

コーパスにおける“了₁”と共起する動詞に関する調査

林 如

1. 調査の背景

中国語における“了₁”¹⁾の意味と機能に関する研究は多方面で行われている。“了₁”は動詞接辞として、動詞に後続するため、“了₁”が伴われる動詞の特徴は“了₁”の研究にとって、重要である。本調査の目的は、コーパスから“了₁”と共起する動詞を集計し、動詞の特徴を調査するところである。

本調査の調査方法はプログラムを使用してコーパスのデータを分析することである。具体的には、コーパスから“了₁”が含まれる例文を収集し、例文から“了₁”と共起する動詞を集計し、出現頻度の高い動詞を配列する。

“了₁”と共起する動詞を集計するために、“了₁”の前の動詞を統計する独自のプログラム“LE Analysis”を設計した²⁾。“LE Analysis”プログラムは大量のデータを分析することが可能である。“LE Analysis”プログラムの中には、“jieba”³⁾と“word2vec”⁴⁾ライブラリを用いたPythonプログラムが含まれており、またこれは幅広いデータを扱っているコーパスを分析するものである。

2. データ源とデバイス

中国語研究でよく使用されているコーパスは“北京大学中国语言学研究中心”（略としてCCLコーパス）と“北京语言大学语料库”（略としてBCCコーパス）がある。BCCコーパス自体は統計機能が付いているため、統計の結果が公式サイトから直接に観察できる。しかし、BCCコーパスではダウンロードできる項目は最大1万件に過ぎず、また重複する項目が多いため、大量の

データを分析するには適していない。本調査は大量のデータの収集を前提としているため、CCLコーパスを用いる。BCCコーパスを主要なデータ源として用いらず、参考データとして使用する。

語学研究のコーパス以外に、本調査では“Blah 书库”⁵⁾にある小説のデータ源も扱う。“Blah 书库”は小説・随筆・漫画・翻訳された洋書を含んだ多様なスタイルの項目をまとめたコーパスである。

分析に使用されたデバイスはAsusのUX331U Notebook PC (Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz, 1.80 GHz, 8.00Gb RAM) である。用いられるプログラム言語はPythonのversion 3.6.4であり、“jieba”のversion 0.42.1であり、“Gensim(word2vec)”のversion 3.8.3である。

3. データの収集

本調査はビッグデータにおいて、“了”を伴う動詞の使用頻度を調査する。そのため、大量のデータを収集しなければならない。データの収集とはコーパスから大量の文をダウンロードすることである。しかし、実際の操作では複数のトラブルが生じた。

第一に、自動ダウンロードの問題である。それぞれのコーパスから助詞“了”を含む例文をダウンロードする際に、CCLコーパスの場合は大量のデータをサイトから直接ダウンロードできるが、“Blah 书库”は一括でダウンロードできないため、“wget”というツールを使って自動的にダウンロードすることになる。

第二に、コーパスからダウンロードできるデータ量の問題である。CCLコーパスのデータを収集する際に、助詞“了”を入力すると、2,591,053個のデータが表示されたが、ダウンロードできる数量は800,000個に過ぎず、すべてのデータをダウンロードすることはできなかった。つまり、CCLコーパスから助詞“了”を含む例文をすべてダウンロードすることはできなかった。例文800,000個の範囲で動詞を集計することができるが、プログラムの使用とビッグデータの背景を考えると、数量不足と見なされている。そこで、データ不足の問題を補うために、“Blah 书库”のデータも取得し、数量の補充を行うことにした。

“Blah 书库”から書籍1,177冊、合計2,000,000個以上の文を収集した。また、BCC コーパスは参考データ源として2,067,458個の例文を収集した。

第三に、“了₁”と“了₂”の区別の問題である。この段階でのデータにおいては、“了₁”と“了₂”を区別せず、すべての“了”が含まれている。この問題の解決方法はデータ分析の段階（4.4節）で説明する。

