

# 信息處理用現代漢語虛詞義類詞典研究和工作單設計<sup>1</sup>

## Study on the Machine Tractable Thesaurus Dictionary of Contemporary Chinese Functional Words for Information Processing and Design Information Terms for Dictionary Entries

陳群秀\*

Qunxiu Chen

### 摘要

目前，世界各國學者都十分重視語言信息處理的知識資源的建設，知識包括詞彙學知識、句法學知識、語義學知識、語用學知識乃至常識方面的知識，核心問題是語義學知識。在語義學知識中，詞彙的語義知識是最基本最重要的語義知識。在過去的 10 年中，我們清華大學和合作者對漢語實詞中主要的語類（例如：動詞、形容詞、名詞）的詞彙語義知識進行了系統、全面的研究，並且研製出“現代漢語述語動詞機器詞典”、“現代漢語述語形容詞機器詞典”、“現代漢語名詞槽關係系統”和“現代漢語語義分類詞典”。但是漢語虛詞的詞彙語義知識的研究特別是面向信息處理用的虛詞的詞彙語義知識的研究至今還是一個空白點。本文首先討論了漢語虛詞研究對漢語信息處理的意義。漢語是孤立語、詞根分析型語言，大多數漢語詞彙本身不能明顯地表達語法意義，句法手段主要靠虛詞和語序，況且虛詞對表示語態、語氣、時態體貌、能願情態、句子關係、程度、範圍、引入對象等句子語義和篇章結構有關鍵的作用，因此漢語虛詞詞彙語義研究對漢語信息處理有著特別重要的意義。其次，本文介紹了一個信息處理用現代漢語虛詞義類詞典的探索研究，是在漢語的語

---

1 本論文有關研究承 863 項目(項目號 2001AA114210)的資助。

\* 智慧技術與系統國家重點實驗室，清華大學計算機科學與技術系，北京 100084，中國  
The State Key Laboratory of Intelligent Technology and System, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China  
E-mail: cqx@s1000e.cs.tsinghua.edu.cn

義層面上研究漢語虛詞的分類的二十一世紀的表達類，亦即研究一個信息處理用現代漢語句子情態表示系統。然後，本文展示了我們初步研究的成果，我們將虛詞義類初步分爲語態範疇、時態體貌範疇、語氣範疇、能願情態範疇、肯定否定範疇、（句子）關係範疇、程度範疇、範圍範疇、對象範疇、其他範疇等十大範疇，每個大範疇根據虛詞表達的意義不同而再分爲若干中範疇和小範疇。再之後，本文提出信息處理用現代漢語虛詞義類詞典工作單的信息項的設計設想。最後，論文還倡議建立現代漢語語義知識庫平台。

**關鍵字：**信息處理用現代漢語虛詞義類詞典、信息處理用現代漢語句子情態表示系統、語態範疇、時態體貌範疇、語氣範疇、能願情態範疇、肯定否定範疇、（句子）關係範疇、程度範疇、範圍範疇、對象範疇、其他範疇、工作單、信息項、現代漢語語義知識庫平台

### Abstract

At present, worldwide scholars from different countries attach great importance to the knowledge resource construction of language information processing. Knowledge includes the aspects of lexicology, syntax, semantics, pragmatics and even general knowledge. The knowledge of semantics is the core, of which the most basic and most important one is the semantic knowledge of vocabulary. During the last 10 years, scholars of Tsinghua University and Cooperators have conducted systematic and complete studies on lexical semantic knowledge of the main categories of Chinese content words (including verb, adjective, noun, etc), and developed and studied “the Machine Tractable Dictionary of Contemporary Chinese Predicate Verbs”, “the Machine Tractable Dictionary of Contemporary Chinese Predicate Adjectives”, “the System of Relations of Slots Centering on Nouns for Contemporary Chinese”, and “the Thesaurus Dictionary of Contemporary Chinese for Information Processing”. However, research on lexical semantic knowledge for information processing in particular of Chinese functional words has so far been limited. Firstly, this paper discusses the value of functional words in Chinese Information Processing. Chinese is a language whose typology involves isolated language and root-word-based analytical language. The vast majority of Chinese words can not themselves clearly express grammatical sense. Chinese syntactical methods mainly depend on the use of functional words and word order. In addition it is important that Chinese functional words express meaning of sentences and structures of texts, e.g. voice, tense & aspect, mood, modality, positive and negative, relation of sentences, degree, range, draw into object, etc. Therefore, the study of lexical semantics of Chinese functional words is great importance to Chinese Information Processing. Secondly, this paper

