資源を参照して選定した動詞補語例文を、外国人学習者用にリライトして収録する。

言語学では、実データから語の意味分析を行う記述的研究や、言語理論ベースの生成語彙意味論、語彙概念構造、フレーム意味論など、語彙意味論研究が多様性を見せながらめざましく発展している。しかし語義を詳細に規定しようとするほど、語義の範疇の確定と意味体系をめぐる意見の対立を引き起こす。従来、辞書の意味の記述には、同義語、類語、反義語、あるいは上下位の概念からその言葉の属性を規定する方法が多く用いられる。これら語義は相互に巨大なネットワークを張り巡らしているはずで、統合的な言語知識に依拠しない勝手な属性の定義は語彙記述間の矛盾となって現れる。近年、Word Net のように高度に形式化した概念構造に依拠した大規模同義語集の実現によって、語彙フレームワークの規定に基づいた語義コーパスの構築が試みられるようになり、これら実データに基づいた語彙のさまざまな語法・語彙属性についての知識を統合、抽出、ルール化、形式化して知識データベースに組み込んだシステムが、検索、機械翻訳、段落ダイジェストなどに威力を発揮している。

一方、自然言語処理(natural language processing)とよばれる応用研究の領域でも、従来の統語分析からより深いテキストの意味の獲得に向かっている。中国語のように有標の語法規則が非顕在的な言語の解釈には形式文法による分析では不足で、個々の語の意味・形態・機能に焦点をあてた記述が多義解消に有効となる。中国語の語義解析は句法(統語規則)ならびに語用と不可分の関係にあり、用途に応じたフレームワークを使って言語事実を濾過し、相互補完的に分析観察する手法が相応しい[図 1]。本辞書の中国語補語動詞の記述方法もこの理念に沿っている。



[図 1]

¹ <http://ccl.pku.edu.cn/vc/>

2 辞典編纂の経緯

本辞書は平成 14 年 CIEC プロジェクト基金を得、仮構築した Web 版中国語補語 CAI 教材を原型に構築してきた。旧版は既刊の数種の書籍版中国語補語辞典の例文を参照し、見出し語約 2500 語、例文約 3 万条を電子版用にリライト後、日訳を付与し、さらに北京大学「現代漢語短語語法信息詞典」の解析データに従って、中国語動詞補語句の統語規則を書き込み、各種検索ツールをつけて、インターネット上で公開を試みた。しかし動詞補語の選定基準が不明瞭で、機械翻訳向け語法属性の転記だけでは学習者の認識・産出を支援できる内容に至らなかった。2006 年 6 月より北京大学漢語語言学研究中心[Center for Chinese Linguistic PKU(以後 CCL と略)]と、同計算語言学研究所(ICL)の協力を得、それぞれの言語コーパスを利用し、統計的データに基づいて補語句を選別しなおし、その上で新たに語義の記述を開始した。

その後、平成 20 年早稲田大学研究推進部、平成 21 年学術振興会研究成果公開補助金にそれぞれ助成を受け、平成 18 年より現在まで北京大学漢語語言学研究中心、早稲田大学総合研究機構ことばの科学研究所など、言語学と情報科学の専門家集団により作成した。構築に当たり代表性を有する現代中国語大規模コーパスと Web 言語資源から中国語動詞補語の基本データを統計的に抽出し、さらに動詞結合価理論により補語動詞の語義特性を正確に記述した。機械翻訳を想定して書き換え処理した豊富な例文 2 万余件、約 100 万華字は自然言語処理や現代漢語語義の解明に不可欠な世界屈指の中国語補語動詞データ集である。最近の編纂経緯は、詹卫东 李超 2009 “现代汉语述补结构用法词典工作阶段总结報告” に詳しい(砂岡 HP [Work Report])²。

3 データ資源

辞書の母資源は北京大学中文系漢語語言学中心(CCL)と同計算語言所(ICL)の言語コーパス[PKU-CCL-Corpus]で[図 2]、これらを検索利用して得られた基幹動詞と補語をさらに Baidu “百度”や Google “谷歌”など Web 中国語データの出現頻度と照合、常用度を判断、さらに第二外国語学習者向け HSK 基準でレベルを統制した[図 4]。

2009 年現在、CCL のデータ量は 4.77 億 Byte (1.06MB)、かつ新聞から小説、日本文書まで網羅するバランスコーパスである[図 3]。



[図 2]



[図 3]