第四に、RAMの容量の問題である。簡単に言えば、高品質のデバイスを使用すると、更に大きなコーパスについての分析が可能であり、簡単に大量のデータを読み込むことができる。さらに大きなRAMの容量を持つデバイスがあれば、計算の時間が早くなる。つまり、RAM容量の大きさはデータ分析の品質と時間に影響を与える。

4. 調査のステップ

4.1 事前処理 (preprocess)

この段階の役割はデータを処理することである。コーパスから収集されたデータには、括弧、記号、外国語などの要素が含まれているため、データ分析に不便である。そのため、具体的な操作方法としては、python言語を使ってプログラムを作成し、データにおける分析に不要な要素を自動的に認識して、削除することになる。例えば、CCLコーパスから収集されたデータは以下のようなものである。

- (1) 教育学的发展大体经历 [了] 萌芽、独立形态、马克思主义教育形成三个阶段。学习教育学的基本方法是理论联系实际。【文件名:\当代CWAC\AEB0001.txt 文章标题：作者：】

上記のデータの中で注目されている要素“了”の外側には角括弧が付いており、文の後ろにはデータの出典などの情報があることが分かる。このような要素は本調査のデータ分析の対象ではないため、取り除く必要がある。しかし、データの量が多いため、手動で1つずつ削除するという操作は調査では効率が悪い。そのため、プログラムを使って先行データを処理することが必要になる。処理されたデータは以下のようなものである。

- (2) 教育学的发展大体经历了萌芽、独立形态、马克思主义教育形成三个阶段。学习教育学的基本方法是理论联系实际。

この方法で処理すると、必要の部分だけを残し、分析するデータ量を減らす。

4.2 区切り (cut)

この段階では主に前の段階から得られたデータを分割する。分割の根拠は主に“。”と“！”と“？”などの句読点である。本調査で文を分割するために使用されたプログラムは“jieba”というモジュール (module) である。“jieba”は Sun Junyi (百度北京) より公開され、“PaddlePaddle”というディープラーニング (deep learning) 法を用いた、MIT ライセンスを持つ Python 言語プログラムである。

“jieba”と同様に文を分割するために使用されるプログラムは pkuseg (北京大学) や THULAC (清華大学) などもある。ただし、他のプログラムと比較すると、“jieba”は文を分割する分野でより広く使用されている。つまり、“jieba”は他のプログラムよりもデータを処理する経験が豊富であり、バグの例をより適切に処理でき、データ処理においてより安定している。そのため、本調査は“jieba”のプログラムを使用する。

具体的な操作は以下のようなものである。

- (3) 教育学的发展大体经历了萌芽、独立形态、马克思主义教育形成三个阶段。

文を分割する際に、文に“了”を含まない部分が削除され、“了”がある文だけを残しておき、この後の“V+了”という構造を抽出することが容易である。

4.3 品詞の付与 (symbolize)

この段階で行うことは、前の段階で取得したデータの各部分に品詞を割り当てることである。品詞を与え、“了”の前後の要素を区別することができ、後続のプログラムの操作を容易にする。

各要素の品詞の基準は、《现代汉语词典》(第7版)⁶⁾、“jieba 辞典”⁷⁾、『白

コーパスにおける“了”と共起する動詞に関する調査

白水社中国語辞典』⁸⁾及び“汉语词典”⁹⁾により、統一されている。それぞれの辞典から取得した情報は、語彙・使用頻度・品詞に限った。1つの要素に複数の品詞がある場合、辞典で先頭に記載している品詞を設定した。各辞典の情報信頼性を考慮した上で、「《现代汉语词典》(第7版) > 白水社中国語辞典 > jieba辞典 > 汉语词典」という順位で品詞を取得した。