introduces our research on the Machine Tractable Thesaurus Dictionary of Contemporary Chinese Functional Words. The results of our initial research are presented. The Chinese functional words are divided into ten main-level categories, e.g. Voice Category, Tense & Aspect Category, Mood Category, Modality Category, Positive and Negative Category, Relation Category of Sentences, Degree Category, Range Category, Object Category, Other Category, etc. Each of main-level category is divided into a number of middle-level categories or minor-level category once and again according to expressing different meaning. This paper also presents some ideas of designing information terms for dictionary entries. Finally, we propose the conception of constructing a base platform for the Contemporary Chinese Semantic Knowledge Bank.

**Keywords:** The Machine Tractable Thesaurus Dictionary of Contemporary Chinese Functional Words, The Modality Representation System of Chinese Sentences for Information Processing, Voice Category, Tense & Aspect Category, Mood Category, Modality Category, Positive and Negative Category, Relation Category of Sentences, Degree Category, Range Category, Object Category, Other Category, Dictionary Entries, Information Terms

## 1. 前言

中文信息處理的研究和進展，依賴于漢語計算語言學的詞彙學、句法學、語義學、語用學的研究和進展，核心問題是語義學。語義學是難度最大、起步較晚的一個薄弱環節。由于漢語缺乏屈折變化，是語義型語言，句法分析對句子的貢獻比英語等要小，因此語義分析對漢語機器理解尤為重要。目前，自然語言理解、漢語信息處理處于一個關鍵時期，處在取得重大突破的前夜，最重要最困難的是語義學的研究和突破。

在計算語言學界，越來越多的專家把機器詞典的規模和質量看作是決定一個自然語言處理系統成敗的關鍵。對於漢語來說，由於缺乏形態變化，漢語的計算機自動分析和處理相對別的语言要困難得多，尤其需要重視語言知識庫特別是語義知識庫的建設。目前中文信息處理領域的語言知識庫有一些，主要是實詞的詞法詞典、實詞的語義詞典、句法規則庫和語料庫，但是至今還沒有一個系統的漢語虛詞詞典。國內外面向人的虛詞的研究也不少，但面向機器自動處理的虛詞研究卻不多，有的話也是零散的個別的，根本沒有系統研究。例如著名的 EDR、WordNet、FrameNet、MindNet 等都是概念詞典，都只有動詞、形容詞、名詞等實詞的概念，都沒有系統地研究虛詞的表達體系。漢語虛詞的語義知識的研究特別是面向信息處理用的虛詞的詞彙語義知識的研究也至今還是一個空白點。

清華大學一直重視和致力於中文信息處理領域的基礎研究，在中文信息處理基礎資源建設方面已經取得了一些成果。在過去的 10 年中，我們對漢語實詞中主要的語類（動詞、形容詞、名詞）的詞彙語義知識進行了系統、全面的研究，並且研製出現代漢語述

語動詞機器詞典、現代漢語述語形容詞機器詞典、現代漢語名詞槽關係系統和現代漢語語義分類詞典。但是對現代漢語虛詞的語義知識特別是面向信息處理用的漢語虛詞的詞彙語義知識一直是我們十分關注而還沒有完成的心願。