²早稲田大学砂岡和子 HP <http://www.f.waseda.jp/ksunaoka/corpus/DCCVRCreport090918.pdf>

| word | comp | comp_type | freq_CCL | freq_Baidu | freq_Google | pos | hsk_level |
|------|------|-----------|----------|------------|-------------|-----|-----------|
| 单调 | | | | | | 形容词 | 乙 |

| | | | 合计 | 1792 | 2415 | | 21072 |
|----|----|----|----|------|------|--|-----------|
| 单调 | 得很 | 程度 | 2 | | | | 537,000 |
| 单调 | 死了 | 程度 | 0 | 915 | | | 409,000 |
| 单调 | 多了 | 程度 | 0 | 1910 | | | 1,630,000 |
| 单调 | 得多 | 程度 | 0 | 1980 | | | 1,620,000 |
| 单调 | 下去 | 方向 | 1 | 947 | | | 378,000 |
| 单调 | 起来 | 方向 | 0 | 8740 | | | 422,000 |
| 单调 | 下来 | 方向 | 0 | 804 | | | 434,000 |

[図 4] 述補語彙リスト選別作業“单调”（部分）

4 収録補語句の範囲

4-1 補語類型

中国語補語の定義や分類には異説が多いが、本辞書は以下の (A) (B) を核に、学説で分類される補語形式を網羅する。

- (A) 述語（動詞、形容詞）＋結果補語
- (B) 述語（動詞、形容詞）＋方向補語
- (C) 述語（動詞、形容詞）＋程度補語
- (D) (A) (B) (C) の派生形：可能補語
- (E) 述語（動詞、形容詞）＋介詞補語

(注) 一般の可能補語は上記 AB の拡張形式とみなし、巻頭で派生規則を説明するに留める。ただし“V+得”“V+得了”“V+不了”など、上記 AB に対応関係にない可能補語は収録。時量・動量・介詞補語は個別記述の対象外とし、程度補語も全般的説明に限定する。

4-2 補語選別過程

外国人学習者の便宜を考え、述語は HSK³ 所収の動詞と形容詞を優先して収録した [図 5]。HSK 甲乙丙丁に含まれる動詞 2900 語と形容詞約 1117 語、計約 4000 語がすべて補語句を取るとは限らず、HSK 語彙が実際の中国語の使用状況を反映しているとは限らない。そこで ICL 各種言語データで動詞と形容詞の出現頻度、およびそれぞれの補語との組み合わせ頻度を調査した。たとえば情報詞典（約 7 万 6 千語収録）中の v+v、v+a、v+p 各補語句の出現数を調査したところ、大半が方向補語動詞で計 6282 回（同形語を含む；以下同様）、場所補語が 1172 回でこれに次ぎ、結果補語は 722 回とさほど多くない。同様に ICL コーパスを利用して、補語を取る形容詞について頻度情報とともに抽出作業を行った。

5 記述事項

平成 21 年 7 月時点で、閲覧用の記述事項には以下の事項の記述がある。

- ① 語と補語の基本情報（語形、ピンイン、頻度、語義特性）、
- ② 当該補語句に関する情報（語彙説明、例文、注釈）

³ 汉语水平考试 (Hànyǔ Shuǐpíng Kǎoshì 略称：HSK)、中国の教育部（日本の文部科学省に相当）が認定する中国語の語学検定試験。2008 年から新 HSK 検定試行開始。

③ 補語の用法に関する情報（結果、方向、可能、程度、介詞の各補語の用法特色、補語と修飾語の対比など）

管理者権限により、以下の検索・編集機能を使用できる。

④ 述語と補語の基本情報（通し番号、ピンイン、語形、来源、頻度、語義特性）、

⑤ 当該補語句に関する情報（語彙説明、例文、注釈）

⑥ 補語の用法に関する情報（結果、方向、可能、程度、介詞の各補語の用法特色、補語と修飾語の対比など）（詳細はマニュアル[検索過程と結果]参照）

| 动词级别 | 不分 义项 词数 | 分义项词 数 | 百分比 | 述补结构条 目数 | 百分比 |
|------------|----------------|-----------|--------|-------------|--------|
| HSK(甲) | 278 | 452 | 15.51% | 4891 | 23.21% |
| HSK(乙) | 478 | 628 | 26.67% | 6721 | 31.90% |
| HSK(丙) | 296 | 361 | 16.52% | 3092 | 14.67% |
| HSK(丁) | 177 | 209 | 9.88% | 1674 | 7.94% |
| HSK(非 HSK) | 563 | 623 | 31.42% | 4705 | 22.33% |

6 入力レコード数

平成 22 年 10 月段階の入力データ数は以下の通り。頻度情報に基づいて選別した中国語主幹動詞 2117 語と補語 581 語につき、それぞれ例文、および注釈データの入力をほぼ 100% 終了し、総計 21,072 件、華字 100 万華字の入力が完了している[図 6]。