上記の辞書に基づき、すべてのデータの要素に品詞を与える。分析のため、一部の品詞を同じ品詞としてまとめ、下の【表1】のように分けている。

記号	品詞		辞典での品詞記号			
			现代汉语词典	白水社中国語辞典	jieba辞典	汉语词典
A	形容词	adjective	形	形容詞	a	形容词
C	连词	conjunction	连	接続詞		并列连词
D	副词	adverb	副	副詞	d	副詞
N	名词	noun	名	名詞	n/f/s/t/nr/ns/nt/nw/nz	名词
P	介词	adposition	介	前置詞	p	介词
Q	量词	quantity	量, 数	量詞	q	量词/数词
R	代词	pronoun	代	代詞	r	代词/疑问词
V	动词	verb	动	動詞	v	动词
U	助词	particle	助	助詞	y/uv/u/uj/df	助词
M	语素	morpheme		付屬形態素	g/ag/rg/tg	
J	感叹词	interjection	叹	感嘆詞		
O	拟声词	onomatopoeia	拟声	擬声擬態語	o	拟声词
S	词缀	affix		接辭	k/x	
X	助动词	auxiliary verb		助動詞		
T	补语	complement		補語		
E	成语	idiome				成语

【表1】品詞対照表

ただし、分析しやすくするために、一部の要素、例えば“了”や“的”のような助詞には品詞を与えず、その要素の品詞は要素自体にする。例えば、“了”の品詞は“了”となる。

品詞を修正した後、下の【図1】のようになる。

0	教育学的发展大体经历了萌芽、独立形态、马克思主义教育形成三个阶段
1	N_ 的 V_ N_ V_ 子 V_ , N_ , N_ N_ N_ V_ O_ N_ N_
2	但数学在中学中提出
3	上述各种解释从不同的角度揭示了教育活动的特点, 虽然说法不一, 但都把教育看成是培养人的一种社会活动
4	A_ R_ V_ P_ A_ 的 N_ V_ 子 N_ V_ 的 N_ , C_ A_ , C_ D_ V_ N_ V_ A_ V_ N_ 的 O_ N_ N_ V_
5	这些新学校的产生具备了条件
6	R_ D_ V_ N_ 的 V_ V_ 子 N_
7	意味, 学校就产生了
8	R_ , N_ D_ V_ 子
9	学校教育一开始便成为政治的产物, 成了他们进行阶级统治的工具, 具有明显的阶级性
10	N_ N_ Q_ V_ D_ N_ N_ V_ , V_ R_ V_ N_ V_ 的 N_ , V_ A_ 的 N_
	在西方, 古希腊的斯巴达, 雅典产生了文法学校, 经学学校, 体操学校以及青年军训团等教育机构
	P_ N_ , N_ 的 N_ , N_ V_ 子 N_ N_ ; N_ N_ , N_ N_ C_ N_ V_ V_ 等 N_ N_
	N_ V_ 子 N_ N_
	到了13、14世纪, 由于手工业和商业的发展, 城市里出现了由手工业联合会的行会学校和商人联合会的基尔特学校, 着重学习生产和业务知识, 为本行业培养人才
	V_ 子 Q_ , Q_ Q_ N_ , P_ N_ A_ N_ 的 V_ , N_ N_ V_ 子 P_ N_ N_ V_ 的 N_ N_ A_ N_ N_ V_ 的 N_ D_ N_ , V_ V_ V_ A_ N_ , V_ N_ N_
	17世纪英国资产阶级革命成功, 建立了世界上第一个资产阶级政府
	O_ Q_ N_ N_ N_ V_ V_ , V_ 子 N_ N_ O_ N_ N_
	大机器代替了手工工场, 科学技术在生产上得以广泛应用, 在不到200年时间内, 社会生产便由机械化到电气化, 直到现在的电子计算机和自动控制普遍应用的航天时代
	A_ N_ V_ 子 N_ N_ , N_ P_ V_ N_ V_ A_ V_ , P_ V_ Q_ Q_ N_ N_ , N_ V_ D_ P_ V_ V_ , A_ 的 N_ A_ N_ A_ V_ 的 V_ N_

【图1】品詞修正結果図

この段階には不足点がある。それは品詞の正確性の検証問題である。品詞が複数ある要素があるため、プログラムが品詞を選択する際にエラーが発生する可能性がある。例えば図の中の3と4では、“产生”(産生)という単語がある。4の“产生”は動詞であるが、3の場合では名詞である。3の“产生”は“了”の直前に置かれていないため、今回の調査には影響がない。しかし、本プログラムは品詞を1つしか認識できないため、文脈による品詞の分析は技術の面で、まだ解決できない。