漢語是孤立語、詞根分析型語言，大多數漢語詞彙本身不能明顯地表達語法意義，句法手段主要靠虛詞和語序。漢語裏的虛詞往往可以顯示詞與語、片語與片語以及句子與句子之間的關係，成為語句組織的脈絡。同時，虛詞對表達語態、時態體貌、語氣、能願情態、肯定否定、句子關係、程度、範圍、引入對象等句子語義和篇章結構有關鍵的作用，是篇章知識的主要來源。因此漢語虛詞詞彙語義研究對中文信息處理有著特別重要的意義。二十多年來，中文信息處理從字處理發展到詞處理進而發展到句處理這個層面，若要取得新的突破，必須進入篇章處理層面。因為一則句子層面要想作深入處理，必須要依靠篇章的知識，不然的話，很多語法歧義、語義歧義、指稱、照應等問題無法解決。二則中文信息處理的一些應用領域（例如機器翻譯、自動文摘、自動問答系統的源文分析和篇章生成）需要對篇章做出有效的分析。因此對漢語虛詞作系統全面的研究不僅是有意義的而且是迫切需要的。但是漢語虛詞的個性很強，運用範圍很廣，運用頻度又高，有的一詞多類兼多義，而且漢語虛詞應用很靈活且缺省現象很嚴重，因此漢語的虛詞特別是信息處理用虛詞詞典研究具有很大難度。

## 2. 信息處理用現代漢語虛詞義類詞典（漢語情態表示系統）的初步研究

兩年多來，我們對現代漢語的虛詞做了大量的調研、分析和初步研究。我們研究的虛詞包括介詞、副詞、連詞、時態助詞、結構助詞、語氣助詞、助動詞、感歎詞、擬聲詞等詞類。我們對漢語虛詞的研究角度，不僅從虛詞的形式、語法作用角度分析，而且還從虛詞所表示的語義角度和語用角度分析，目的在於研究一個信息處理用現代漢語虛詞義類表達，是在漢語的語義層面上研究漢語虛詞的分類的二十世紀的表達類，亦即研究一個信息處理用現代漢語句子情態表示系統。經過初步研究，我們將漢語虛詞義類分為語態範疇、時態體貌範疇、語氣範疇、能願情態範疇、肯定否定範疇、（句子）關係範疇、程度範疇、範圍範疇、對象範疇、其他範疇十大範疇，每個大範疇根據虛詞表達的意義不同再分為若干中範疇和小範疇。下面將虛詞義類的大、中、小範疇分類例示如下：

### 1 語態範疇：指說話者選擇主體還是選擇客體作話題。

#### 1.1 主動態

#### 1.2 被動態

形態/准形態標誌：被、讓、由

#### 1.3 使役態

形態/准形態標誌：動詞為“讓、請、使、勸、叫、教…”等使役動詞；

2 時態體貌範疇：指因說話者不同的觀察點而表達的事件的時間進程軸上所處的特定階段或與時間無關而與動量有關的運動狀態和情貌

2.1 時態範疇

2.1.1 現在時

2.1.2 過去時

2.1.3 將來時

2.2 體貌範疇

2.2.1 預期體 框架標誌：將 V

2.2.2 即始體 框架標誌：即將 V

2.2.3 開始體 框架標誌：V 起來；V 起；V 開；V 上；

例如：戰爭打起來了；奏起軍歌；議論開了；議論上了；

2.2.4 剛始體 框架標誌：剛 V 起來；剛剛開始

例如：戰爭剛打起來；

2.2.5 進行體 框架標誌：V 著；正 V 著（呢）；正在 V 著；在 V 著；V<sub>1</sub> 著 V<sub>2</sub>；  
V<sub>1</sub> 著 V<sub>1</sub> 著 V<sub>2</sub>；V 著 N；V 著點兒！V 著！