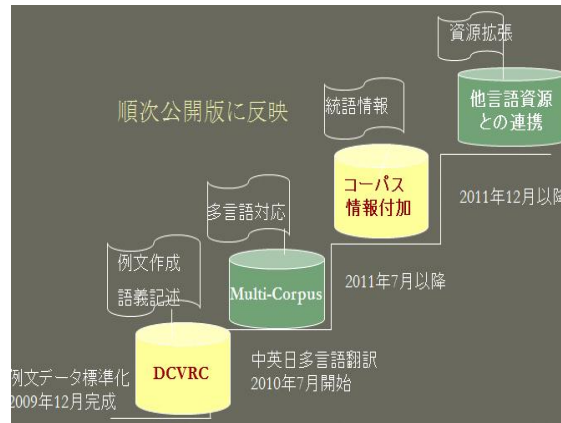
| | | |
|----------------|------------|------------|
| 述补条目总数 | 21072 | |
| 结果补语 | 7974 | 37.84% |
| 趋向补语 | 7271 | 34.51% |
| 可能补语 | 3390 | 16.09% |
| 程度补语 | 1341 | 6.36% |
| 介补补语 | 1096 | 5.20% |
| 述语词数 | 1639(不分义项) | 2117(分义项) |
| 补语词数 | 499(不分义项) | 581(分义项) |
| 带补语次数最多的 述语 | 烧 | 127 个述补实例 |
| 作补语次数最多的 补语 | 起来 | 1122 个述补实例 |
| 已填写释义条目数 | 20954 | 99.44% |
| 已填写例句条目数 | 21072 | 100.00% |

7 検索機能

Web 辞書の特性を生かし、A 単純検索と B 高次検索の 2 種のオンライン検索機能を備える。管理者用検索、編集機能を使えば、主幹動詞ごとの補語付帯率、付帯平均値、同じく類型別補語付帯率など、最多の基幹動詞や補語の頻度情報を動態で可視化できる[詳細はマニュアル：検索過程と結果]参照)。

8 今後の拡張計画

本辞書を活用した研究教育成果を砂岡 HP で公開する(砂岡 2009、詹卫东 2010、王秋萍、覃士娟 2009 など)。平成 23 年以降、東京外国語大学望月圭子研究室、天理大学中川裕三研究室の支援を得て、中国語補語句と例文に英語訳と日本語訳を付け、中日英対照マルチリンガル・オンライン辞典に拡張し、インターネット上で継続公開予定。既存の辞書に欠ける自然言語中の中国語補語動詞を、出現・結合頻度情報とともに自在に検索が可能となるだけでなく、中日英 3 言語の動詞文比較対照研究にとり、極めて実用価値の高い Web 言語資源となろう[図 12]。



参考文献：

北京大学中文系漢語語言学研究中心 (CCL) HP <http://ccl.pku.edu.cn/>

早稲田大学砂岡和子 HP <http://www.f.waseda.jp/ksunaoka/corpus/index.html>

詹卫东 2001 确立语义范畴的原则及语义范畴的相对性，《世界汉语教学》第 2 期

詹卫东 2003 一个汉语语义知识表达框架：广义配价模式《第 5 届全国计算语言学联合学术会议论文集》

詹卫东，汉语语义分类系统及语义关系描述基本框架（设计草案）

http://ccl.pku.edu.cn/doubtfire/semantics/973_Beida/index.htm

詹卫东，「補語辞典編集大綱草稿（中文）」<http://ccl.pku.edu.cn/vc/>

俞士汶他 2005 北京大学『現代漢語語法信息詞典』

侯精一等編著 2001『中国語補語例解』商務印書館

孟琮等 2003『漢語動詞用法詞典』商務印書館

王硯農等編 1987『漢語動詞—結果補語搭配詞典』北京語言学院出版社

繆錦安 1987《汉语的语义结构和补语形式》上海外语教育出版社

Zhu Hong, Yang Liu October, 2006; “MCD: A Chinese-Korean-Japanese Multilingual Concept Dictionary” 2nd International Symposium on Knowledge Processing and Service for China, Japan and Korea: METADATA and ONTOLOGY”, Beijing, China

尹明, 砂岡和子, 成田誠之助 2003 「多国語 Windows 作業系統下基于 Corpus 的中国語教学課件的開發」 The 3rd International Conference on Internet Chinese Education

張国憲 2006 『中国語の補語』《补语的句位探索-关于非可控义》日中対照言語学会白帝社
申亜敏・望月圭子・TerenceSEAH 2007 年中国語の結果を表す複合動詞の分類と英語・日本語との対照コーパス、多言語社会に貢献する言語教育学研究者養成プログラム報告集 5
申亜敏・望月圭子 2009 年「中国語の結果複合動詞-日本語の結果複合動詞・英語の結果構文との比較から」小野尚之編『結果構文のタイポロジー』ひつじ書房所収

砂岡、尹明 2003 「コーパス利用による中国語教育(1) 中国語コーパス利用とデータ変換」
早稲田大学政治経済学部紀要『諸学教養』紀要 114 号

砂岡和子、詹衛東 2007 「言語コーパス利用の中国語電子補語辞典編纂とその課題」日本中国語学会於関東支部拡大例会発表

砂岡和子、劉揚、朱虹 2007 [中韓日 Multilingual Concept Dictionary 構築の現況]
言語処理学会第 13 回年次大会(NLP2007)発表論文