さらに、特殊な文型では品詞は確定している。例えば、“教育”(教育)という要素には名詞と動詞の2つの品詞があるが、図の中の例2は“把”の構文であるため、“把”の後の位置に置かれる要素は名詞しかなく、プログラムが選択した品詞は名詞である。したがって、プログラムの検証を容易にするために、品詞の多義性問題を解決する方法は、プログラムと組み合わせる専門の辞書を使用することである。

中国語学研究において、孟琮らが編集した《汉语动词用法词典》が動詞の検証に有利であるが、電子データが見つからないため、より正確な検証を行うことができない。辞書の問題が解決できれば、実験結果はより正確になると考えている。

4.4 簡略化 (clean)

この段階ではプログラムを利用して、“V+了”という動詞句を含む文だけを

残し、つまり“了_i”が存在する文だけを残した。この操作の目的は、“了₂”によるデータへの影響を大幅に軽減し、次のステップにおいて動詞を抽出するために、利便性を高めることである。

プログラムを使って文字を1つずつ検索し、“了_i”が検索できた際に前の単語が動詞であるかどうかを確認し、動詞であればその文を保留し、動詞でない場合その文を削除する。そうすることによって、“了_i”の意味の正確さを最大限保証する。その結果は以下の【図2】の通りである。

```

0  教育学的发展大体经历了萌芽
   N__的V__N__V__了V__
1  但数字早在甲骨文中就出现了
   CVND__N__NDV__了
2  上述各种解释从不同的角度揭示了教育活动的特点
   A__R__V__PA__的N__V__了N__V__的N__
3  这些都为学校的产生准备了条件
   R__DVN__的V__V__了N__
4  学校就产生了
   N__DV__了
5  成了他们进行阶级统治的工具
   V了R__V__N__V__的N__
6  雅典产生了文法学校
   N__V__了N__N__
7  古埃及产生了宫廷学校
   N__V__了N__N__
8  城市里出现了由手工业联合会办的行会学校和商人联合会办的基尔特学校
   N__N__V__了PN__N__V__的N__N__AN__N__V__的N__DN__
9  建立了历史上第一个资产阶级校
   V__了N__NO__N__N__
10 大机器代替了手工工场
    AN__V__了N__N__
    
```

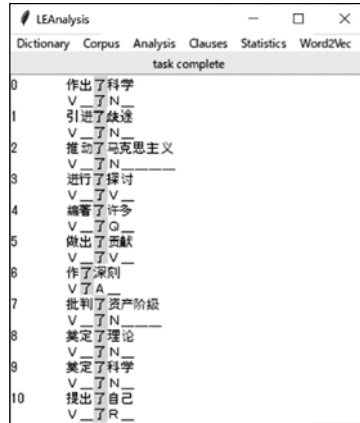
【図2】結果簡略化図

この段階の欠点は前の段階と同じく、品詞の弁別には有効な専門辞書の電子データが公開されていないため、多義性や誤ったデータが解釈されることである。

4.5 文型とフォーカス (make clauses and focus on)

この段階では主に前の段階のデータの文型構造をまとめ、簡略化された文型を統計する。つまり、前の段階のデータから扱った動詞句を単独の構造にする。これは文型ごとに出てくる頻度を集計するのを容易にするためである。それから動詞と後の要素との関係によって、初期の統計をして、プログラムを使用して同様な構造を合併し、基本的な文型構造を得た。このうち“V了N”という構造は最も出現回数の多い構造であり、それに基づいた“了_i”の前後の要素の分析は正確である可能性が高い。次に、このステップでは“了_i”の前後の要素に

着目し、前の動詞と後の名詞句や数量詞句をすべて抽出した。つまり、前の段階の文をより小さい動詞句の単位に単純化することであり、1つの長文に複数の動詞句が含まれている場合、それを1つとして見なすのではなく、2つに分割することである。その結果は下の【図3】の通りである。