例如：他正寫著報告呢；門開著；她笑著說；說著說著哭了；牆上掛著軍用地圖；聽著點兒！你聽著！

2.2.6 繼續體 框架標誌：V 下去；V 下來；

例如：這種武器還要生產下去；戰士們堅持下來了。

2.2.7 完成體 框架標誌：V 了<sub>1</sub>；V 過<sub>2</sub>(了<sub>1</sub>)；

例如：吃過飯就去。

2.2.8 結果完成體 框架標誌：V 好；V 完；V 成；V 上；V 上來；V 到；V 著(zháo)；

V 住；V 掉；V 下；V 下來；V 去；V 了<sub>3</sub>(lou)；

例如：穿好軍裝；穿上軍裝；穿完軍裝；抓住敵人；抓著敵人；抓到敵人；  
脫下軍裝；脫了<sub>3</sub>軍裝；脫去軍裝；脫掉軍裝；

2.2.9 剛歷體 框架標誌：V…來著；

例如：河對岸剛才打槍來著。

2.2.10 經歷體 框架標誌：V 過<sub>1</sub>；曾經 V 過<sub>1</sub>；曾 V 過<sub>1</sub>；

例如：執行過偵察任務；

2.2.11 試量體 框架標誌：VV；VV…看；V 了<sub>1</sub>V；

例如：把坦克修修；讓他來修修看；

2.2.12 小量體 框架標誌：V 一 V；V 一下；V 了<sub>1</sub>一 V；V 了<sub>1</sub>一下；

- 例如：稍微把準星調一調；
- 2.2.13 反復體 框架標誌：V 來 V 去；V 過來 V 過去；  
例如：跑來跑去；在天上飛過來飛過去；
- 2.2.14 多發體 框架標誌：V<sub>1</sub>V<sub>1</sub>V<sub>2</sub>V<sub>2</sub>；  
例如：說說笑笑；蹦蹦跳跳；
- 2.2.15 已然體 框架標誌：已經 V；已 V；早已 V；(是)N<sub>處</sub>V 的 N；  
例如：是西單上的車；已經打響；
- 2.2.16 未然體 框架標誌：沒有 V；沒 V；尚未 V；  
例如：他沒打槍；敵人尚未進攻；
- 2.2.17 經常體 框架標誌：經常 V；常 V；時常 V；  
例如：經常騷擾敵人；
- 2.2.18 漸進體 框架標誌：漸漸（表示程度或數量的漸變增減）
- 2.1.19 瞬時體 框架標誌：忽然，倏然
- 2.2.20 一般體 框架標誌：(無)  
例如：上前線；觀察敵情；

### 3 語氣範疇：指說話者表示交際意圖的語氣。

#### 3.1 功能語氣

- 3.1.1 陳述語氣 句終標點：。
- 3.1.2 疑問語氣 句終標點：？
- 3.1.3 反詰語氣 形態/准形態標誌+句終標點：{豈|難道|何嘗|不成|不行}+？
- 3.1.4 祈使語氣 句終標點：(吧)+{。|!}
- 3.1.5 感歎語氣 形態/准形態標誌+句終標點：  
{太|多|多麼|真|特|特別}+（{了|啊}）+!
- 3.1.6 確認語氣 框架標誌：({當然|自然})+(是)……+的；
- 3.1.7 非確認語氣 形態/准形態標誌：吧

#### 3.2 情感語氣

- 3.2.1 詫異語氣 形態/准形態標誌：{竟|竟然|居然}
- 3.2.2 料定語氣 形態/准形態標誌：{果然|果真|敢情}
- 3.2.3 領悟語氣 形態/准形態標誌：{難怪|怪不得|敢情|噢|原來}
- 3.2.4 僥倖語氣 形態/准形態標誌：{幸虧|幸好|幸而|省得}

- 3.2.5 索性反正語氣 形態/准形態標誌：{索性|簡真|反正|偏偏|偏不}
- 3.2.6 不滿語氣 形態/准形態標誌：什麼；怎麼；
- 3.2.7 追究語氣 形態/准形態標誌：究竟；到底；
- 3.2.8 肯定承諾語氣 形態/准形態標誌：一定；肯定；
- 3.2.9 加強堅決語氣 形態/准形態標誌：就；