【図3】文構造図

このように処理すると、多様な文構造を記号化し、次の計算にとって簡単になる。そして、最後の結果（例えば文構造の現れる頻度）を配列する際にも便利になる。

4.6 計算 (statistics)

この段階では主に前段階のデータを集計する。各文において、“了”に関するパターン（例えば“V了N”）を得る。コーパス全体にプログラムを繰り返し使用してそれぞれのパターンの発生頻度を数える。前の動詞の出現回数と後の名詞や数量詞などの出現状況を計算して統計表を作成する。統計表は以下の【表2】、【表3】の通りである。

コーパスにおける“了_i”と共起する動詞に関する調査

word	type	了。	了・	了N	了Q	了A	了C	了D	了P	了R	了V	了U	了M	了J	了O	了S	了X	了T	了E
成	V/A/N	14	7640	4011	942	746	20	139	13	694	1061	2	7	0	1	0	0	0	0
到	V/A/N	225	6817	3339	1289	423	19	130	12	582	996	3	12	0	0	0	0	0	0
参加	V	23	7331	3187	377	233	11	16	48	747	2692	1	7	0	0	0	0	0	11
有	V/N	115	11380	3084	1022	4073	21	511	144	858	1653	1	4	0	0	1	1	0	1
介绍	V	6	3927	2636	247	84	3	47	12	567	323	1	6	0	0	0	0	0	0
促进	V	0	3369	2522	62	276	1	7	6	249	245	0	1	0	0	0	0	0	0
进行	V	17	12562	2314	1492	2789	15	325	50	61	5487	1	1	0	0	0	0	0	26
建立	V	9	5217	2221	954	777	1	76	126	141	914	2	2	0	0	0	0	0	0
形成	V	26	6484	2006	1452	985	16	123	504	347	1041	1	8	0	0	0	0	0	0
得到	V	10	5161	1846	277	1119	5	286	54	160	1410	3	1	0	0	0	0	0	0
取得	V	0	9589	1715	645	4879	2	297	283	81	1684	0	0	0	0	0	0	0	0
成立	V	35	3111	1705	385	145	3	78	308	46	438	1	1	0	1	0	0	0	0
实现	V	56	3417	1676	230	367	5	72	92	178	792	1	1	0	0	2	0	0	0
反映	V	4	2773	1630	220	233	5	9	14	430	229	1	1	0	0	1	0	0	0
提出	V	4	7039	1584	1362	1445	11	282	137	342	1871	0	3	0	0	0	0	0	0
提供	V	1	7366	1564	1089	2929	3	458	106	99	1113	0	0	0	0	0	0	0	2

【表2】CCLコーパス“了_i”の使用状況統計表

word	type	了。	了・	了N	了Q	了A	了C	了D	了P	了R	了V	了U	了M	了J	了O	了S	了X	了T	了E
到	V/A/N	1613	37080	19836	5056	2055	109	705	57	5855	3109	39	135	8	50	11	7	0	0
在	V/P/D	24	13238	7576	2078	646	56	118	4	2143	543	10	54	0	3	0	0	0	0
成	V/A/N	86	15815	6559	3430	1672	37	322	18	1711	1963	12	50	2	11	2	0	0	11
出	V	54	10015	5322	1819	664	32	403	6	1272	450	3	31	0	5	3	0	0	0
有	V/N	281	14412	4747	2552	1965	60	388	33	1845	2748	9	26	5	2	3	10	0	2
进	V/Q/N	19	6219	4304	755	242	9	39	2	583	258	3	17	3	3	0	0	0	0
变成	V	8	7655	2938	2015	1010	17	162	20	500	945	4	15	0	17	1	5	0	6
离开	V	629	4046	2620	92	129	13	26	1	921	191	21	15	0	5	3	0	0	0
来到	V	48	4147	2410	585	200	19	52	4	574	274	2	22	1	2	0	0	0	0
回到	V	0	3715	2398	291	185	8	29	1	598	189	4	10	0	0	0	0	0	1
看	V	558	15024	2341	4093	449	31	1066	50	1589	5265	97	22	1	6	2	3	1	0
失去	V	74	4600	2297	273	387	12	57	15	485	1042	2	5	0	2	1	0	0	1
发生	V	398	6859	2029	972	897	14	97	4	1825	969	11	2	0	7	3	6	0	0
去	V	3835	4771	1968	899	206	19	149	21	827	241	291	21	17	58	19	5	0	2
做	V	561	8533	1966	3575	638	39	268	20	1055	904	23	6	1	12	5	6	0	2
起	V/P/N	29	5108	1850	1058	740	10	133	6	249	1023	0	23	0	4	3	2	0	2

【表3】Blahコーパス“了_i”の使用状況表

コーパスから得られた動詞は1万語以上あるが、出現頻度で配列すると、配列の後半の語は動詞と判別しかねたため、本調査は前半の配列1-100の動詞を対象とし、それらの特徴を考察する。

5. 調査結果

本調査では、CCLコーパスと“Blah 书库”の中から、“V了N”構造に当てはまる動詞を集計し、再頻出動詞をそれぞれ100語選んで考察する。集計表は以下の通りである。

rank	Blah	CCL	rank	Blah	CCL	rank	Blah	CCL	rank	Blah	CCL
1	到	成	26	给	完成	51	当	设立	76	受到	开展
2	在	到	27	发现	影响	52	死	发表	77	写	签订
3	成	参加	28	开	推动	53	知道	做	78	穿	实行
4	出	有	29	吃	发生	54	占	加强	79	昏	走上
5	有	介绍	30	杀	赢得	55	想到	做出	80	挡住	充满
6	进	促进	31	没	奠定	56	完	找到	81	丢	留下
7	变成	进行	32	恢复	保证	57	放	失去	82	达到	显示
8	离开	建立	33	换	受到	58	入	来到	83	找	选择
9	来到	形成	34	打开	解决	59	引起	决定	84	停下	回到
10	回到	得到	35	走进	获得	60	生	产生	85	动	构成
11	看	取得	36	得到	增强	61	产生	讨论	86	断	打破
12	失去	成立	37	看见	增加	62	接受	给	87	掌握	表现
13	发生	实现	38	忘	进	63	参加	发	88	破	当
14	去	反映	39	不见	制定	64	买	回答	89	脱	加入
15	做	提出	40	听到	作	65	变	访问	90	错	调动
16	起	提供	41	走出	变成	66	发出	召开	91	带来	损害
17	见	会见	42	说	带来	67	获得	达到	92	遇到	看
18	充满	出现	43	喝	在	68	犯	发现	93	尽	去
19	走	出	44	来	改变	69	传来	开始	94	挤满	表达
20	听	创造	45	露出	举行	70	留下	给予	95	加入	担任
21	受	引起	46	改变	进入	71	交给	展示	96	送到	保持
22	找到	讲	47	住	作出	72	作	造成	97	响起	经历
23	进入	提高	48	吸	看到	73	皱起	离开	98	选择	掌握
24	出现	体现	49	没有	接受	74	布满	起	99	提供	阐述
25	想起	通过	50	拿	签署	75	走到	带动	100	占据	规定

【表4】コーパス1-100動詞集計表

2つのコーパスから、同じ動詞が39個、異なる動詞が122個見られ、合計161個の動詞を集計した。

コーパスにおける“了”と共起する動詞に関する調査

Vendler (1967) は動詞の場面を、state (状態動詞)、activity (活動動詞)、accomplishment (達成動詞)、achievement (到達動詞) 4種類に分けた。達成動詞は活動動詞から生じ、有界的な範囲が設けられている。例えば、runは活動動詞であるが、有界的な距離があれば、run a mileは達成動詞になる。本調査はコーパスにおける動詞の特徴を考察し、目的語や修飾語などの有界的な範囲を考えない。そのため、本調査の動詞の分類はVendler (1967) を援用し、state (状態動詞)、activity (活動動詞)、achievement (到達動詞) の3種類を用いる。

“Blah 书库”とCCLコーパスにおける動詞の状況をまとめて【表5】で表す。【表5】では、動詞分類の例とコーパスによって現れる数量を表す。“Blah 书库”の数量は前に示し、CCLコーパスの数量は後に示している。