4 能願情態範疇：指因說話者固有的觀點而引發的評價、告誡等態度和強烈程度。

4.1 評價型能願情態

- 4.1.1 “能”類評價 形態/准形態標誌：能；能夠；只能；才能；
- 4.1.2 “要”類評價 形態/准形態標誌：要；想；
- 4.1.3 “會”類評價 形態/准形態標誌：會；必然；必定；不會；一定；
- 4.1.4 “敢”類評價 形態/准形態標誌：敢；
- 4.1.5 “可能”類評價 形態/准形態標誌：可能；不可能；
- 4.1.6 “也許”類評價 形態/准形態標誌：也許；或許；大概；大約；多半；不一定；
- 4.1.7 “沒法”類評價 形態/准形態標誌：沒法；無法；
- 4.1.8 “肯”類評價 形態/准形態標誌：肯；願意；情願；樂意；願；巴不得；想；
- 4.1.9 “可以”類評價 形態/准形態標誌：可以；可；
- 4.1.10 “準備”類評價 形態/准形態標誌：準備；預備；打算；計畫；
- 4.1.11 “得”類評價 形態/准形態標誌：得；不得；“一定得”；“准得”；

4.2 告誡型能願情態

- 4.2.1 “別”類告誡 形態/准形態標誌：別；勿；萬萬不要；
- 4.2.2 “要”類告誡 形態/准形態標誌：要；
- 4.2.3 “應該”類告誡 形態/准形態標誌：應；應該；應當；該；不該；不應；
- 4.2.4 “只有”類告誡 形態/准形態標誌：只有；
- 4.2.5 “必須”類告誡 形態/准形態標誌：必須；一定；務必；
- 4.2.6 “用不著”類告誡 形態/准形態標誌：用不著；不必；無庸；
- 4.2.7 “得”類告誡 形態/准形態標誌：得；

5 肯定否定範疇：指說話者對事件的肯定或否定態度。

- 5.1 全盤肯定 形態/准形態標誌：全；都；通通；全部；統統；

- 5.2 全盤否定 形態/准形態標誌：全不；全沒；都不；都沒；
- 5.3 部分肯定 形態/准形態標誌：並非…都；未必都；
- 5.4 部分否定 形態/准形態標誌：
- 5.5 堅決否定 形態/准形態標誌：統統不；通通不；
- 5.6 雙重否定 形態/准形態標誌：未嘗不可；未嘗沒有；未必不是；
- 5.7 不定 形態/准形態標誌：不 A；不 A（反義）；例：不上不下；不方不圓；不死不活；
- 6 關係範疇：指說話者在用複句表達事情時對事件間(分句間)的語義、時間、邏輯上的關係的認定。
- 6.1 並列關係：各分句在意義上是平排並列的。  
常用的關聯詞語有：既…又…；又…又…；不但…同時…；…同時…；及；
- 6.2 承接關係：各分句間具有先後相繼關係。  
有時用“首先…，其次…，然後…”，“一…就…”；“既而”；
- 6.3 遞進關係：後面分句的意思比前面分句更進一層。  
常用的關聯詞語有：不但…而且…；不但…並且…；…並且…；不但…同時…；…同時…；不但…尤其…；…尤其…；不但…甚至…；…甚至…；不但…也…；…也…；不但…還…；不但…更…；不但…又…；不但…簡直…；不但…反而…；不但…竟然…；尚且…何況…；不但…連…也…；不但…既使…也…；不但…就是…也…；不但…哪怕…也…；不但…而且…甚至…；不但…而且…簡直…；不但…而且…尤其…；不但…哪怕…也…，甚至…也…；且不說…，也不說…，只說說…；要…還要…；
- 6.4 選擇關係：分句間有選擇關係。  
常用的關聯詞語有：是…還是…；不是…就是…；或者…或者…；…，或者…；與其…不如…；與其…寧可…；不是…而是…；毋寧；
- 6.5 轉折關係：分句間意思上有所轉折。  
常用的關聯詞語有：雖然…但是…；…但…；…可是…；雖然…可…；雖然…還…；…，雖然…；…卻…；…可(卻)…；…還…；…就…；…唯獨…；儘管…還是…；
- 6.6 因果關係：分句間具有原因和結果的關係。  
常用的關聯詞語有：因為…，所以…；…，因為…；…，所以…；之所以…，是因為…；因為…，…；…，因此…；…，因而…；…，於是…；…，一則…，二則…；由於…，因此…；既然…，(就)…；…，反正…；…，結果…；到底；