結果を表す動詞 63, 60	述補構造 51, 58	動方式 离开, 带来, 走出, 加入 21, 12		
		動結式 30, 46	典型的な動結式 看见, 打破, 赢得, 回到 24, 26	
			述補式複合詞 6, 20	瞬間的な動詞 出现, 改变 5, 9
				瞬間的でない動詞 恢复, 推动 1, 11
	単音節動詞 到, 成, 完, 变 12, 2			
状態動詞 在, 有, 知道 5, 5				
活動動詞 走, 听, 选择, 保持 37, 38				

【表5】動詞分類表

【表4】によると、“Blah 书库”の場合、活動動詞37語（“走、听、选择”など）

と状態動詞5語(“在、有、忘、当、知道”)がある他、結果を表す動詞が多く見られた。結果を表す動詞は合計63語あり、そのうち単音節の瞬間動詞12語と「述語-補語類の動詞」(以下、述補動詞とする)51語が含まれている。さらに、述補動詞は「動詞-結果補語式」(以下、動結式とする)と「動詞-方向補語式」(以下、動方式とする)に分けられ¹⁰⁾、動結式は30語、動方式は21語見られた。また、動結式はさらに典型的な結果補語動詞24語と述補式複合詞¹¹⁾6語が分けられた。述補式複合詞は瞬間的な動詞5語(“发生、出现、发现、改变、产生”)と瞬間的でない動詞1語(“恢复”)に分けている。

CCLコーパスの場合も同様、活動動詞38語(“走、听、选择”など)と状態動詞5語(“有、在、当、担任、保持”)、結果を表す動詞は60語見られた。そのうち単音節の瞬間動詞2語(“成、到”)と述補動詞58語が見られ、さらに述補動詞において、動結式は46語、動方式は26語ある。動結式はさらに典型的な結果補語動詞は26語あり、述補式複合詞は20語がある。述補式複合詞はさらに瞬間的な動詞は9語(“建立、实现、设立”など)、瞬間的でない動詞は11語(“推动、奠定、保证”など)に分けられる。

本調査はコーパスを利用して、大量のデータを集計した。“了”を伴う動詞の中に、述補構造の動詞または結果を表す動詞が多い。それは“了_i”の意味機能に関わると推測する。具体的な理由と“了_i”の意味機能は今後の研究課題とする。

〈注〉

- 1) 中国語における助詞“了”を大きく分類すると、“了₁”と“了₂”の2種類がある。“了₁”は動詞接辞として動詞の後に現れ、“了₂”は文末助詞として文末に現れる。本調査の調査対象は“了₁”と“了₂”が伴われる動詞である。
- 2) プログラムの作成は慶応義塾大学LEVE VICTORの協力である。
- 3) 公式サイト: <https://github.com/fxsjy/jieba>
- 4) 公式サイト: <https://code.google.com/archive/p/word2vec/>
- 5) 公式サイト: <https://blah.me/>
- 6) 公式サイト: <https://github.com/CNMan/XDHYCD7th>

- 7) 公式サイト：https://github.com/fxsjy/jieba/raw/master/extra_dict/dict.txt.big
- 8) 公式サイト：<https://cjjc.weblio.jp/>
- 9) 公式サイト：<https://cidian.911cha.com/>
- 10) 朱德熙 (1982) は、述補構造を粘着性述補構造と結合性述補構造に分けたと指摘している。補語が述語の後に直接付いている構造は粘着性述補構造であり、“得”が付いている述補構造は結合性述補構造である。
- 11) 述補式複合詞という分類方法は施春宏 (2003/2008) で提唱されている。これは述語動詞と補語動詞は意味上の関連が緊密であるため、述補構造は複合詞になることを指す。

〈参考文献〉

- 施春宏 (2003/2008) : 《汉语动结式的句法语义研究》, 北京: 北京语言大学出版社。
- 朱德熙 (1982) : 《语法讲义》, 北京: 商务印书馆。
- Vendler Z. (1967) : *Linguistics in philosophy*. Ithaca: Cornell University Press.