例如：因難年還好辦，反正買不到菜，沒有什麼三盤四樣。

6.7 假設關係：分句間有假設和結果的關係。

常用的關聯詞語有：如果…，則…；…，如果…；要是…就…；…吧…吧；假如…，那末…；假如…就…；如果…，那末…；如果…，那麼…；若是…，那末…；如果…，就…；(如果)…的話，…就…；要不然的話，…；…就…；萬一…，那…；…一…，(那)…；

例如：冰一化，那魚是會發臭的。/不管吧，先進人物發展成道德敗壞，說明制藥廠的政治思想工作不得力；管吧，弄不好又是對先進人物的誹謗和妒忌。

6.8 條件關係：分句間具有條件和結果的關係。

常用的關聯詞語有：只有…才(能)…；不管…也要…；不管…都…；無論…一定要…；只要…就…；只要…(就)一定(能)…；除非…才…；除非…否則…；…一…就…；

例如：除非他去，否則這場球贏不了。/她一聞見煙味就咳嗽。

6.9 讓步關係：一個分句對另一個分句表示退讓一步。

常用的關聯詞語有：即使…也還要…；…，即使…，即使…也…；縱然…都…；…則…，但…；…是…，就是…；

例如：好是好，就是太貴了。/巧則巧矣，但恐不合于說漢話人的心理。

6.10 目的關係：前後分句分別表示行動和目的的關係。

常用的關聯詞語有：爲了…；…，爲的是…；…，也好…；…，以便…；…，免得…；…，以免…；

例如：爲啥不早點來，也好讓大媽把你當作女兒來疼你。/學好本領，以便更好地爲人民服務。/事先要準備好，免得臨時手忙腳亂。

6.11 發展關係：分句間表示程度隨著條件的發展而發展。一般第一個分句爲條件，后面的分句表示隨條件的發展而發展的程度。

常用的關聯詞語有：越…越…；愈…愈…；

例如：環境越是艱苦，他的革命意志越是堅定。

6.12 比擬關係：表示主要內容說完后意猶未盡，又比擬之，或是使未竟之形象化，或是使未竟之意理性化。

常用的關聯詞語有：好象；就象；象；

例如：他嚇得從長椅上跳起來，就象書生幻想美人，美人來了卻是狐狸變的，嚇得他魂不附體。/徐麗莎恍恍惚惚地好幾天，象被狂風卷上了天，在空氣中翻著筋斗。

6.13 詮釋句：后段是直接解釋前段末某個成分的。

前一段末一般都可以加“：”號，或是後段加“說是”之類的詞語。

例如：當時的工作組也曾有過懷疑：這樣的人家能不能稱作城市貧民呢？/我的男朋友只愛我一點，說是跟我在一起時感到很快樂。/南京的新名勝，不用說首推中山陵。

6.14 時同關係(伴隨關係)：表示在做某件事同時或伴隨著發生某件事的關係。

常用的關聯詞語有：一邊…一邊…；邊…邊…；一面…一面…；連…帶…；隨…隨…；

例如：他一邊走一邊嗑著瓜子。/她一路上連哭帶罵。/兒子總是邊看電視邊吃飯。

7 程度範疇：指說話者對程度大小、程度高低、程度深淺的認定態度。

例如，表示悲痛的程度有：很；十分；十二分；相當；太；極其；萬分；無比，等等。又如，表示不滿程度的有：很；非常；大為；極為，等。

8 範圍範疇：指說話者對物體數量多少、時間和處所的範圍大小的認定態度。

例如，表示整體、全體或部分的有：統統；全都；無一；無一不；都；大都；大半；大體；多半；僅僅；只；才不過；惟獨；除了；除去；除開，等等。表示處所的範圍大小有：處處；到處；無處不；大多；只；太凡；僅，等等。表示時間的範圍大小有：時時；時時刻刻；時常；有時；偶爾，等等。

9 對象範疇：指說話者對涉及的個體、時間、地點、依據的引入。

9.1 引進對象範疇：指說話者對引進對象的表示。

例如，由“給”、“于”、“對”、“爲”、“由”、“替”、“向”、“沖”、“沖著”、“奔”、“奔著”、“當著”等引進的人或物對象。

9.2 來源去向經由範疇：表示時間、地點的來源、去向或經由的範疇。

例如，“打”、“從”、“來自”、“由”、“向”、“打從”、“從打”、“經”、“奔”、“朝”等。

9.3 依據憑藉範疇：表示說話、做事的依據和憑藉的範疇。

例如，“根據”、“據”、“憑”、“憑藉”等。

9.4 目的範疇：表示說話或做事的目的對象。

例如，“爲”、“爲了”等。

10 其他範疇：不屬於 1~9 範疇的其他範疇。

例如，象“啾啾”、“噤哩呱啦”、“淅淅瀝瀝”、“啞當”、“嗷嗷”、“吧唧”、“吧嗒”這樣的象聲詞，還有表示並列名詞或名詞性片語的“和”、“與”、“跟”、“及”、“以及”等。

### 3. 信息處理用現代漢語虛詞義類詞典表示信息項目設計

對於機器詞典來說，所表示的信息項目的設立最為重要。我們的信息處理用現代漢語虛詞義類詞典（現代漢語情態表示系統）信息表示項目設立原則是虛詞的語法形式、語法意義、語義意義、語用用法相結合，期望為大家提供最精確完整和細緻的信息。我們的工作單中對每一個虛詞描述信息項目有：詞形、拼音、詞類、義項數目、義項序號、釋義、情態範疇（包括大範疇、中範疇、小範疇）、表示意義、表示意義的範疇值、形態/准形態標誌、框架標誌、常用的近義/同義關聯詞語、句例、備註。工作單樣式如下：

詞條序號				製作者				工作單號				
詞形			拼音			詞類			義項數目			義項序號
釋義												
情態範疇	大範疇				中範疇				小範疇			
表示意義												
表示意義的範疇值												
形態/准形態標誌												
框架標誌												
常用的近義/同義關聯詞語												
句例												
備註												

其中“拼音”填寫的是該虛詞在該範疇下的拼音讀寫，填寫的拼音必須帶聲調，用1、2、3、4、5分別表示漢語的四聲和輕聲。“詞類”填寫的是該虛詞在該拼音和該範疇下的詞類。“義項數目”指的是該虛詞在該拼音該詞類下共有的義項數目，“義項序號”指的是該虛詞在該拼音該詞類該義項數目下的第幾個義項。“情態範疇”填寫的是該虛詞在該拼音該詞類該義項序號下屬於什麼大範疇、什麼中範疇、什麼小範疇。“表示意義”填寫的是該虛詞表示的語法意義、語義意義，包括語體。此項填寫有助於辨析該虛詞該義項與該虛詞其他義項的區別。“表示意義的範疇值”表示的是“程式範

疇”、“範圍範疇”等範疇的屬性值。例如，表示程度範疇的副詞有“很、十分、十分、十二分、萬分、極、相當、挺、無比”，填寫程度範圍值時“十分”一詞填寫“10”，“很”填寫值“10”。“十二分”填寫“12”、萬分填寫“10000”。“相當”填寫“8”、“挺”填寫“9”，“極”填寫“1000”，“無比”填寫“+∞”等。這樣將來機器理解時可以對它們進行比較和計算。“框架標誌”填寫的是該虛詞在使用環境中可以多少種模式框架，與什麼類詞性的詞搭配使用等等。“常用的近義/同義關聯詞語”顧名思義是填寫與之同義或相近意義的關聯詞語或與之成對使用、前後呼應的關聯詞語。“句例”填寫的是該拼音該詞類該義項序號該範疇下的句子各種用法的真實例子。“備註”中可填寫在以上各信息項中尚未包括的信息內容或作必要的說明。

#### 4. 現代漢語語義知識庫平台建設的構想

目前我們正在編寫和修改“信息處理用現代漢語虛詞義類詞典”填寫規範，並且正在進行詞典工作單的試填寫實驗。信息處理用現代漢語虛詞義類詞典的管理軟件系統也初步設計和實現了，準備在試填寫實驗后進行工作單的正式填寫和錄入校對，以期儘快構建這個漢語虛詞義類詞典（亦即信息處理用現代漢語句子情態表示系統）。同時，我們也正在對已構建的“現代漢語述語動詞機器詞典”、“現代漢語述語形容詞機器詞典”、“現代漢語名詞槽關係系統”、“信息處理用現代漢語語義分類機器詞典”等四個實詞的語義詞典進行整合集成和機器學習功能的研究。待“信息處理用現代漢語虛詞義類詞典”構建後，我們準備將這五個實詞、虛詞的語義詞典整合集成一個現代漢語語義知識庫平台，此平台可為中文信息處理提供豐富、全面、可靠的語義知識支持（包括詞彙語義知識和句子語義知識），可為現代漢語語言學、語義學研究、對外漢語教學、中小學語文教學計算機輔助教學提供有力的工具和資源。目前我們正在進行現代漢語述語動詞機器詞典、現代漢語述語形容詞機器詞典在漢語信息處理方面應用的探索，也正在研究和設計這些語義詞典在對外漢語教學輔助教學方面的應用軟件。

#### 參考文獻

- 林杏光，《詞彙語義和計算語言學》。北京：語文出版，1999。
- 林杏光，《複句與表達》。北京：中國物資出版社，1986。
- 陳群秀，“信息處理用現代漢語句型系統的初步研究”，第二十屆東方語言計算機處理國際學術會議（20<sup>th</sup> ICCPOL' 2003）論文集《Advances in Computation of Oriental Language》，2003年8月，205-212。北京：清華大學出版社。
- 申小龍，《漢語句型研究》。海南：海南人民出版社，1989。
- 賀陽，“漢語‘語氣’(Modality)及其標誌簡表”，1991。
- 史有為，羅建林，“漢語‘體’及其標誌簡表”，1991。
- 陳群秀，“漢語自然語言理解研究概況、前景及難點討論”，1990年國際中文與東方語言計算機處理學術會議論文，長沙，1990。
- 陸儉明，馬真，《現代漢語虛詞散論》。北京：北京大學出版社，1985。

- 朱德熙，《語法問答》。北京：商務印書館，1985。
- 呂叔湘，《漢語語法分析問題》。北京：商務印書館，1979。
- 胡裕樹主編，《現代漢語（增訂本）》。上海：上海教育出版社，1981。
- 房玉清，《實用漢語語法》。北京：北京語言學院出版社，1984。
- 姚殿芳，潘兆明，《實用漢語修辭》。北京：北京大學出版社，1987。
- 黃伯榮，廖序東，《現代漢語（下冊）》。蘭州：甘肅人民出版社，1980。
- 陳群秀，“信息處理用現代漢語虛詞義類研究初步構想”，香港：第四屆漢語辭彙語義學研討會（4thCLSW），2003年6月。

